



UNIVERSIDADE FEDERAL DO PAMPA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

CARLOS ALEXANDRE SILVA DOS SANTOS

**RECONHECIMENTO DE IMAGENS DE MARCAS DE GADO UTILIZANDO REDES
NEURAS CONVOLUCIONAIS E MÁQUINAS DE VETORES DE SUPORTE**

ALEGRETE

2017

CARLOS ALEXANDRE SILVA DOS SANTOS

**RECONHECIMENTO DE IMAGENS DE MARCAS DE GADO UTILIZANDO REDES
NEURAIS CONVOLUCIONAIS E MÁQUINAS DE VETORES DE SUPORTE**

Dissertação apresentada ao programa de Pós-graduação Stricto Sensu em Engenharia Elétrica da Universidade Federal do Pampa, como requisito parcial para obtenção do Título de Mestre em Engenharia Elétrica.

Orientador: Daniel Welfer

ALEGRETE

2017

Ficha catalográfica elaborada automaticamente com os dados fornecidos
pelo(a) autor(a) através do Módulo de Biblioteca do
Sistema GURI (Gestão Unificada de Recursos Institucionais) .

S284r Santos, Carlos Alexandre Silva dos
Reconhecimento de Imagens de Marcas de Gado utilizando
Redes Neurais Convolucionais e Máquinas de Vetores de Suporte
/ Carlos Alexandre Silva dos Santos.

135 p.

Dissertação(Mestrado)-- Universidade Federal do Pampa,
MESTRADO EM ENGENHARIA ELÉTRICA, 2017.

"Orientação: Daniel Welfer".

1. Aprendizagem Profunda. 2. Redes Neurais Convolucionais.
3. Máquinas de Vetores de Suporte. 4. Reconhecimento de
imagens. 5. Marcas de gado. I. Título.

Carlos Alexandre Silva dos Santos

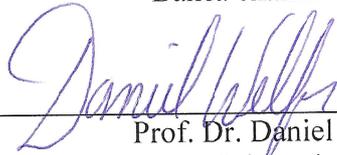
RECONHECIMENTO DE IMAGENS DE MARCAS DE GADO UTILIZANDO REDES NEURAIAS CONVOLUCIONAIS E MÁQUINAS DE VETORES DE SUPORTE

Dissertação apresentada ao programa de Pós-graduação Stricto Sensu em Engenharia Elétrica da Universidade Federal do Pampa, como requisito parcial para obtenção do Título de Mestre em Engenharia Elétrica.

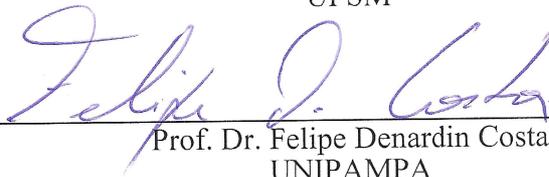
Área de concentração: Sistemas de Energia

Dissertação defendida e aprovada em 26 de setembro de 2017.

Banca examinadora:



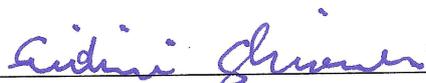
Prof. Dr. Daniel Welfer
Orientador
UFSM



Prof. Dr. Felipe Denardin Costa
UNIPAMPA



Prof.^a Dr.^a Patricia Della Mèa Plentz
UFSC



Prof. Dr. Sidinei Ghissoni
UNIPAMPA

Dedico esta dissertação a todos aqueles que têm um sonho e jamais desistem dele. Que são curiosos e sempre buscam alavancar novos conhecimentos. Que lutam todas as batalhas com força e honra, sem jamais entregar-se.

AGRADECIMENTOS

A Deus por iluminar meus caminhos. Sempre senti sua presença, tanto nos bons momentos quanto nas horas mais difíceis de minha vida.

À minha família que sempre esteve ao meu lado e que me ensinou todos os princípios necessários para que me tornasse um homem digno.

À minha noiva que foi compreensiva em relação ao tempo de dedicação que dispensei ao Mestrado.

Aos meus professores, do ensino fundamental ao mestrado. Meus mestres, que sempre me apoiaram durante essa jornada, que são fontes de inspiração para mim, em especial: à Professora Ana da 4ª série do Ensino Fundamental; às Professoras Joscelaine (Português), Nara (Matemática) e Hilda (História/Geografia), da 5ª à 8ª série do Ensino Fundamental; à Professora Ediles (Português) e ao Professor Ciro (Matemática), do Ensino Médio; à Professora Josiane do Ensino Técnico e também da Graduação; à Professora Janaina da Especialização; ao Professor Roberto da Graduação e ao Professor Alessandro, da Graduação e também do Mestrado.

Ao meu orientador, prof. Daniel Welfer, ao qual tenho muito respeito e admiração. Agradeço eternamente pela oportunidade dada a mim de poder realizar esse sonho. Agradeço por estar sempre ao meu lado, corrigindo-me quando necessário, apontando os melhores caminhos para a pesquisa e sempre agindo de forma ética e profissional.

Ao meu grande amigo e incentivador Sérgio Bortolin Júnior, ex-colega de graduação e atualmente Analista de Tecnologia da Informação da Unipampa, Campus Alegrete. Obrigado pelo apoio e dicas sempre oportunas. Homem íntegro, competente, inteligente e sempre disponível aos amigos.

Ao grupo de pesquisa de Arquitetura de Computadores e Microeletrônica (GAMA) da UNIPAMPA, como também aos colegas do PPGEE, Nathalie, Alian e Anderson, pelo apoio e incentivo durante o curso.

Por fim, à Prefeitura Municipal de São Francisco de Assis, pela concessão de horário especial para acompanhamento das disciplinas do curso de mestrado e pelo fornecimento do conjunto de imagens utilizado no trabalho.

*“Quando tudo o resto falhar ou faltar resta-nos
a imaginação.”
Albert Einstein*

RESUMO

O reconhecimento automático de imagens de marca de gado é uma necessidade para os órgãos governamentais responsáveis por esta atividade. Para auxiliar neste processo, este trabalho propõe uma arquitetura que seja capaz de realizar o reconhecimento automático dessas marcas. Nesse sentido, uma arquitetura foi implementada e experimentos foram realizados com dois métodos: *Bag-of-Features* e Redes Neurais Convolucionais (CNN). No método *Bag-of-Features* foi utilizado o algoritmo SURF para extração de pontos de interesse das imagens e para criação do agrupamento de palavras visuais foi utilizado o clustering *K-means*. O método *Bag-of-Features* apresentou acurácia geral de 86,02% e tempo de processamento de 56,705 segundos para um conjunto de 12 marcas e 540 imagens. No método CNN foi criada uma rede completa com 5 camadas convolucionais e 3 camadas totalmente conectadas. A 1ª camada convolucional teve como entrada imagens transformadas para o formato de cores RGB. Para ativação da CNN foi utilizada a função ReLU, e a técnica de *maxpooling* para redução. O método CNN apresentou acurácia geral de 93,28% e tempo de processamento de 12,716 segundos para um conjunto de 12 marcas e 540 imagens. O método CNN consiste de seis etapas: a) selecionar o banco de imagens; b) selecionar o modelo de CNN pré-treinado; c) pré-processar as imagens e aplicar a CNN; d) extrair as características das imagens; e) treinar e classificar as imagens utilizando SVM; f) avaliar os resultados da classificação. Os experimentos foram realizados utilizando o conjunto de imagens de marcas de gado de uma prefeitura municipal. Para avaliação do desempenho da arquitetura proposta foram utilizadas as métricas de acurácia geral, *recall*, precisão, coeficiente *Kappa* e tempo de processamento. Os resultados obtidos foram satisfatórios, nos quais o método CNN apresentou os melhores resultados em comparação ao método *Bag-of-Features*, sendo 7,26% mais preciso e 43,989 segundos mais rápido. Também foram realizados experimentos com o método CNN em conjuntos de marcas com número maior de amostras, o qual obteve taxas de acurácia geral de 94,90% para 12 marcas e 840 imagens, e 80,57% para 500 marcas e 22.500 imagens, respectivamente.

Palavras-chave: Aprendizagem Profunda. Redes Neurais Convolucionais. Máquinas de Vetores de Suporte. Reconhecimento de imagens. Marcas de gado.

ABSTRACT

The automatic recognition of cattle branding is a necessity for government agencies responsible for this activity. In order to improve this process, this work proposes an architecture which is able of performing the automatic recognition of these brandings. The proposed software implements two methods, namely: Bag-of-Features and CNN. For the Bag-of-Features method, the SURF algorithm was used in order to extract points of interest from the images. We also used K-means clustering to create the visual word cluster. The Bag-of-Features method presented a overall accuracy of 86.02% and a processing time of 56.705 seconds in a set containing 12 brandings and 540 images. For the CNN method, we created a complete network with five convolutional layers, and three layers fully connected. For the 1st convolutional layer we converted the input images into the RGB color format. In order to activate the CNN, we performed an application of the ReLU, and used the maxpooling technique for the reduction. The CNN method presented 93.28% of overall accuracy and a processing time of 12.716 seconds for a set containing 12 brandings and 540 images. The CNN method includes six steps: a) selecting the image database; b) selecting the pre-trained CNN model; c) pre-processing the images and applying the CNN; d) extracting the features from the images; e) training and classifying the images using SVM; f) assessing the classification results. The experiments were performed using the cattle branding image set of a City Hall. Metrics of overall accuracy, recall, precision, Kappa coefficient, and processing time were used in order to assess the performance of the proposed architecture. Results were satisfactory. The CNN method showed the best results when compared to Bag-of-Features method, considering that it was 7.26% more accurate and 43.989 seconds faster. Also, some experiments were conducted with the CNN method for sets of brandings with a greater number of samples. These larger sets presented a overall accuracy rate of 94.90% for 12 brandings and 840 images, and 80.57% for 500 brandings and 22,500 images, respectively.

Keywords: Deep Learning. Convolutional Neural Networks. Support Vector Machines. Image recognition. Cattle branding.

LISTA DE ILUSTRAÇÕES

Figura 1 – Fluxograma do processo de registro de marca de gado realizado atualmente no município de São Francisco de Assis, Brasil.	19
Figura 2 – Livros de marcas de gado no município de São Francisco de Assis, Brasil.	20
Figura 3 – Ficha catalográfica de marcas de gado do município de São Francisco de Assis, Brasil.	20
Figura 4 – Sistema especialista para auxílio no processo de reconhecimento de imagens de marcas de gado.	21
Figura 5 – Etapas de um sistema de processamento de imagens.	24
Figura 6 – Etapas que compõem o método <i>Bag-of-Features</i>	26
Figura 7 – Núcleos de Convolução do SURF (boxfiler ou filtro caixa).	28
Figura 8 – A figura mostra o ponto central marcado por um X, o qual é comparado com seus 8 vizinhos da mesma escala e seus 9 vizinhos da escala superior e inferior.	29
Figura 9 – Fluxograma do algoritmo <i>K-means</i>	30
Figura 10 – Modelo de um neurônio artificial.	32
Figura 11 – Arquitetura de uma RNA Multilayer Perceptron.	33
Figura 12 – Mapas de características de uma arquitetura convolucional.	36
Figura 13 – Visão geral da arquitetura de uma rede neural convolucional.	36
Figura 14 – Arquitetura de rede neural convolutiva com kernels de tamanho 5x5.	37
Figura 15 – Hiperplano de decisão.	41
Figura 16 – Imagens de marcas de gado utilizadas nos experimentos.	53
Figura 17 – Fluxograma sumarizado do método proposto.	57
Figura 18 – Extração de 6.751 pontos de interesse da marca “1093” usando o algoritmo SURF.	58
Figura 19 – Histograma com o número de ocorrências de palavras visuais e tamanho do dicionário.	60
Figura 20 – Método Bag of Features aplicado no trabalho proposto.	61
Figura 21 – Fluxograma sumarizado do método proposto.	63
Figura 22 – Modelo geral de uma arquitetura CNN.	64
Figura 23 – Arquitetura da rede neural convolucional proposta.	66
Figura 24 – Filtros da 1ª camada convolucional do experimento realizado.	67
Figura 25 – Ilustração do algoritmo de extração de características desenvolvido.	68
Figura 26 – Matriz de confusão obtida na etapa de validação - Experimento I	71
Figura 27 – Percentual de marcas classificadas corretamente e percentual de erros - Experimento I.	72
Figura 28 – Percentual de marcas corretamente classificadas - Experimento I.	73

Figura 29 – Tempo de processamento do algoritmo em função do número de amostras - Experimento I.	75
Figura 30 – Matriz de confusão obtida na etapa de validação - Experimento II. .	77
Figura 31 – Percentual de marcas classificadas corretamente e percentual de erros - Experimento II.	78
Figura 32 – Proporção de marcas corretamente classificadas em função do número de amostras - Experimento II.	79
Figura 33 – Tempo de processamento do algoritmo em função do número de amostras - Experimento II.	81
Figura 34 – Matriz de confusão obtida na etapa de validação - Experimento III. .	82
Figura 35 – Matriz de confusão obtida na etapa de validação - Experimento IV. .	84
Figura 36 – Recall comparado dos métodos CNN e BoF - Experimentos I e II. .	86
Figura 37 – Precisão comparada dos métodos CNN e BoF - Experimentos I e II.	87
Figura 38 – Análise de tempo de processamento dos métodos CNN e BoF - Experimentos I e II.	88

LISTA DE TABELAS

Tabela 1 – Métricas de avaliação para um classificador	43
Tabela 2 – Escala para interpretação do coeficiente Kappa	44
Tabela 3 – Trabalhos relacionados	50
Tabela 4 – Marcas, proprietários e total de amostras por marca	53
Tabela 5 – Sensitividade (TPR), Especificidade (TNR) e Precisão (PPV) - Experimento I	74
Tabela 6 – Acurácia Geral, Erro e Tempo de Processamento - Método BoF - Experimento I	76
Tabela 7 – Sensitividade (TPR), Especificidade (TNR) e Precisão (PPV) - Experimento II	80
Tabela 8 – Acurácia Geral, Erro e Tempo de Processamento - Método CNN - Experimento II	81
Tabela 9 – Acurácia por marca - Método CNN - Experimento III	83
Tabela 10 – Acurácia Geral e Tempo de processamento - Método CNN - Experimento IV	85
Tabela 11 – Resultados comparativos dos métodos propostos para a arquitetura desenvolvida - Experimentos I e II.	88
Tabela 12 – Resultados comparativos de Acurácia Geral e Taxa de erro dos Experimentos I, II, III e IV.	90
Tabela 13 – Classificações Corretas versus Classificações Preditas - Método BoF	105
Tabela 14 – Classificações Corretas versus Classificações Preditas - Método CNN	106
Tabela 15 – Acurácia por marca do melhor resultado obtido nas simulações realizadas no Experimento IV	107

LISTA DE ABREVIATURAS E SIGLAS

2D	Objetos e entidades com duas dimensões
Acc	Accuracy
BoF	Bag-of-Features
BoVW	Bag of Visual Words
BoW	Bag of Words
CBIR	Content-Based Image Retrieval
CNN	Convolutional Neural Network
CUDA	Compute Unified Device Architecture
Err	Erro
FAO	Food and Agriculture Organization
FN	Falso Negativo
FP	Falso Positivo
GPU	Graphics Processing Unit
k-NN	k-Nearest Neighbors
LCN	Local Contrast Normalization
MatLab	Matrix Laboratory
MLP	Rede neural Multilayer Perceptron
MSER	Maximally Stable Extremal Regions
OA	Overall Accuracy
PCA	Principal Component Analysis
PNG	Portable Network Graphics
PPV	Positive Predictive Value
ReLU	Rectified Linear Unit
RGB	Espaço de cor que utiliza, na formação de cores, as componentes Vermelha (Red), Verde (Green) e Azul (Blue)

RNA	Redes Neurais Artificiais
SIFT	Scale Invariant Feature Transform
SURF	Speeded Up Robust Features
SVM	Support Vector Machines
TNR	True Negative Rate
TPR	True Positive Rate
VLAD	Vector of Locally Aggregated Descriptors
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo

LISTA DE SÍMBOLOS

k	Coeficiente Kappa
ϕ	Mapeamento do conjunto de entrada

SUMÁRIO

1	INTRODUÇÃO	18
1.1	MOTIVAÇÃO	18
1.2	OBJETIVOS	21
1.3	ORGANIZAÇÃO DA DISSERTAÇÃO	22
2	FUNDAMENTAÇÃO TEÓRICA	23
2.1	ETAPAS DE UM SISTEMA DE PROCESSAMENTO DE IMAGENS	23
2.2	BAG-OF-FEATURES	25
2.2.1	DETECÇÃO E DESCRIÇÃO DE PONTOS DE INTERESSE	27
2.2.2	CONSTRUÇÃO DO VOCABULÁRIO OU DICIONÁRIO VISUAL	29
2.2.2.1	ALGORITMO K-MEANS	30
2.3	REDES NEURAIIS ARTIFICIAIS	31
2.3.1	REDE NEURAL MULTILAYER PERCEPTRON	33
2.3.2	ALGORITMO BACKPROPAGATION	34
2.4	APRENDIZAGEM PROFUNDA (DEEP LEARNING)	34
2.4.1	REDES NEURAIIS CONVOLUCIONAIS	35
2.4.1.1	ARQUITETURA	38
2.4.1.2	PRÉ-TREINAMENTO NÃO SUPERVISIONADO	39
2.5	MÁQUINAS DE VETORES DE SUPORTE	40
2.6	AVALIAÇÃO DO CLASSIFICADOR	42
2.6.1	MATRIZ DE CONFUSÃO	42
2.6.2	MÉTRICAS DE AVALIAÇÃO	42
2.7	CONSIDERAÇÕES	45
3	TRABALHOS RELACIONADOS	46
3.1	BAG-OF-FEATURES	47
3.2	REDES NEURAIIS CONVOLUCIONAIS	48
3.3	CONSIDERAÇÕES	51
4	DESENVOLVIMENTO DA ARQUITETURA	52
4.1	MATERIAIS	52
4.1.1	CONJUNTO DE IMAGENS	52
4.1.2	EQUIPAMENTOS UTILIZADOS	54
4.1.3	MATRIX LABORATORY (MATLAB)	54
4.1.3.1	PACOTE NEURAL NETWORK	55
4.1.3.2	PACOTE PARALLEL COMPUTING	55
4.1.3.3	PACOTE STATISTICS AND MACHINE LEARNING	55
4.1.4	BIBLIOTECA OPEN SOURCE VLFEAT	55

4.2	MÉTODO BOF	56
4.2.1	ETAPAS DO MÉTODO PROPOSTO	56
4.2.1.1	SELEÇÃO DO BANCO DE IMAGENS	57
4.2.1.2	EXTRAÇÃO DOS PONTOS DE INTERESSE DAS MARCAS ATRAVÉS DO ALGORITMO SURF	57
4.2.1.3	CRIAÇÃO DO DICIONÁRIO DE PALAVRAS UTILIZANDO K-MEANS	58
4.2.1.4	CRIAÇÃO DO HISTOGRAMA E VETOR DE CARACTERÍSTICAS DAS PALAVRAS VISUAIS	59
4.2.1.5	TREINAMENTO E CLASSIFICAÇÃO DAS IMAGENS COM A UTILIZAÇÃO DE MÁQUINAS DE VETORES DE SUPORTE	61
4.2.1.6	AVALIAÇÃO DOS RESULTADOS DA CLASSIFICAÇÃO - MATRIZ DE CONFUSÃO	62
4.3	MÉTODO CNN	62
4.3.1	ETAPAS DO MÉTODO PROPOSTO	62
4.3.1.1	SELEÇÃO DO BANCO DE IMAGENS	64
4.3.1.2	SELEÇÃO DO MODELO DE CNN PRÉ-TREINADO	64
4.3.1.3	PRÉ-PROCESSAMENTO DAS IMAGENS E APLICAÇÃO DA CNN	64
4.3.1.4	EXTRAÇÃO DE CARACTERÍSTICAS DAS IMAGENS	67
4.3.1.5	TREINAMENTO E CLASSIFICAÇÃO DAS IMAGENS COM A UTILIZAÇÃO DE SVM	68
4.3.1.6	AVALIAÇÃO DOS RESULTADOS DA CLASSIFICAÇÃO - MATRIZ DE CONFUSÃO	68
4.4	CONSIDERAÇÕES	69
5	EXPERIMENTOS, RESULTADOS E DISCUSSÕES	70
5.1	EXPERIMENTO I: MÉTODO BOF APLICADO À 12 MARCAS EM UM CONJUNTO DE 540 IMAGENS	70
5.1.1	ACURÁCIA, SENSITIVIDADE, PRECISÃO E COEFICIENTE KAPPA	70
5.1.2	DESEMPENHO DE PROCESSAMENTO	74
5.2	EXPERIMENTO II: MÉTODO CNN APLICADO À 12 MARCAS EM UM CONJUNTO DE 540 IMAGENS	76
5.2.1	ACURÁCIA, SENSITIVIDADE, PRECISÃO E COEFICIENTE KAPPA	76
5.2.2	DESEMPENHO DE PROCESSAMENTO	80
5.3	EXPERIMENTO III: MÉTODO CNN APLICADO À 12 MARCAS EM UM CONJUNTO DE 840 IMAGENS	82
5.3.1	ACURÁCIA GERAL	82
5.4	EXPERIMENTO IV: MÉTODO CNN APLICADO À 500 MARCAS EM UM CONJUNTO DE 22.500 IMAGENS	83
5.4.1	ACURÁCIA GERAL	83
5.5	DISCUSSÕES	85

5.5.1	LIMITAÇÕES DO MÉTODO PROPOSTO	90
6	CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS	92
	REFERÊNCIAS	94
	APÊNDICES	101
	APÊNDICE A - EXEMPLOS DE AMOSTRAS DAS MARCAS DE GADO UTILIZADAS PARA TREINAMENTO E VALIDAÇÃO	102
	APÊNDICE B - PSEUDOCÓDIGO DO MÉTODO <i>BAG-OF-FEATURES</i>	103
	APÊNDICE C - PSEUDOCÓDIGO DO MÉTODO REDES NEURAIAS CONVOLUCIONAIS	104
	APÊNDICE D - CLASSIFICAÇÕES CORRETAS <i>VERSUS</i> CLASSIFICAÇÕES PREDITAS - EXPERIMENTOS I e II	105
	APÊNDICE E - ACURÁCIA POR MARCA - EXPERIMENTO IV	107
	APÊNDICE F - ARTIGO PUBLICADO NA REVISTA IEEE LATIN AMERICA TRANSACTIONS	112
	APÊNDICE G - ARTIGO SUBMETIDO NA REVISTA IBEROAMERICANA DE INTELIGÊNCIA ARTIFICIAL	119
	ANEXOS	132
	ANEXO A - CONJUNTO DE IMAGENS DE MARCAS DE GADO UTILIZADO NO TRABALHO	133

1 INTRODUÇÃO

1.1 MOTIVAÇÃO

O desenvolvimento de ferramentas computacionais para auxiliar a análise e reconhecimento de imagens é alvo de interesse dos mais renomados centros de pesquisa do mundo. O uso da computação para análise e reconhecimento de imagens está em constante desenvolvimento, gerando vários benefícios à sociedade nas mais diversificadas áreas do conhecimento.

No que se refere particularmente ao reconhecimento de marcas de gado, atividade esta tão tradicional e de grande relevância socioeconômica para países da América Latina, incluindo o Brasil, não há um método para este fim. Para se ter uma dimensão da importância desta atividade, de acordo com a *Food and Agriculture Organization* – FAO, entre os países produtores, o Brasil e a Índia possuem os maiores rebanhos, sendo que o Brasil encontra-se em 1º lugar, com uma média de 209.215.666 cabeças (SEPLAN, 2015).

A produção pecuária possui um papel relevante na formação social, e que ainda hoje se mantém como uma atividade de grande importância nas expressões culturais associadas a ela, pois está associada à cultura e ao modo de vida do campo, além do papel na afirmação ou construção de identidades individuais ou de grupos (ARNONI, 2013). A prática de registrar a posse sobre algo é relevante para o indivíduo ou família a qual o gado pertence, pois é dessa forma que os animais são identificados quando ficam misturados aos de outras pessoas ou de outros grupos. Essa atividade assume ainda mais importância sociocultural na Região Platina, formada pela região que abrange grande parte do pampa do estado do Rio Grande do Sul no Brasil, pelo Uruguai, e a parte da Mesopotâmia argentina, formada pelas províncias de Misiones, Corrientes e Entre Ríos.

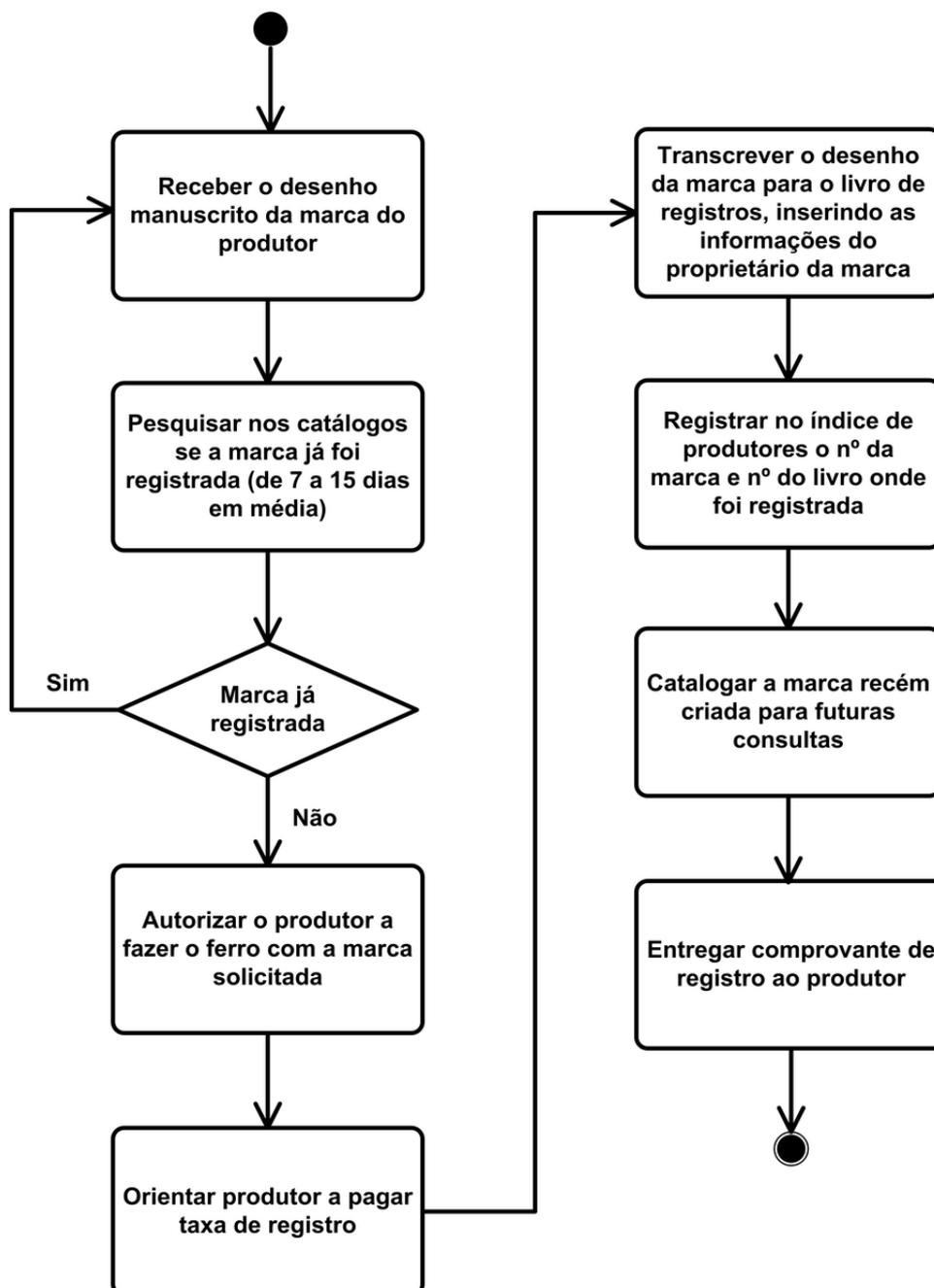
A utilização das marcas ou sinais no gado pressupõe o reconhecimento público de sua propriedade por um indivíduo ou grupo. Utilizadas desde o início da colonização ibérica na América, o início de sua institucionalização se deu a partir do registro em órgãos, sendo oficiais, reconhecidamente portadores de legitimação pública (ARNONI, 2013).

No Brasil as tentativas e os investimentos de aprimoramento no sistema de registro de marcas de gado sempre foi alvo de polêmica, em função da resistência dos pecuaristas. O grande receio está associado ao temor de perda das marcas familiares e o significado que elas adquiriram com o tempo. Atualmente, os registros de marcas no Brasil são realizados nos municípios, em geral sem uma sistematização mais efetiva e sem a necessidade de serem renovadas.

A esses registros seguem-se as regulamentações que procuram, além de oficializar a marcação em si, orientar a forma e o período a ser feita, discriminar a forma

do registro, instituir valores a este, a construção dos ferros e a taxação governamental. O fluxograma do processo de registro de marca de gado realizado atualmente no município de São Francisco de Assis, Brasil, é apresentado na Figura 1.

Figura 1 – Fluxograma do processo de registro de marca de gado realizado atualmente no município de São Francisco de Assis, Brasil.



Fonte: do próprio autor.

Em geral, os registros das marcas de gado constituem-se de livros com os desenhos das marcas e com a identificação de seu proprietário. A Figura 2 apresenta os livros de marcas de gado no município de São Francisco de Assis, Brasil.

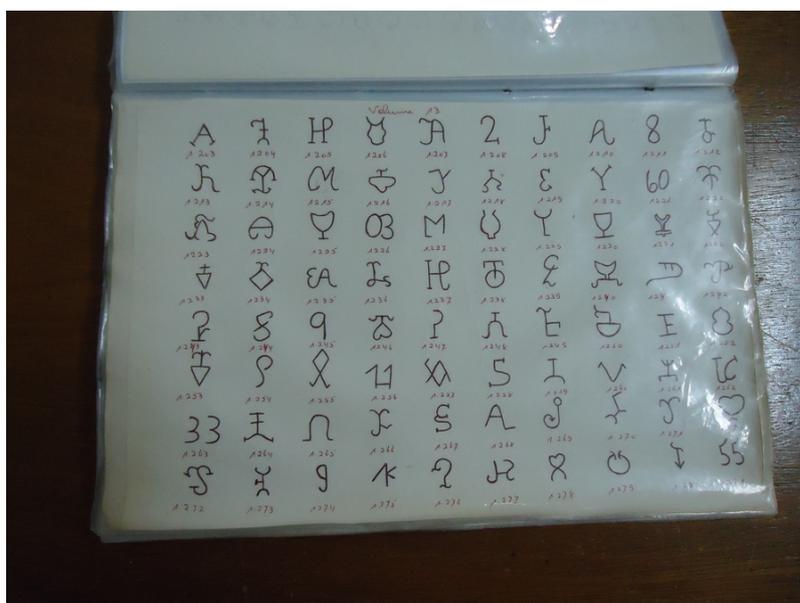
Figura 2 – Livros de marcas de gado no município de São Francisco de Assis, Brasil.



Fonte: Setor de Registro de Marcas de Gado - São Francisco de Assis (2016).

A Figura 3 demonstra uma ficha catalográfica de marcas de gado do município de São Francisco de Assis, Brasil, utilizada para simples consulta.

Figura 3 – Ficha catalográfica de marcas de gado do município de São Francisco de Assis, Brasil.



Fonte: Setor de Registro de Marcas de Gado - São Francisco de Assis (2016).

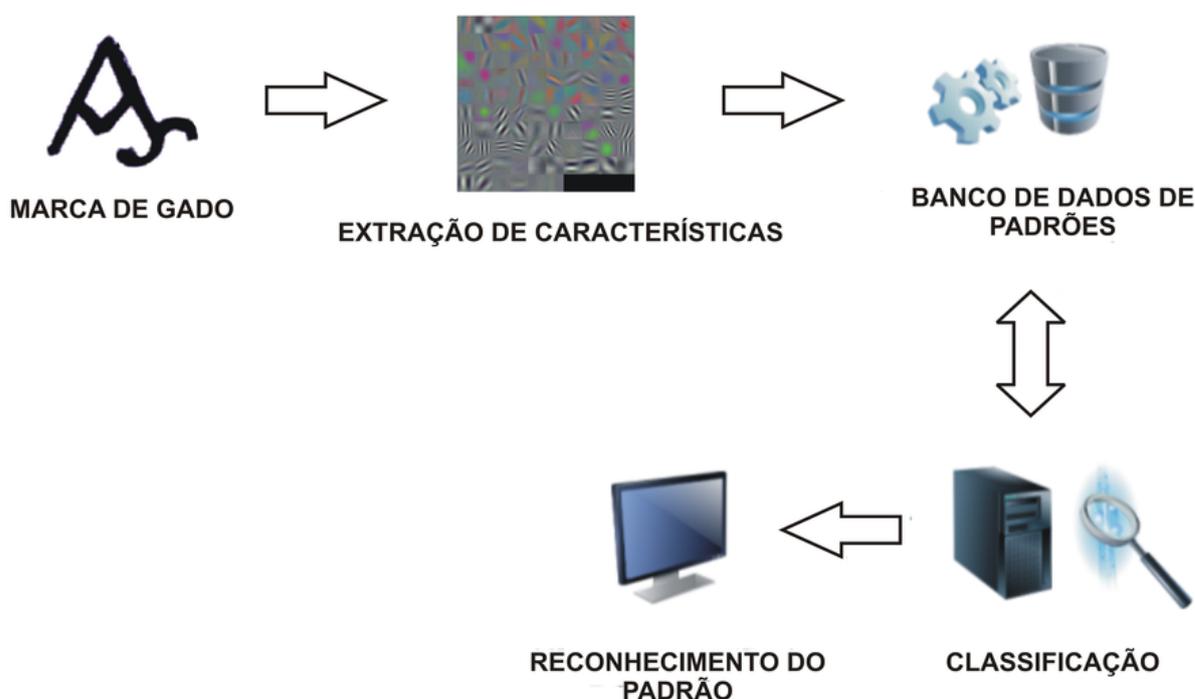
De acordo com as Figuras 2 e 3, é possível observar que o processo de registro, classificação, catalogação e consulta das marcas de gado no município de São Fran-

cisco de Assis é realizado através de livros físicos, sem o auxílio de qualquer tipo de sistema especialista para reconhecimento das imagens das marcas.

Esse processo produz uma demora acentuada no registro de novas marcas, uma vez que é necessário realizar uma verificação em todo o acervo de marcas do município, além disso, também eleva o risco de falha humana no processo de consulta, uma vez que a conferência é realizada somente por inspeção visual em um conjunto de imagens com aproximadamente 11.000 registros.

A Figura 4 apresenta uma ilustração de um sistema especialista para auxílio no processo de reconhecimento de imagens de marcas de gado pretendido para o município de São Francisco de Assis.

Figura 4 – Sistema especialista para auxílio no processo de reconhecimento de imagens de marcas de gado.



Fonte: Adaptado de Conci et al (2007).

1.2 OBJETIVOS

Perante o contexto apresentado, este trabalho tem como objetivo geral apresentar e avaliar uma arquitetura que realize o reconhecimento automático de marcas de gado, auxiliando os órgãos governamentais responsáveis pelo registro dessas marcas, reduzindo o risco de falha humana e o tempo de espera para registro de novas marcas, além de contribuir com as autoridades de segurança no combate a crimes de abigeato. Entre os objetivos específicos, podem-se citar:

- Utilizar um conjunto de imagens de marcas de gado fornecido pela Prefeitura de São Francisco de Assis-RS para realizar os experimentos;
- Implementar uma arquitetura capaz de realizar o reconhecimento automático das marcas de gado;
- Apresentar um método que alcance bom desempenho computacional mesmo quando aplicado a grandes volumes de dados;
- Realizar o treinamento e validação da arquitetura proposta;
- Avaliar a eficácia e eficiência da arquitetura através de métricas que permitam verificar a performance do classificador desenvolvido.

1.3 ORGANIZAÇÃO DA DISSERTAÇÃO

Esta dissertação está estruturada como segue. O Capítulo 2 apresenta a fundamentação teórica, descrevendo as técnicas de processamento de imagens para extração de características, classificação e reconhecimento das imagens de marcas de gado. O Capítulo 3 apresenta os trabalhos relacionados ao tema. O Capítulo 4, por sua vez, descreve a implementação da arquitetura desenvolvida, apresentando os materiais utilizados e o detalhamento dos métodos propostos. O Capítulo 5, apresenta a discussão dos resultados de forma quantitativa, demonstrando a tabulação dos dados comparativos dos experimentos realizados. No Capítulo 6, por fim, são realizadas as considerações finais do trabalho, dificuldades encontradas e trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta os conceitos básicos de um sistema de processamento de imagens, que pode ser dividido nas etapas de aquisição, pré-processamento, segmentação, representação e descrição, e reconhecimento e interpretação. O reconhecimento de padrões de imagens digitais pode envolver diferentes técnicas e métodos computacionais que têm como objetivo extrair características numéricas ou simbólicas e construir um esquema de classificação para classificar ou descrever uma informação de acordo com as características extraídas. Logo, este capítulo apresenta métodos de visão computacional para a representação e descrição, assim como para reconhecimento e interpretação, que foram empregados neste trabalho.

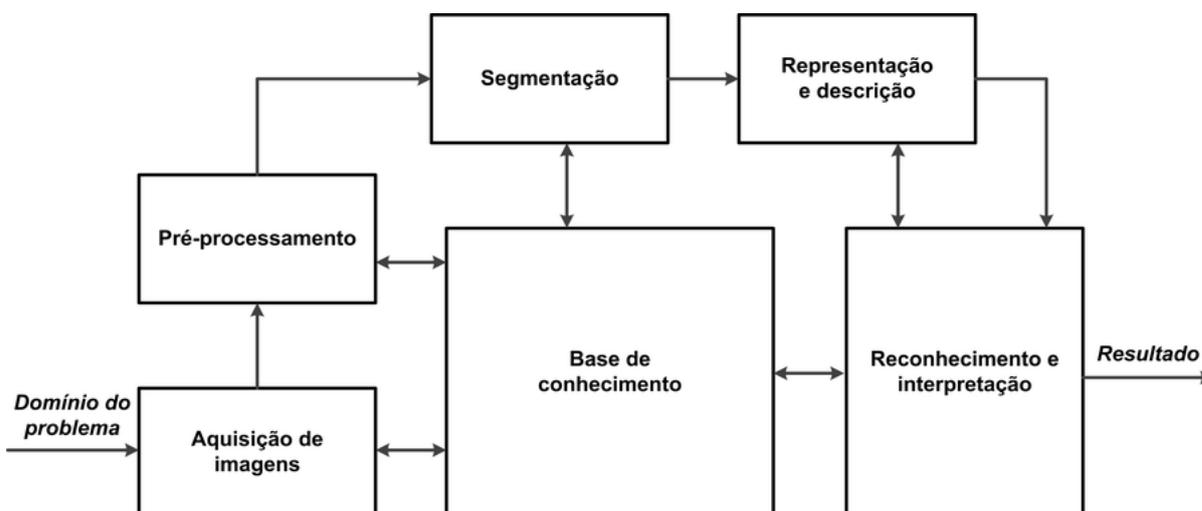
Na seção 2.1 são apresentadas as etapas de um sistema de processamento de imagens; Na seção 2.2 é apresentado o método *Bag-of-Features*; Na seção 2.3 são mostrados conceitos básicos de Redes Neurais Artificiais; Na seção 2.4 é indicado o método de Aprendizagem Profunda, que consiste no aprendizado de características hierárquicas através de características mais simples; Na seção 2.5 é apresentado o modelo de aprendizagem estatística de Máquinas de Vetores de Suporte; Na seção 2.6, por fim, são indicadas métricas encontradas na literatura para avaliação de classificadores.

2.1 ETAPAS DE UM SISTEMA DE PROCESSAMENTO DE IMAGENS

Pode-se definir o processamento de imagens digitais como um conjunto de técnicas de captura, representação e transformação computacional de imagens e que tem por objetivos a extração e identificação de informações contidas na imagem, como também o realce de atributos contidos na mesma. Essa análise é possível a partir da utilização de recursos que vão desde a aquisição da imagem até o reconhecimento e interpretação de seu conteúdo, de acordo com a necessidade imposta pela área de estudo que se utiliza dessa ferramenta (PEDRINI; SCHWARTZ, 2008).

Os recursos utilizados em uma rotina de processamento de uma dada imagem constituem um sistema típico de processamento de imagens, representado pela Figura 5, e estão divididas nas etapas de aquisição, pré-processamento, segmentação, representação e descrição, reconhecimento e interpretação, e resultado.

Figura 5 – Etapas de um sistema de processamento de imagens.



Fonte: Adaptado de Pedrini e Schwartz (2008).

A etapa de aquisição consiste da obtenção e digitalização da imagem que se deseja analisar (GONZALEZ; WOODS, 2010). O passo seguinte trata do pré-processamento da imagem obtida. Essa etapa consiste na compensação de distorções específicas que são adquiridas na etapa de aquisição de maneira que seja possível aplicar as etapas seguintes da sequência de processamento (PEDRINI; SCHWARTZ, 2008). A segmentação consiste da partição da imagem em regiões disjuntas, que serão analisadas de acordo com o objetivo proposto. A etapa de representação é aquela que define de que maneira os dados devem ser convertidos para uma forma adequada ao processamento computacional. Este processo envolve, basicamente, duas escolhas: representar a região em função de suas características externas (fronteiras) ou representar a região em termos de suas características internas (os *pixels* que compõem a região). O processo de descrição tem como objetivo extrair as características que resultem em informação quantitativa de interesse ou que sejam importantes para a diferenciação entre classes de objetos.

A última etapa do processamento de imagens digitais trata do reconhecimento e também da interpretação. O reconhecimento é definido como a etapa em que ocorre a atribuição de um rótulo (ou *label*) a um objeto em função da informação que é fornecida por seu descritor, que nada mais é do que a extração de atributos que serão responsáveis pela identificação desse objeto (por exemplo, “marca de gado”). Nesse caso, é feito o reconhecimento de padrões na imagem para que as informações contidas na imagem possam ser corretamente atribuídas. Conforme (GONZALEZ; WOODS, 2010), um padrão é um arranjo de descritores, que frequentemente é denotado com o termo “característica”. Uma classe de padrões é uma família de padrões que compartilham propriedades em comum.

O reconhecimento de padrões classifica dados visuais numéricos ou simbólicos

baseados em informações contidas em bancos de dados de padrões. Muitas técnicas estatísticas e sintáticas têm sido desenvolvidas para classificação de padrões (CONCI; AZEVEDO; LETA, 2007).

A interpretação é vista como a atribuição de significado a um conjunto de objetos reconhecidos. O algoritmo necessário para a realização dessa etapa envolve a aplicação de sistemas lógicos e redes semânticas para a interpretação das informações obtidas das imagens (GONZALEZ; WOODS, 2010).

2.2 BAG-OF-FEATURES

O método *Bag-of-Features* (BoF), também descrito na literatura como *Bag of Words* (BoW), ou *Bag of Visual Words* (BoVW), foi originalmente proposto por (SIVIC; ZISSERMAN, 2003). O método baseia-se em quantização das características locais obtidas por meio de algoritmos auxiliares como os de detecção de pontos de interesse de imagens, com o objetivo de construir palavras visuais, em conformidade com um dicionário visual pré-construído, gerando um vetor composto por essas palavras que representam uma imagem.

De acordo com (O'HARA; DRAPER, 2011), um vocabulário visual é construído para representar o dicionário, agrupando recursos extraídos de um conjunto de imagens de treinamento. Os recursos de imagem representam áreas locais da imagem, assim como as palavras são características locais de um documento ou imagem.

Segundo CAICEDO, CRUZ e GONZALEZ (2009), o *Bag-of-Features* é uma adaptação do *Bag of Words* utilizado para categorização e recuperação de texto. A ideia fundamental é a construção de uma tabela de codificação, ou seja, um vocabulário visual, em que os padrões mais representativos são codificados como palavras visuais. Em seguida, a representação da imagem é gerada por meio de uma análise de frequência simples de cada palavra-chave dentro de uma imagem.

Este método pode ser classificado como um Sistema de Recuperação de Imagens por Conteúdo (*Content-Based Image Retrieval* – CBIR). Para exemplificar o funcionamento do método, TORRES e FALCÃO (2006) mostram algumas abordagens comuns em sistemas CBIR onde um vetor é extraído das imagens de acordo com características como forma, textura, distribuição de cores. Em uma nova consulta, um vetor com as mesmas características é extraído da imagem consultada e comparado com os demais vetores de características existentes na base através de uma função de distância.

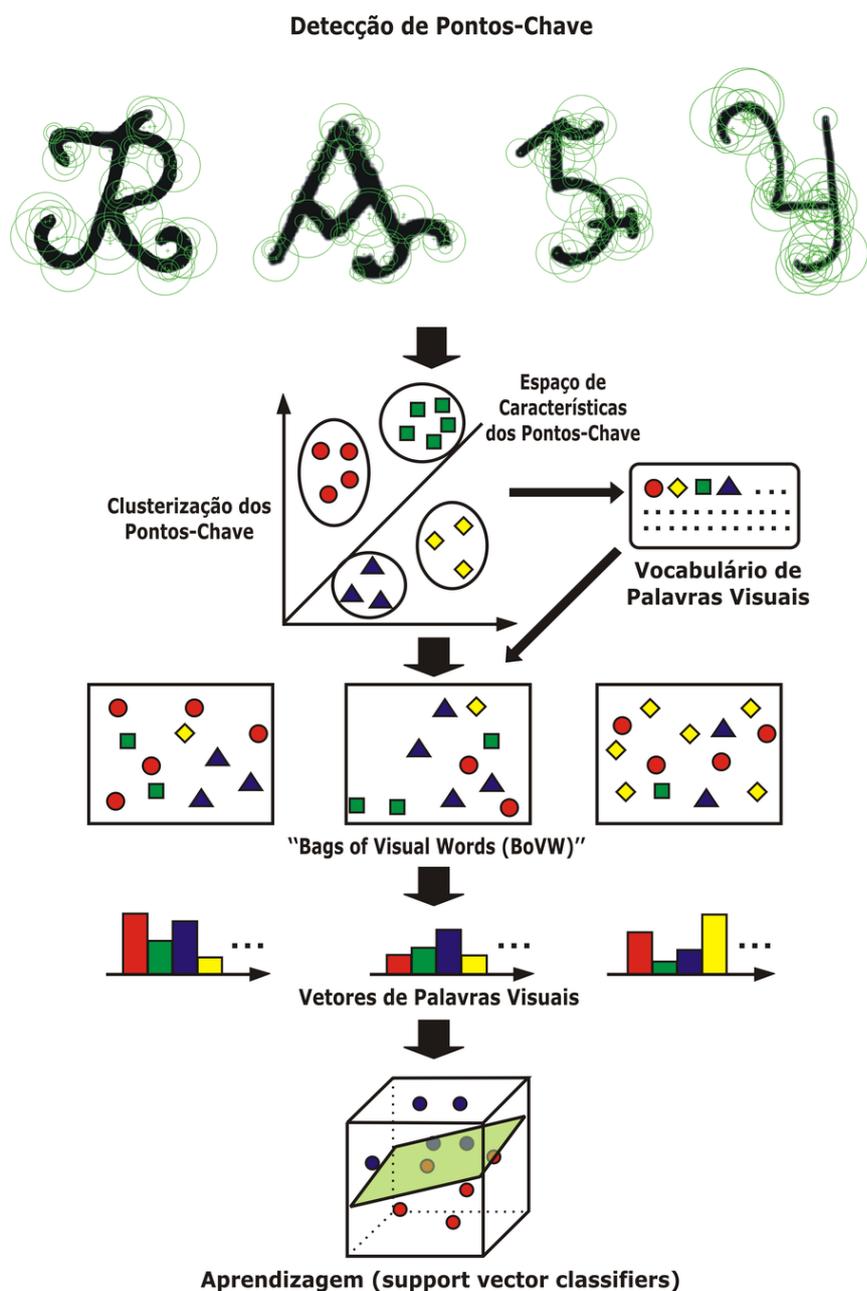
CSURKA et al. (2004) descrevem quatro etapas para classificar imagens utilizando o método *Bag-of-Features*:

- 1) Detecção e descrição de pontos de interesse (*feature detection and feature description*);

- 2) Construção do dicionário visual (*codebook construction*);
- 3) Construção do vocabulário ou dicionário visual (*image representation*);
- 4) Formação dos algoritmos de aprendizagem (*support vector classifiers*).

Uma ilustração das etapas que compõem o método *Bag-of-Features* é apresentada na Figura 6.

Figura 6 – Etapas que compõem o método *Bag-of-Features*.



Fonte: do próprio autor.

Na Figura 6, primeiramente um conjunto de marcas de gado é selecionado. Em seguida são selecionados pontos-chaves dessas imagens, para na sequência ser

realizada a descrição dos pontos de interesse de cada marca. O próximo passo é a realização da aglomeração (*clustering*) desses vetores com as descrições das imagens, com o intuito de criar um vocabulário de palavras visuais que representem as marcas. Após a criação desse histograma de palavras visuais é possível aplicar um algoritmo de aprendizagem supervisionada, como por exemplo uma máquina de vetor de suporte para realizar o reconhecimento de novas marcas de gado.

O *Bag-of-Features* consiste em uma estrutura adaptável e flexível, em que cada etapa permite ser determinada por meio de diferentes técnicas conforme as necessidades do domínio que será aplicado. A seguir serão detalhadas as etapas do processo.

2.2.1 DETECÇÃO E DESCRIÇÃO DE PONTOS DE INTERESSE

A detecção é a etapa na qual são identificados os pontos de interesse que se destacam de uma imagem e que, posteriormente, podem ser trabalhados em etapas de reconhecimento.

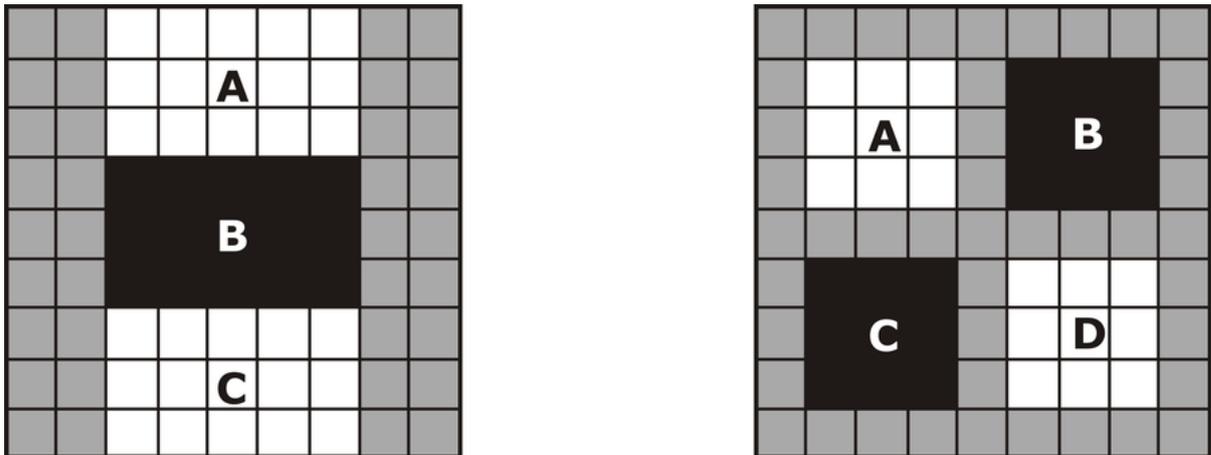
Geralmente, o objetivo é identificar uma região limitada espacialmente da imagem que é saliente ou proeminente. Esses pontos são regiões da imagem que apresentam uma alteração intensa dos *pixels* da vizinhança em diversas direções.

São apresentadas diversas técnicas na literatura para detectar as características locais, que são implementadas por distintas propriedades visuais, tais como bordas, saliências ou, até mesmo, cantos. Assim que são detectadas as características locais, o passo subsequente é descrever o conteúdo das referidas regiões locais.

O algoritmo *Speeded Up Robust Features* (SURF), proposto por Herbert Bay em 2006 é um exemplo de método capaz de realizar a detecção e descrição de pontos de interesse de uma imagem eficientemente. O SURF usa um modelo aproximativo do espaço de escala, baseado em imagens integrais.

Além disso, o detector de pontos usado pelo SURF não necessita que o tamanho original da imagem seja alterado. A detecção dos pontos de interesse é baseada na supressão de não máximos do determinante da matriz Hessiana (matriz cujas entradas são as segundas derivadas parciais de uma função dada). Em consequência disto, o SURF tende a detectar cantos e regiões com textura. O algoritmo SURF utiliza *boxfilter*, como ilustrado na Figura 7, que são os núcleos de convolução, também denominados filtros caixa. Portanto, cada filtro caixa corresponde a uma aproximação das derivadas parciais da função Gaussiana, que por sua vez, compõem as posições da matriz Hessiana.

Figura 7 – Núcleos de Convolução do SURF (boxfiler ou filtro caixa).



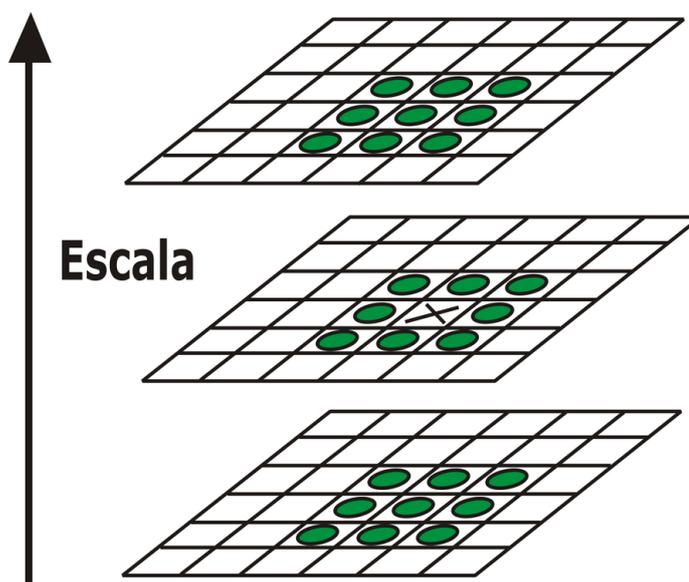
Fonte: Bay et al (2006).

Contudo, existe um problema, se para cada ponto precisar percorrer toda a região para obter a soma de seus *pixels*. Este problema é resolvido utilizando a integral da imagem, que tem a propriedade de obter a soma de todos os *pixels* de uma região retangular de qualquer tamanho da imagem de entrada com uma fórmula, como mostrado na Equação 2.1, onde A, B, C e D são os valores dos pontos na imagem integrada.

$$\Sigma = A - B - C + D \quad (2.1)$$

Na etapa de detecção dos pontos terá como entrada todas as escalas das oitavas e buscará pelos máximos locais. Essa busca utiliza a mesma técnica do algoritmo SIFT (*Scale Invariant Feature Transform*), proposto por LOWE (2004), ou seja, o ponto será comparado com os seus oito vizinhos da mesma escala e também com seus 9 vizinhos da escala anterior e posterior, como mostrado na Figura 8. Se o valor desse ponto for maior que todos, então é colocado como ponto de interesse.

Figura 8 – A figura mostra o ponto central marcado por um X, o qual é comparado com seus 8 vizinhos da mesma escala e seus 9 vizinhos da escala superior e inferior.



Fonte: Adaptado de Lowe (2004).

2.2.2 CONSTRUÇÃO DO VOCABULÁRIO OU DICIONÁRIO VISUAL

O vocabulário ou dicionário visual é construído usando um algoritmo de quantização de agrupamento ou vetor (CAICEDO; CRUZ; GONZALEZ, 2009). O vocabulário é uma maneira de construir um vetor de característica de classificação que se relaciona com novos descritores da imagem consultada para descritores previamente vistos em treinamento (CSURKA et al., 2004). Este vocabulário ou histograma de palavras visuais compõe as características que serão utilizadas no processo de recuperação da imagem.

O algoritmo de análise de agrupamento *K-means* ou algoritmo das *k*-médias, proposto por MACQUEEN, J. (1967), é um dos algoritmos mais conhecidos e utilizados na literatura e tem por objetivo localizar um conjunto de centroides dentro do conjunto de dados das características locais.

Um ponto muito importante que deve ser considerado para a construção do vocabulário é a definição do seu tamanho, ou seja, quantos vocábulos são necessários para representar o conteúdo da imagem. CSURKA et al. (2004) relata que o desempenho (acurácia) melhora progressivamente à medida que cresce o vocabulário. Um pequeno vocabulário pode não ter um bom poder discriminatório, uma vez que duas imagens podem ser atribuídos a um mesmo vocábulo, mesmo que essas não sejam iguais. Por outro lado, um grande vocabulário é menos generalizável, menos sensível a ruídos.

Contudo, um tamanho elevado de dicionário pode acarretar uma sobrecarga de

processamento, elevando conseqüentemente o custo computacional (JIANG; YANG; NGO, 2009). O que se almeja é um tamanho de vocabulário balanceado, ou seja, que proporcione resultados desejáveis a um custo computacional aceitável.

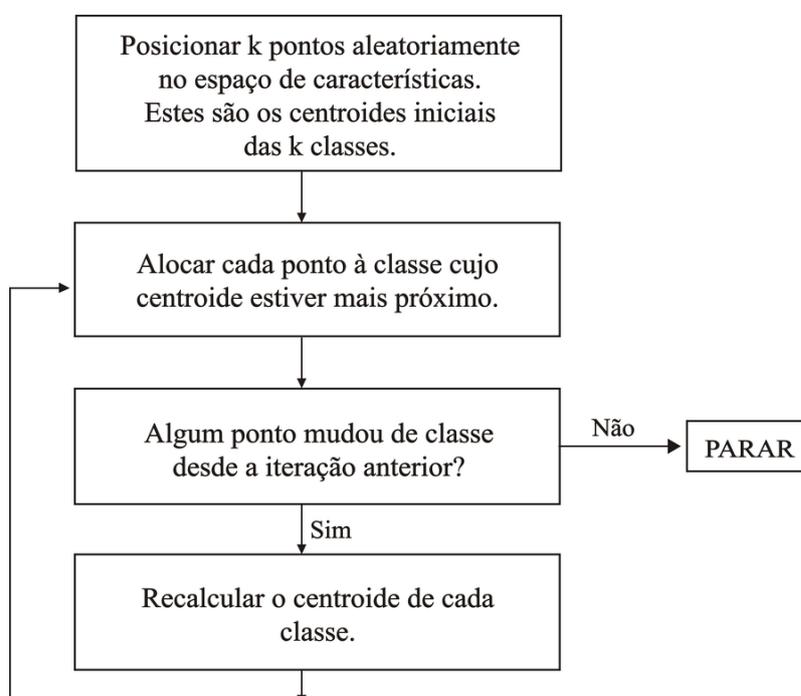
2.2.2.1 ALGORITMO K-MEANS

Segundo SOLOMON e BRECKON (2013), o algoritmo *K-means* pertence a categoria dos classificadores não supervisionados, e por conseguinte não funciona com exemplos de classes rotuladas para treinamento. O algoritmo explora os dados e busca padrões ou aglomerações (*clusters*) que ocorrem naturalmente no conjunto de dados analisado.

Uma vez que aglomerações tenham sido encontradas, fronteiras de decisão podem ser construídas para classificar dados não vistos usando métodos muito similares aos empregados em classificadores supervisionados. Este procedimento é resumido na Figura 9, e introduz uma forma conceitualmente simples de dividir um conjunto de dados em um dado número k de aglomerações (*clusters*).

A principal desvantagem do algoritmo *K-means* é a necessidade de fixar o número de aglomerações *a priori*, uma vez que na maioria dos casos não há como saber quantas aglomerações existem. A Figura 9 apresenta o fluxograma do algoritmo *K-means*.

Figura 9 – Fluxograma do algoritmo K-means.



Fonte: Solomon e Breckon (2013).

Não existe uma solução teórica geral para determinar o número ótimo de aglomerações para certo conjunto de dados. Uma abordagem simples consiste em comparar

os resultados de múltiplas execuções do algoritmo com diferentes valores de k e escolher o melhor segundo algum critério adequado; todavia, é necessário cuidado na escolha do tamanho do dicionário, uma vez que o aumento do *cluster* eleva o risco de sobre-ajuste (*overfitting*) dos dados, que ocorre quando o modelo estatístico se ajusta em demasiado ao conjunto de dados/amostra (SOLOMON; BRECKON, 2013).

Nesse trabalho, a busca do valor ideal de aglomerações para o algoritmo *K-means* é obtido através da comparação dos resultados de acurácia e tempo de processamento alcançados por múltiplas execuções do algoritmo com diferentes valores de k .

2.3 REDES NEURAS ARTIFICIAIS

Redes Neurais Artificiais (RNAs) podem ser utilizadas para realizar funções complexas em vários campos, incluindo reconhecimento, identificação e classificação de padrões, aplicações envolvendo voz, visão e em sistemas de controle (HAGAN et al., 1995). Problemas difíceis para computadores convencionais ou para seres humanos também podem ser resolvidos através de redes neurais treinadas para este fim (GIL, 2014).

A evolução dos computadores ajudou as pesquisas sobre RNAs, possibilitando o desenvolvimento de novas arquiteturas, com maior capacidade e com algoritmos mais sofisticados. As RNAs vêm sendo cada vez mais utilizadas devido a características como (HAYKIN, 2001):

- Capacidade de trabalhar com problemas não-lineares;
- Capacidade de aprendizado, obtido através do treinamento com um conjunto de amostras contendo exemplos de entrada e saída (aprendizado supervisionado);
- Adaptabilidade;
- Resposta a evidências;
- Informação contextual;
- Generalização; e
- Tolerância a falhas.

Redes neurais tem a capacidade de receber ao mesmo tempo várias entradas e distribuí-las de maneira organizada. Geralmente, as informações armazenadas por uma rede neural são compartilhadas por todas as suas unidades de processamento.

Em um sistema de rede neural, a informação pode parecer ter representação redundante, porém, o fato de que ela se encontre distribuída por todos os elementos da

rede significa que mesmo que parte da rede seja destruída, a informação contida nesta parte ainda estará presente na rede, e poderá ser recuperada. Portanto, a redundância na representação de informações em uma rede neural, diferente de outros sistemas, transforma-se em uma vantagem, que torna o sistema tolerante a falhas.

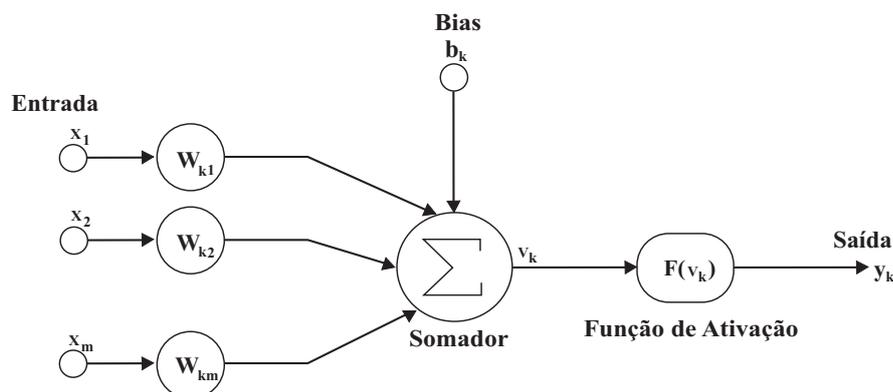
Os atributos de uma RNA, tais como aprender através de exemplos, generalizações redundantes, e tolerância a falhas, proporcionam fortes incentivos para a escolha de redes neurais como uma escolha apropriada para aproximação para a modelagem de sistemas biológicos.

Estruturas de redes neurais possuem características como adaptação e aprendizagem. Essas habilidades significam que modelos de RNAs podem lidar com dados imprecisos e situações não totalmente definidas. Uma rede treinada de maneira razoável tem a habilidade de generalizar quando é apresentada a entradas que não estão presentes em dados já conhecidos por ela.

A característica mais significante de redes neurais está no fato de aproximar qualquer função contínua não linear de um grau de correção desejado. Esta habilidade das redes neurais as tem tornado útil para modelar sistemas não lineares na combinação de controladores não lineares. RNAs podem ter várias entradas e várias saídas, portanto são aplicáveis à sistemas com muitas variáveis.

As RNAs são sistemas computacionais distribuídos compostos de unidades de processamento simples, densamente interconectadas, onde essas unidades são conhecidas como neurônios artificiais (FACELI et al., 2011). Assim como no sistema nervoso, o neurônio artificial é a unidade fundamental para o processamento de informação de uma RNA. A Figura 10 apresenta o modelo de um neurônio artificial.

Figura 10 – Modelo de um neurônio artificial.



Fonte: Oliveira (2005).

No neurônio artificial os valores são recebidos pelas entradas (x_1, x_2, \dots, x_m) e são ponderados pela ação dos respectivos pesos sinápticos $(W_{k1}, W_{k2}, \dots, W_{km})$ e combinados por uma função matemática V_k . Em seguida, a saída gerada pelo somador é aplicada a uma função de ativação $F(V_k)$, que restringe a amplitude do sinal de

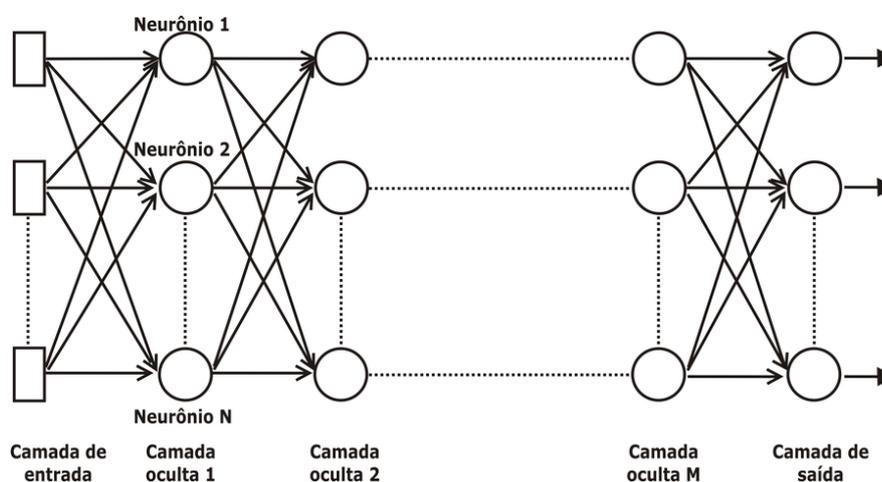
saída ao condicionador de ativação do sinal à ultrapassagem de um determinado limiar pelo valor da soma ponderada das entradas. A saída gerada pela função de ativação y_k é a resposta do neurônio para as entradas que lhes foram apresentadas. Complementarmente o bias b_k pode ser utilizado para aumentar ou diminuir a entrada da função de ativação (OLIVEIRA, 2005).

2.3.1 REDE NEURAL MULTILAYER PERCEPTRON

Definir a arquitetura de uma RNA é uma tarefa de fundamental importância, pelo fato da sua organização depender do problema a ser tratado pela rede. Outro fator que também está ligado à arquitetura da rede é o tipo de algoritmo utilizado para o treinamento da RNA. Dentre os vários tipos de arquiteturas existentes, a rede *Multilayer Perceptron* (MLP) é uma das mais utilizadas.

As redes MLPs são formadas por grupos de neurônios organizados em forma de camadas. A Figura 11 apresenta a estrutura de uma rede MLP. Na primeira camada estão as entradas, unidades sensoriais que não possuem poder de processamento, seguida pela camada oculta (que pode possuir uma ou várias camadas), formada por grupos de nós computacionais (*perceptrons*), e, por fim, uma camada de saída, também formada por nós computacionais. O sinal recebido pela entrada se propaga por todas as camadas da rede (PRADO, 2014).

Figura 11 – Arquitetura de uma RNA Multilayer Perceptron.



Fonte: Prado (2014).

A rede apresentada na Figura 11 é considerada como um modelo totalmente conectado, uma vez que todos os neurônios de uma camada estão ligados a todos os neurônios da camada seguinte. Caso esteja faltando alguma conexão na rede, dizemos que a mesma está parcialmente conectada.

Em uma rede MLP, cada neurônio realiza uma função específica. A função implementada por um neurônio de uma dada camada é uma combinação das funções

realizadas pelos neurônios da camada anterior que estão conectados a ele.

O processamento se torna mais complexo à medida em que o processamento avança nas camadas. A combinação das funções desempenhadas por cada neurônio da rede define a função associada à RNA como um todo (FACELI et al., 2011). Para que uma rede MLP seja capaz de solucionar um determinado problema, é necessário realizar o treinamento da mesma.

2.3.2 ALGORITMO BACKPROPAGATION

O *Backpropagation* é o principal algoritmo utilizado para o treinamento de redes MLP (ALENCAR, 2015). Sua técnica consiste em realizar iterações em duas fases, uma fase para frente (propagação) e outra fase para trás (retropropagação).

Na fase de propagação, cada objeto de entrada é apresentado à rede. O objeto é primeiramente recebido por cada um dos neurônios da primeira camada oculta. Os objetos são ponderados pelos pesos associados às conexões de entrada correspondentes ao neurônio e, em seguida, o valor total recebido é aplicado à função de ativação e a saída produzida pela função é utilizada como entrada para o neurônio da próxima camada.

Esse processo é repetido até os neurônios da camada de saída. Os valores gerados pelos neurônios da camada de saída são comparados com a saída desejada. O cálculo da diferença entre esses valores indica o erro cometido pela rede (FACELI et al., 2011). O valor do erro de cada neurônio da camada de saída é então utilizado na fase de retropropagação para ajustar os pesos de entrada.

2.4 APRENDIZAGEM PROFUNDA (DEEP LEARNING)

O termo "*Deep Learning*", que pode ser traduzido literalmente para "Aprendizagem Profunda", representa um conjunto de métodos para o aprendizado de características hierárquicas através da composição de características mais simples. Esse uso de múltiplos níveis de abstração é particularmente similar ao funcionamento do sistema visual dos primatas Serre et al. (2007), com uma sequência de estágios de processamento: detecção de bordas, formas primitivas até gradualmente identificando formatos visuais mais complexos. Esse processo se repete através das regiões até que os neurônios da região final têm a atribuição de detectar características de mais alto nível de abstração, tais como faces ou objetos específicos (BENGIO; COURVILLE; VINCENT, 2013).

O aprendizado de características em múltiplos níveis de abstração permite ao um sistema aprender complexas funções de mapeamento dos dados de entrada para saída, de maneira independente de características criadas manualmente. Ou seja, é uma forma de automatizar a geração de características que sejam mais representativas

de um determinado problema de reconhecimento de padrões (GIL, 2014).

Redes de aprendizagem profunda estão associadas a um método de aprendizado de máquina que utiliza modelos de redes neurais com vários níveis de representação de dados, os quais permitem aprender dados mais complexos. Os níveis hierárquicos de representação de dados são organizados por abstrações, características e conceitos.

Aprendizagem profunda é o termo usado para denotar o problema de treinar redes neurais artificiais que realizam o aprendizado de características de forma hierárquica, de tal forma que características nos níveis mais altos da hierarquia sejam formadas pela combinação de características de mais baixo nível (SCHMIDHUBER, 2015; DENG; DONG, 2014; ITAMAR; ROSE; KARNWSKI, 2010; GOODFELLOW; BENGIO; COURVILLE, 2016; LECUN; BENGIO; HINTON, 2015; BENGIO, 2009).

Quanto mais alto o nível, mais abstrato e não linear será a representação dos dados. Esses níveis hierárquicos são representados pelas camadas da RNA. Desse modo, a profundidade da rede neural diz respeito ao número de níveis de composição das operações não lineares perante os dados treinados, ou seja, quanto mais camadas, mais não linear e profunda será a RNA (OLIVEIRA; BARBAR; SOARES, 2015).

2.4.1 REDES NEURAS CONVOLUCIONAIS

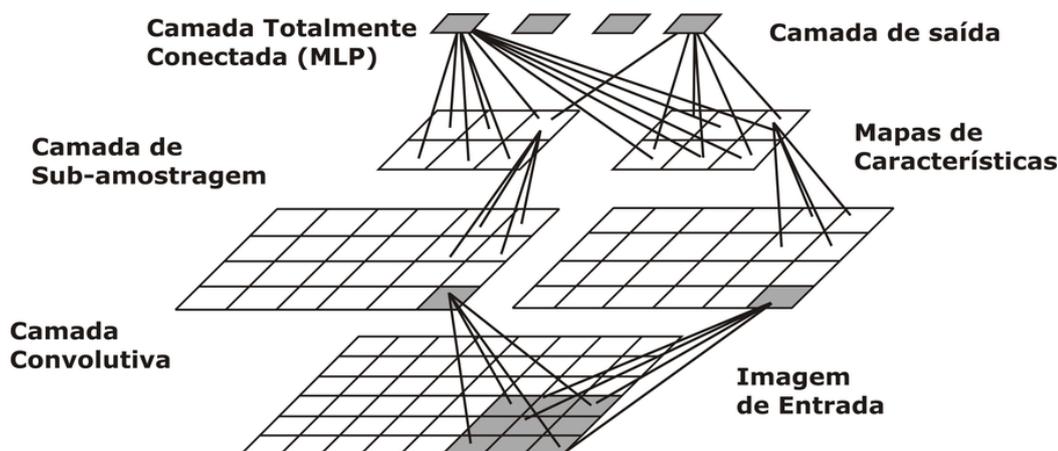
Redes Neurais Convolucionais (Convolutional Neural Network, CNN) é um método de *Deep Learning* especialmente projetado para lidar com a variabilidade em dados bidimensionais (2D), como por exemplo, imagens em sua forma matricial (LECUN et al., 1998).

Redes neurais convolucionais se inspiram no funcionamento do córtex visual (ZEILER; FERGUS, 2014). Objetivando invariância a deslocamento, escala e distorção, as CNN combinam as seguintes ideias arquiteturas:

- Mapeamento de características, ou campos receptivos locais, como um dos pontos de maior similaridade com as redes neurais biológicas que garante uma maior robustez a distorções locais.
- Pesos compartilhados que permitem uma redução dos parâmetros livres e invariância geométrica.
- Sub-amostragem temporal ou espacial, o que reduz o tamanho total dos mapas de características a cada camada, chegando assim na última camada apenas como valores unidimensionais, o que nesse ponto em diante as torna equivalentes a uma rede neural MLP.

A Figura 12 ilustra os mapas de características de uma arquitetura convolucional (LI, 1999).

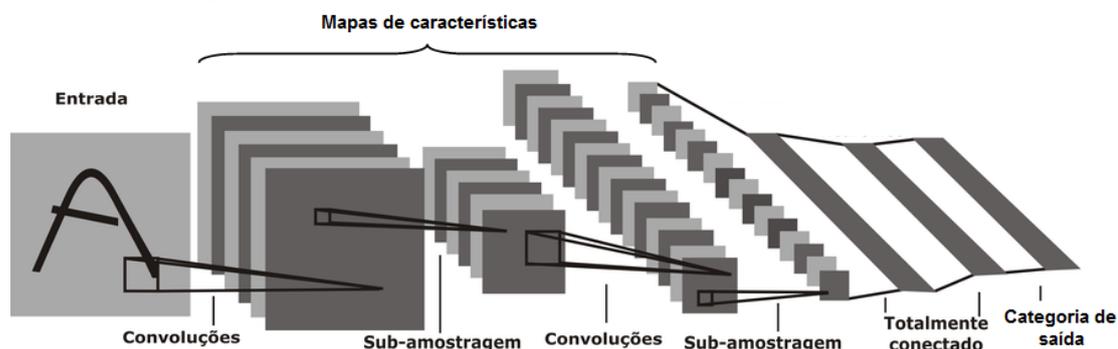
Figura 12 – Mapas de características de uma arquitetura convolucional.



Fonte: Li (1999).

A Figura 13 apresenta uma visão geral da arquitetura de uma rede neural convolucional (LECUN et al., 1998)

Figura 13 – Visão geral da arquitetura de uma rede neural convolucional.

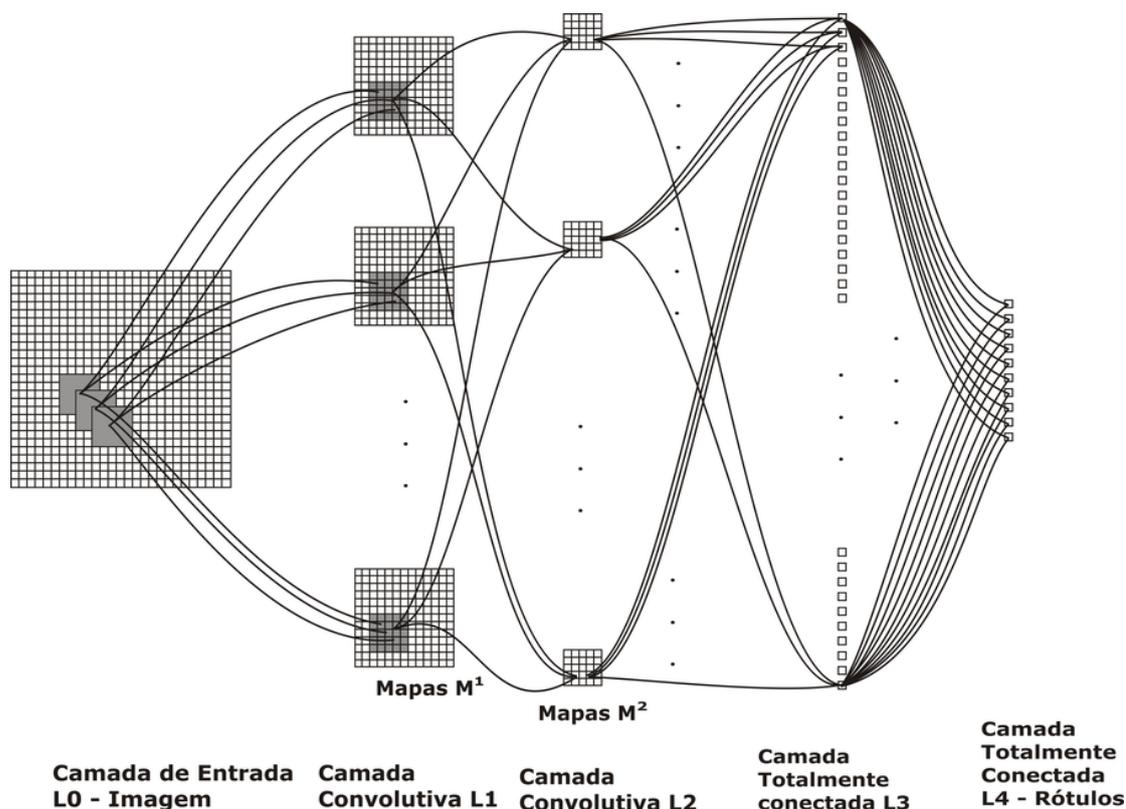


Fonte: Adaptado de LeCun et al (1998).

A partir de uma imagem como elemento de entrada, cada uma das camadas seguintes mapeia uma determinada região da matriz de entrada em uma região específica da matriz de saída. Por este comportamento, a matriz resultante de cada camada é chamada de mapa de características, conforme ilustrado na Figura 12.

Essa implementação sequencial do mapa de características contempla uma passagem total da imagem de entrada por meio dos correspondentes locais no mapa de características. Esta operação é equivalente a uma convolução, sendo esse portanto o motivo do nome das redes convolucionais. O *Kernel* da convolução, dessa forma, seria os pesos da conexão usados pelas unidades do mapa de características. A Figura 14 ilustra uma arquitetura de rede neural convolucional com matriz de convolução (*kernel*), de tamanho 5x5 (CIREGAN; MEIER; SCHMIDHUBER, 2012).

Figura 14 – Arquitetura de rede neural convolutiva com kernels de tamanho 5x5.



Fonte: Ciregan et al (2012).

Através do algoritmo *backpropagation* é processado o gradiente da função de perda com respeito a todos os pesos em todas as camadas da rede convolucional. Porém, o algoritmo *backpropagation* padrão usado em redes MLP deve ser modificado para lidar com o compartilhamento de pesos.

Uma maneira rápida de implementar esse ajuste é primeiro calcular as derivadas parciais da função de perda com respeito a todas as conexões, como se a rede fosse uma rede neural multicamada convencional. Então, as derivadas parciais de todas as conexões que compartilham o mesmo parâmetro são adicionadas para formar o derivativo com respeito a esse parâmetro (GIL, 2014).

Redes neurais convolucionais tradicionalmente são completamente conectadas, ou seja, cada unidade na camada está conectada a todas as unidades das camadas anteriores. Em vez da conectividade global, uma rede convolucional utiliza conectividade local (BEZERRA, 2016).

Em uma CNN as unidades de uma determinada camada são organizadas em conjuntos disjuntos, cada um dos quais é denominado mapa de característica (*feature map*), também conhecido como filtro.

As unidades contidas em um mapa de características são únicas na medida em que cada um delas está ligada a um conjunto de unidades diferente na camada anterior. Além disso, todas as unidades de um mapa compartilham os mesmos parâmetros. O

resultado disso é que essas unidades dentro de um mapa servem como detectores de uma mesma característica, mas cada uma delas está conectada a uma região diferente da imagem.

Portanto, em uma CNN, uma camada oculta é segmentada em diversos mapas de características, em que cada unidade de um mapa tem o objetivo realizar a mesma operação sobre a imagem de entrada, com cada unidade aplicando essa operação em uma região específica dessa imagem.

A seguir, apresenta-se a arquitetura típica de uma Rede Neural Convolutacional.

2.4.1.1 ARQUITETURA

Em geral, uma CNN possui diversos tipos de camadas: camadas de convolução, camadas de subamostragem, camadas de normalização de contraste e camadas completamente conectadas. Na forma mais comum de arquitetar uma CNN, a rede é organizada em estágios. Cada estágio é composto por uma ou mais camadas de convolução em sequência, seguidas por uma camada de subamostragem, que por sua vez é seguida (opcionalmente) por uma camada de normalização.

Uma CNN pode conter vários estágios empilhados após a camada de entrada (que corresponde à imagem). Após o estágio final da rede, são adicionadas uma ou mais camadas completamente conectadas (BEZERRA, 2016).

A camada de convolução é composta por diversos neurônios, cada um responsável por aplicar um filtro em um segmento específico da imagem. A operação de convolução consiste em deslizar um filtro bidimensional, conhecido como *kernel*, sobre as camadas da imagem de entrada e calcular o produto escalar entre as entradas do filtro e as da imagem.

A camada de subamostragem (*subsampling layer* ou *pooling layer*) é composta por um conjunto de mapas de características e tem por objetivo realizar a agregação das saídas (ativações) de um conjunto de unidades da camada anterior. As camadas de subamostragem resultam em uma rede mais robusta a transformações espaciais, ou seja, mais invariantes a eventuais translações dos objetos em uma imagem.

Redes neurais convolucionais também podem conter as denominadas camadas de normalização de contraste local (Local Contrast Normalization, LCN) (JARRETT; KAVUKCUOGLU; LECUN, 2009). Quando utilizada, essa camada é posicionada na saída da camada de subamostragem. Uma camada LCN normaliza o contraste de uma imagem de forma não linear. Em vez de realizar uma normalização da imagem como um todo, a camada LCN aplica a normalização sobre regiões locais da imagem, considerando cada *pixel* por vez.

A normalização pode corresponder a subtrair a média da vizinhança de um *pixel* particular e dividir pela variância dos valores de *pixel* dessa vizinhança. Esta transformação equipa a CNN com a invariância de brilho, propriedade útil no contexto

de reconhecimento de imagens (BEZERRA, 2016).

Outro tipo de camada em uma CNN é a completamente conectada. Em arquiteturas de CNNs mais atuais, é comum encontrar uma ou duas camadas desse tipo antes da camada de saída. Juntamente com as camadas de convolução e subamostragem, as camadas totalmente conectadas geram descritores de características da imagem que podem ser mais facilmente classificados pela camada de saída, que normalmente usa a função *Softmax*.

A função *Softmax* é descrita na Equação 2.2 que garante que a soma de todas as probabilidades de saída da MLP resulta em 1, onde Y corresponde ao vetor de saída da MLP. A resposta final de qual categoria pode ser obtida através da seleção do item com maior probabilidade descrita na Equação 2.4 (JURASZEK, 2014).

$$\text{softmax}(y_i) = \frac{e^{z_i}}{\sum_{j \in Y} e^{z_j}} \quad (2.2)$$

$$\text{logsoftmax}(y_i) = \log(\text{softmax}(y_i)) \quad (2.3)$$

$$y^n = \text{argmax} P(Y = i | x^n, W, b) \quad (2.4)$$

Os pesos dos mapas de características em cada camada de convolução são aprendidos durante o treinamento. Por conta de as operações de convolução e subamostragem serem diferenciáveis, uma CNN pode ser também treinada por meio do algoritmo *Backpropagation*, com a diferença de que os pesos são atualizados considerando a média dos gradientes dos pesos compartilhados. Durante esse treinamento, é comum aplicar a técnica de desligamento (*Dropout*) às camadas totalmente conectadas para evitar o sobre-ajuste (BEZERRA, 2016).

2.4.1.2 PRÉ-TREINAMENTO NÃO SUPERVISIONADO

Apesar das vantagens computacionais e de ser um modelo mais aproximado do cérebro, o treinamento de redes neurais profundas era uma tarefa de difícil execução até uma década atrás. Com o intuito de se minimizar esse complicador e propor um modelo de aprendizado mais adequado para redes neurais profundas, diminuindo o tempo de treinamento pode ser utilizado o método de pré-treinamento não supervisionado (*unsupervised pre-training*) (HINTON, 2007).

Esse procedimento permite posicionar os pesos da rede em uma região do espaço de parâmetros tal que seja mais fácil para métodos de otimização baseados em gradientes encontrarem um valor ótimo para o objetivo durante o treinamento supervisionado (ERHAN et al., 2010).

Durante a aplicação desse procedimento, camadas intermediárias são adicionadas à rede iterativamente. Em cada iteração, uma nova camada intermediária é

adicionada à rede. A nova camada intermediária adicionada deve capturar regularidades nos padrões de ativação das unidades contidas na camada anterior. Esse procedimento pode ser aplicado para a construção de redes com um número arbitrário de camadas intermediárias.

Após a fase de pré-treinamento não supervisionado (que substitui a iniciação aleatória dos pesos), a camada de saída (a que produz uma distribuição de probabilidades sobre os rótulos da tarefa de classificação em questão), é então adicionada à rede. Os pesos que conectam essa camada com a camada anterior podem ser iniciados de forma aleatória).

Finalmente, a rede completa é treinada da forma supervisionada usual (com sequências de propagações de sinais adiante intercaladas por retropropagações do erro). Essa segunda fase é denominada sintonia fina (*fine-tuning*).

Uma vantagem do pré-treinamento é que é possível tirar proveito da existência de dados não rotulados, que em geral são mais comuns do que dados rotulados (BEZERRA, 2016). Outra vantagem é que normalmente o aprendizado com pré-treinamento converge mais rapidamente nos trabalhos que propuseram originalmente o uso de pré-treinamento (HINTON, 2007).

O trabalho desenvolvido utilizou uma rede neural pré-treinada com o objetivo de reduzir o tempo para treinamento da arquitetura de CNN proposta e otimizar a utilização dos recursos computacionais disponíveis, além de convergir mais rapidamente na classificação e reconhecimento das marcas de gado.

2.5 MÁQUINAS DE VETORES DE SUPORTE

Máquinas de Vetores de Suporte (*Support Vector Machines*, SVM), são modelos de aprendizagem estatística que estão na classe dos classificadores supervisionados, ou seja, aqueles que utilizam um conjunto de dados que já são mapeados para as categorias desejadas. Foram originalmente introduzidos por Vapnik na década de 90 (VAPNIK, 1995).

O objetivo do classificador, considerando o problema mais básico de apenas duas classes, é a construção de um hiperplano de margem máxima que separa objetos pertencentes a classes diferentes. Resumidamente as SVMs consistem num método para treino de amostras que se baseia num princípio de minimização do risco estrutural, que minimiza, assim, o erro de generalização (OSUNA; FREUD; GIROSI, 1997).

O classificador SVM é largamente usado devido a sua alta acurácia, habilidade de lidar com dados de alta dimensionalidade e flexibilidade em modelar diversas fontes de dados (GIL, 2014).

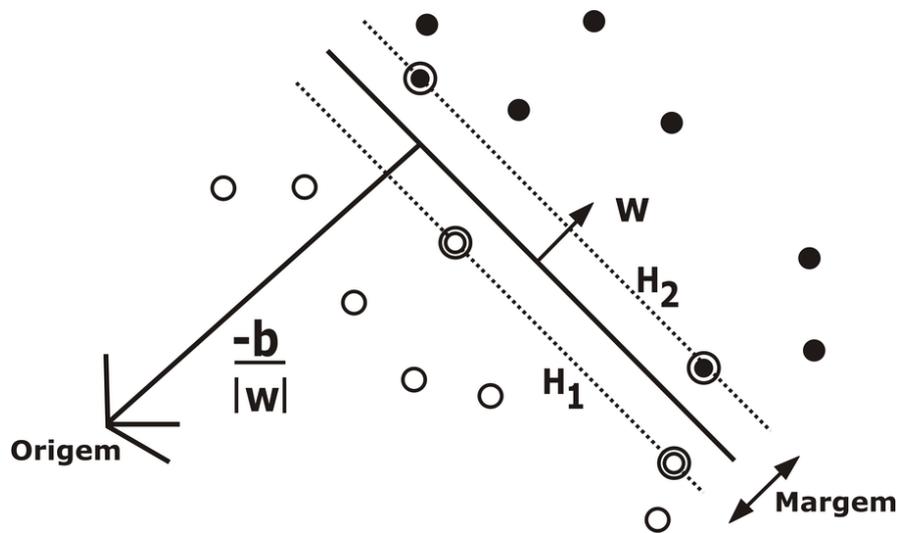
Definido por HARRINGTON (2012) como um método de destaque em aprendizagem de máquina (*Machine Learning*), SVM é basicamente voltado para classificação binária, ou seja, problemas de reconhecimento entre duas classes, apesar de existirem

versões modificadas para lidar com problemas multiclasse.

Essa teoria estabelece uma série de princípios que devem ser seguidos na obtenção de classificadores com boa generalização, definida como a sua capacidade de prever corretamente a classe de novos dados do mesmo domínio em que o aprendizado ocorreu (LORENA; CARVALHO, 2007).

Segundo HARRINGTON (2012), para um dado hiperplano, pode-se denotar por x_+ (x_-) o ponto mais próximo do hiperplano entre as amostras positivas e negativas. A norma do vetor w denotado por $\|w\|$ é o seu comprimento, e é dado por $w^T w$.

Figura 15 – Hiperplano de decisão.



Fonte: Burges (1998)

Dado um conjunto de amostras de treinamento em pares com suas saídas desejadas $(x_i, y_i), i = 1 \dots m$, onde $x_i \in R^n$ e $y_i \in \{1, -1\}^m$, o SVM HARRINGTON (2012) então seleciona o hiperplano de decisão ótimo que maximiza a margens cujo resultado resolve a Equação 2.5:

$$\min_{w,b} (1/2w^T w) \text{ Sujeito a } y_i(w^T x_i + b) \geq 1 \quad (2.5)$$

Onde w é o vetor de pesos e b a polarização. Quando os pontos de treinamento são não-linearmente separáveis, a função de custo é reformulada ao introduzir uma variável de folga $\xi_i \geq 0, i = 1, 2, \dots, m$

$$\min_{w,b} (1/2w^T w + C \sum_{i=1}^m \xi_i) \text{ Sujeito a } y_i(w^T x_i + b) \geq 1 - \xi_i \xi_i \geq 0 \quad (2.6)$$

Onde $C > 0$ é um parâmetro de penalidade do termo de erro. Embora, quando a função decisão não é linear, o esquema acima não pode ser usado diretamente e o

SVM requer que a solução para o problema de otimização seja definida da seguinte maneira:

$$\min_{w,b} (1/2w^T w + C \sum_{i=1}^m \xi_i) \text{ Sujeito a } y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i, \xi_i \geq 0 \quad (2.7)$$

2.6 AVALIAÇÃO DO CLASSIFICADOR

As métricas de avaliação são utilizadas para medir o desempenho de um classificador utilizado nos experimentos realizados, sendo comumente empregadas em problemas de aprendizado de máquina e recuperação da informação (RIBEIRO-NETO; BAEZA-YATES, 1999). A literatura apresenta diferentes métricas de avaliação para classificadores, sendo as mais comumente usadas as métricas de acurácia, sensibilidade, especificidade e precisão.

2.6.1 MATRIZ DE CONFUSÃO

A matriz de confusão contém informações relativas a classificações efetuadas através da aplicação de um classificador. O desempenho dos classificadores é frequentemente avaliado através dos dados retirados desta matriz (KOHAVI; PROVOST, 1998). Segundo RASCHKA (2014), a matriz de confusão, também conhecida como matriz de erro, é uma maneira de resumir o desempenho de um classificador para tarefas de classificação binária.

De acordo com COLGATON (1991), matriz de confusão é uma das técnicas mais utilizadas para avaliação da acurácia da classificação de dados. Além disso, contém informações sobre a classificação realizada por um determinado sistema, onde seus dados demonstram o desempenho de tais sistemas.

A matriz de confusão é formada por linhas que representam o nome das classes e colunas que apresentam a quantidade de imagens classificadas corretamente, formando um quadrado, que apresenta os números resultantes do desempenho do classificador testado, atribuídos a uma determinada classe ou categoria.

Através da matriz de confusão obtida é possível identificar o número total de imagens corretamente classificadas, valor este obtido pela soma dos valores da diagonal principal da matriz. Os restantes elementos representam erros de classificação, ou seja, imagens incorretamente classificadas (TEIXEIRA, 2011).

2.6.2 MÉTRICAS DE AVALIAÇÃO

As métricas de avaliação de classificadores geralmente tomam por base os resultados gerados a partir de uma matriz de confusão. Alguns exemplos de métricas encontradas na literatura são apresentadas na Tabela 1 (MATOS, 2009).

Tabela 1 – Métricas de avaliação para um classificador

Métrica	Equação	Comentários
VP (Verdadeiro Positivo)	-	Número de exemplos positivos classificados corretamente.
FP (Falso Positivo)	-	Número de exemplos negativos classificados incorretamente (como sendo positivos).
FN (Falso Negativo)	-	Número de exemplos positivos classificados incorretamente (como sendo negativos).
VN (Verdadeiro Negativo)	-	Número de exemplos negativos classificados corretamente.
Sensitividade (Recall)	$TPR = \frac{VP}{VP+FN}$	Taxa com que classifica como positivos todos os exemplos que são positivos. Nenhum exemplo positivo é deixado de fora. Apresenta uma indicação do quanto do total de informação relevante foi recuperada.
Especificidade	$TNR = \frac{VN}{VN+FP}$	Taxa com que uma instância verdadeiramente negativa é classificada como negativa.
Precisão	$PPV = \frac{VP}{VP+FP}$	Taxa com que todos os exemplos classificados como positivos são realmente positivos. Nenhum exemplo negativo é incluído.
Acurácia	$Acc = \frac{VP+VN}{VP+VN+FP+FN}$	Proporção de classificações corretas para o total de elementos classificados.
Erro	$Err = 1 - Acc$	Proporção de classificações incorretas para o total de elementos classificados.
Coeficiente Kappa	$k = \frac{P_O - P_E}{1 - P_E}$	Medida de concordância estatística. Adequada para tarefas de classificação realizada por vários anotadores/juízes.

Fonte: do próprio autor.

Precisão e Sensitividade são medidas amplamente utilizadas para avaliar a qualidade dos resultados em diversas áreas do conhecimento. Precisão é uma medida de fidelidade, enquanto a sensibilidade (conhecida também como revocação, cobertura, sensibilidade ou *recall*) é uma medida de completude (MATOS, 2009).

A acurácia é uma medida frequentemente utilizada para avaliação de problemas de classificação de aprendizado de máquina. Quanto maior o valor resultante para a acurácia, mais eficiente será o desempenho do classificador (REIS, 2015).

O coeficiente *Kappa* é uma medida de concordância estatística formulada por Jacob Cohen (COHEN, 1960). Adequada para tarefas de classificação realizada por vários anotadores/juízes. Os anotadores podem ser humanos ou algoritmos de classificação (FIOL; HAUG, 2009).

O resultado obtido com a medida *Kappa* define um limiar de concordância entre anotadores, seja em tarefas de classificação ou de extração de informação. Esse resultado depende do número de anotadores, número de classes e número de exemplos a serem classificados. Esta medida de concordância assume valor máximo igual a 1, que representa total concordância ou ainda pode assumir valores próximos e até abaixo de 0, os quais indicam nenhuma concordância.

LANDIS e KOCH (1977) propuseram uma escala de seis níveis para interpretar os valores de *Kappa* como mostrado na Tabela 2.

Tabela 2 – Escala para interpretação do coeficiente Kappa

Concordância	Escala
Ruim	Abaixo de 0
Leve	0,00 a 0,20
Sofrível	0,21 a 0,40
Moderada	0,41 a 0,60
Boa	0,61 a 0,80
Quase perfeita	Acima de 0,81

Fonte: Landis et al (1977).

Em geral a medida *Kappa* acima de 0,8 é considerada como um bom acordo, entre 0,67 e 0,8 é tido como justo acordo e abaixo de 0,67 é considerado como duvidosa para avaliação, embora a interpretação dos resultados da medida dependa dos fins para os quais os dados serão utilizados (MANNING; RAGHAVAN; SCHÜTZE, 2008).

2.7 CONSIDERAÇÕES

Um sistema de processamento de imagens digitais geralmente é constituído das etapas de aquisição, pré-processamento, segmentação, representação e descrição, reconhecimento e interpretação e resultado.

A etapa de aquisição consiste da obtenção e digitalização de imagem que se deseja analisar. O pré-processamento consiste na compensação de distorções específicas que são adquiridas na etapa de aquisição. A segmentação consiste da partição da imagem em regiões disjuntas, que serão analisadas de acordo com o objetivo proposto. A etapa de representação é aquela que define de que maneira os dados devem ser convertidos para uma forma adequada ao processamento computacional. O processo de descrição tem como objetivo extrair as características que resulte em informação quantitativa de interesse ou que sejam importantes para a diferenciação entre classes de objetos. A última etapa trata do reconhecimento e também da interpretação. O reconhecimento é definido como a atribuição de um rótulo a um objeto em função da informação que é fornecida por seu descritor. O reconhecimento de padrões classifica dados visuais numéricos ou simbólicos baseados em informações contidas em bancos de dados de padrões. A interpretação é a atribuição de significado a um conjunto de objetos reconhecidos.

De modo geral, os métodos existentes para reconhecimento de imagens são robustos e eficientes, porém a escolha de um método em detrimento de outro depende das especificidades de cada problema, assim como da realização de experimentos a fim de se analisar qual método produz resultados mais satisfatórios, sendo portanto mais adequado para solução do problema investigado.

Os métodos *Bag-of-Features* e Redes Neurais Convolucionais são citados em diversos trabalhos encontrados na literatura devido à elevada taxa de reconhecimento e ao baixo custo computacional no que diz respeito ao processamento de imagens digitais quando comparados aos demais métodos.

O Capítulo 3, a seguir, apresenta trabalhos relacionados ao reconhecimento de imagens digitais utilizando os métodos *Bag-of-Features* e Redes Neurais Convolucionais.

3 TRABALHOS RELACIONADOS

Neste capítulo é apresentada uma revisão do estado da arte de métodos computacionais utilizados para o reconhecimento de imagens. O objetivo desta seção, portanto, é investigar os métodos já existentes, descrevendo as abordagens e procedimentos utilizados por cada autor, identificando assim características em comum e os resultados alcançados nesses trabalhos.

Nas seções 3.1 e 3.2 são apresentadas publicações atuais da área utilizando os métodos *Bag-of-Features* e Redes Neurais Convolucionais, respectivamente, com uma descrição das características principais dos trabalhos e os resultados obtidos.

De maneira geral, não foram encontrados trabalhos na revisão da literatura que reportem especificamente a utilização dos métodos *Bag-of-Features* ou de Redes Neurais Convolucionais para o reconhecimento de imagens de marcas de gado.

SANCHEZ e RODRIGUEZ (2016) apresentam uma arquitetura para reconhecimento de marcas de gado utilizando momentos de Hu e Legendre para extração de características de imagens em escala de cinza, e um classificador de k-Vizinhos mais Próximos (k-NN).

Os autores utilizaram momentos de Hu e Legendre com o propósito de extrair características que não fossem suscetíveis a transformações de rotação, translação e escala. O percentual máximo de classificação correta apresentado pelos autores foi de 99,3 %, porém com uma diminuição significativa de acurácia à medida que aumentava o número de imagens classificadas.

Outro resultado apresentado foi o tempo de processamento da classificação. Pelo fato de ter sido utilizado um classificador k-NN, a cada novo objeto que se quer classificar, utiliza-se os dados de treinamento para verificar quais são os objetos nesta base de dados que mais se assemelham ao novo objeto que se quer classificar. O objeto é classificado dentro da classe mais comum a que pertencem os objetos mais similares a ele. Assim, a classificação é feita por analogia. Nenhum modelo de classificação é criado. Ao invés disto, a cada novo objeto a ser classificado, os dados de treinamento são escaneados, dessa forma o classificador proposto se torna computacionalmente dispendioso.

Diferentemente do trabalho apresentado por SANCHEZ e RODRIGUEZ (2016), o trabalho aqui proposto pretende apresentar resultados que possam ser generalizados ou replicados, utilizando técnicas no estado da arte no que diz respeito à extração de características e classificação estatística de imagens digitais, tais como *Bag-of-Features* (BoF), Redes Neurais Convolucionais (CNN) e Máquinas de Vetores de Suporte (SVM), objetivando criar uma arquitetura responsável pela classificação e recuperação das imagens de marca de gado, porém com resultados eficientes quando aplicadas a grandes bases de dados. O método descrito em SANCHEZ e RODRIGUEZ

(2016) apresenta uma perda de eficiência (acurácia e velocidade) significativa quando aplicado a grandes quantidades de imagens.

3.1 BAG-OF-FEATURES

De acordo com a literatura podemos encontrar vários trabalhos relacionados à técnica *bag-of-features*. O trabalho proposto por SIVIC e ZISSERMAN (2003) apresenta a técnica como uma abordagem para recuperar todas as ocorrências de um objeto em cenas (*frames*) de um determinado vídeo. Para isso, os objetos são representados como um conjunto de descritores invariantes à escala, rotação, translação, iluminação e oclusão parcial. Os resultados apresentados pelos autores confirmaram que o método demonstrou recuperação imediata na tarefa de reconhecimento de objetos em tempo de execução ao longo de uma base de dados de filmes, apesar das mudanças significativas que ocorriam em vários *frames*.

CSURKA et al. (2004) utiliza a técnica com o objetivo de encontrar um processo que seja genérico para lidar com diversos tipos de objetos e ao mesmo tempo tratar as variações de iluminação, visualização, rotação e oclusão, típicos de cenas do mundo real. Os autores propõem e compararam duas implementações usando os classificadores Naive Bayes e SVM. As principais vantagens apresentadas foram: simplicidade na implementação, eficiência computacional, e inerentemente invariante. Foram classificadas sete categorias simultaneamente. Os resultados demonstraram que o método é robusto e produz boa precisão de categorização mesmo sem explorar informações geométricas.

O modelo BoF tem apresentado bom desempenho em uma ampla variedade de tarefas, tais como: reconhecimento de ações, ULLAH, PARIZI e LAPTEV (2010); textura, ZHANG et al. (2007); gestos, DARDAS, CHEN e GEORGANAS (2010) e classificação de imagens, NOWAK, JURIE e TRIGGS (2006).

No trabalho de LOPES et al. (2009) o método é utilizado para detecção de nudez em vídeos. Já BATISTA, LOPES e ARAÚJO (2009) foi utilizada a metodologia BoF para identificação automática de imagens que contêm fachadas e edifícios no acervo digitalizado do Arquivo Público Mineiro.

LI, IMAI e KANEKO (2010) utilizaram BoF baseado em blocos para reconhecimento de faces. WANG et al. (2011b) aplicou BoF pesando as palavras visuais na recuperação de imagens médicas. Já em WANG et al. (2011a), os autores desenvolveram um algoritmo baseado em BoF para classificação de imagens de densidade do tecido mamário em mamografias. BARATA, MARQUES e MENDONÇA (2013) propuseram dois sistemas para a detecção de melanomas em imagens dermatológicas, onde o primeiro sistema usou métodos globais para classificar lesões de pele, e o segundo utilizou características locais e o método BoF para classificação das imagens. LI, WANG e ZHANG (2016) utilizaram o descritor SURF e pirâmide espacial em uma metodologia

bag-of-features para melhorar o reconhecimento e classificação de imagens.

3.2 REDES NEURAI CONVOLUCIONAIS

Segundo JARRETT, KAVUKCUOGLU e LECUN (2009), os trabalhos encontrados na literatura para reconhecimento e classificação de imagens que utilizam descritores ou filtros para extração de características, seguida de uma etapa de quantização e agrupamento e, por último, uma etapa de classificação são divididos em algoritmos de estágio único de extração de características, e algoritmos de dois ou mais estágios.

O treinamento de redes neurais com múltiplos estágios (camadas intermediárias), desperta o surgimento de diversos algoritmos agrupados em uma área conhecida como *deep learning*.

O objetivo principal dos algoritmos que utilizam dois ou mais estágios é aprender não apenas a distinguir as classes com base em descritores artificiais, mas aprender os próprios descritores com base nos dados brutos, no caso de imagens, os próprios valores dos *pixels* (JURASZEK, 2014).

JURASZEK (2014) apresentou diferentes técnicas para o problema de reconhecimento de produtos a partir da análise de imagens. Foram avaliadas a acurácia, tempo de processamento total, tempo de processamento para agrupamento, tempo de processamento para classificação das técnicas nas bases de imagens *Caltech 101* e em uma base de imagens de produtos criada pelo autor (RecogProd). Os resultados mostraram que a utilização de uma CNN, previamente treinada em uma base de imagens, como um descritor natural para extração de um vetor de características e treinamento de um classificador SVM, apresentou a melhor acurácia com 0,855 na base *Caltech 101* e 0,905 na base criada, RecogProd, em uma escala de 0 a 1.

Redes neurais convolucionais têm sido frequentemente utilizadas no reconhecimento de imagens, tendo obtido grande sucesso no reconhecimento de caracteres no trabalho de LECUN et al. (1990). Nesse trabalho, os autores apresentaram uma aplicação de redes com *backpropagation* para reconhecimento de dígitos manuscritos. Para a realização desta tarefa foi necessário realizar um pré-processamento dos dados utilizados nos experimentos. As imagens de entrada utilizadas eram constituídas de dígitos isolados. O método alcançou taxa de erro de 1%, e aproximadamente 9% de taxa de rejeição para dígitos de código postal fornecidos pelo serviço postal dos EUA.

De forma geral as CNN são treinadas de forma supervisionada, mas trabalhos sugerem que o pré-treinamento da CNN com filtros obtidos de forma não supervisionada apresentam um melhor resultado (KAVUKCUOGLU et al., 2010).

No trabalho realizado por SERMANET et al. (2014) é apresentado um *framework* utilizando CNN para efetuar o reconhecimento, localização e detecção de imagens. Os autores apresentaram uma estrutura integrada para o uso de redes convolutivas, como uma abordagem de janela multi escala e deslizante, podendo ser implemen-

tada eficientemente dentro de um ConvNet. Também introduziram uma abordagem inovadora de aprendizagem profunda para a localização, aprendendo a prever os limites de objetos. Os autores mostraram que diferentes tarefas podem ser aprendidas simultaneamente usando uma única rede compartilhada. Foi proposto um *pipeline* integrado que pode executar tarefas diferentes ao compartilhar uma base de extração de recursos comum, inteiramente aprendida diretamente dos pixels. Este trabalho foi o vencedor do campeonato *ImageNet Large Scale Visual Recognition Challenge* 2013 (ILSVRC2013). A base ILSVRC2013 consiste em 1,2 milhões de imagens divididas em 1000 categorias.

Uma característica importante da CNN é a capacidade de serem reutilizadas e refinadas para diferentes bases de imagens. No trabalho realizado por RAZAVIAN et al. (2014) foi utilizada uma CNN pré-treinada chamada *Overfeat* SERMANET et al. (2014), para realizar extração de um descritor de diferentes bases de imagens na qual a CNN não foi originalmente treinada. Nesse caso, os descritores são então classificados utilizando um classificador linear SVM. Os resultados demonstram um desempenho compatível com o estado da arte, mesmo comparando com algoritmos que utilizam imagens segmentadas manualmente procedimento que não é necessário quando utilizada a CNN, e treinados especificamente na base analisada.

A utilização de *deep learning* também é descrita no trabalho de CONSTANTE et al. (2016), os quais utilizaram uma rede neuronal de três camadas com entrada usando *backpropagation*. Nesse trabalho, o método foi utilizado para realizar classificação de morangos, e chegou a resultados de reconhecimento de 92,5% na categoria “Extra”; 90% na categoria “Consumo”; 90% na categoria “Matéria prima”; e 100% na categoria “Objetos estranhos”.

A Tabela 3 apresenta um resumo cronológico dos trabalhos encontrados na revisão da literatura que reportam a utilização de métodos computacionais para o reconhecimento de imagens.

Tabela 3 – Trabalhos relacionados

Método	Autores	Características principais
Redes Neurais Convolucionais	(LECUN et al., 1990)	Reconhecimento de caracteres.
<i>Bag-of-Features</i>	(SIVIC; ZISSERMAN, 2003)	Recuperação de ocorrências de um objeto em cenas (<i>frames</i>) de vídeo.
<i>Bag-of-Features</i>	(NOWAK; JURIE; TRIGGS, 2006)	Classificação de imagens.
<i>Bag-of-Features</i>	(ZHANG et al., 2007)	Classificação de texturas de objetos.
<i>Bag-of-Features</i>	(LOPES et al., 2009)	Detecção de nudez em vídeos.
<i>Bag-of-Features</i>	(BATISTA; LOPES; ARAÚJO, 2009)	Identificação de edifícios em fotografias históricas.
<i>Bag-of-Features</i>	(ULLAH; PARIZI; LAPTEV, 2010)	Reconhecimento de ações em vídeos.
<i>Bag-of-Features</i>	(DARDAS; CHEN; GEORGANAS, 2010)	Reconhecimento de gestos.
<i>Bag-of-Features</i>	(LI; IMAI; KANEKO, 2010)	Reconhecimento de faces.
<i>Bag-of-Features</i>	(WANG et al., 2011a)	Classificação de imagens de tecido mamário.
Redes Neurais Convolucionais	(CIREGAN; MEIER; SCHMIDHUBER, 2012)	Reconhecimento e classificação de imagens.
<i>Bag-of-Features</i>	(BARATA; MARQUES; MENDONÇA, 2013)	Detecção de melanomas em imagens dermatológicas.
Redes Neurais Convolucionais	(SERMANET et al., 2014)	<i>Framework</i> para efetuar reconhecimento, localização e detecção de imagens.
Redes Neurais Convolucionais	(JURASZEK, 2014)	Reconhecimento de produtos por imagem.
Redes Neurais Convolucionais	(RAZAVIAN et al., 2014)	Extração de características em diferentes bases de imagens usando CNN pré-treinada.
Momentos Hu e Legendre	(SANCHEZ; RODRIGUEZ, 2016)	Reconhecimento de imagens de marca de gado com a extração de características de imagens em escala de cinza.
<i>Bag-of-Features</i>	(LI; WANG; ZHANG, 2016)	Reconhecimento e classificação de imagens.
Redes Neurais Convolucionais	(CONSTANTE et al., 2016)	Classificação industrial de morangos.

Fonte: do próprio autor.

3.3 CONSIDERAÇÕES

De maneira geral, não foram encontrados trabalhos na revisão da literatura que reportem especificamente a utilização dos métodos *Bag-of-Features* ou de Redes Neurais Convolucionais para o reconhecimento de imagens de marcas de gado.

O método encontrado na literatura aplicado para reconhecimento de marcas de gado utilizou momentos de Hu e Legendre, juntamente com um classificador não supervisionado *k-nearest neighbors* (k-NN), porém o método apresentado não alcançou bom desempenho quando aplicado a grandes volumes de registros de marcas, além da considerável diminuição da taxa de reconhecimento.

Em contrapartida, o presente trabalho apresenta os resultados obtidos com os métodos *Bag-of-Features* e Redes Neurais Convolucionais utilizados para o desenvolvimento de uma arquitetura capaz de realizar o reconhecimento de imagens de marcas de gado, cujo objetivo é alcançar boas taxas de reconhecimento (acurácia), mesmo quando aplicados a grandes volumes de dados.

O estudo sobre os trabalhos já existentes contribuiu à definição da proposta da arquitetura, assunto detalhado no Capítulo 4.

4 DESENVOLVIMENTO DA ARQUITETURA

Neste capítulo são apresentados os materiais utilizados para a realização desta pesquisa, os quais são descritos na seção 4.1. Também são apresentadas as etapas do método *Bag-of-Features*, descritas na seção 4.2, e as etapas do método de Redes Neurais Convolucionais implementadas neste trabalho, as quais são apresentadas na seção 4.3. A seção 4.4, por fim, realiza o fechamento do capítulo.

4.1 MATERIAIS

4.1.1 CONJUNTO DE IMAGENS

As imagens de marcas de gado utilizadas nesse trabalho foram fornecidas pela Prefeitura de São Francisco de Assis, Brasil. O conjunto de imagens é composto por um total de 500 marcas de gado.

Algoritmos utilizados em aprendizado de máquina necessitam de uma grande quantidade de exemplos para atingir resultados satisfatórios. Não há como fazer uma estimativa numérica acerca da quantidade ideal de imagens para o treinamento de um classificador tendo em vista atingir bons resultados, pois a quantidade de exemplos necessários para esse propósito pode variar de acordo com a complexidade do problema, além outros fatores, tais como a qualidade do conjunto de imagens e os algoritmos de visão computacional utilizados.

Como a tarefa de reconhecimento das marcas é supervisionada, os exemplos de treino precisam ser pré-classificados por humanos, o que torna esses dados escassos. Uma alternativa menos trabalhosa que criar novos exemplos é modificar os exemplos já existentes. Esse aumento artificial dos dados visa explorar as variações dos dados que podem ocorrer em casos reais. A maior quantidade de dados diminui a chance do modelo se adaptar a valores muito específicos dos exemplos de treino e perder a capacidade de generalização.

Quando o modelo se adapta bem ao treino porém não interpreta bem os dados fora desse conjunto o modelo está sofrendo *overfitting*. Um dos maiores problemas de modelos com muitos parâmetros, como as CNN, é o *overfitting* (ROCHA, 2015).

Nesse sentido, foram utilizadas 12 imagens de marca de gado, e dessas, foram criadas artificialmente 45 subimagens para cada marca (com variações de escala, rotação e translação), totalizando 540 amostras oriundas das imagens originais, com o objetivo de explorar possíveis variações das imagens em situações reais de forma a aumentar o poder de identificação dos padrões e consequente acurácia do método proposto. As imagens foram disponibilizadas em alta resolução no formato *Portable Network Graphics* (PNG), com tamanho de 600 x 600 *pixels*.

As marcas utilizadas nos experimentos são ilustradas na Figura 16. O código

das marcas, os proprietários e a quantidade de amostras (subimagens) de cada marca são ilustrados na Tabela 4.

Figura 16 – Imagens de marcas de gado utilizadas nos experimentos.

					
802	803	804	805	811	812
					
813	814	815	821	822	1093

Fonte: Setor de Registro de Marcas - São Francisco de Assis (2016).

Tabela 4 – Marcas, proprietários e total de amostras por marca

Marca	Proprietário	Total de amostras
802	Proprietário "A"	45
803	Proprietário "B"	45
804	Proprietário "C"	45
805	Proprietário "D"	45
811	Proprietário "E"	45
812	Proprietário "F"	45
813	Proprietário "G"	45
814	Proprietário "H"	45
815	Proprietário "I"	45
821	Proprietário "J"	45
822	Proprietário "K"	45
1093	Proprietário "L"	45

Fonte: do próprio autor.

4.1.2 EQUIPAMENTOS UTILIZADOS

Para implementação da arquitetura proposta, assim como o armazenamento do conjunto de imagens utilizado, a realização dos experimentos e geração dos resultados foi utilizado um computador com processador Intel Core i5-3330 3 GHz, 6 MB L2 *cache*, com 8 GB de memória DDR3 1600 MHz e placa de vídeo Nvidia GeForce GTX 750 Ti, com suporte à plataforma de computação paralela CUDA (*Compute Unified Device Architecture*), e *compute capability version* 5.0.

CUDA é uma extensão para a linguagem de programação C, a qual possibilita o uso de computação paralela, cujo objetivo é proporcionar a programadores a utilização dos recursos da unidade de processamento gráfico (*Graphics Processing Unit*, GPU), para realizar algumas operações mais rapidamente (VASCONCELLOS, 2009).

A empresa Nvidia apresentou a tecnologia CUDA em 2006 (NVIDIA, 2016). Essa tecnologia desperta o interesse de usuários nos campos científico, biomédico, da computação, da análise de risco e da engenharia devido a característica de aplicações altamente paralelizáveis nesses campos e também ao seu grande potencial computacional.

Com a tecnologia CUDA é possível enviar código em C, C++ e Fortran diretamente à GPU, sem precisar usar uma linguagem de compilação. Empresas como Adobe, Autodesk e MathWorks utilizam cada vez mais a GPU para executar computação científica de propósito geral e de engenharia em uma ampla variedade de plataformas.

Atualmente a computação com GPU possibilita muito mais que processar imagens, pois lida com um *teraflop* de performance de ponto flutuante, o que permite processar tarefas de aplicativos projetados para diversos fins, desde finanças e medicina até *Deep Learning*.

4.1.3 MATRIX LABORATORY (MATLAB)

MatLab, sigla de *Matrix Laboratory*, é um software interativo de alta performance voltado para cálculos numéricos. O MatLab integra análise numérica, cálculo com matrizes, processamento de sinais e construção de gráficos em ambiente intuitivo, onde problemas e soluções são expressos como são escritos matematicamente, ao contrário da programação tradicional (UFMG, 2016).

É um sistema interativo cujo elemento básico de informação é uma matriz que não requer dimensionamento. Permite a resolução de muitos problemas numéricos em apenas uma fração do tempo que se gastaria para escrever um programa semelhante em linguagem Fortran, Basic ou C. Além disso, alguns pacotes para cálculos específicos podem ser incorporados a rotinas pré-definidas no MatLab.

Para a codificação dos algoritmos apresentados nesse trabalho foi utilizado o software MatLab na versão R2016a, com os pacotes *Neural Network*, *Parallel Computing* e *Statistics and Machine Learning*.

4.1.3.1 **PACOTE NEURAL NETWORK**

O pacote de rede neural fornece algoritmos, funções e aplicativos para criar, treinar, visualizar e simular redes neurais. Também é possível executar a classificação, regressão, *clustering*, redução de dimensionalidade, previsão de séries temporais e modelagem de sistema dinâmico e controle.

Este pacote inclui algoritmos *deep learning*, tais como redes neurais convolucionais, utilizadas para classificação de imagens e tarefas de aprendizagem. Para acelerar o processamento de grandes conjuntos de dados e cálculos, os mesmos podem ser distribuídos para processadores multicore, GPUs e *clusters* de computadores, com a utilização do pacote *Parallel Computing Toolbox*.

4.1.3.2 **PACOTE PARALLEL COMPUTING**

O pacote de computação paralela permite resolver problemas de uso intensivo de dados usando processadores multicore, GPUs e *clusters* de computadores. Permite a construção de algoritmos em alto nível de *loops* paralelos, tipos numéricos especiais e matrizes paralelizadas. Este pacote permite paralelizar aplicações interativamente ou em *batch*.

4.1.3.3 **PACOTE STATISTICS AND MACHINE LEARNING**

O pacote de aprendizagem de máquina e estatística fornece funções e aplicativos para descrever, analisar e modelar dados. Através desse pacote é possível realizar algoritmos de regressão e classificação que permitem fazer inferências a partir de dados e construir modelos preditivos.

Para a análise de dados multidimensionais o pacote fornece seleção de recursos, regressão, análise de componentes principais (PCA), regularização, e outros métodos de redução de dimensionalidade que permitem identificar variáveis ou características que impactam o modelo criado.

Além disso, o pacote fornece algoritmos de aprendizado de máquina supervisionado e não supervisionado, incluindo máquinas de vetor de suporte (SVM), k-vizinhos mais próximos (k-NN), *K-means*, modelos ocultos de Markov, dentre outros.

4.1.4 **BIBLIOTECA OPEN SOURCE VLFEAT**

O modelo de rede convolucional pré-treinado foi obtido junto à biblioteca VLFeat.org. VLFeat é uma biblioteca *open source* que implementa algoritmos de visão

computacional especializados na extração de características locais, análise e reconhecimento de imagens digitais.

Estão inclusos na biblioteca os algoritmos *Fisher Vector*, VLAD, SIFT, MSER, *K-means*, *hierarchical K-means*, SVM, dentre outros. Os algoritmos são escritos em linguagem C, possuem compatibilidade com o software MatLab e suporte aos sistemas operacionais Windows, Mac OS X e Linux (VLFEAT, 2016).

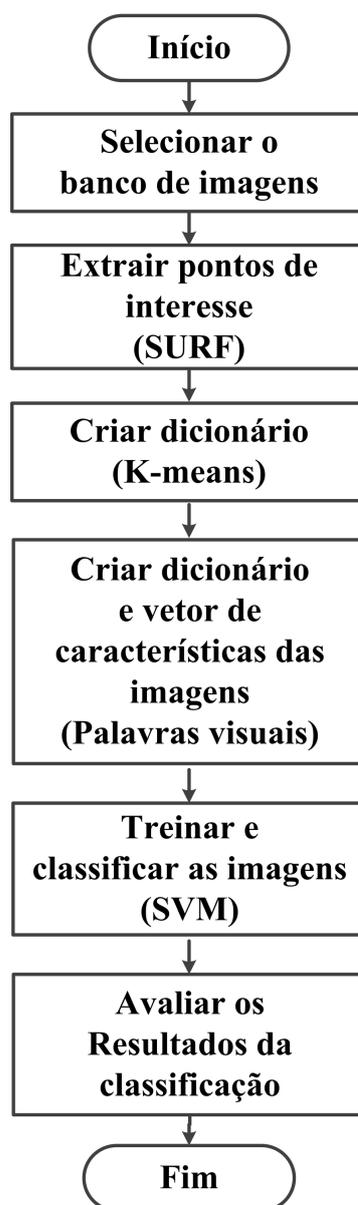
Além disso, a biblioteca disponibiliza o modelo MatConvNet que pode ser utilizado em redes neurais convolucionais (CNN), no qual o modelo pré-treinado é utilizado para classificação de imagens, segmentação, reconhecimento facial, detecção de texto, dentre outras aplicações.

4.2 MÉTODO BOF

4.2.1 ETAPAS DO MÉTODO PROPOSTO

O método proposto é composto por seis etapas, que são: seleção do banco de imagens; extração de pontos de interesse com utilização do algoritmo SURF; criação do dicionário de palavras visuais com o *clustering K-means*; criação dos histogramas e vetores de características das imagens; treinamento e classificação das imagens por meio de Máquinas de Vetores de Suporte; e, por fim, avaliação dos resultados da classificação. A Figura 17 ilustra o fluxograma sumarizado do método proposto.

Figura 17 – Fluxograma sumarizado do método proposto.



Fonte: do próprio autor.

4.2.1.1 SELEÇÃO DO BANCO DE IMAGENS

As marcas utilizadas nos experimentos realizados no método BoF foram ilustradas na Figura 16, assim como o código das marcas, os proprietários e a quantidade de amostras de cada marca foram apresentados na Tabela 4.

4.2.1.2 EXTRAÇÃO DOS PONTOS DE INTERESSE DAS MARCAS ATRAVÉS DO ALGORITMO SURF

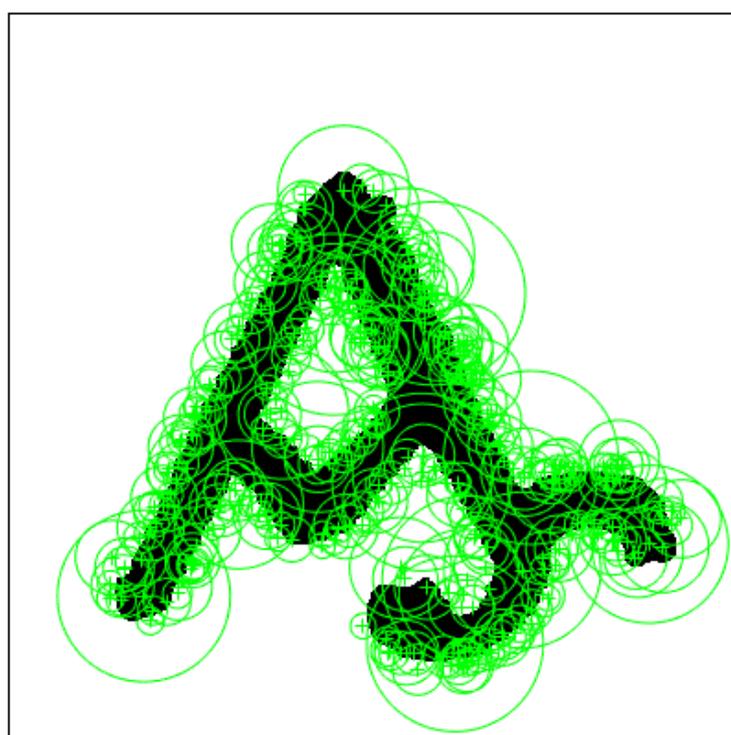
Após a seleção do banco de imagens é realizada a extração de pontos de interesse das imagens através da utilização do algoritmo SURF. Esse algoritmo baseia-se nas somas das respostas da *wavelet* Haar 2-D e no uso de imagens integrais

para detecção dos pontos de interesse, sendo dessa forma um descritor e detector de características locais robusto.

O algoritmo SURF é utilizado para diversas tarefas de visão computacional, como por exemplo o reconhecimento de objeto para reconstrução 3D. Embora, o SURF seja inspirado no algoritmo SIFT, ainda sim é mais rápido e robusto, e com a vantagem de ser um descritor de região invariante.

A Figura 18 ilustra a extração de 6.751 pontos de interesse (pontos-chave), da marca “1093” usando o algoritmo SURF.

Figura 18 – Extração de 6.751 pontos de interesse da marca “1093” usando o algoritmo SURF.



Fonte: do próprio autor.

O algoritmo SURF foi aplicado ao conjunto de imagens de marcas de gado das amostras. Inicialmente, o algoritmo detecta os pontos de interesse para em seguida calcular as características nesses pontos. O método procura encontrar as características que possuem posições salientes na imagem. A detecção dessas saliências baseia-se em um detector multi-escala e multi-orientação *Fast-Hessian* (matriz Hessiana), cujo descritor é baseado na distribuição da mudança de nível de cinza.

4.2.1.3 CRIAÇÃO DO DICIONÁRIO DE PALAVRAS UTILIZANDO K-MEANS

O agrupamento (*clustering*) é um método utilizado para agrupar objetos de acordo com características similares entre si. O método adotado para *clustering* no trabalho proposto foi o *K-means*.

A definição do tamanho do dicionário de palavras quando aplicado o método *K-means* é um dos pontos críticos para a criação do dicionário. A escolha do tamanho do dicionário (número de agrupamentos) é realizada de forma empírica, geralmente obtido após sucessivas execuções do algoritmo à procura da melhor sensibilidade (FERRAZ, 2016).

Esta escolha é primordial, pois além de influenciar no poder discriminativo do dicionário, tem influência direta na eficiência computacional e na utilização de memória durante os processos de geração do dicionário e de classificação das imagens. Além disso, os vetores das características (palavras visuais) que compõem o dicionário possuem alta dimensionalidade, dificultando então a aplicação do processamento do algoritmo.

Após realização de experimentos em que se avaliou a taxa de reconhecimento e desempenho do algoritmo, o tamanho de agrupamento proposto no trabalho apresentado foi de 500 aglomerações (*clusters*). Uma vez definido o tamanho do dicionário de palavras visuais, foi possível associar a cada descritor de ponto-chave à palavra visual mais próxima. Dessa forma, cada imagem foi representada por um histograma o qual indica com que frequência cada palavra visual ocorre no dicionário (de forma análoga ao que se faz em recuperação de informações textuais). O tamanho do dicionário pode ser verificado no eixo x da Figura 19.

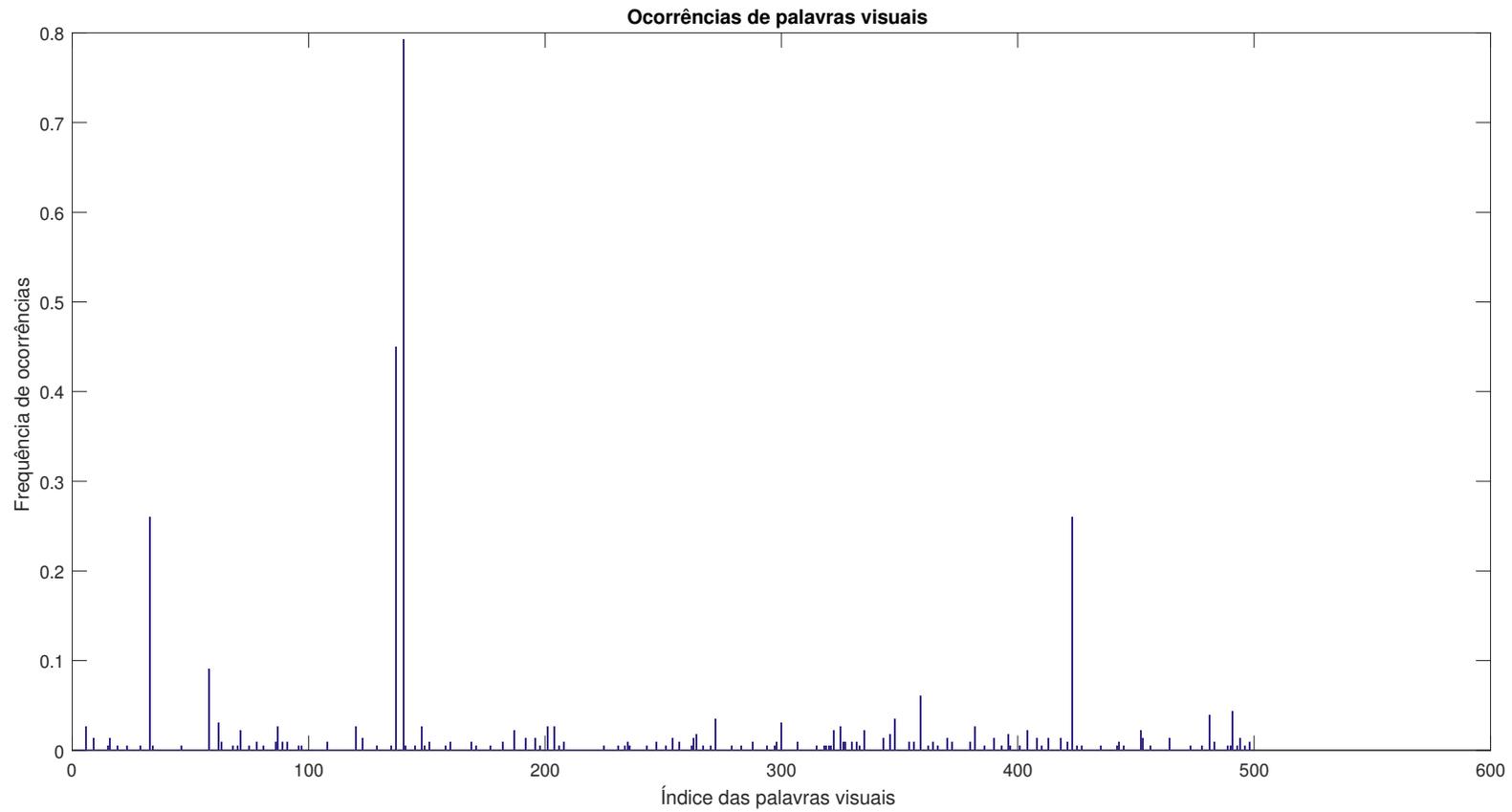
4.2.1.4 **CRIAÇÃO DO HISTOGRAMA E VETOR DE CARACTERÍSTICAS DAS PALAVRAS VISUAIS**

Com o dicionário de palavras visuais criado foram extraídos todos os descritores de todas as imagens da base de treinamento, identificando a qual palavra visual o descritor pertence e criando um histograma contendo a quantidade de cada palavra visual na imagem analisada.

Este histograma é um vetor onde cada posição corresponde a uma palavra visual e seu valor corresponde a quantidade de palavras visuais daquele tipo de imagem. O processamento foi realizado em todas as imagens de treinamento e o resultado é um histograma para cada imagem.

No método proposto, os vetores com os histogramas das imagens são criados utilizando o método *encode* da classe *bagOfFeatures* do MatLab. A Figura 19 apresenta o histograma com o número de ocorrências de palavras visuais e o tamanho do dicionário.

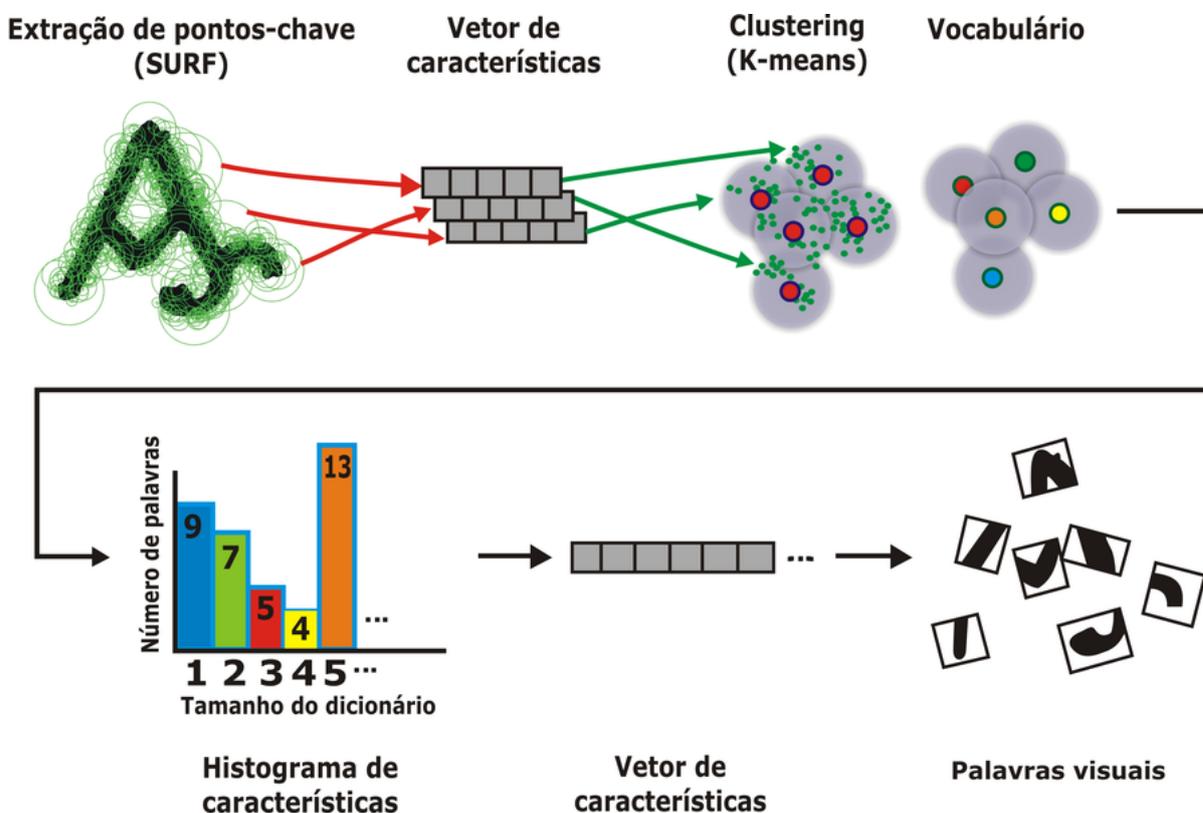
Figura 19 – Histograma com o número de ocorrências de palavras visuais e tamanho do dicionário.



Fonte: do próprio autor.

A Figura 20 ilustra o funcionamento do método *Bag-of-Features* aplicado no trabalho proposto.

Figura 20 – Método Bag of Features aplicado no trabalho proposto.



Fonte: do próprio autor.

4.2.1.5 TREINAMENTO E CLASSIFICAÇÃO DAS IMAGENS COM A UTILIZAÇÃO DE MÁQUINAS DE VETORES DE SUPORTE

A quinta etapa do método consiste no treinamento de um classificador linear para que através do histograma de palavras visuais seja possível determinar a que categoria a imagem pertence. O modelo de aprendizagem automática adotado foi o classificador supervisionado *Support Vector Machine* (SVM).

Máquinas de Vetores de Suporte é um método de classificação conhecido por ser bem sucedido em uma grande variedade de aplicações. As SVM são uma das abordagens mais populares para modelagem e classificação de dados.

As suas vantagens incluem a excelente capacidade de generalização, que diz respeito à capacidade de classificar corretamente as amostras que não estão dentro do espaço da característica usado para o treino (TEIXEIRA, 2011).

Dadas duas classes e um conjunto de pontos a essas classes, o SVM determina o hiperplano que separa os pontos de forma a colocar o maior número de pontos da mesma classe do mesmo lado, maximizando a distância de cada classe a esse hiperplano, sendo conseqüentemente denominado classificador de margem máxima (LU

et al., 2007). Com efeito, uma larga margem entre os valores correspondentes aos pontos dos dois subconjuntos de dados implica um risco de generalização minimizado do classificador.

As SVMs são utilizadas para classificar e reconhecer padrões em diversos tipos de dados, sendo utilizadas em diversas aplicações, tais como reconhecimento de faces, diagnósticos clínicos, supervisão de processos industriais, processamento e análise de imagens (TCHANGANI, 2005).

O classificador SVM foi utilizado após a extração de características das marcas pertencentes a diferentes conjuntos de amostras. Na aprendizagem supervisionada dado um conjunto de exemplos (X_1, X_2) em X_1 representa um exemplo e X_2 a sua classificação, deve-se reproduzir um classificador capaz de prever a classe a que pertencem novos dados, efetuando assim o processo de treino.

Na pesquisa proposta, separamos aleatoriamente o conjunto de imagens em duas partes, onde uma dessas partes foi utilizada para a fase de treinamento e a outra para a fase de validação, eliminando assim a polarização nos resultados. O resultado final é a média do resultado obtido na validação. A divisão da percentagem utilizada foi de 30% para treinamento e 70% para validação.

4.2.1.6 **AVALIAÇÃO DOS RESULTADOS DA CLASSIFICAÇÃO - MATRIZ DE CONFUSÃO**

Para avaliação do método *Bag-of-Features* foi gerada uma matriz de confusão, a qual indicou o número de classificações corretas versus as previsões efetuadas para cada caso, sobre um conjunto de exemplos.

Nesta matriz as linhas representam os casos reais e as colunas as previsões efetuadas pelo modelo. Através da matriz de confusão é possível obter informação relativa ao número de imagens corretamente classificadas e incorretamente classificadas, para cada conjunto de amostras.

Esta matriz é do tipo $A \times A$, sendo A o número de categorias ao qual se aplica o classificador, no caso do experimento efetuado existem 12 marcas, sendo portanto a matriz de confusão de tamanho 12×12 .

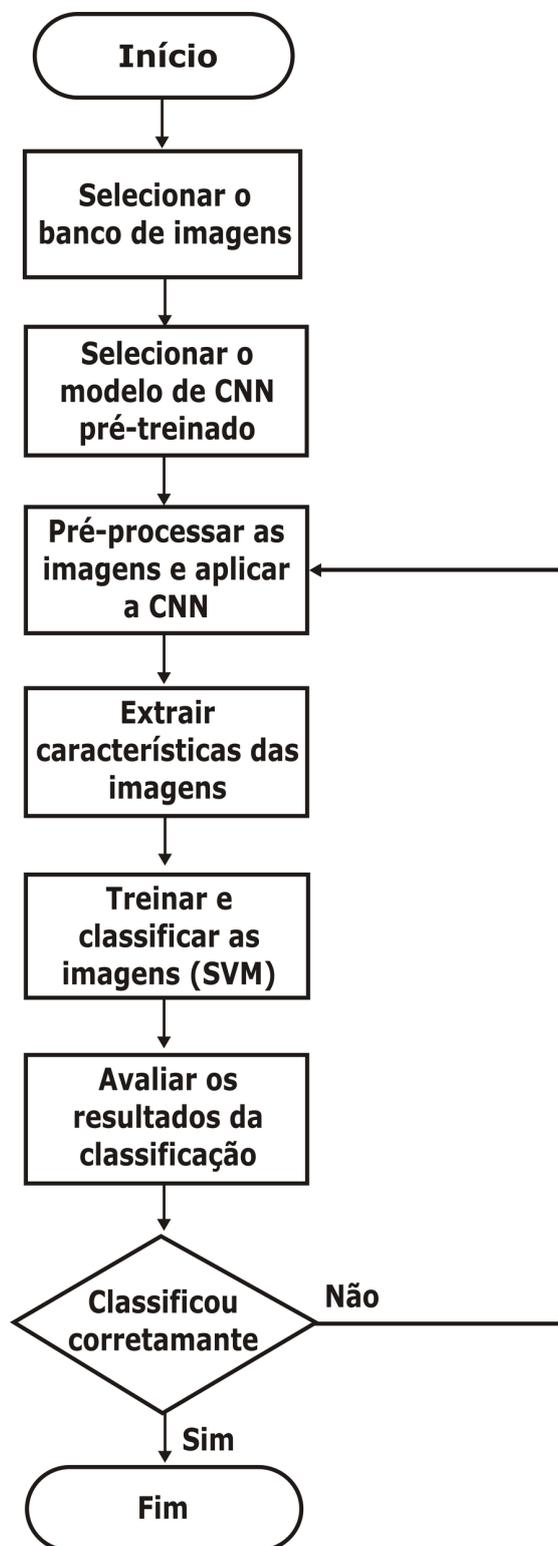
4.3 **MÉTODO CNN**

4.3.1 **ETAPAS DO MÉTODO PROPOSTO**

O método proposto é composto por seis etapas, que são: seleção do banco de imagens; seleção do modelo de CNN pré-treinado; pré-processamento das imagens e aplicação da CNN; extração de características das imagens; treinamento e classificação das imagens por meio de Máquinas de Vetores de Suporte; e, por fim, avaliação dos

resultados da classificação. A Figura 21 ilustra o fluxograma sumarizado do método proposto.

Figura 21 – Fluxograma sumarizado do método proposto.



Fonte: do próprio autor.

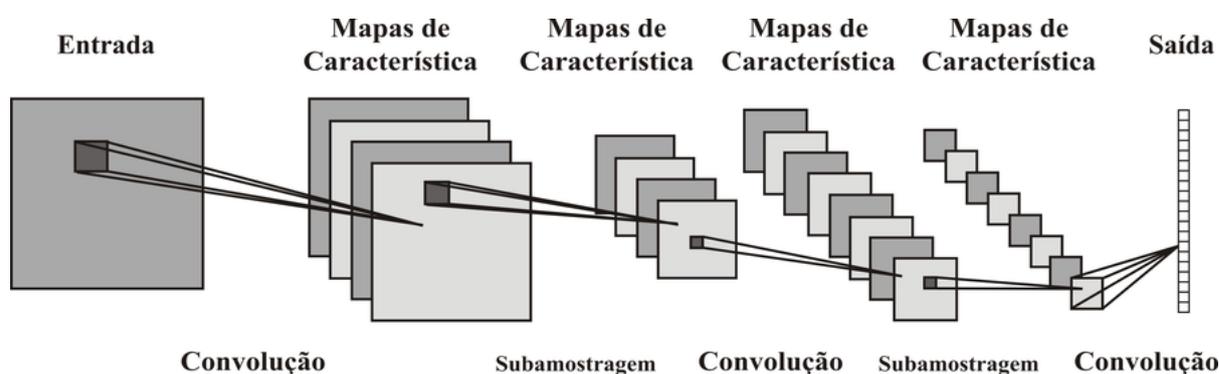
4.3.1.1 SELEÇÃO DO BANCO DE IMAGENS

As marcas utilizadas nos experimentos realizados no método Redes Neurais Convolucionais foram ilustradas na Figura 16, assim como o código das marcas, os proprietários e a quantidade de amostras de cada marca foram apresentados na Tabela 4.

4.3.1.2 SELEÇÃO DO MODELO DE CNN PRÉ-TREINADO

Por definição, uma arquitetura profunda é uma estrutura hierárquica de múltiplas etapas, onde cada etapa é formada por uma rede neuronal de pelo menos 3 camadas, e cada etapa é treinada por *backpropagation* (CONSTANTE et al., 2016). A Figura 22 ilustra o modelo geral de uma arquitetura CNN.

Figura 22 – Modelo geral de uma arquitetura CNN.



Fonte: do próprio autor.

As redes neurais convolucionais dispõem de uma propriedade ainda pouco explorada conhecida como transferência de conhecimento. Esta propriedade remete ao fato de uma CNN poder ser treinada em um conjunto de imagens “A” (sendo assim, seus pesos são ajustados para classificação do conjunto “A”) sendo os pesos de aprendizado (e filtros nas CNN) considerados genéricos o suficiente para serem usados no treinamento de uma nova base “B”.

Utilizando-se desse conceito foi selecionado o modelo CNN da biblioteca *open source* VLFeat, cujos dados utilizados para pré-treinamento da rede neural convolucional utilizada no trabalho apresentado foram obtidos junto ao arquivo “imagenet-cafe-alex.mat”. A utilização do modelo de CNN pré-treinado supracitado não influenciou diretamente na taxa de reconhecimento das imagens de marcas de gado.

4.3.1.3 PRÉ-PROCESSAMENTO DAS IMAGENS E APLICAÇÃO DA CNN

O método adotado consiste em uma rede neural com cinco camadas convolucionais. Caso a imagem seja em tons de cinza é realizado um pré-processamento, no

qual a imagem é replicada 3 vezes para criar uma imagem no padrão RGB, ou seja, vermelho, verde e azul, do inglês *Red, Green, Blue*.

A 1ª camada convolucional tem como entrada os 3 canais de cores da imagem (RGB). Cada convolução realiza a aplicação da função de ativação não linear *ReLU* e a redução através do *Maxpooling*. As últimas camadas são compostas por neurônios totalmente conectados.

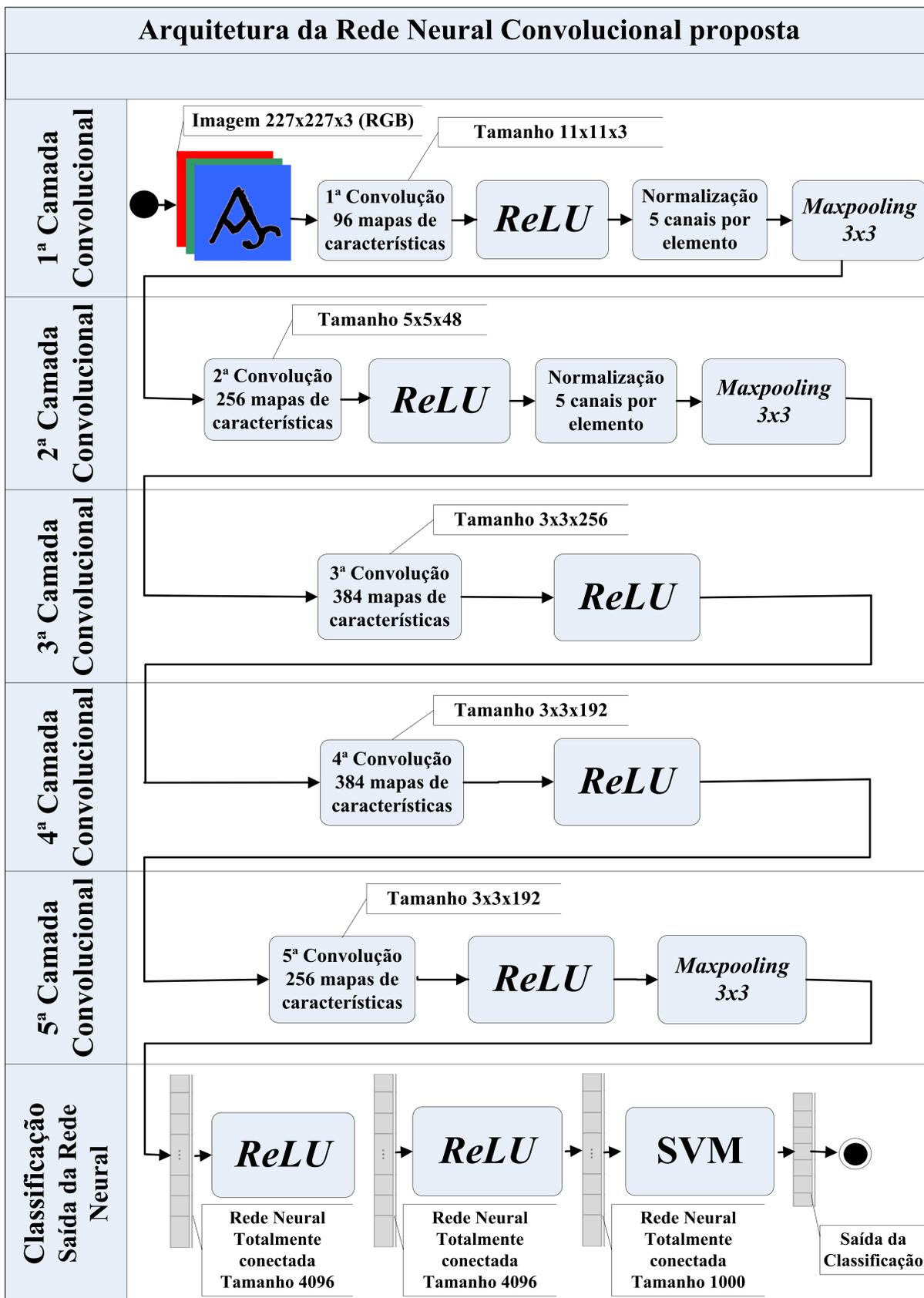
A Figura 23 apresenta a arquitetura da rede neural convolucional proposta. Caso a imagem seja em tons de cinza é necessário realizar um pré-processamento, no qual a imagem é replicada 3 vezes para criar uma imagem RGB.

O número de camadas convolucionais, o número de camadas totalmente conectadas, a quantidade de filtros aprendidos em cada camada, assim como os demais parâmetros utilizados na arquitetura proposta foram baseados no modelo “AlexNet” proposto por (KRIZHEVSKY; SUTSKEVER; HINTON, 2012). A disposição das diversas camadas da arquitetura implementada, assim como os parâmetros utilizados são descritos a seguir.

A 1ª camada convolucional tem como entrada os 3 canais de cores da imagem (RGB) e tamanho de 227x227, redimensionadas das imagens originais. Nessa camada é realizada uma convolução com 96 filtros. Após a convolução ser realizada é aplicada a função *ReLU*. Em seguida é realizada a normalização do canal transversal com 5 canais por elemento. Na sequência, é realizada a redução (*maxpooling*) através do valor máximo de vizinhança 3x3.

O processo se repete de forma análoga nas próximas camadas convolucionais, porém com a 2ª camada com 256 filtros, a 3ª camada com 384 filtros, a 4ª camada com 384 filtros e a 5ª camada com 256 filtros. Além disso, a arquitetura implementada é composta por 3 camadas totalmente conectadas.

Figura 23 – Arquitetura da rede neural convolucional proposta.



Fonte: do próprio autor.

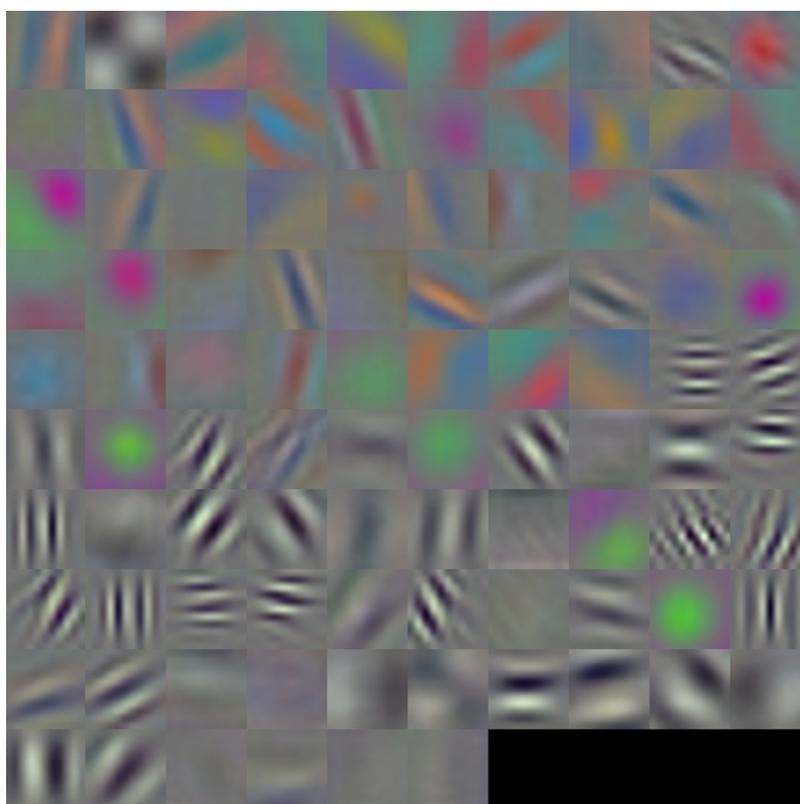
4.3.1.4 **EXTRAÇÃO DE CARACTERÍSTICAS DAS IMAGENS**

O conjunto de filtros aprendidos pela CNN durante o treinamento é responsável pela detecção das características na nova imagem no momento de uma consulta. No primeiro nível de filtros é possível observar algumas linhas e orientações utilizadas para essa detecção.

A Figura 24 mostra os filtros aprendidos na primeira camada convolucional utilizando espaço de cores RGB. Existem 96 conjuntos individuais que representam os 96 filtros utilizados nessa camada.

É possível observar como as áreas com saliências horizontais, verticais e diagonais são destacadas após a realização da primeira convolução.

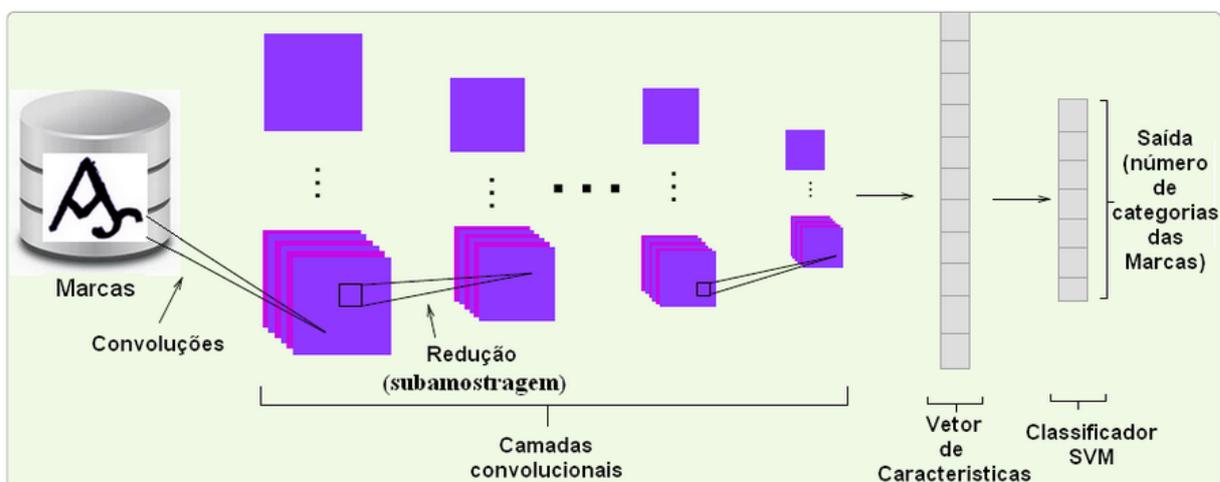
Figura 24 – Filtros da 1ª camada convolucional do experimento realizado.



Fonte: do próprio autor.

Após a realização das convoluções foi possível realizar a extração das características das imagens de forma que o classificador pudesse ser treinado. A Figura 25 apresenta uma ilustração do algoritmo de extração de características desenvolvido.

Figura 25 – Ilustração do algoritmo de extração de características desenvolvido.



Fonte: do próprio autor.

4.3.1.5 TREINAMENTO E CLASSIFICAÇÃO DAS IMAGENS COM A UTILIZAÇÃO DE SVM

O modelo de aprendizagem automática adotado para treinamento e classificação após a implementação da Rede Neural Convolucional foi o classificador supervisionado *Support Vector Machine* (SVM). O classificador foi utilizado após a extração das características do conjunto de marcas de gado selecionadas para a realização do experimento.

Para treinamento do classificador separamos aleatoriamente o conjunto de imagens em duas partes, onde uma dessas partes foi utilizada para a fase de treinamento e a outra para a fase de validação, eliminando assim a polarização nos resultados. O resultado final é a média do resultado obtido na validação. A divisão da porcentagem utilizada foi de 30% para treinamento e 70% para validação.

4.3.1.6 AVALIAÇÃO DOS RESULTADOS DA CLASSIFICAÇÃO - MATRIZ DE CONFUSÃO

Para avaliação do método de Redes Neurais Convolucionais foi gerada uma matriz de confusão, a qual indicou o número de classificações corretas versus as previsões efetuadas para cada caso, sobre um conjunto de exemplos. Nesta matriz as linhas representam os casos reais e as colunas as previsões efetuadas pelo modelo.

Através da matriz de confusão é possível obter informação relativa ao número de imagens corretamente classificadas e incorretamente classificadas, para cada conjunto de amostras. Esta matriz é do tipo $A \times A$, sendo A o número de categorias ao qual se aplica o classificador, no caso do experimento efetuado existem 12 marcas, sendo portanto a matriz de confusão de tamanho 12×12 .

4.4 CONSIDERAÇÕES

Neste capítulo apresentou-se os materiais e métodos utilizados no trabalho. As características necessárias para a arquitetura proposta foram baseadas nas especificidades do problema proposto assim como nos trabalhos relacionados.

Nos materiais, foram apresentados os conjuntos de imagens e os equipamentos utilizados nos experimentos. Trabalhou-se com o software MatLab para implementação dos algoritmos e o modelo de rede convolucional pré-treinado da biblioteca VLFeat para treinamento da Rede Neural Convolucional proposta.

No método *Bag-of-Features* aplicou-se o algoritmo de extração de pontos de interesse SURF e o método *K-means* para clusterização do dicionário de palavras visuais.

A classificação em ambos os métodos apresentados foi realizada utilizando o método supervisionado de Máquinas de Vetores de Suporte, e a avaliação foi realizada mediante a análise dos resultados oriundos das matrizes de confusão geradas nos experimentos realizados.

Para medição e comparação dos resultados obtidos foram utilizadas as métricas de acurácia, sensibilidade (*recall*), precisão, coeficiente *Kappa* e tempo de processamento, conforme detalhado no próximo capítulo.

5 EXPERIMENTOS, RESULTADOS E DISCUSSÕES

Este capítulo visa apresentar a discussão dos resultados de forma quantitativa, demonstrando a tabulação dos dados comparativos dos experimentos realizados. Para avaliação dos resultados foram utilizadas as métricas de acurácia, sensibilidade (*recall*), especificidade, precisão, coeficiente *Kappa* e tempo de processamento.

A escolha das métricas utilizadas para avaliar a arquitetura proposta levou em conta o fato dessas possuírem a capacidade de serem adaptadas tanto para classificação binária quanto de múltiplas classes, cujo objetivo é prever qual é a categoria a que uma marca de gado pertence. Além disso, tais métricas são amplamente utilizadas na literatura para avaliar se um sistema de Recuperação de Informações se comporta da maneira na qual ele foi proposto, ou seja, não avalia somente se o sistema recupera as informações corretas, mas também o quanto o sistema abstrai as informações erradas.

O capítulo se encerra com a apresentação das limitações da arquitetura proposta observadas durante a fase de avaliação dos experimentos realizados.

5.1 EXPERIMENTO I: MÉTODO BOF APLICADO À 12 MARCAS EM UM CONJUNTO DE 540 IMAGENS

5.1.1 ACURÁCIA, SENSITIVIDADE, PRECISÃO E COEFICIENTE KAPPA

Através dos resultados obtidos foi possível avaliar o método proposto. A avaliação dos resultados do experimento foi realizada com base na taxa de reconhecimento obtida da matriz de confusão gerada a partir da classificação realizada na etapa de validação. Além disso, o tempo total de processamento do método proposto também foi verificado.

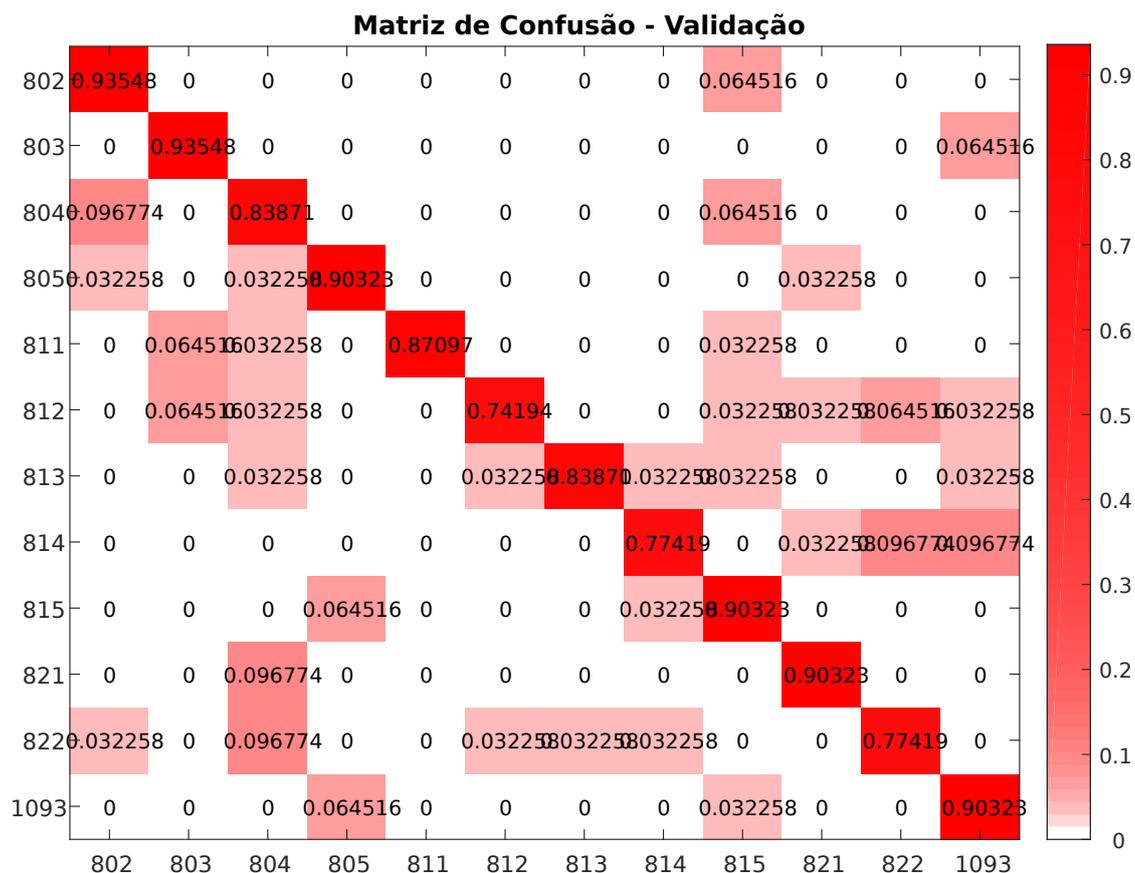
A Figura 26 apresenta a matriz de confusão para o melhor resultado obtido no experimento, cuja taxa de reconhecimento (acurácia geral, ou *overall accuracy*, OA), chegou a 86,02%. Percebe-se através da análise da diagonal principal que a taxa de classificação correta se destaca em duas marcas, “802” e “803”, nas quais o percentual de acerto chega a 93,55%. É possível observar também que as marcas que tiveram as menores taxas de classificação correta foram a “812”, “814” e a “822”, com percentual de 74,19%, 77,42% e 77,42%, nessa ordem. As demais marcas, “804”; “805”; “811”; “813”; “815”; “821” e “1093”, alcançaram taxa de classificação correta de 83,87%; 90,32%; 87,09%; 83,87%; 90,32%; 90,32% e 90,32%, respectivamente.

A hipótese de classificação errada das marcas de gado conforme mostrada na matriz de confusão está associada à complexidade das amostras e ao tamanho do dicionário adotado para agrupamento. Porém, é importante destacar que um dicionário de tamanho maior impacta diretamente no desempenho do algoritmo.

Em geral, as amostras de imagens de marcas com maior poder descritivo

e melhor qualidade classificaram corretamente mais imagens, pois abrangem mais características quando comparadas às marcas com menor qualidade das amostras, e, conseqüentemente, menos características extraídas. A capacidade de reconhecer padrões de uma imagem sobre um conjunto de imagens depende da quantidade de informações que se conhece *a priori* do objeto em questão.

Figura 26 – Matriz de confusão obtida na etapa de validação - Experimento I



Fonte: do próprio autor.

A Figura 27 mostra o percentual de marcas classificadas corretamente, assim como o percentual de falsos positivos e falsos negativos com base no experimento realizado na fase de validação.

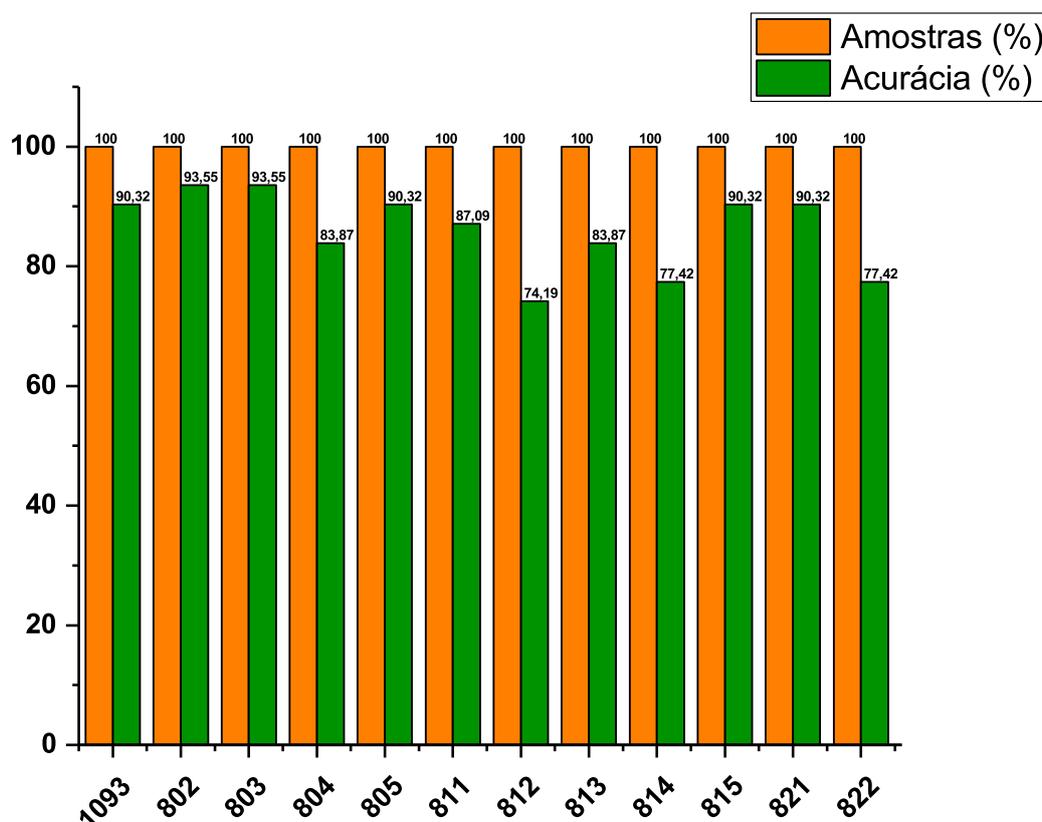
Figura 27 – Percentual de marcas classificadas corretamente e percentual de erros - Experimento I.



Fonte: do próprio autor.

A Figura 28 mostra um gráfico comparativo entre o percentual de marcas classificadas corretamente em função das amostras utilizadas no experimento.

Figura 28 – Percentual de marcas corretamente classificadas - Experimento I.



Fonte: do próprio autor.

As barras em laranja apresentam o percentual de amostras de cada marca e as barras em verde apresentam o percentual das marcas de gado classificadas corretamente (acurácia). Das 12 marcas analisadas, 2 apresentaram 93,55% de classificações corretas, no caso as marcas “802” e “803”. As marcas de gado que tiveram o menor percentual de acerto foram a “812”, “814” e “822”, com 74,19%, 77,42% e 77,42%, respectivamente.

Com base na análise da matriz de confusão apresentada na Figura 26, é possível observar que o método apresentou um valor relativamente baixo de Falsos Negativos (FN), ou seja, classificou corretamente quem deveria ser classificado, portanto teve uma taxa de Sensibilidade alta. Além disso, é possível verificar também que o número de Falsos Positivos (FP) foi baixo, o que proporcionou uma taxa de Precisão significativa para todas as marcas classificadas.

O Coeficiente *Kappa*, que serve como métrica para avaliar o desempenho do método proposto em função das medidas de precisão e sensibilidade, obteve um índice de 0,848.

A Tabela 5 apresenta os resultados obtidos pelo classificador com a utilização do Método *Bag-of-Features*. Foram utilizadas as métricas de Sensibilidade, Especificidade

e Precisão, respectivamente.

Tabela 5 – Sensitividade (TPR), Especificidade (TNR) e Precisão (PPV) - Experimento I

Marca	TPR	TNR	PPV
802	0,85294	0,99408	0,93548
803	0,87879	0,99407	0,93548
804	0,72222	0,98512	0,83871
805	0,87500	0,99118	0,90323
811	1,00000	0,98841	0,87097
812	0,92000	0,97695	0,74194
813	0,96296	0,98551	0,83871
814	0,88889	0,97971	0,77419
815	0,77778	0,99107	0,90323
821	0,90323	0,99120	0,90323
822	0,82759	0,97959	0,77419
1093	0,80000	0,99110	0,90323

Fonte: do próprio autor.

A sensibilidade reflete o quanto o método foi eficaz em identificar corretamente, dentre todas as imagens avaliadas, aquelas que realmente pertencem a uma determinada marca. Já a especificidade refletiu o quanto o método foi eficaz em identificar corretamente as imagens que não pertencem a uma determinada marca.

Dessa forma, quando o método apresenta baixa sensibilidade está mais propenso a fornecer resultados chamados de falso-negativos, enquanto que ao apresentar baixa especificidade, o classificador está mais propenso a fornecer resultados falso-positivos. É importante que exista um equilíbrio entre os valores de sensibilidade e especificidade obtidos, assim como observado na Tabela 5.

5.1.2 DESEMPENHO DE PROCESSAMENTO

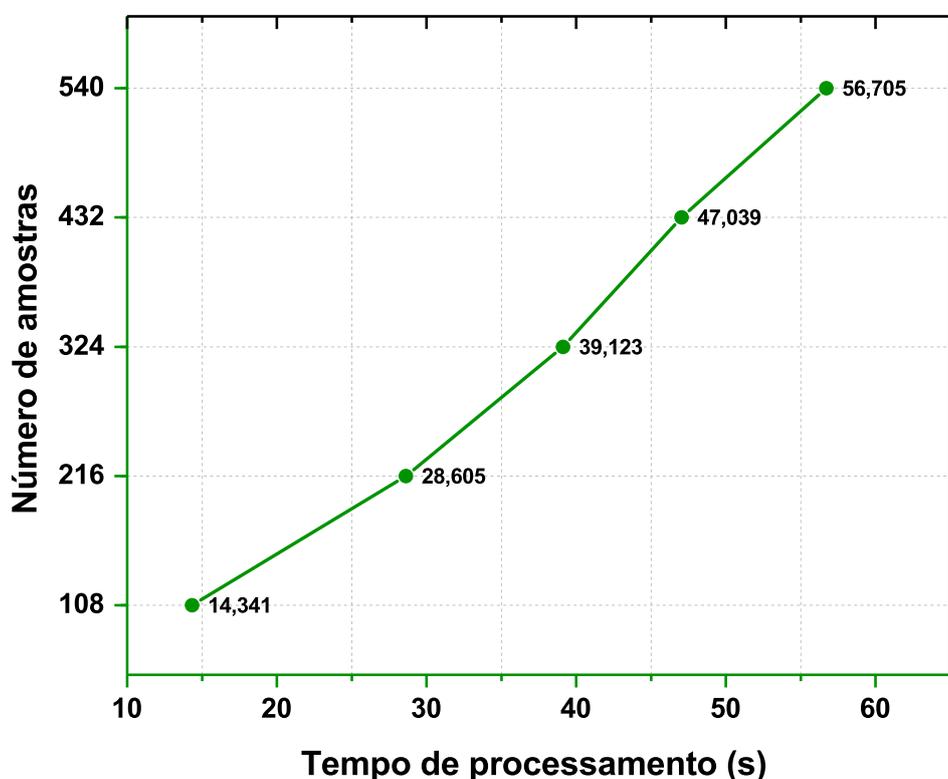
O tempo de processamento do algoritmo em função do número de amostras de marcas de gado é apresentado na Figura 29.

Foram medidos os tempos de processamento do método proposto para a classificação em cinco grupos de amostras. Cada grupo contendo 108, 216, 324,

432 e 540 imagens. Os tempos de processamento para classificação das imagens em cada grupo foram de 14,341s; 28,605s; 39,123s; 47,039s e 56,705s, nessa ordem.

Ao analisar o gráfico ilustrado na Figura 29 é possível perceber que o tempo de processamento observado no eixo y, varia de forma diretamente proporcional ao aumento do número de amostras de marcas de gado classificadas do eixo x, ou seja, observa-se um padrão de crescimento linear da função, mesmo que a taxa de crescimento não seja exatamente um valor constante.

Figura 29 – Tempo de processamento do algoritmo em função do número de amostras - Experimento I.



Fonte: do próprio autor.

Os resultados obtidos com o experimento realizado com 12 marcas de gado e 540 amostras de imagens, utilizadas tanto para treinamento quanto para validação, alcançaram taxa de reconhecimento (acurácia geral) de 86,02%, taxa de erro de 13,98% e tempo total de processamento de 56,705 segundos.

A acurácia geral foi obtida mediante o cálculo da média aritmética das marcas classificadas corretamente da matriz de confusão, e o tempo total de processamento foi obtido mediante a utilização do software MatLab, que ao final do processamento do código realiza o detalhamento da velocidade de processamento do algoritmo.

Com relação ao tempo de processamento é importante destacar que o mesmo está associado à arquitetura e configuração dos equipamentos utilizados nos experimentos, e que poderá variar de acordo com os dispositivos de *hardware* empregados, podendo alcançar desempenho diferente em comparação ao apresentado nesse trabalho.

Na Tabela 6 são apresentados os resultados obtidos com o experimento utilizando o Método *Bag-of-Features*.

Tabela 6 – Acurácia Geral, Erro e Tempo de Processamento - Método BoF - Experimento I

Marcas	Amostras	Acurácia geral (%)	Taxa de erro (%)	Tempo de processamento (s)
12	540	86,02	13,98	56,705

Fonte: do próprio autor.

5.2 EXPERIMENTO II: MÉTODO CNN APLICADO À 12 MARCAS EM UM CONJUNTO DE 540 IMAGENS

5.2.1 ACURÁCIA, SENSITIVIDADE, PRECISÃO E COEFICIENTE KAPPA

Através dos resultados obtidos foi possível avaliar o método proposto. A avaliação dos resultados do experimento foi realizada com base na taxa de reconhecimento obtida da matriz de confusão gerada a partir da classificação realizada na etapa de validação. Além disso, o tempo total de processamento do método proposto também foi verificado.

A Figura 30 apresenta a matriz de confusão para o melhor resultado obtido no experimento, cuja taxa de acurácia geral chegou a 93,28%. Percebe-se através da análise da diagonal principal que a taxa de classificação correta se destaca em quatro marcas, “802”, “812”, “815” e “821”, nas quais o percentual de acerto chega a 100%, correspondendo a 31 acertos. É possível observar também que as marcas que tiveram as menores taxas de classificação correta foram a “813” e a “805”, com percentual de 77,42% e 80,64%, nessa ordem. As demais marcas, “803”; “804”; “811”; “814”; “822” e “1093”, alcançaram taxa de classificação correta de 96,77%; 93,55%; 90,32%; 90,3%; 96,77% e 93,55%, respectivamente.

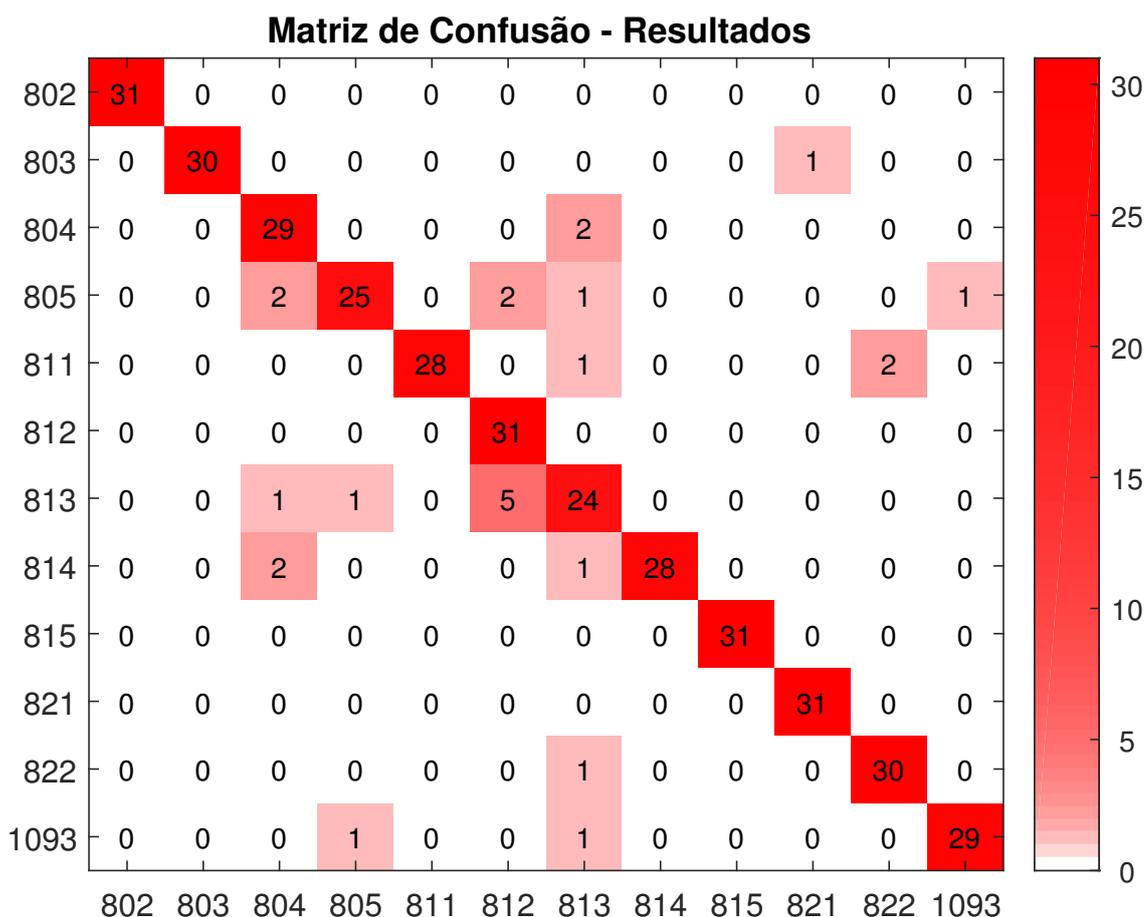
A hipótese de classificação errada das marcas de gado conforme mostrada na matriz de confusão está associada à complexidade das amostras, pois algumas imagens possuem características similares entre si.

Em geral, as amostras de imagens de marcas com maior poder descritivo

e melhor qualidade classificaram corretamente mais imagens, pois abrangem mais características quando comparadas às marcas com menor qualidade das amostras, e, conseqüentemente, menos características extraídas. A capacidade de reconhecer padrões de uma imagem sobre um conjunto de imagens depende da quantidade de informações que se conhece *a priori* do objeto em questão.

Outro fato importante a ser apresentado, fruto da análise da Figura 30, é o número baixo de falsos positivos e falsos negativos, observados fora da diagonal principal da matriz de confusão, o que se deve principalmente à capacidade do método proposto de extrair características de imagens, mesmo em situações adversas, em que as imagens apresentem diferentes tamanhos, formas, escalas, orientações, distorções, ruídos, cores diferentes e diferente contexto de fundo. Nesse tipo de experimento os ruídos nas imagens podem prejudicar a acurácia da classificação.

Figura 30 – Matriz de confusão obtida na etapa de validação - Experimento II.



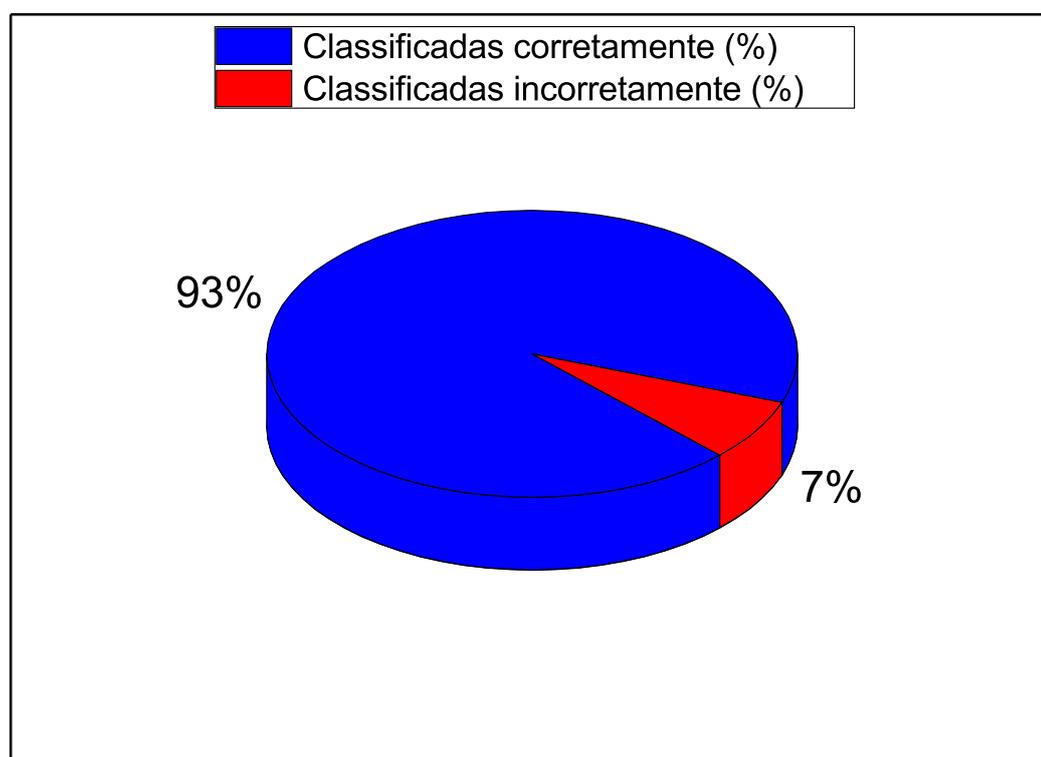
Fonte: do próprio autor.

Tendo em consideração a matriz de confusão apresentada na Figura 30, observa-se que através do somatório da diagonal principal obtém-se o total de marcas de gado

classificadas corretamente, mais precisamente 347, ao passo que a soma dos demais valores equivale às marcas classificadas incorretamente, ou seja, 25.

A Figura 31 mostra o percentual de marcas classificadas corretamente assim como o percentual de falsos positivos e falsos negativos com base no experimento realizado na fase de validação.

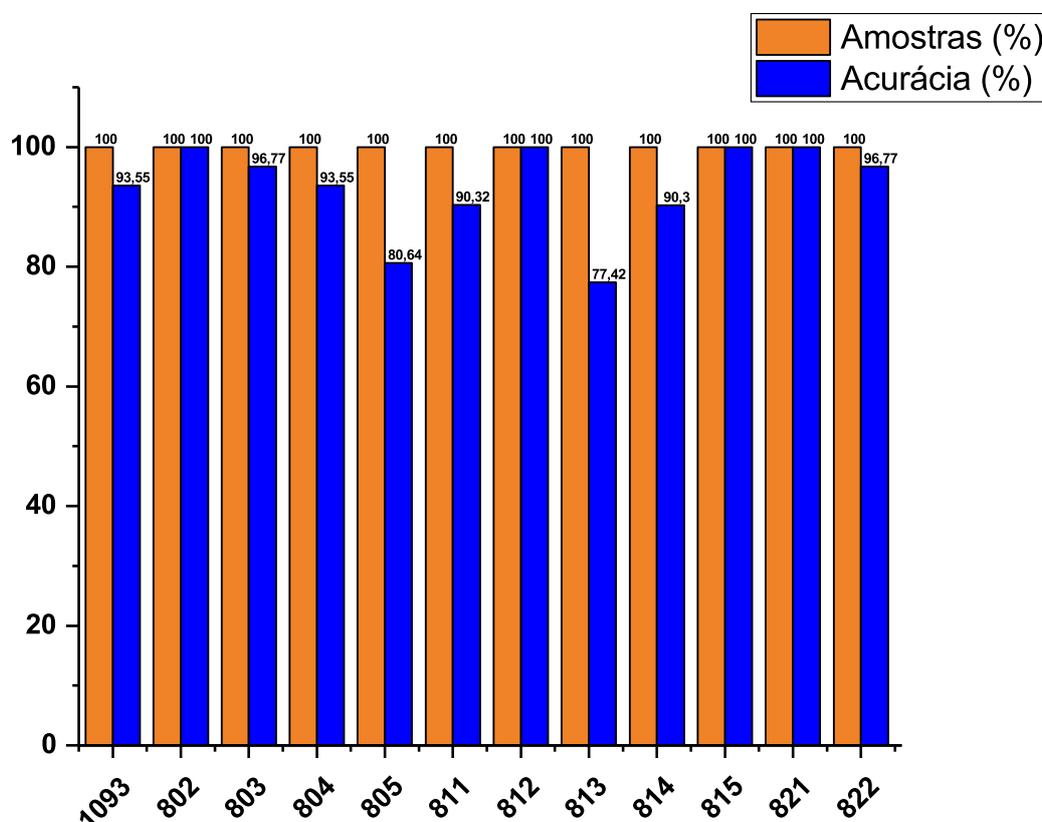
Figura 31 – Percentual de marcas classificadas corretamente e percentual de erros - Experimento II.



Fonte: do próprio autor.

A Figura 32 mostra um gráfico comparativo entre as marcas classificadas corretamente em função do número de amostras utilizadas na validação.

Figura 32 – Proporção de marcas corretamente classificadas em função do número de amostras - Experimento II.



Fonte: do próprio autor.

As barras em azul apresentam as marcas de gado classificadas corretamente (acurácia), e as barras em laranja apresentam o total de amostras de cada marca. Das 12 marcas analisadas, 4 apresentaram 100% de classificações corretas, no caso as marcas “802”; “812”; “815” e “821”. As marcas de gado que tiveram a menor taxa de acerto foram a “813” e “805”, com 24 e 25 imagens classificadas corretamente, respectivamente. Observa-se também o elevado índice de classificações corretas, mesmo nas marcas que não alcançaram 100% de acerto, como no caso das marcas “803”; “804”; “811”; “814”; “822” e “1093”.

Com base na análise da matriz de confusão apresentada na Figura 30, é possível observar que o método apresentou um valor relativamente baixo de Falsos Negativos (FN), ou seja, classificou corretamente quem deveria ser classificado, portanto teve uma taxa de Sensitividade alta. Além disso, é possível verificar também que o número de Falsos Positivos (FP) foi baixo, o que proporcionou uma taxa de Precisão significativa para todas as marcas classificadas.

O Coeficiente *Kappa*, que serve como métrica para avaliar o desempenho do método proposto em função das medidas de precisão e sensibilidade, obteve um índice

de 0,927.

A Tabela 7 apresenta os resultados obtidos pelo classificador com a utilização do Método de Redes Neurais Convolucionais. Foram utilizadas as métricas de Sensitividade, Especificidade e Precisão, respectivamente.

Tabela 7 – Sensitividade (TPR), Especificidade (TNR) e Precisão (PPV) - Experimento II

Marca	TPR	TNR	PPV
802	1,00000	1,00000	1,00000
803	1,00000	0,99708	0,96774
804	0,85294	0,99408	0,93548
805	0,92593	0,98261	0,80645
811	1,00000	0,99128	0,90323
812	0,81579	1,00000	1,00000
813	0,77419	0,97947	0,77419
814	1,00000	0,99128	0,90323
815	1,00000	1,00000	1,00000
821	0,96875	1,00000	1,00000
822	0,93750	0,99706	0,96774
1093	0,96667	0,99415	0,93548

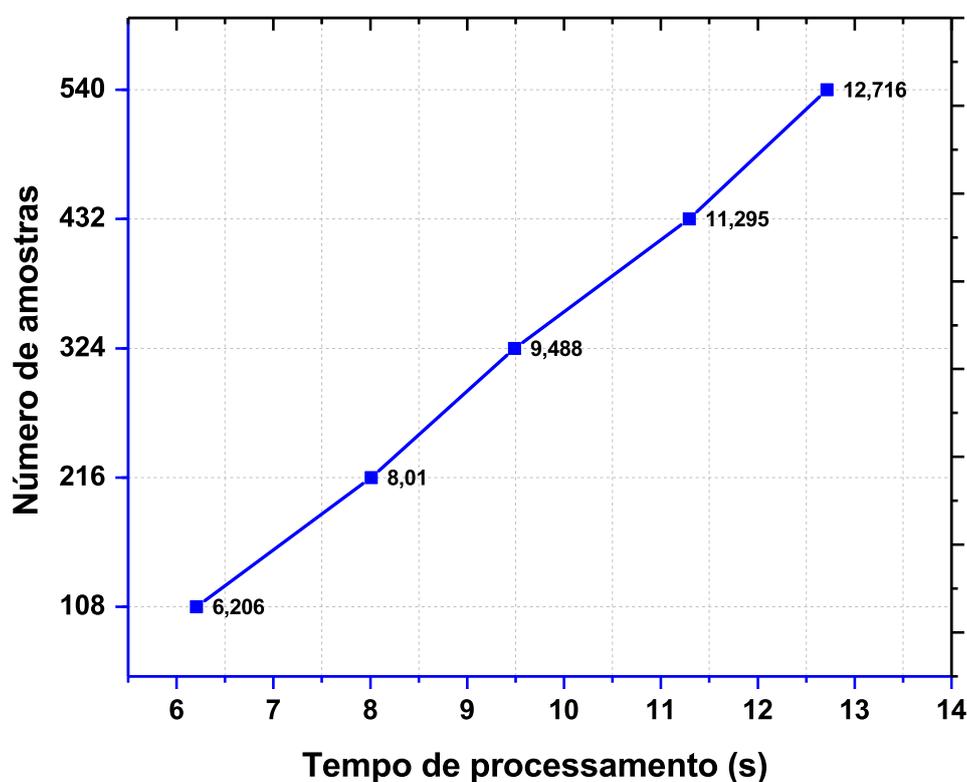
Fonte: do próprio autor.

5.2.2 DESEMPENHO DE PROCESSAMENTO

O tempo de processamento do algoritmo em função do número de amostras de marcas de gado é apresentado na Figura 33.

Foram medidos os tempos de processamento do método proposto para a classificação em cinco grupos de amostras. Cada grupo contendo 108, 216, 324, 432 e 540 imagens. Os tempos de processamento para classificação das imagens em cada grupo foram de 6,206s; 8,01s; 9,488s; 11,295s e 12,716s, nessa ordem.

Figura 33 – Tempo de processamento do algoritmo em função do número de amostras - Experimento II.



Fonte: do próprio autor.

Os resultados obtidos com o experimento realizado com 12 marcas de gado e 540 amostras de imagens, utilizadas tanto para treinamento quanto para validação, alcançaram taxa de reconhecimento (acurácia geral) de 93,28%, taxa de erro de 6,72% e tempo total de processamento de 12,716 segundos.

Na Tabela 8 são apresentados os resultados obtidos com o experimento utilizando o Método de Redes Neurais Convolucionais.

Tabela 8 – Acurácia Geral, Erro e Tempo de Processamento - Método CNN - Experimento II

Marcas	Amostras	Acurácia geral (%)	Taxa de erro (%)	Tempo de processamento (s)
12	540	93,28	6,72	12,716

Fonte: do próprio autor.

5.3 EXPERIMENTO III: MÉTODO CNN APLICADO À 12 MARCAS EM UM CONJUNTO DE 840 IMAGENS

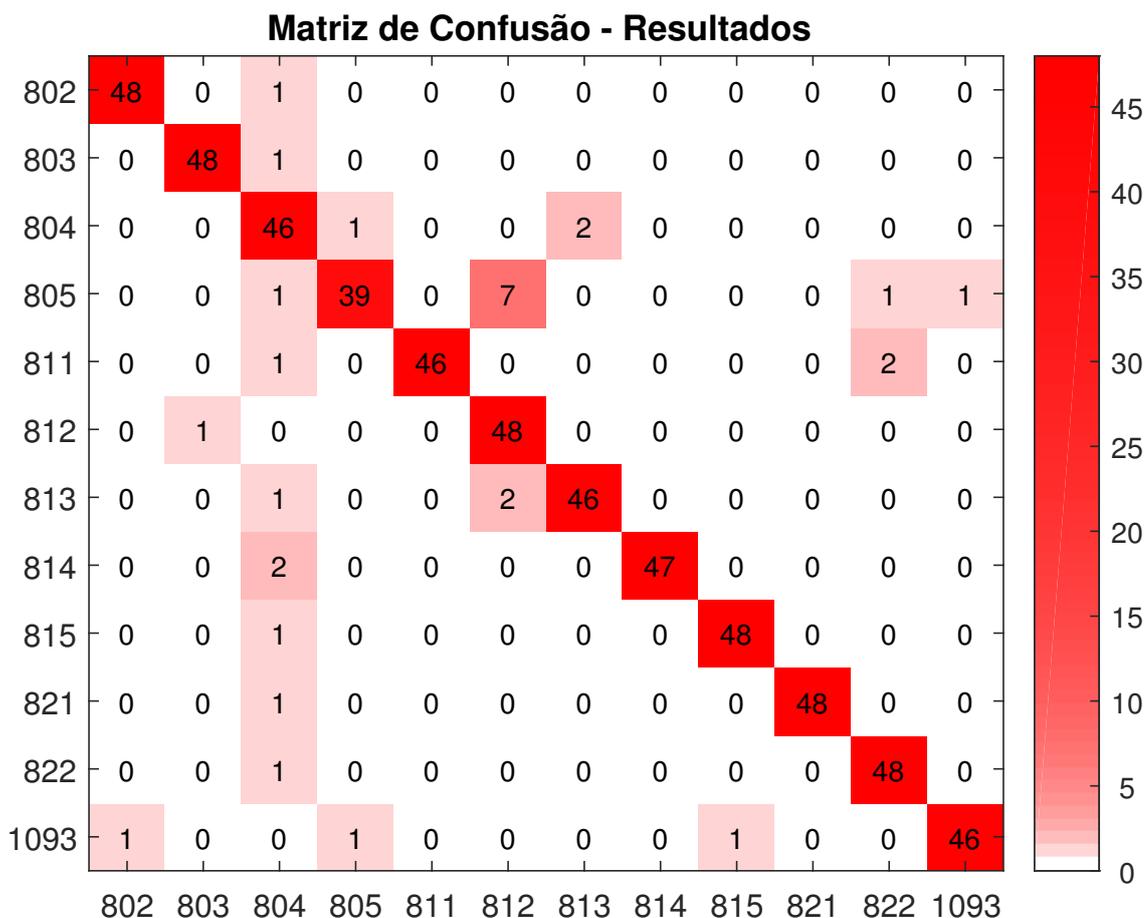
5.3.1 ACURÁCIA GERAL

Nesse experimento, foram utilizadas as mesmas 12 marcas de gado utilizadas no segundo experimento, porém utilizando um conjunto de 70 subimagens por marca, sendo 30% dessas utilizadas para treinamento, diferentemente das 45 subimagens utilizadas nos experimentos anteriores.

O objetivo desse experimento é demonstrar que com o aumento de número de subimagens utilizadas para extrair as características de uma marca na fase de treinamento contribui significativamente na performance da acurácia do método de Redes Neurais Convolucionais.

A Figura 34 apresenta a matriz de confusão para o melhor resultado obtido no experimento, cuja taxa de acurácia geral chegou a 94,90%.

Figura 34 – Matriz de confusão obtida na etapa de validação - Experimento III.



Fonte: do próprio autor.

A Tabela 9 apresenta a taxa de acurácia alcançada por cada uma das marcas avaliadas no experimento, na qual é possível verificar que 6 marcas obtiveram 97,959% de precisão, enquanto que somente uma marca ficou com percentual muito abaixo das demais, com 79,592% de acerto.

Tabela 9 – Acurácia por marca - Método CNN - Experimento III

Marca	Amostras	Acurácia (%)
802	70	97,959
803	70	97,959
804	70	93,878
805	70	79,592
811	70	93,878
812	70	97,959
813	70	93,878
814	70	95,918
815	70	97,959
821	70	97,959
822	70	97,959
1093	70	93,878

Fonte: do próprio autor.

O tempo de processamento do algoritmo para o conjunto de 840 imagens examinadas no experimento foi de 18,946 segundos.

5.4 EXPERIMENTO IV: MÉTODO CNN APLICADO À 500 MARCAS EM UM CONJUNTO DE 22.500 IMAGENS

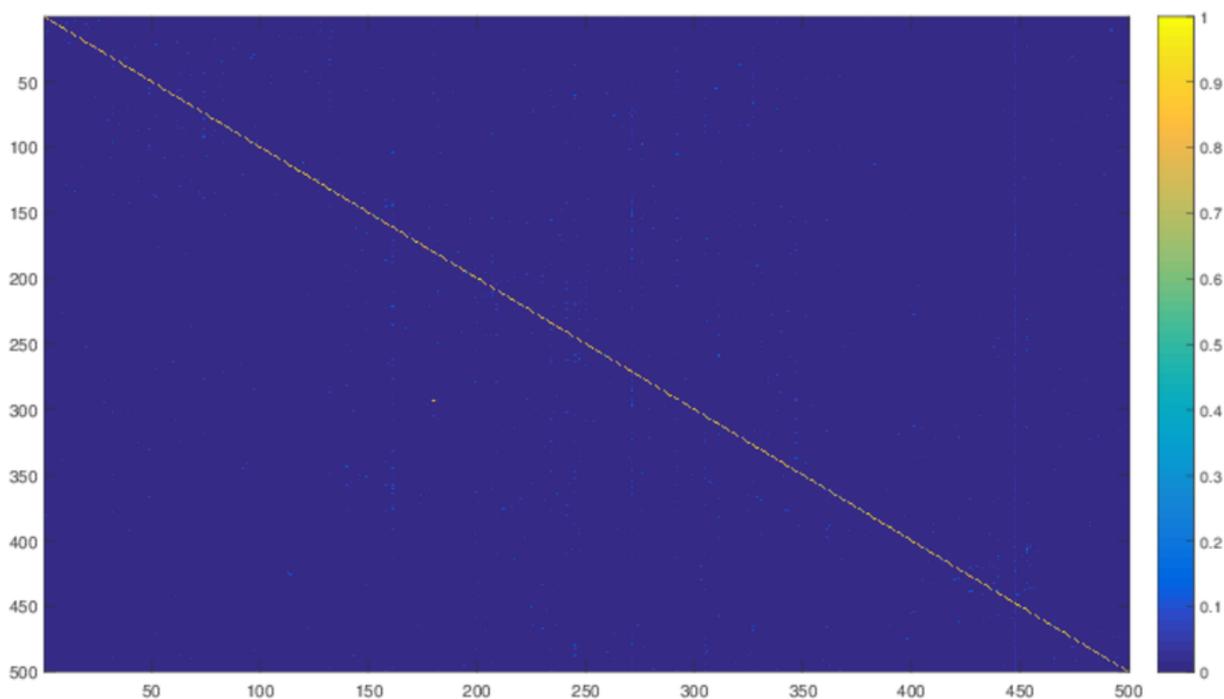
5.4.1 ACURÁCIA GERAL

No Experimento IV foram utilizadas 500 marcas as quais são apresentadas no Anexo A. Desse conjunto de marcas foram criadas artificialmente do mesmo modo dos Experimentos I e II, 45 subimagens para cada marca, aplicando efeitos de rotação, translação e escala, produzindo um total de 22.500 imagens que foram utilizadas para treinamento e validação.

A avaliação dos resultados do experimento foi realizada com base na taxa de acurácia geral obtida da matriz de confusão gerada a partir da classificação realizada na etapa de validação. Além disso, o tempo total de processamento do método proposto também foi aferido.

A Figura 35 apresenta a matriz de confusão para o melhor resultado obtido nas simulações, cujo percentual de acurácia geral alcançou 80,57% de precisão.

Figura 35 – Matriz de confusão obtida na etapa de validação - Experimento IV.



Fonte: do próprio autor.

Devido ao tamanho da matriz de confusão ser de ordem 500, a melhor maneira de ilustrá-la graficamente foi através da função *colorbar* do software MatLab, a qual realiza a plotagem de gráficos em 2D no qual é mostrado um mapa de cores na forma de uma barra (escala) de cores.

Na Figura 35 a diagonal principal da matriz, que corresponde aos percentuais de acurácia de cada marca, está representada pela cor “amarela”, que segundo a escala da *colorbar* vai até o número “1”. Já os pontos em “azul mais claro” observados na matriz correspondem aos erros de classificação. Quanto mais “claros” esses pontos em azul maiores são os erros cometidos na classificação das marcas, em contrapartida, segundo a colocar, quanto mais próximo do “0”, mais escuro o tom de azul será, significando que naquela entrada da matriz não houve erro de classificação.

Os eixos “x” e “y” dessa matriz correspondem às marcas utilizadas no experimento, no entanto como é possível apresentar nominalmente os códigos da 500

marcas pelos motivos já apresentados, as quantidades de marcas examinadas são apresentadas sob a forma de uma escala que vai de “0” até “500”.

A Tabela 10 apresenta as taxas de acurácia geral e tempos de processamento do algoritmo obtidos em 15 simulações realizadas no Experimento IV.

Tabela 10 – Acurácia Geral e Tempo de processamento - Método CNN - Experimento IV

Simulação	Acurácia Geral (%)	Tempo de processamento (s)
1 ^a	79,60	543,466
2 ^a	79,49	550,505
3 ^a	78,74	480,648
4 ^a	79,26	554,832
5 ^a	79,36	547,823
6 ^a	79,47	547,501
7 ^a	79,72	539,290
8 ^a	79,93	550,629
9 ^a	79,30	481,262
10 ^a	79,35	479,826
11 ^a	79,46	480,780
12 ^a	79,83	480,261
13 ^a	80,57	478,995
14 ^a	80,09	470,105
15 ^a	79,66	468,108

Fonte: do próprio autor.

5.5 DISCUSSÕES

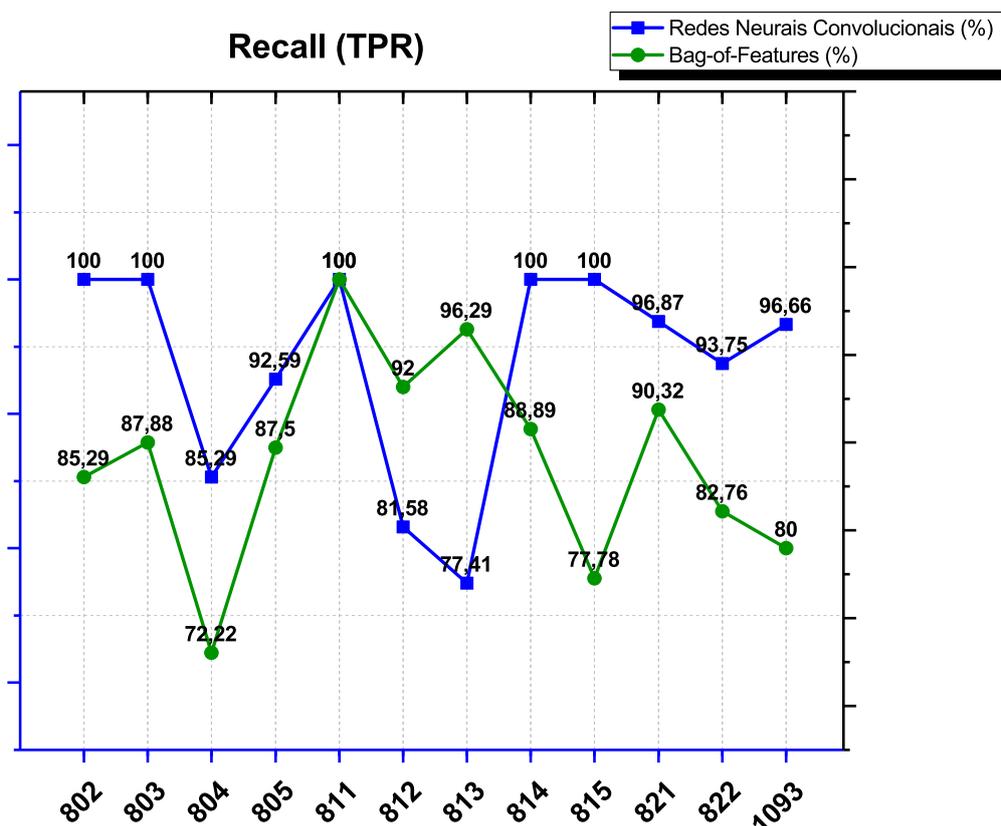
Neste capítulo, são apresentadas as análises dos dados coletados nos experimentos realizados. Percebe-se que ambos os métodos propostos para reconhecimento das marcas obtiveram resultados satisfatórios tanto no percentual de reconhecimento das marcas de gado quanto no tempo de processamento levado para processamento do algoritmo.

Na comparação das métricas adotadas para avaliação dos métodos, observou-se o método de Redes Neurais Convolucionais conseguiu alcançar resultados superiores em relação ao método *Bag-of-Features*. Isto acontece principalmente quando se analisa a acurácia geral, taxa de erro, coeficiente *Kappa* e tempo de processamento dos métodos, conforme apresentado na Tabela 9.

Métricas de processamento de imagens, como sensibilidade (*recall*) e precisão também foram analisadas para definir a confiabilidade dos métodos utilizados para a arquitetura proposta.

A Figura 36 apresenta o gráfico comparativo do *recall* (TPR) dos métodos CNN e BoF, no qual é possível observar que o primeiro obteve os melhores índices, ficando com desempenho inferior somente nas marcas “813”; com 77,41%, e “812”; com 81,58%.

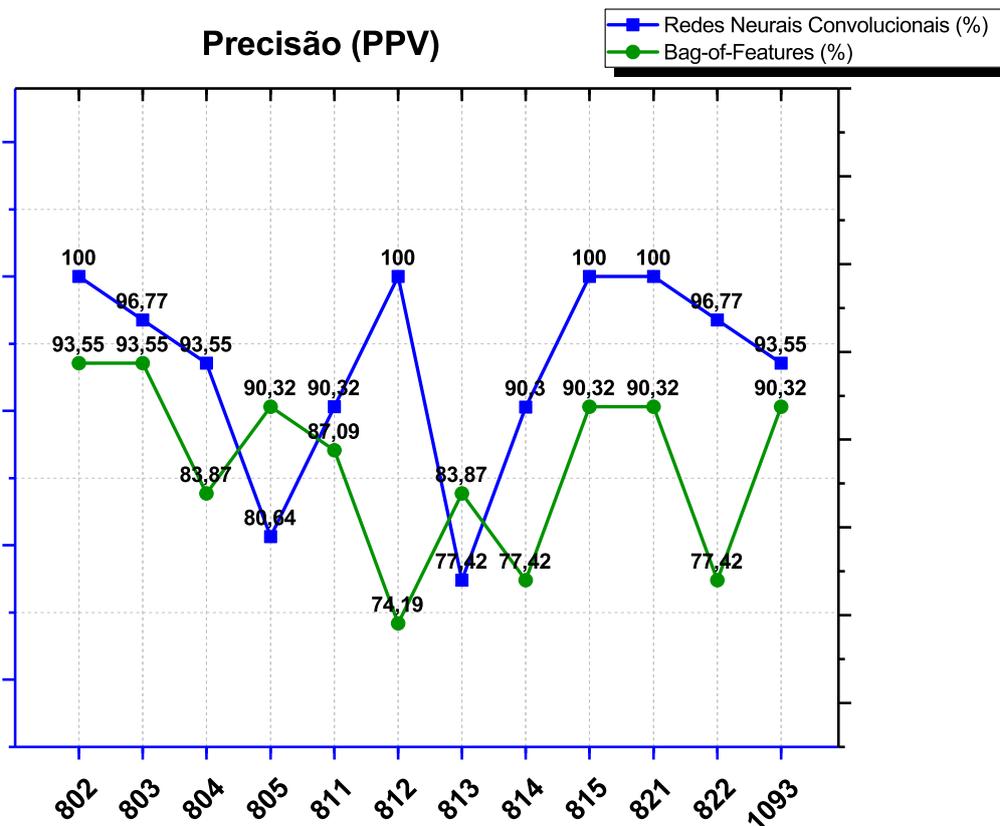
Figura 36 – Recall comparado dos métodos CNN e BoF - Experimentos I e II.



Fonte: do próprio autor.

A Figura 37 apresenta o gráfico comparativo da precisão (PPV) dos métodos CNN e BoF. Analisando o gráfico é possível verificar que o método CNN obteve os melhores índices, assim como no *recall*, porém ficando com desempenho inferior nas marcas “813”; com 77,42%, e “805”; com 80,64%.

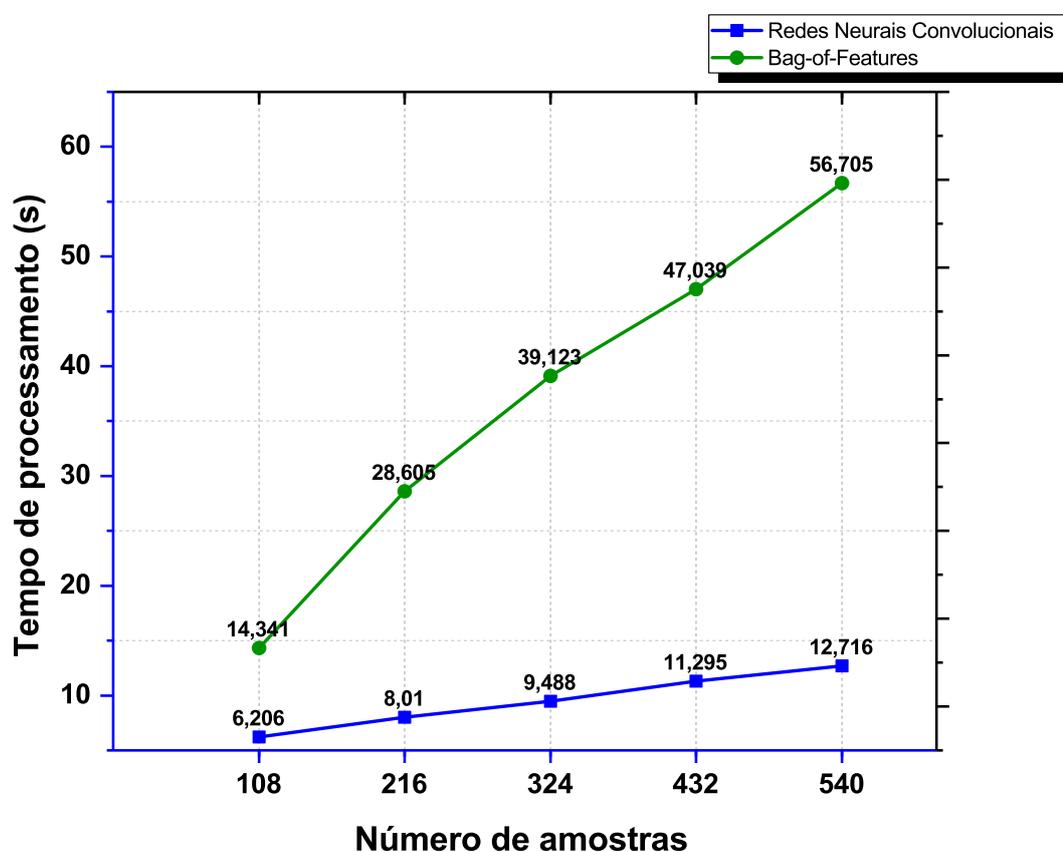
Figura 37 – Precisão comparada dos métodos CNN e BoF - Experimentos I e II.



Fonte: do próprio autor.

A Figura 38 ilustra o gráfico comparativo do tempo de processamentos dos métodos CNN e BoF. Com ele, é possível constatar que o método CNN obteve os melhores tempos em todos os experimentos realizados, chegando a uma média de 27,6196s mais rápido que o método BoF, levando em consideração os experimentos realizados com cinco grupos de amostras, com cada grupo contendo 108, 216, 324, 432 e 540 imagens, nessa ordem.

Figura 38 – Análise de tempo de processamento dos métodos CNN e BoF - Experimentos I e II.



Fonte: do próprio autor.

A acurácia geral, a taxa de erro, o coeficiente *Kappa* e o tempo de processamento da arquitetura de acordo com os métodos utilizados são apresentados na Tabela 11.

Tabela 11 – Resultados comparativos dos métodos propostos para a arquitetura desenvolvida - Experimentos I e II.

	Acurácia geral (%)	Taxa de erro (%)	Coeficiente <i>Kappa</i>	Tempo de processamento (s)
BoF	86,02	13,98	0,848	56,705
CNN	93,28	6,72	0,927	12,716

Fonte: do próprio autor.

A acurácia geral obtida pela arquitetura proposta foi significativa, uma vez que os métodos BoF e CNN alcançaram, respectivamente, as taxas de 86,02% e 93,28% no

reconhecimento das marcas de gado.

O Coeficiente *Kappa* dos métodos BoF e CNN atingiu os índices de 0,848 e 0,927, nessa ordem, que segundo a escala proposta por LANDIS e KOCH (1977), corresponde a uma concordância quase perfeita, o que confirma o desempenho e confiabilidade das classificações realizadas pela arquitetura proposta.

O tempo de processamento é outro fator fundamental, pois está associado à eficiência da arquitetura no que se refere ao reconhecimento das marcas, e ambos os métodos alcançaram resultados satisfatórios, principalmente considerando o número total de marcas de gado utilizadas nos experimentos, o que comprova a capacidade dos métodos de processar grandes volumes de registros a um baixo custo computacional.

Com base nos resultados apresentados na Tabela 11, é possível verificar que o método de Redes Neurais Convolucionais obteve os melhores resultados, tanto na taxa geral de reconhecimento das marcas de gado (7,26% mais preciso), quanto na velocidade de processamento do algoritmo (43,989s mais rápido).

Com relação ao Experimento III, houve uma melhoria na taxa geral de acurácia da arquitetura proposta, alcançando 94,90%. Conforme apresentado na Tabela 9 houve um aumento no percentual de acerto de sete marcas em relação ao Experimento II. Através da análise desses experimentos é possível constatar que o acréscimo de subimagens para representação das marcas contribuiu significativamente quanto a extração de características que melhor representassem essas marcas, dessa forma gerando resultados mais precisos e percentuais de acurácia geral mais elevados.

No Experimento IV, em que foram utilizadas 500 marcas e 45 subimagens por marca, totalizando 22.500 imagens para treinamento e validação, a acurácia geral atingiu 80,57%. A razão pela qual houve uma diminuição dessa taxa foi pelo acréscimo do número de marcas tomadas para análise e à complexidade das mesmas, pois com o incremento de imagens aumento também a probabilidade de marcas com características análogas entre si, o que causou uma dificuldade maior da arquitetura proposta em extrair e classificar essas marcas corretamente.

Outro fator determinante para redução da habilidade da arquitetura em reconhecer corretamente as marcas foi a qualidade das novas marcas tomadas para o experimento, e a presença de ruídos na mesmas, pois não foi foco desse trabalho realizar qualquer tipo de pré-processamento nessas imagens antes da aplicação da CNN para extração das características das marcas.

As marcas que apresentaram menor poder descritivo e qualidade inferior, como explicado anteriormente, tiveram percentuais de classificações muito baixos, geralmente entre 60% a 70%, conforme apresentado no Anexo E, o que impactou diretamente na redução do percentual de acurácia geral do Experimento IV.

Os resultados de acurácia geral e Taxa de erro obtidos nos quatro experimentos são apresentados na Tabela 12.

Tabela 12 – Resultados comparativos de Acurácia Geral e Taxa de erro dos Experimentos I, II, III e IV.

	Acurácia geral (%)	Taxa de erro (%)
Experimento I	86,02	13,98
Experimento II	93,28	6,72
Experimento III	94,90	5,10
Experimento IV	80,57	19,43

Fonte: do próprio autor.

De acordo com os resultados apresentados na Tabela 12 é possível verificar que nos Experimentos I e II, em que foram utilizadas 12 marcas e um total de 540 imagens, o método de Redes Neurais Convolucionais obteve melhores resultados em relação ao método *Bag-of-Features*.

Também é possível concluir que um número maior de subimagens para representar a marca aumenta a taxa de acurácia geral, pois melhor o poder descritivo dessa marca na fase treinamento, como no caso do Experimento III, em que foram utilizadas 70 subimagens para cada marca, em contraposição ao Experimento I e II que foram utilizadas apenas 45 subimagens.

Por fim, o Experimento IV apresentou uma redução considerável na taxa de acurácia da arquitetura proposta o que pode ser explicado pelo aumento do número de marcas analisadas, a qualidade dessas novas marcas, à similaridade de algumas marcas, e principalmente à limitação própria do método CNN, que será discutida na Seção a seguir.

5.5.1 LIMITAÇÕES DO MÉTODO PROPOSTO

O principal motivo da redução da taxa de reconhecimento das marcas por parte da arquitetura proposta é a limitação inerente do método de Redes Neurais Convolucionais, que necessita de grandes quantidades de amostras de imagens para extração de características e correta classificação dessas marcas.

Mesmo utilizando uma rede pré-treinada a melhoria da precisão do método está condicionada ao número de imagens utilizadas para extração de atributos que serão utilizadas posteriormente na fase de treinamento, além da qualidade das imagens também ser uma particularidade relevante para se alcançar resultados mais satisfatórios.

Quanto maior o número de marcas examinadas maior deve ser o conjunto de amostras para cada marca, caso contrário a acurácia tende a cair drasticamente. Porém,

esse incremento de imagens de amostra também pode acarretar outras dificuldades, pois no Experimento IV foram utilizadas 500 marcas e um total de 22.500 imagens para treinamento e validação da arquitetura proposta, no entanto, em situações reais esse número é significativamente maior, por exemplo, somente a Prefeitura de São Francisco de Assis são 11 mil marcas.

Levando em conta que para cada marca seria necessário um conjunto de subimagens as quais seriam utilizadas para extração de características pela CNN, o número total de imagens para que o método atingisse bons resultados seria muito alto, o que elevaria o custo computacional para o problema em questão, ou até mesmo impactar o tempo para consulta de uma marca em um banco de dados com muitas imagens.

No próximo capítulo, são realizadas as considerações finais e trabalhos futuros.

6 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

Este trabalho surgiu com o intuito de desenvolver uma arquitetura capaz de reconhecer imagens de marcas de gado para órgãos governamentais responsáveis pelo registro e controle dessas marcas, visto que atualmente este processo é realizado de forma manual e sem nenhum sistema computacional especializado para realização desta tarefa.

Para atingir este objetivo, foi realizado um levantamento bibliográfico acerca de processamento e reconhecimento de imagens digitais, assim como um estudo sobre os trabalhos relacionados na literatura, de forma que se averiguasse quais métodos eram mais adequados para a solução do problema de pesquisa.

Após esse levantamento foram selecionados dois métodos para extração de características, o *Bag-of-Features* e Redes Neurais Convolucionais, com o intuito de se propor um método para a arquitetura proposta que tivesse a melhor taxa de reconhecimento das marcas, com menor tempo de processamento e baixo custo computacional. Para a classificação das marcas, em ambos os métodos, utilizou-se a técnica supervisionada de Máquinas de Vetores de Suporte.

A pesquisa foi desenvolvida entre duas instituições; a Prefeitura Municipal de São Francisco de Assis e Universidade Federal do Pampa. No capítulo de materiais e métodos foi apresentado o conjunto de imagens utilizado nos experimentos realizados nesta pesquisa. Foram avaliadas inicialmente 12 marcas de gado e utilizado um total de 540 amostras para treino e validação.

Para o método *Bag-of-Features* o algoritmo SURF foi utilizado para extração de pontos de interesse das imagens e para criação do agrupamento de palavras visuais foi utilizado o *clustering K-means*. O método apresentou uma acurácia geral (86,02%) e tempo de processamento do algoritmo de (56,705s) para um conjunto de 540 imagens.

No método de Rede Neural Convolucional foi criada uma CNN completa com 5 camadas convolucionais, e 3 camadas totalmente conectadas. A 1ª camada convolucional tem como entrada imagens transformadas para o formato de cores RGB. Nessa CNN foi utilizada a função de ativação ReLU e a técnica de *maxpooling* para redução. Os resultados foram expressivos, acurácia geral (93,28%) e tempo de processamento do algoritmo (12,716s) quando aplicado a 12 marcas em um conjunto de 540 imagens.

Além da acurácia geral e do tempo de processamento, os métodos também foram avaliados sob as métricas de *recall*, precisão e coeficiente *Kappa*. Com base nos resultados dos Experimentos I e II, o método que utiliza Redes Neurais Convolucionais obteve os melhores resultados, tanto na taxa geral de reconhecimento das marcas de gado (7,26% mais preciso), quanto na velocidade de processamento do algoritmo (43,989s mais rápido).

O método CNN realizou eficaz e eficientemente o reconhecimento de diferentes marcas de gado, mesmo utilizando uma CNN pré-treinada, porém sua principal limitação é a necessidade de uma grande quantidade de imagens de amostra para treinamento do classificador, uma vez que a Rede Neural Convolucional necessita dessas imagens para extração de características nas camadas convolucionais. Essa limitação foi evidenciada nos Experimentos III e IV, que alcançaram taxas de acurácia geral de 94,90% e 80,57%, respectivamente.

Para apresentar à comunidade científica as ideias e resultados que foram obtidos no decorrer deste trabalho, alguns artigos científicos foram produzidos e submetidos para revisão em periódicos e simpósios internacionais. Foram publicados dois artigos, sendo um em periódico e um em congresso internacional. Estes artigos são listados a seguir:

- 1) SILVA, C. ; WELFER, D. ; GIODA, F. P. ; DORNELLES, C. . *Cattle Brand Recognition using Convolutional Neural Network and Support Vector Machines*. IEEE Latin America Transactions, v. 15, n.2, p. 310-316, 2017.
- 2) SILVA, C. ; WELFER, D. ; WEBER, J. . *Deep Learning Techniques Applied to the Cattle Brand Recognition*. In: 22nd Iberoamerican Congress on Pattern Recognition (CIARP), 2017, Valparaíso, Chile. Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 22nd Iberoamerican Congress on Pattern Recognition, CIARP 2017, (Lecture Notes in Computer Science), 2017.

Como trabalho futuro objetiva-se a ampliação do conjunto de imagens tomado para treinamento e validação, com o intuito de se obter novas medições de acurácia e desempenho da arquitetura proposta, visando ao desenvolvimento de um protótipo e a disponibilização de um sistema devidamente validado e consolidado para os órgãos governamentais responsáveis pelo registro e controle das marcas de gado. Em adição, sugere-se também o desenvolvimento de uma interface gráfica amigável para a ferramenta, de forma a facilitar a operacionalização por parte dos usuários que farão o uso da mesma.

A arquitetura de CNN proposta nesse trabalho apresentou a capacidade de trabalhar com problemas não-lineares, boa adaptabilidade e generalização, alto desempenho, tanto em nível de velocidade de execução quanto em nível de acerto dos resultados, de maneira que é possível aplicá-la na resolução de problemas em outras áreas do conhecimento, como no processamento e análise de imagens médicas, em que são desenvolvidos métodos visando auxiliar a elaboração de diagnósticos de forma a contribuir para a detecção precoce de doenças, como no reconhecimento de lesões na retina, ou também na robótica móvel, em que a arquitetura de redes neural convolucional pode ser utilizada para extração de informações de imagens para realizar o controle de um robô através do aprendizado por reforço.

REFERÊNCIAS

- ALENCAR, F. *Desenvolvimento de um Sistema para Classificação Automática de Imagens Dermatoscópicas para Dispositivos Móveis*. 2015. Dissertação (Mestrado em Ciência da Computação) — Universidade do Estado do Rio Grande do Norte e a Universidade Federal Rural do Semi-Árido. Disponível em: <<https://ppgcc.ufersa.edu.br/wp-content/uploads/sites/42/2014/09/francisco-eduardo-silva-alencar.pdf>>. Acesso em: 23/10/2016. Citado na página 34.
- ARNONI, R. Os Registros e Catálogos de Marcas de Gado da Região Platina. Pelotas. *Revista Memória em Rede da UFPEL*, Programa de Pós-Graduação em Memória Social e Patrimônio Cultural, v. 5, n. 9, 2013. ISSN 2177-4129. Disponível em: <<https://periodicos.ufpel.edu.br/ojs2/index.php/Memoria/article/view/9460/6208>>. Acesso em: 10/10/2015. Citado na página 18.
- BARATA, C.; MARQUES, J.; MENDONÇA, T. Bag-of-features classification model for the diagnose of melanoma in dermoscopy images using color and texture descriptors. In: *10th International Conference, ICIAR 2013, Póvoa do Varzim, Portugal, June 26-28. Proceedings*. [S.l.: s.n.], 2013. v. 7950, p. 555 – 555. Citado 2 vezes nas páginas 47 e 50.
- BATISTA, N.; LOPES, A.; ARAÚJO, A. Vocabulários visuais aplicados à detecção de edifícios em fotografias históricas. In: *XXXV Conferência Latinoamericana de Informática (Latin-American Conference on Informatics), CLEI, Pelotas, RS, Brazil*. [S.l.: s.n.], 2009. Citado 2 vezes nas páginas 47 e 50.
- BENGIO, Y. Learning deep architectures for AI. *Foundations and Trends® in Machine Learning*, Now publishers inc., v. 2, n. 1, p. 1 – 127, January 2009. Disponível em: <<http://www.nowpublishers.com/article/Details/MAL-006>>. Acesso em: 05/11/2016. Citado na página 35.
- BENGIO, Y.; COURVILLE, A.; VINCENT, V. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, p. 35(8):1798–1828 –, 2013. Citado na página 34.
- BEZERRA, E. Introdução à Aprendizagem Profunda. In: UNIVERSIDADE FEDERAL DA BAHIA, 2016, Salvador. *31º Simpósio Brasileiro de Banco de Dados*. 2016. Disponível em: <<http://sbbd2016.fpc.ufba.br/sbbd2016/minicursos/minicurso3.pdf>>. Acesso em: 07/01/2017. Citado 4 vezes nas páginas 37, 38, 39 e 40.
- CAICEDO, J.; CRUZ, A.; GONZALEZ, F. Histopathology image classification using bag of features and kernel functions. *Springer*, Berlin/Heidelberg, v. 5651, p. 126 – 135, 2009. Citado 2 vezes nas páginas 25 e 29.
- CIREGAN, D.; MEIER, U.; SCHMIDHUBER, J. Multi-Column Deep Neural Networks for Image Classification. In: *Proceedings of the 25th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2012)*. [S.l.: s.n.], 2012. p. 3642 – 3649. Citado 2 vezes nas páginas 36 e 50.
- COHEN, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, v. 20, n. 1, p. 37 – 46, 1960. Citado na página 44.

- COLGATON, R. A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sensing of Environment*, v. 49, n. 12, p. 1671 – 1678, 1991. Citado na página 42.
- CONCI, A.; AZEVEDO, E.; LETA, F. R. *Computação Gráfica: teoria e prática*. 1. ed. Rio de Janeiro: Elsevier, 2007. v. 2. ISBN 978-85-352-2329-3. Citado na página 25.
- CONSTANTE, P. et al. Artificial Vision Techniques for Strawberry's Industrial Classification. *IEEE Latin America Transactions*, v. 14, n. 6, p. 2576 – 2581, 2016. Citado 3 vezes nas páginas 49, 50 e 64.
- CSURKA, G. et al. Visual categorization with bags of keypoints. In: *Workshop on Statistical Learning in Computer Vision, ECCV*. [S.l.: s.n.], 2004. p. 1 – 22. Citado 3 vezes nas páginas 25, 29 e 47.
- DARDAS, N.; CHEN, Q.; GEORGANAS, N. Hand gesture recognition using bag-of-features and multi-class support vector machine. In: *Proceedings of IEEE International Symposium on Haptic Audio-Visual Environments and Games*. [S.l.: s.n.], 2010. p. 1 – 5. Citado 2 vezes nas páginas 47 e 50.
- DENG, L.; DONG, Y. Deep learning: Methods and applications. *Foundations and Trends in Signal Processing*, v. 7, n. 3-4, p. 197 – 387, 2014. Citado na página 35.
- ERHAN, D. et al. Why Does Unsupervised Pre-training Help Deep Learning? *The Journal of Machine Learning Research*, JMLR.org, v. 11, p. 625 – 660, 2010. ISSN 1532-4435. Disponível em: <<http://dl.acm.org/citation.cfm?id=1756006.1756025>>. Citado na página 39.
- FACELI, K. et al. *Inteligência Artificial: uma abordagem de aprendizagem de máquina*. 1. ed. Rio de Janeiro: LTC, 2011. ISBN 978-8521618805. Citado 2 vezes nas páginas 32 e 34.
- FERRAZ, C. *Novos descritores de textura para localização e identificação de objetos em imagens usando Bag-of-Features*. 2016. Tese (Doutorado em Processamento de Sinais de Instrumentação) — Escola de Engenharia de São Carlos, Universidade de São Paulo. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/18/18152/tde-28092016-141219/pt-br.php>>. Acesso em: 21/12/2016. Citado na página 59.
- FIOL, G. D.; HAUG, P. J. Classification models for the prediction of clinicians' information needs. *Journal of Biomedical Informatics*, v. 42, n. 1, p. 82 – 89, 2009. Citado na página 44.
- GIL, A. *Reconhecimento de Dígitos Manuscritos: busca de um Classificador com Máxima Taxa de Acerto*. 2014. Dissertação (Mestrado em Engenharia Elétrica) — Universidade Federal do Amazonas. Disponível em: <<http://tede.ufam.edu.br/handle/tede/4428>>. Acesso em: 01/12/2016. Citado 4 vezes nas páginas 31, 35, 37 e 40.
- GONZALEZ, R.; WOODS, R. *Processamento digital de imagens*. 3. ed. São Paulo: Pearson Prentice Hall, 2010. Citado 2 vezes nas páginas 24 e 25.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Deep learning. *Book in preparation for MIT Press*, 2016. Disponível em: <www.deeplearningbook.org>. Acesso em: 30/12/2016. Citado na página 35.

HAGAN, M. T. et al. *Neural Network Design*. 2. ed. PWS PUB., 1995. ISBN 978-0-9717321-1-7. Disponível em: <<http://hagan.okstate.edu/NNDesign.pdf>>. Acesso em: 05/11/2016. Citado na página 31.

HARRINGTON, P. *Machine Learning in Action*. 1. ed. [S.l.]: Shelter Island: Manning Publications, 2012. ISBN 9781617290183. Citado 2 vezes nas páginas 40 e 41.

HAYKIN, S. *Redes Neurais: Princípios e Práticas*. 2. ed. Porto Alegre: Bookamn, 2001. Citado na página 31.

HINTON, G. To Recognize Shapes, First Learn to Generate Images. In: *Progress in Brain Research*. [S.l.]: Elsevier, 2007. v. 165, p. 535 – 547. Citado 2 vezes nas páginas 39 e 40.

ITAMAR, A.; ROSE, D.; KARNWSKI, T. Research frontier: Deep machine learning—a new frontier in artificial intelligence research. *Comp. Intell. Mag*, v. 5, n. 4, p. 13 – 18, Novembro 2010. Citado na página 35.

JARRETT, K.; KAVUKCUOGLU, K.; LECUN, Y. What Is The Best Multi-Stage Architecture for Object Recognition? In: IEEE, 2009. *12th International Conference on Computer Vision (ICCV)*. [S.l.], 2009. p. 2146 – 2153. Citado 2 vezes nas páginas 38 e 48.

JIANG, Y. G.; YANG, J.; NGO, C. W. Representations of keypoint-based semantic concept detection: a comprehensive study. *IEEE Transactions on Multimedia*, IEEE, v. 12, n. 1, p. 42 – 53, 2009. Citado na página 30.

JURASZEK, G. *Reconhecimento de Produtos por Imagem Utilizando Palavras Visuais e Redes Neurais Convolucionais*. 2014. Dissertação (Mestrado em Computação Aplicada) — Joinville: UDESC. Disponível em: <<http://tede.udesc.br/handle/handle/2045>>. Acesso em: 03/03/2015. Citado 3 vezes nas páginas 39, 48 e 50.

KAVUKCUOGLU, K. et al. Learning Convolutional Feature Hierarchies for Visual Recognition. In: *Advances in Neural Information Processing Systems*. [S.l.]: ed. by J.D Lafferty and C.K.I. Williams and J. Shawe-Taylor and R.S. Zemel and A. Culotta, 2010. v. 23, p. 1090 – 1098. Citado na página 48.

KOHAVI, R.; PROVOST, F. Glossary of Terms. *Machine Learning*, v. 30, n. (2-3), p. 271 – 274, 1998. Citado na página 42.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. Imagenet Classification with Deep Convolutional Neural Networks. In: PEREIRA C. J. C. BURGESS, L. B. F.; WEINBERGER, K. Q. (Ed.). *Neural Information Processing Systems (NIPS)*. [S.l.]: Curran Associates, Inc., 2012. p. 1097 – 1105. Citado na página 65.

LANDIS, J.; KOCH, G. The Measurement of Observer Agreement for Categorical Data. *International Biometric Society*, v. 33, n. 1, p. 159 – 174, 1977. Citado 2 vezes nas páginas 44 e 89.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, n. 521, p. 436 – 444, 2015. Citado na página 35.

- LECUN, Y. et al. Handwritten Digit Recognition with a Back-Propagation Network. In: *Advances in Neural Information Processing Systems*. [S.l.]: Morgan Kaufmann, 1990. p. 396 – 404. Citado 2 vezes nas páginas 48 e 50.
- LECUN, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, v. 86, n. 11, p. 2278 – 2324, November 1998. Citado 2 vezes nas páginas 35 e 36.
- LI, B. Building pattern classifiers using convolutional neural networks. In: *Neural Networks. IJCNN '99. International Joint Conference on*. [S.l.: s.n.], 1999. p. 3081 – 3085. Citado na página 35.
- LI, K.; WANG, F.; ZHANG, L. A new algorithm for image recognition and classification based on improved bag of features algorithm. In: *Optik*. [S.l.: s.n.], 2016. v. 127, p. 4736 – 4740. Citado 2 vezes nas páginas 47 e 50.
- LI, Z.; IMAI, J.; KANEKO, M. Robust face recognition using block-based bag of words. In: 20TH INTERNATIONAL CONFERENCE, 2010. *Pattern Recognition (ICPR)*. [S.l.]: IEEE, 2010. p. 1285 – 1288. Citado 2 vezes nas páginas 47 e 50.
- LOPES, A. et al. Nude detection in vídeo using bag-of-visual-features. In: SIBGRAPI, 2009. *Proceedings of the 2009 XXII Brazilian Symposium on Computer Graphics and Image Processing*. [S.l.]: IEEE Computer Society, 2009. p. 224 – 231. Citado 2 vezes nas páginas 47 e 50.
- LORENA, A.; CARVALHO, A. Uma Introdução às Support Vector Machines. *Revista de Informática Teórica e Aplicada*, v. 14, n. 2, p. 43 – 67, 2007. Citado na página 41.
- LOWE, D. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, v. 60, n. 2, p. 91 – 110, 2004. Citado na página 28.
- LU, H. et al. Real-Time Facial Expression Recognition Based on Pixel Pattern-Based Texture Feature. *Proc. Electronic Letters*, p. 916 – 918, 2007. Citado na página 62.
- MACQUEEN, J. Some methods for classification and analysis of multivariate observations. In: *Fifth Berkeley Symposium on Mathematical Statistics and Probabilities. Proceedings*. [S.l.]: FIFTH BERKELEY SYMPOSIUM ON MATHEMATICAL STATISTICS AND PROBABILITIES. Proceedings, 1967. p. 281 – 296. Citado na página 29.
- MANNING, C. D.; RAGHAVAN, P.; SCHÜTZE, H. *Introduction to information retrieval*. 2008. Disponível em: <<http://www.math.unipd.it/~aiolli/corsi/0910/IR/irbookprint.pdf>>. Acesso em: 31/04/2016. Citado na página 44.
- MATOS, P. et al. *Projeto “Um Ambiente para Análise de Dados da Doença Anemia Falciforme”*. 2009. Disponível em: <<http://conteudo.icmc.usp.br/pessoas/tasparado/TechReportUFSCar2009a-MatosEtAl.pdf>>. Acesso em: 08/01/2017. Citado 2 vezes nas páginas 42 e 44.
- NOWAK, E.; JURIE, F.; TRIGGS, B. Sampling strategies for bag-of-features image classification. In: *Computer Vision*. [S.l.]: 9th european conference on computer vision. [S.l.: s.n.], 2006. v. 3954, p. 490 – 503. Citado 2 vezes nas páginas 47 e 50.

NVIDIA. *Histórico da computação com GPU*. 2016. Disponível em: <http://www.nvidia.com.br/object/cuda_home_new_br.html>. Acesso em: 20/12/2016. Citado na página 54.

O'HARA, S.; DRAPER, B. Introduction to the bag of features paradigm for image classification and retrieval. *ArXiv*, p. 1101 – 3354, 2011. Citado na página 25.

OLIVEIRA, A. *Redes Neurais Artificiais Aplicadas na Detecção, Classificação e Localização de Defeitos em Linhas de Transmissão*. 2005. Dissertação (Mestrado em Engenharia Elétrica) — Universidade Federal de Juiz de Fora, Juiz de Fora. Disponível em: <<http://www.ufjf.br/ppee/files/2008/12/211047.pdf>>. Acesso em: 25/04/2016. Citado na página 33.

OLIVEIRA, T.; BARBAR, J.; SOARES, A. Predição do tráfego de rede de computadores usando redes neurais tradicionais e de aprendizagem profunda. *Revista de Informática Teórica e Aplicada - RITA*, Porto Alegre, v. 22, n. 1, 2015. Disponível em: <<http://www.seer.ufrgs.br/rita/article/view/RITA-VOL22-NR1-10>>. Acesso em: 25/03/2016. Citado na página 35.

OSUNA, E.; FREUD, R.; GIROSI, F. Training support vector machines: an application to face detection. In: *Computer Vision and Pattern Recognition*. [S.l.]: IEEE, 1997. p. 130 – 136. Citado na página 40.

PEDRINI, H.; SCHWARTZ, W. R. *Análise de Imagens Digitais: Princípios, Algoritmos e Aplicações*. São Paulo: THOMSON LEARNING, 2008. Citado 2 vezes nas páginas 23 e 24.

PRADO, R. *Implementação de uma Arquitetura Fuzzy Neural com Treinamento Online em Hardware Field Programmable Gate Array – FPGA*. 2014. Tese (Doutorado em Engenharia Elétrica) — Universidade Federal do Rio Grande do Norte, Natal. Disponível em: <<https://repositorio.ufrn.br/jspui/handle/123456789/19280>>. Acesso em: 05/03/2016. Citado na página 33.

RASCHKA, S. An Overview of General Performance Metrics of Binary Classifier Systems. *arXiv preprint*, p. 1 – 5, 2014. Citado na página 42.

RAZAVIAN, A. S. et al. CNN Features Off-the-Shelf: An Astounding Baseline for Recognition. *Computer Vision and Pattern Recognition Workshops (CVPRW)*, IEEE, p. 512 – 519, 2014. Citado 2 vezes nas páginas 49 e 50.

REIS, M. *Detecção de Caminhões em Monitoramento Automático de Trânsito*. 2015. Dissertação (Mestrado Profissional em Sistemas de Informação e Gestão do Conhecimento) — Universidade FUMEC, Belo Horizonte. Disponível em: <<http://www.fumec.br/revistas/sigc/article/view/3418>>. Acesso em: 03/06/2016. Citado na página 44.

RIBEIRO-NETO, B.; BAEZA-YATES, R. *Modern Information Retrieval*. 1. ed. [S.l.]: Addison-Wesley Professional, 1999. ISBN 0-201-39828-X. Citado na página 42.

ROCHA, R. *Reconhecimento de Objetos por Redes Neurais*. 2015. Dissertação (Mestrado em Informática) — PUC-Rio, Rio de Janeiro. Disponível em: <<http://www-di.inf.puc-rio.br/~milidui/inf2979/aulas/RafaelTese.pdf>>. Acesso em: 05/07/2016. Citado na página 52.

- SANCHEZ, G.; RODRIGUEZ, M. Cattle Marks Recognition by Hu and Legendre Invariant Moments. *ARPN Journal of Engineering and Applied Sciences*, v. 11, n. 1, p. 607 – 614, 2016. Citado 3 vezes nas páginas 46, 47 e 50.
- SCHMIDHUBER, J. Deep Learning in Neural Networks: an Overview. *Neural Networks*, v. 61, p. 85 – 117, 2015. Citado na página 35.
- SEPLAN. *Secretaria do Planejamento e Desenvolvimento Regional - Governo do Estado do Rio Grande do Sul, Brasil*. 2015. Disponível em: <http://www.atlassocioeconomico.rs.gov.br/conteudo.asp?cod_menu_filho=819&cod_menu=817&tipo_menu=ECONOMIA&cod_conteudo=1580>. Acesso em: 23/06/2015. Citado na página 18.
- SERMANET, P. et al. Overfeat: Integrated Recognition, Localization and Detection Using Convolutional Networks. In: *International Conference on Learning Representations (ICLR2014)*. [s.n.], 2014. Disponível em: <<http://arxiv.org/abs/1312.6229>>. Acesso em: 03/06/2016. Citado 3 vezes nas páginas 48, 49 e 50.
- SERRE, T. et al. A quantitative theory of immediate visual recognition. *Progress in Brain Research*, Elsevier, v. 165, p. 33 – 56, 2007. Citado na página 34.
- SIVIC, J.; ZISSERMAN, A. Video google: A text retrieval approach to object matching in videos. In: *Proceedings of the Ninth IEEE International Conference Computer Vision*. [S.l.: s.n.], 2003. v. 2, p. 1470 –. Citado 3 vezes nas páginas 25, 47 e 50.
- SOLOMON, C.; BRECKON, T. *Fundamentos de processamento digital de imagens : uma abordagem prática com exemplos em Matlab*. 1. ed. Rio de Janeiro: LTC, 2013. Citado 2 vezes nas páginas 30 e 31.
- TCHANGAN, A. Support Vector Machines: A Tool for Pattern Recognition and Classification. *Studies. Informatics & Control Journal*, v. 14, n. 2, p. 99 – 110, 2005. Citado na página 62.
- TEIXEIRA, A. *Desenvolvimento de uma Interface Gráfica para Classificadores de Imagem*. 2011. Dissertação (Mestre em Desenvolvimento de Software e Sistemas Interativos) — Instituto Politécnico de Castelo Branco. Disponível em: <<https://repositorio.ipcb.pt/bitstream/10400.11/1155/1/disserta~AT1\textsectionao.pdf>>. Acesso em: 09/10/2016. Citado 2 vezes nas páginas 42 e 61.
- TORRES, R.; FALCÃO, A. Content-based image retrieval: Theory and applications. *Revista de Informática Teórica e Aplicada*, v. 13, p. 161 – 185, 2006. Citado na página 25.
- UFMG. *Introdução ao MATLAB e ao pacote GAAL*. 2016. Disponível em: <<http://www.mat.ufmg.br/gaal/matlab/matlab1.html>>. Acesso em: 21/12/2016. Citado na página 54.
- ULLAH, M.; PARIZI, S.; LAPTEV, I. Improving bag-of-features action recognition with non-local cues. In: LABROSSE, F. et al. (Ed.). *Proceedings of the British Machine Vision Conference*. [S.l.]: BMVA Press, 2010. p. 1 – 11. Citado 2 vezes nas páginas 47 e 50.

VAPNIK, V. *The Nature of Statistical Learning Theory*. New York: Springer, 1995. Citado na página 40.

VASCONCELLOS, F. *Programando com GPUs: Paralelizando o Método Lattice-Boltzmann com CUDA*. 2009. Monografia (Ciência da Computação) — Universidade Federal do Rio Grande do Sul. Disponível em: <http://www.inf.ufrgs.br/~nicolas/pdf/tcc_vasconcellos.pdf>. Acesso em: 01/09/2016. Citado na página 54.

VLFEAT. *Biblioteca Open Source VLFeat*. 2016. Disponível em: <<http://www.vlfeat.org/matconvnet/models/beta16/imagenet-caffe-alex.mat>>. Acesso em: 03/06/2016. Citado na página 56.

WANG, J. et al. Bag of features based classification of breast parenchymal tissue in the mammogram via jointly selecting and weighting visual words. In: *ICIG '11, 2011. Proceedings of the 2011 Sixth International Conference on Image and Graphics*. [S.l.]: IEEE, 2011. p. 622 – 627. Citado 2 vezes nas páginas 47 e 50.

WANG, J. et al. Bag-of-features based medical image retrieval via multiple assignment and visual words weighting. *IEEE Trans. Med. Imaging*, v. 30, n. 11, p. 1996 – 2011, 2011. Citado na página 47.

ZEILER, M.; FERGUS, R. Visualizing and Understanding Convolutional Networks. *Springer International Publishing*, p. 818 – 833, 2014. Citado na página 35.

ZHANG, J. et al. Local features and kernels for classification of texture and object categories: A comprehensive study. *International Journal of Computer Vision*, Springer, v. 73, n. 2, p. 213 – 238, 2007. Citado 2 vezes nas páginas 47 e 50.

Apêndices

**APÊNDICE A - EXEMPLOS DE AMOSTRAS DAS MARCAS DE GADO
UTILIZADAS PARA TREINAMENTO E VALIDAÇÃO**

802	802_33	802_34	802_36	802_48	802_52	802_61	802_68
803	803_33	803_34	803_36	803_48	803_52	803_61	803_68
804	804_33	804_34	804_36	804_48	804_52	804_61	804_68
805	805_33	805_34	805_36	805_48	805_52	805_61	805_68
811	811_33	811_34	811_36	811_48	811_52	811_61	811_68
812	812_33	812_34	812_36	812_48	812_52	812_61	812_68
813	813_33	813_34	813_36	813_48	813_52	813_61	813_68
814	814_33	814_34	814_36	814_48	814_52	814_61	814_68
815	815_33	815_34	815_36	815_48	815_52	815_61	815_68
821	821_33	821_34	821_36	821_48	821_52	821_61	821_68
822	822_33	822_34	822_36	822_48	822_52	822_61	822_68
1093	1093_33	1093_34	1093_36	1093_48	1093_52	1093_61	1093_68

Fonte: do próprio autor.

APÊNDICE B - PSEUDOCÓDIGO DO MÉTODO *BAG-OF-FEATURES*

- 1 Defina o caminho da pasta temporaria e as categorias de marcas
- 2 Construa uma matriz de conjuntos de imagens com base em categorias
- 3 Determine o numero de imagens das categorias de marcas
- 4 Randomize as imagens separando-as em 30% para treinamento e 70% para validacao
- 5 Selecione o metodo K-means para clusterizacao e defina o tamanho 500 para o cluster
- 6 Defina do algoritmo SURF para extracao de pontos de interesse das imagens
- 7 Construa o vocabulario de palavras visuais utilizando o metodo bagOfFeatures
- 8 Armazene o resultado em um vetor utilizando o metodo Encode
- 9 Mostre o histograma de ocorrencias das palavras visuais
- 10 Treine o classificador SVM utilizando os vetores gerados pelo metodo Encode
- 11 Tabule os resultados do SVM em uma matriz de confusao
- 12 Calcule a precisao media obtida e apresente o resultado
- 13 Faça a leitura de uma nova imagem de marca
- 14 Aplique o metodo bagOfFeatures
- 15 Faça a predicao (reconhecimento) da imagem aplicando o classificador SVM
- 16 Apresente o resultado da classificacao para a nova marca

APÊNDICE C - PSEUDOCÓDIGO DO MÉTODO REDES NEURAIAS CONVOLUCIONAIS

- 17 Defina o caminho da pasta temporaria e as categorias de marcas
- 18 Determine o numero de imagens das categorias de marcas
- 19 Selecione o modelo de CNN pre-treinado
- 20 Carregue a rede MatConvNet do modelo CNN pre-treinado e salve em SeriesNetwork
- 21 Inspeção a primeira camada Convolutiva
- 22 Inspeção a ultima camada Convolutiva
- 23 Leia o conjunto de imagens da pasta temporaria
- 24 Realize o Pre-processamento das imagens, transformando-as para o padrao RGB e redimensionando-as para o tamanho definido na CNN
- 25 Randomize as imagens separando-as em 30% para treinamento e 70% para validacao
- 26 Aplique a CNN implementada para extrair as caracteristicas das imagens
- 27 Apresente os mapas de carateristicas obtidos na primeira camada convolutiva
- 28 Treine o classificador SVM utilizando os resultados obtidos pela CNN
- 29 Tabule os resultados do SVM em uma matriz de confusao
- 30 Calcule a precisao media obtida e apresente o resultado
- 31 Faça a leitura de uma nova imagem de marca
- 32 Realize o pre-processamento da imagem
- 33 Extraia as caracteristicas da imagem aplicando a CNN
- 34 Faça a predicao (reconhecimento) da imagem aplicando o classificador SVM
- 35 Apresente o resultado da classificacao para a nova marca

**APÊNDICE D - CLASSIFICAÇÕES CORRETAS *VERSUS* CLASSIFICAÇÕES
PREDITAS - EXPERIMENTOS I E II**

**Tabela 13 – Classificações Corretas versus Classificações Preditas - Método
BoF**

Marca	VP	FP	FN	VN
802	29	2	5	336
803	29	2	4	335
804	26	5	10	331
805	28	3	4	337
811	27	4	0	341
812	23	8	2	339
813	26	5	1	340
814	24	7	3	338
815	28	3	8	333
821	28	3	3	338
822	24	7	5	336
1093	28	3	7	334

Tabela 14 – Classificações Corretas versus Classificações Preditas - Método CNN

Marca	VP	FP	FN	VN
802	31	0	0	341
803	30	1	0	341
804	29	2	5	336
805	25	6	2	339
811	28	3	0	341
812	31	0	7	334
813	24	7	7	334
814	28	3	0	341
815	31	0	0	341
821	31	0	1	340
822	30	1	2	339
1093	29	2	1	340

APÊNDICE E - ACURÁCIA POR MARCA - EXPERIMENTO IV

Tabela 15 – Acurácia por marca do melhor resultado obtido nas simulações realizadas no Experimento IV

Marca	Acc								
802	96,77	1404	83,87	2332	83,87	2505	87,10	2612	83,87
803	77,42	1405	74,19	2333	100	2506	87,10	2613	90,32
804	93,55	1406	80,65	2334	67,74	2507	67,74	2614	83,87
805	77,42	1407	80,65	2384	77,42	2509	87,10	2615	90,32
811	77,42	1408	70,97	2386	80,65	2510	90,32	2616	80,65
812	77,42	1409	83,87	2387	48,39	2511	74,19	2617	74,19
813	74,19	1450	77,42	2388	83,87	2511A	96,77	2618	90,32
814	77,42	1451	77,42	2390	83,87	2512	64,52	2619	61,29
815	100	1452	74,19	2391	90,32	2513	83,87	2620	77,42
821	93,55	1454	77,42	2392	80,65	2514	80,65	2621	83,87
822	87,1	1455	87,10	2393	67,74	2515	100	2622	87,10
823	64,52	1456	90,32	2394	77,42	2516	93,55	2623	80,65
824	96,77	1457	74,19	2396	77,42	2517	93,55	2624	74,19
825	90,32	1458	83,87	2397	87,10	2518	77,42	2625	80,65
831	83,87	1459	90,32	2398	64,52	2519	70,97	2626	64,52
832	93,55	1460	80,65	2399	77,42	2520	90,32	2627	83,87
833	96,77	1461	96,77	2400	96,77	2521	77,42	2628	70,97
834	83,87	1462	100	2404	83,87	2523	80,65	2629	93,55
835	80,65	1463	96,77	2405	96,77	2524	77,42	2630	83,87

Marca	Acc								
841	93,55	1464	70,97	2406	70,97	2525	74,19	2631	74,19
842	77,42	1465	100	2408	74,19	2526	74,19	2632	83,87
843	96,77	1466	87,10	2410	61,29	2527	83,87	2633	77,42
844	87,1	1467	96,77	2411	77,42	2528	74,19	2634	87,10
845	74,19	1468	93,55	2412	80,65	2529	67,74	2635	70,97
851	77,42	1469	74,19	2414	64,52	2530	74,19	2636	87,10
853	77,42	1471	80,65	2415	80,65	2531	77,42	2637	77,42
854	90,32	1472	87,10	2416	93,55	2532	83,87	2638	90,32
855	87,1	1473	93,55	2417	80,65	2533	100	2639	83,87
861	83,87	1474	90,32	2418	64,52	2534	83,87	2640	83,87
862	67,74	1475	90,32	2420	77,42	2535	100	2641	93,55
863	67,74	1476	74,19	2421	87,10	2536	80,65	2642	67,74
864	77,42	1477	93,55	2422	83,87	2537	83,87	2643	90,32
865	80,65	1478	100	2423	61,29	2538	83,87	2644	74,19
942	80,65	1479	96,77	2424	87,10	2539	80,65	2645	83,87
943	87,1	1480	61,29	2426	87,10	2540	96,77	2646	77,42
944	70,97	1481	80,65	2427	83,87	2541	54,84	2648	96,77
1081	96,77	1482	77,42	2428	80,65	2542	80,65	2649	83,87
1082	64,52	1483	93,55	2429	83,87	2543	74,19	2650	61,29
1083	80,65	1484	87,10	2430	83,87	2544	87,10	2651	83,87
1084	83,87	1485	90,32	2432	80,65	2545	83,87	2652	67,74
1085	93,55	1486	77,42	2433	74,19	2546	83,87	2653	96,77

Marca	Acc								
1086	93,55	1487	77,42	2434	74,19	2547	80,65	2654	80,65
1087	80,65	1488	70,97	2435	80,65	2548	93,55	2665	90,32
1088	80,65	1489	70,97	2436	67,74	2549	77,42	2666	80,65
1089	67,74	1770	67,74	2437	70,97	2550	96,77	2667	100
1090	80,65	1771	64,52	2438	93,55	2551	70,97	2668	83,87
1091	74,19	1772	77,42	2439	77,42	2552	77,42	2669	93,55
1092	74,19	1773	83,87	2440	87,10	2553	77,42	2670	70,97
1093	74,19	1774	74,19	2441	90,32	2554	67,74	2671	100
1094	90,32	1775	87,10	2443	93,55	2556	80,65	2672	100
1095	93,55	1776	80,65	2444	80,65	2557	70,97	2673	80,65
1096	80,65	1777	80,65	2445	83,87	2558	70,97	2674	74,19
1097	96,77	1778	70,97	2446	74,19	2559	80,65	2675	96,77
1098	87,10	1779	77,42	2447	100	2560	80,65	2676	90,32
1099	80,65	1780	64,52	2448	90,32	2561	90,32	2677	80,65
1100	64,52	1782	83,87	2449	83,87	2562	67,74	2678	83,87
1101	77,42	1783	67,74	2450	74,19	2563	74,19	2679	100
1102	67,74	1784	77,42	2451	74,19	2564	64,52	2680	77,42
1103	70,97	1785	77,42	2452	80,65	2565	67,74	2681	70,97
1104	77,42	1786	83,87	2453	58,06	2566	87,10	2682	87,10
1105	74,19	1787	0	2455	64,52	2567	67,74	2683	74,19
1106	74,19	1788	83,87	2456	77,42	2568	90,32	2684	93,55
1107	83,87	1789	74,19	2457	77,42	2569	80,65	2685	77,42

Marca	Acc								
1364	74,19	1790	90,32	2458	83,87	2570	74,19	2687	96,77
1365	67,74	1791	64,52	2459	74,19	2571	64,52	2690	61,29
1368	90,32	1793	83,87	2460	80,65	2572	74,19	2691	80,65
1369	74,19	1794	74,19	2462	77,42	2573	77,42	2692	77,42
1370	80,65	1795	77,42	2463	93,55	2574	74,19	2693	74,19
1371	93,55	1796	93,55	2464	87,1	2575	74,19	2694	74,19
1372	67,74	1797	64,52	2465	77,42	2576	70,97	2695	93,55
1373	77,42	1798	70,97	2466	93,55	2577	67,74	2696	93,55
1374	77,42	1799	70,97	2467	83,87	2578	77,42	2697	83,87
1375	80,65	1800	96,77	2469	77,42	2579	77,42	2698	96,77
1376	77,42	1802	80,65	2470	93,55	2580	74,19	2699	83,87
1377	87,10	1803	87,10	2471	77,42	2581	93,55	2700	70,97
1378	80,65	1804	83,87	2472	90,32	2582	74,19	2701	74,19
1379	61,29	2306	87,10	2473	100	2585	70,97	2702	90,32
1380	70,97	2307	87,10	2474	74,19	2586	61,29	2703	96,77
1381	87,10	2308	83,87	2476	80,65	2587	80,65	2704	77,42
1382	93,55	2310	74,19	2477	70,97	2588	96,77	2705	83,87
1383	80,65	2311	77,42	2478	90,32	2590	96,77	2706	77,42
1384	93,55	2312	70,97	2479	74,19	2591	77,42	2707	77,42
1385	83,87	2313	87,10	2480	90,32	2592	80,65	2708	74,19
1386	74,19	2314	70,97	2481	87,10	2593	93,55	2709	77,42
1387	70,97	2315	90,32	2483	80,65	2594	70,97	2710	80,65

Marca	Acc								
1388	70,97	2316	83,87	2484	87,10	2596	90,32	2711	74,19
1389	74,19	2317	74,19	2485	70,97	2597	93,55	2712	87,10
1390	77,42	2318	77,42	2486	80,65	2598	83,87	2713	96,77
1391	67,74	2319	70,97	2487	96,77	2599	93,55	2714	70,97
1392	70,97	2320	83,87	2488	90,32	2600	67,74	2715	70,97
1393	80,65	2321	67,74	2490	77,42	2602	96,77	2716	74,19
1394	74,19	2322	80,65	2491	87,1	2603	64,52	2717	74,19
1395	87,10	2323	90,32	2492	74,19	2604	83,87	2718	74,19
1396	74,19	2324	83,87	2494	0	2605	77,42	2719	87,10
1397	90,32	2325	77,42	2495	80,65	2606	87,10	2720	67,74
1398	74,19	2326	58,06	2496	83,87	2607	64,52	2721	74,19
1399	70,97	2327	67,74	2498	74,19	2608	93,55	2722	77,42
1400	100	2329	77,42	2499	67,74	2609	83,87	2723	87,10
1401	70,97	2330	90,32	2500	87,10	2610	74,19	2724	77,42
1402	87,10	2331	93,55	2502	74,19	2611	96,77	2725	70,97

APÊNDICE F - ARTIGO PUBLICADO NA REVISTA IEEE LATIN AMERICA TRANSACTIONS

310

IEEE LATIN AMERICA TRANSACTIONS, VOL. 15, NO. 2, FEB. 2017

Cattle Brand Recognition using Convolutional Neural Network and Support Vector Machines

C. Silva, *Member, IEEE*, D. Welfer, F. P. Gioda and C. Dornelles

Abstract— The recognition images of cattle brand in an automatic way is a necessity to governmental organs responsible for this activity. To help this process, this work presents a method that consists in using Convolutional Neural Network for extracting of characteristics from images of cattle brand and Support Vector Machines for classification. This method consists of six stages: a) select database of images; b) select pre-trained CNN; c) pre-process the images and apply CNN; d) extract images of features; e) train and sort images (SVM); f) evaluate the classification results. The accuracy of the method was tested on database of municipal city hall, where it achieved satisfactory results, comparable to other methods from the literature, reporting 93.28% of accuracy and 12.716 seconds of processing time, respectively.

Keywords— Recognition Images, Cattle Brands, Convolutional Neural Network, Support Vector Machines, Deep Learning.

I. INTRODUÇÃO

O DESENVOLVIMENTO de ferramentas computacionais para auxiliar a análise e reconhecimento de imagens é alvo de interesse dos mais renomados centros de pesquisa do mundo. O uso da computação para análise e reconhecimento de imagens está em constante desenvolvimento, gerando vários benefícios à sociedade nas mais diversificadas áreas do conhecimento. No que se refere particularmente ao reconhecimento de marcas de gado, atividade esta tão tradicional e de grande relevância socioeconômica para países da América Latina, incluindo o Brasil, não há um método específico e devidamente consolidado para este fim. Para se ter uma dimensão da importância desta atividade, de acordo com a *Food and Agriculture Organization* – FAO, entre os países produtores, o Brasil e a Índia possuem os maiores rebanhos, sendo que o Brasil encontra-se em 1º lugar, com uma média de 209.215.666 cabeças [1].

A produção pecuária possui um papel relevante na formação social, e que ainda hoje se mantém como uma atividade de grande importância nas expressões culturais associadas a ela, uma vez que é associada à cultura e ao modo de vida do campo, além do papel na afirmação ou construção

de identidades individuais ou de grupos [2].

A utilização das marcas ou sinais no gado pressupõe o reconhecimento público de sua propriedade por um indivíduo ou grupo. Utilizadas desde o início da colonização ibérica na América, o início de sua institucionalização se deu a partir do registro em órgãos, sendo oficiais, reconhecidamente portadores de legitimação pública [2]. A esses registros seguem-se as regulamentações que procuram, além de oficializar a marcação em si, orientar a forma e o período a ser feita, discriminar a forma do registro, instituir valores a este, a construção dos ferros e a taxaço governamental. Em geral, os registros das marcas de gado constituem-se de livros com os desenhos das marcas e com a identificação de seu proprietário. No Brasil as tentativas e os investimentos de aprimoramento no sistema de registro de marcas de gado sempre foi alvo de polêmica, em função da resistência dos pecuaristas. Boa parte desse receio está associado ao temor de perda das marcas familiares e o significado que elas adquiriram com o tempo. Atualmente, os registros de marcas no Brasil são realizados nos municípios, em geral sem uma sistematização mais efetiva e sem a necessidade de serem renovadas.

Perante o contexto apresentado, este trabalho tem como objetivo apresentar e avaliar uma ferramenta que realiza o reconhecimento automático de marcas de gado, com o propósito de substituir o controle manual de marcas de gado que é realizado atualmente, de forma a diminuir potencialmente a possibilidade de registros duplicados, reduzir o tempo de espera para registro de novas marcas, aprimorar a gestão governamental no que diz respeito ao acervo de marcas sob sua responsabilidade e auxiliar as autoridades de segurança no combate a crimes de abigeato.

O trabalho concentrou-se no desenvolvimento dos algoritmos responsáveis pelo reconhecimento das marcas de gado da Prefeitura Municipal de São Francisco de Assis, Rio Grande do Sul, Brasil. Desta maneira, os funcionários do Setor de Registro de marcas de gado e do Centro de Processamento de Dados do município validaram a ferramenta proposta.

O restante do trabalho está organizado da seguinte maneira: na Seção 2 são descritos os trabalhos relacionados. Na Seção 3 são apresentados os materiais e métodos utilizados para este trabalho. A Seção 4 descreverá os resultados e discussões obtidos através da aplicação dos métodos propostos. Por fim, na Seção 5 serão descritas as considerações finais.

II. TRABALHOS RELACIONADOS

De maneira geral, não foram encontrados trabalhos na revisão da literatura que reportem a utilização de redes neurais

C. Silva, Universidade Federal do Pampa (UNIPAMPA), Alegrete, Rio Grande do Sul, Brasil, carlossilva@ieee.org

D. Welfer, Universidade Federal de Santa Maria (UFSM), Departamento de Computação Aplicada – DCOM, Santa Maria, Rio Grande do Sul, Brasil, daniel.welfer@ufsm.br

F. P. Gioda, Prefeitura Municipal de São Francisco de Assis, Rio Grande do Sul, Brasil, administracao@saofranciscodeassis.rs.gov.br

C. Dornelles, Prefeitura Municipal de São Francisco de Assis, Rio Grande do Sul, Brasil, administracao@saofranciscodeassis.rs.gov.br

convolucionais para reconhecimento de imagens de marcas de gado.

Sanchez *et al* [3] apresenta uma ferramenta para reconhecimento de marcas de gado utilizando momentos de Hu e Legendre para extração de características de imagens em escala de cinza, e um classificador de k-Vizinhos mais Próximos (k-NN). Os autores utilizaram momentos de Hu e Legendre com o propósito de extrair características que não fossem suscetíveis a transformações de rotação, translação e escala. O percentual máximo de classificação correta apresentado pelos autores foi de 99,3 %, porém com uma diminuição significativa de acurácia à medida que aumentava o número de imagens classificadas. Outro resultado apresentado foi o tempo de processamento da classificação. Pelo fato de ter sido utilizado um classificador k-NN, a cada novo objeto que se quer classificar, utiliza-se os dados de treinamento para verificar quais são os objetos nesta base de dados que mais se assemelham ao novo objeto que se quer classificar. O objeto é classificado dentro da classe mais comum a que pertencem os objetos mais similares a ele. Assim, a classificação é feita por analogia. Nenhum modelo de classificação é criado. Ao invés disto, a cada novo objeto a ser classificado, os dados de treinamento são escaneados, dessa forma o classificador proposto se torna computacionalmente dispendioso.

Diferentemente do trabalho apresentado por Sanchez *et al* [3], o trabalho aqui proposto pretende apresentar resultados que possam ser generalizados ou replicados, utilizando técnicas no estado da arte no que diz respeito à extração de características e classificação estatística de imagens digitais, tais como Redes Neurais Convolucionais (CNN) e Máquinas de Vetores de Suporte (SVM), objetivando desempenhos superiores no reconhecimento de imagens em bases de dados com número maior de registros, porém com custo computacional e tempo de processamento menores.

Os trabalhos encontrados na literatura para reconhecimento e classificação de imagens que utilizam descritores ou filtros para extração de características, seguida de uma etapa de quantização e agrupamento e, por último, uma etapa de classificação são divididos em duas categorias: algoritmos de estágio de extração de características único e algoritmos de dois ou mais estágios [4].

A possibilidade de treinar redes neurais com múltiplas camadas intermediárias desperta o surgimento de diversos algoritmos agrupados em um área conhecida como *deep learning*.

O objetivo principal dos algoritmos que utilizam dois ou mais estágios é aprender não apenas a distinguir as classes com base em descritores artificiais, mas aprender os próprios descritores com base nos dados brutos, no caso de imagens, os próprios valores dos *pixels* [5].

Redes neurais convolucionais vêm sendo utilizadas a vários anos no reconhecimento de imagens, tendo obtido grande sucesso no reconhecimento de caracteres no trabalho de Cun *et al* [6].

Estudos mais recentes utilizando redes neurais convolucionais conhecidas também como *Deep Convolutional*

Neural Networks (CNN) obtiveram o novo estado da arte no reconhecimento de objetos em bases CIFAR-10 e NORB [7].

De forma geral as CNN são treinadas de forma supervisionada, mas trabalhos sugerem que o pré-treinamento da CNN com filtros obtidos de forma não supervisionada apresentam um melhor resultado [8].

No trabalho realizado por Sermanet *et al* [9] é apresentado um *framework* utilizando CNN para efetuar o reconhecimento, localização e detecção de imagens sendo o vencedor do campeonato *ImageNet Large Scale Visual Recognition Challenge* 2013 (ILSVRC2013). A base ILSVRC2013 consiste em 1,2 milhões de imagens divididas em 1000 categorias.

Uma característica importante da CNN é a capacidade de serem reutilizadas e refinadas para diferentes bases de imagens. No trabalho realizado por Razavian *et al* [10] foi utilizada uma CNN pré-treinada chamada *Overfeat* [9] para realizar extração de um descritor de diferentes bases de imagens na qual a CNN não foi originalmente treinada. Nesse caso, os descritores são então classificados utilizando um classificador linear SVM. Os resultados demonstram um desempenho compatível com o estado da arte, mesmo comparando com algoritmos que utilizam imagens segmentadas manualmente procedimento que não é necessário quando utilizada a CNN, e treinados especificamente na base analisada.

A utilização de *deep learning* também é descrita no trabalho de Constante *et al* [11], os quais utilizaram uma rede neuronal de três camadas com entrada usando *backpropagation*. Nesse trabalho, o método foi utilizado para realizar classificação de morangos, e chegou a resultados de reconhecimento de 92,5% na categoria “Extra”; 90% na categoria “Consumo”; 90% na categoria “Matéria prima”; e 100% na categoria “Objetos estranhos”.

III. MATERIAIS E MÉTODOS

As imagens de marcas de gado utilizadas nesse trabalho foram fornecidas pela Prefeitura de São Francisco de Assis – RS. Foram utilizadas 12 imagens de marca de gado, sendo cada uma dessas constituídas por 45 subimagens (amostras), totalizando 540 amostras oriundas das imagens originais, porém com variações de tamanho e orientação, com o objetivo de identificar padrões com a maior independência possível desses fatores. As imagens foram disponibilizadas em alta resolução no formato *Portable Network Graphics* com tamanho 600 x 600 *pixels*.

Para implementação da ferramenta proposta, assim como o armazenamento do banco de imagens, processamento dos algoritmos e visualização dos resultados foi utilizado um computador pessoal com placa de vídeo com suporte à plataforma de computação paralela CUDA com *compute capability version* 5.0. Além disso, foi utilizado o *software* MatLab com as bibliotecas *Neural Network*, *Parallel Computing* e *Statistics and Machine Learning* e o modelo de rede convolucional pré-treinado obtido junto à biblioteca *open source* da VLFeat.org [12].

312

O método proposto é composto por seis etapas, que são: seleção do banco de imagens; seleção do modelo de CNN pré-treinado; pré-processamento das imagens e aplicação da CNN; extração de características das imagens; treinamento e classificação das imagens por meio de Máquinas de Vetores de Suporte; e, por fim, avaliação dos resultados da classificação. A Fig. 1 ilustra o fluxograma sumarizado do método proposto.

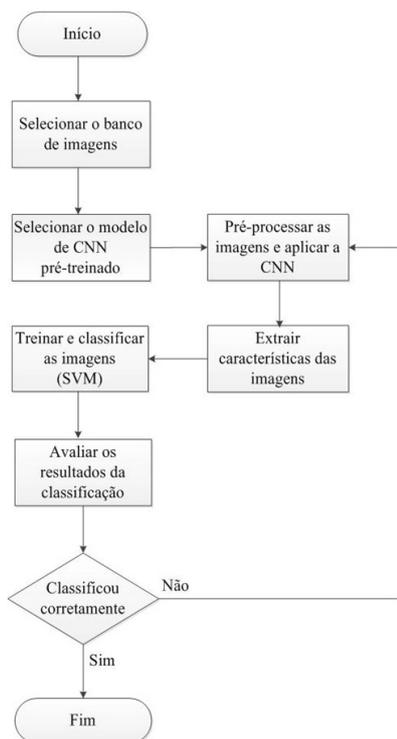


Figura 1. Fluxograma sumarizado do método proposto.

A. Seleção do banco de imagens

As marcas utilizadas no trabalho são ilustradas na Fig. 2. O código das marcas, os proprietários e a quantidade de amostras de cada marca podem ser obtidos na Tabela I.

IEEE LATIN AMERICA TRANSACTIONS, VOL. 15, NO. 2, FEB. 2017

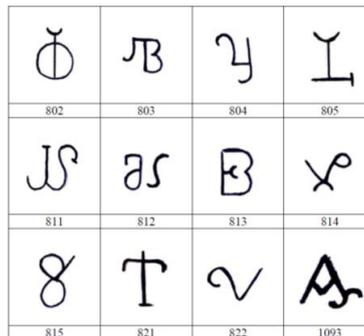


Figura 2. Imagens de marcas de gado utilizadas no trabalho.

TABELA I
MARCAS, PROPRIETÁRIOS E TOTAL DE AMOSTRAS POR MARCA

Marca	Proprietário	Total de amostras
802	Proprietário "A"	45
803	Proprietário "B"	45
804	Proprietário "C"	45
805	Proprietário "D"	45
811	Proprietário "E"	45
812	Proprietário "F"	45
813	Proprietário "G"	45
814	Proprietário "H"	45
815	Proprietário "I"	45
821	Proprietário "J"	45
822	Proprietário "K"	45
1093	Proprietário "L"	45

B. Seleção do modelo de CNN pré-treinado

Redes neurais convolucionais (CNN) são arquiteturas biologicamente inspiradas capazes de serem treinadas e aprenderem representações invariantes à escala, translação, rotação e transformações afins [13]. As CNN compõem um dos tipos de algoritmos da área conhecida como *deep learning* e são projetadas para uso com dados em duas dimensões tornando-as uma boa candidata para a solução de problemas envolvendo reconhecimento de imagens [14]. Por definição, uma arquitetura profunda é uma estrutura hierárquica de múltiplas etapas, onde cada etapa é formada por uma rede neuronal de pelo menos 3 camadas, e cada etapa é treinada por *backpropagation*.

A Fig. 3 ilustra o modelo geral de uma arquitetura CNN.

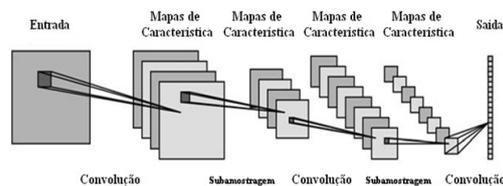


Figura 3. Modelo geral de uma arquitetura CNN.

As redes neurais convolucionais dispõem de uma propriedade ainda pouco explorada conhecida como transferência de conhecimento. Esta propriedade remete ao fato de uma CNN poder ser treinada em uma base de imagens A (sendo assim, seus pesos são ajustados para classificação da base A) sendo os pesos de aprendizado (e filtros nas CNN) considerados genéricos o suficiente para serem usados no treinamento de uma nova base B. Utilizando-se desse conceito foi selecionado o modelo CNN da biblioteca *open source* VLFFeat, cujos dados utilizados para pré-treinamento da rede neural convolucional utilizada no trabalho apresentado foram obtidos junto ao arquivo “imagenet-caffe-alex.mat”. A utilização do modelo de CNN pré-treinado supracitado não influenciou diretamente na taxa de reconhecimento das imagens de marcas de gado.

C. Pré-processamento das imagens e aplicação da CNN

O método adotado consiste em uma rede neural com cinco camadas convolucionais. Caso a imagem seja em tons de cinza é realizado um pré-processamento, no qual a imagem é replicada 3 vezes para criar uma imagem RGB. A 1ª camada convolucional tem como entrada os 3 canais de cores da imagem (RGB). Cada convolução realiza a aplicação da função de ativação não linear *ReLU* e a redução através do *Maxpooling*. As últimas camadas são compostas por neurônios totalmente conectados. Para o treinamento da CNN foi utilizada a função *Softmax* e o algoritmo *backpropagation*.

A Fig. 4 apresenta a arquitetura da rede neural convolucional proposta.

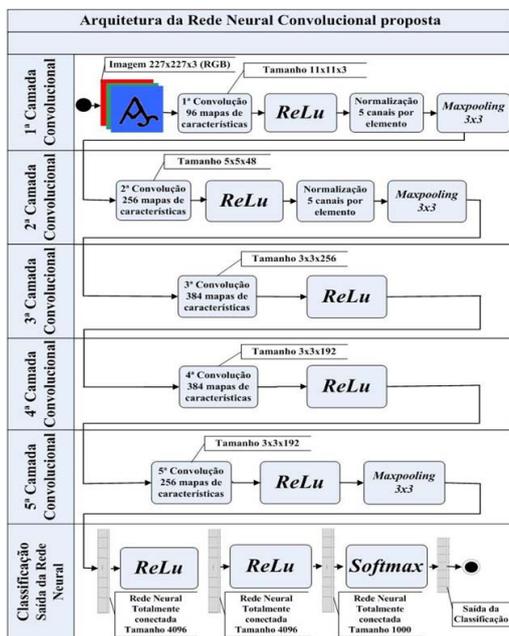


Figura 4. Arquitetura da rede neural convolucional proposta.

D. Extração de características das imagens

O conjunto de filtros aprendidos pela CNN durante o treinamento é responsável pela detecção das características na nova imagem no momento de uma consulta. No primeiro nível de filtros é possível observar algumas linhas e orientações utilizadas para essa detecção. A Fig. 5 mostra os filtros aprendidos na primeira camada convolucional utilizando espaço de cores RGB. Existem 96 conjuntos individuais que representam os 96 filtros utilizados nessa camada. É possível observar como as áreas com saliências horizontais, verticais e diagonais são destacadas após a realização da primeira convolução.

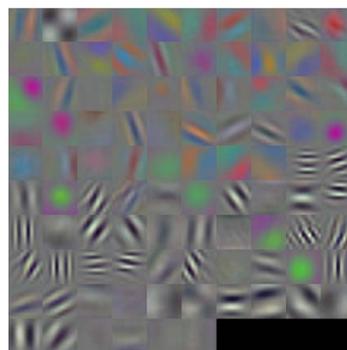


Figura 5. Filtros da 1ª camada convolucional do experimento realizado.

Após a realização das convoluções foi possível realizar a extração das características das imagens de forma que o classificador pudesse ser treinado. A Fig. 6 apresenta uma ilustração do algoritmo de extração de características desenvolvido.

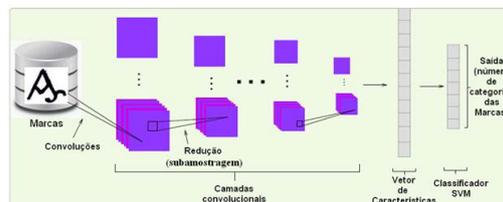


Figura 6. Ilustração do algoritmo de extração de características desenvolvido.

E. Treinamento e classificação das imagens com a utilização de Máquinas de Vetores de Suporte

O modelo de aprendizagem automática adotado no trabalho apresentado foi o classificador supervisionado *Support Vector Machine* (SVM). *Support Vector Machine* é um algoritmo de classificação conhecido por ser bem sucedido numa grande variedade de aplicações. As SVM são uma das abordagens mais populares para modelagem e classificação de

dados. As suas vantagens incluem a excelente capacidade de generalização, que diz respeito à capacidade de classificar corretamente as amostras que não estão dentro do espaço da característica usado para o treino [15]. Dadas duas classes e um conjunto de pontos a essas classes, o SVM determina o hiperplano que separa os pontos de forma a colocar o maior número de pontos da mesma classe do mesmo lado, maximizando a distância de cada classe a esse hiperplano, sendo consequentemente denominado classificador de margem máxima [16]. Com efeito, uma larga margem entre os valores correspondentes aos pontos dos dois subconjuntos de dados implica um risco de generalização minimizado do classificador.

As SVMs são utilizadas para classificar e reconhecer padrões em diversos tipos de dados, sendo utilizadas em diversas aplicações, tais como reconhecimento de faces, diagnósticos clínicos, supervisão de processos industriais, processamento e análise de imagens [17].

Na ferramenta proposta nesse trabalho o classificador foi utilizado após a extração de características das marcas pertencentes a diferentes conjuntos de amostras. Na aprendizagem supervisionada dado um conjunto de exemplos (X_1, X_2) em X_1 representa um exemplo e X_2 a sua classificação, deve-se reproduzir um classificador capaz de prever a classe a que pertencem novos dados, efetuando assim o processo de treino. Nesse contexto, utilizou-se o método *cross-validation* [15], na qual as imagens foram divididas aleatoriamente em 2 partes, onde uma dessas partes foi utilizada para treinamento e a outra para validação, de forma que não houvesse polarização dos resultados. O resultado final é a média do resultado obtido na validação. A divisão da porcentagem utilizada foi de 30% para treinamento e 70% para validação.

F. Avaliação dos resultados da classificação - Matriz de Confusão

A matriz de confusão contém informações relativas a classificações efetuadas através da aplicação de um classificador. O desempenho dos classificadores é frequentemente avaliado através dos dados retirados desta matriz [18].

A matriz de confusão de um classificador indica o número de classificações corretas versus as previsões efetuadas para cada caso, sobre um conjunto de exemplos. Nesta matriz as linhas representam os casos reais e as colunas as previsões efetuadas pelo modelo. Através da matriz de confusão é possível obter informação relativa ao número de imagens corretamente classificadas e incorretamente classificadas, para cada conjunto de amostras. Esta matriz é do tipo $A \times A$, sendo A o número de categorias ao qual se aplica o classificador, no caso do experimento efetuado existem 12 marcas, sendo portanto a matriz de confusão de tamanho 12×12 .

IV. RESULTADOS E DISCUSSÕES

Através dos resultados obtidos foi possível avaliar o método proposto. A avaliação dos resultados dos experimentos foi realizada com base na taxa de reconhecimento obtida da matriz de confusão gerada a partir

da classificação realizada na etapa de validação. Além disso, o tempo total de processamento do método proposto também foi verificado.

A Fig. 7 apresenta a matriz de confusão para o melhor resultado obtido nos experimentos, cuja taxa de reconhecimento chegou a 93,28%. Percebe-se através da análise da diagonal principal que a taxa de classificação correta se destaca em quatro marcas, “802”, “812”, “815” e “821”, nas quais o percentual de acerto chega a 100%, correspondendo a 31 acertos. É possível observar também que as marcas que tiveram as menores taxas de classificação correta foram a “813” e a “805”, com percentual de 77,42% e 80,64%, nessa ordem. As demais marcas, “803”; “804”; “811”; “814”; “822” e “1093”, alcançaram taxa de classificação correta de 96,77%; 93,55%; 90,32%; 90,3%; 96,77% e 93,55%, respectivamente.

A hipótese de classificação errada das marcas de gado conforme mostrada na matriz de confusão pode estar também associada à complexidade das amostras, pois algumas imagens possuem características similares entre si.

Em geral, as amostras de imagens de marcas com maior poder descritivo e melhor qualidade classificaram corretamente mais imagens, pois abrangem mais características quando comparadas às marcas com menor qualidade das amostras, e, consequentemente, menos características extraídas. A capacidade de reconhecer padrões de uma imagem sobre um conjunto de imagens depende da quantidade de informações que se conhece a priori do objeto em questão.

Outro fato importante a ser apresentado, fruto da análise da Fig. 7, é o número baixo de falsos positivos e falsos negativos, observados fora da diagonal principal da matriz de confusão, o que se deve principalmente à capacidade do método proposto de extrair características de imagens, mesmo em situações adversas, em que as imagens apresentem diferentes tamanhos, formas, escalas, orientações, distorções, ruídos, cores diferentes e diferente contexto de fundo. Nesse tipo de experimento os ruídos nas imagens podem prejudicar a acurácia da classificação.

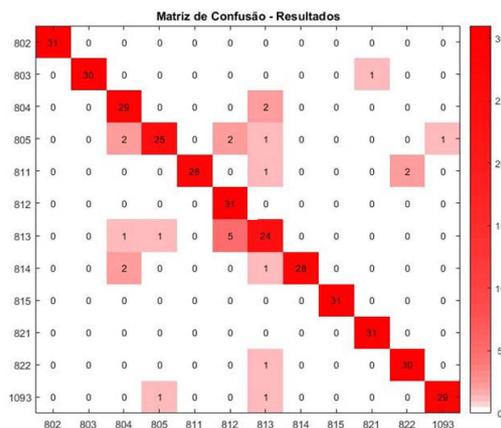


Figura 7. Matriz de confusão obtida na etapa de validação.

Tendo em consideração a matriz de confusão apresentada na Fig. 7, observa-se que através do somatório da diagonal principal obtém-se o total de marcas de gado classificadas corretamente, mais precisamente 347, ao passo que a soma dos demais valores equivale às marcas classificadas incorretamente, ou seja, 25.

A Fig. 8 mostra um gráfico comparativo entre as marcas classificadas corretamente em função do número de amostras utilizadas na validação.

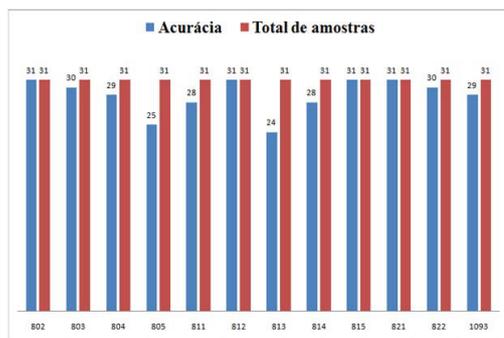


Figura 8. Proporção de marcas corretamente classificadas em função do número de amostras.

As barras em azul apresentam as marcas de gado classificadas corretamente (acurácia), e as barras em vermelho apresentam o total de amostras de cada marca. Das 12 marcas analisadas, 4 apresentaram 100% de classificações corretas, no caso as marcas “802”; “812”; “815” e “821”. As marcas de gado que tiveram a menor taxa de acerto foram a “813” e “805”, com 24 e 25 imagens classificadas corretamente, respectivamente. Observa-se também o elevado índice de classificações corretas, mesmo nas marcas que não alcançaram 100% de acerto, como no caso das marcas “803”; “804”; “811”; “814”; “822” e “1093”.

O tempo de processamento do algoritmo em função do número de amostras de marcas de gado é apresentado na Fig. 9.

Foram medidos os tempos de processamento do método proposto para a classificação em cinco grupos de amostras. Cada grupo contendo 108, 216, 324, 432 e 540 imagens. Os tempos de processamento para classificação das imagens em cada grupo foram de 6,206s; 8,01s; 9,488s; 11,295s e 12,716s, nessa ordem.

Ao analisar o gráfico ilustrado na Fig. 9 é possível perceber que o tempo de processamento observado no eixo y, varia de forma diretamente proporcional ao aumento do número de amostras de marcas de gado classificadas do eixo x, ou seja, observa-se um padrão de crescimento linear da função, mesmo que a taxa de crescimento não seja exatamente um valor constante.

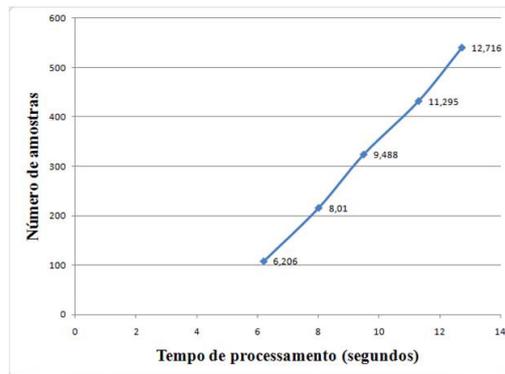


Figura 9. Tempo de processamento do algoritmo em função do número de amostras.

Os resultados obtidos com os experimentos realizados com 12 marcas de gado e 540 amostras de imagens, utilizadas tanto para treinamento quanto para validação, alcançaram taxa de reconhecimento de 93,28%, taxa de erro de 6,72% e tempo total de processamento de 12,716 segundos.

A taxa de reconhecimento foi obtida mediante o cálculo da média aritmética das marcas classificadas corretamente da matriz de confusão, e o tempo total de processamento foi obtido mediante a utilização do software MatLab, que ao final do processamento do código realiza o detalhamento da velocidade de processamento do algoritmo.

V. CONCLUSÃO

Nesse trabalho foi apresentado um método automatizado para reconhecimento de marcas de gado. O projeto foi desenvolvido entre duas instituições; a Prefeitura Municipal de São Francisco de Assis e UNIPAMPA.

Os experimentos realizados neste trabalho utilizaram uma Rede Neural Convolutiva (CNN) para extração de características e um classificador supervisionado SVM. Na CNN foi criada uma rede convolutiva completa utilizando como entrada imagens transformadas para o formato de cores RGB. Todos os experimentos foram realizados na base de marcas de gado fornecida pela Prefeitura Municipal de São Francisco de Assis.

A utilização do método proposto apresentou uma acurácia média de 93,28% e tempo de processamento do algoritmo de 12,716 segundos para 12 marcas avaliadas, num total de 540 amostras utilizadas para treino e validação.

O método utilizado realizou eficaz e eficientemente o reconhecimento de diferentes marcas de gado, mesmo utilizando uma CNN pré-treinada, porém sua principal limitação foi a necessidade de uma grande quantidade de imagens de amostra para treinamento do classificador, uma vez que a CNN necessita dessas imagens para extração de características nas camadas convolucionais.

Como trabalho futuro objetiva-se a ampliação da base de imagens tomadas para treinamento e validação, com o intuito de se obter novas medições de acurácia e desempenho do método proposto, visando a disponibilização de uma

ferramenta devidamente validada e consolidada para os órgãos governamentais responsáveis pelo registro e controle das marcas de gado.

REFERÊNCIAS

- [1] SEPLAN. *Secretaria do Planejamento e Desenvolvimento Regional – Governo do Estado do Rio Grande do Sul, Brasil*. 2015. URL: http://www.scp.rs.gov.br/atlas/conteudo.asp?cod_menu_filho=819&cod_menu=817&tipo_menu=ECONOMIA&cod_conteudo=1580
- [2] ARNONI, R. “Os Registros e Catálogos de Marcas de Gado da Região Platina.” Pelotas: *Revista Memória em Rede da UFPEL*, 2013.
- [3] SANCHEZ, G.; RODRIGUEZ, M. “Cattle Marks Recognition by Hu and Legendre Invariant Moments”. *ARPN Journal of Engineering and Applied Sciences*, Vol. 11, Nº 1, 2016.
- [4] JARRETT, K.; KAVUKCUOGLU, K.; LECUN, Y. “What Is The Best Multi-Stage Architecture for Object Recognition?” *IEEE 12th International Conference on Computer Vision*, 2009.
- [5] JURASZEK, G. “Reconhecimento de Produtos por Imagem Utilizando Palavras Visuais e Redes Neurais Convolucionais”. Joinville: *UDESC*, 2014.
- [6] CUN, L.; BOSER, B.; DENKER, J. S.; HENDERSON, D.; HOWARD, R. E.; HUBBARD, W.; JACKEL, L. D. “Handwritten Digit Recognition with a Back-Propagation Network”. In: *Advances in Neural Information Processing Systems*. [S.l.]: Morgan Kaufmann, 1990. p. 396-404.
- [7] CIRESAN, D.; MEIER, U.; SCHMIDHUBER, J. “Multi-Column Deep Neural Networks for Image Classification”. In: *Proceedings of the 25th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2012)*. [S.l.: s.n.], 2012. p. 3642-3649.
- [8] KAVUKCUOGLU, K.; SERMANET, P.; BOUREAU, Y. Ian; GREGOR, K.; MATHIEU, M.; LECUN, Y. *Learning Convolutional Feature Hierarchies for Visual Recognition*. 2012.
- [9] SERMANET, P.; EIGEN, D.; ZHANG, X.; MATHIEU, M.; FERGUS, R.; LECUN, Y. *Overfeat: Integrated Recognition, Localization and Detection Using Convolutional Networks*. CoRR, abs/1312.6229, 2013.
- [10] RAZAVIAN, A. S.; AZIZPOUR, H.; SULLIVAN, J.; CARLSSON, S. *CNN Features Off-the-Shelf: An Astounding Baseline for Recognition*. CoRR, abs/1403.6382, 2014.
- [11] CONSTANTE, P.; GORDÓN, A.; CHANG, O.; PRUNA, E.; ESCOBAR, I.; ACUÑA, F. “Artificial Vision Techniques for Strawberry’s Industrial Classification”. *IEEE Latin America Transactions*, Vol. 14, Nº 6, 2016.
- [12] VLFEAT. *Biblioteca Open Source VLFeat*. 2016. URL: <http://www.vlfeat.org/matconvnet/models/beta16/imagenet-cafe-alex.mat>
- [13] LECUN, Y.; KAVUKCUOGLU, K.; FARABET, C. “Convolutional Networks and Applications in Vision”. In: *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*. [S.l.: s.n.], 2010. p. 253-256.
- [14] AREL, I.; ROSE, D.; KARNOWSKI, T. “Deep Machine Learning - A New Frontier in Artificial Intelligence Research [research frontier]”. *Computational Intelligence Magazine, IEEE*, v. 5, n. 4, p. 13-18, 2010. ISSN 1556-603X.
- [15] TEIXEIRA, A. “Desenvolvimento de uma Interface Gráfica para Classificadores de Imagem”. 2016. URL: <https://repositorio.ipcb.pt/bitstream/10400.11/1155/1/disserta%C3%A7ao.pdf>
- [16] LU, H.; HUANG, Y. CHEN, Y.; YANG, D. “Real-Time Facial Expression Recognition Based on Pixel Pattern-Based Texture Feature”. In: *Proc. Electronic Letters*, pp. 916-918, 2007.
- [17] TCHANGANI, A. “Support Vector Machines: A Tool for Pattern Recognition and Classification”. *Studies in Informatics & Control Journal* 14: 2. 99-110, 2005.
- [18] KOHAVI, R.; PROVOST, F. “Glossary of Terms”. *Machine Learning*, 30(2-3), 271-274, 1998.
- [19] VAPINIK, N. “The Nature of Statistical Learning Theory”. New York: *Springer*, 1995.
- [20] KIM, K. I.; JUNG, K.; PARK, S. H.; KIM, H. J. “Support Vector Machines for Texture Classification”. *IEEE Trans, PAMI*, 2002.
- [21] LU, H.; HUANG, Y. CHEN, Y.; YANG, D. “Real-Time Facial Expression Recognition Based on Pixel Pattern-Based Texture Feature”. In: *Proc. Electronic Letters*, pp. 916-918, 2007.
- [22] LOVELL, B.; WALDER, C. “Support Vector Machines for Business Applications”. *Business Applications and Computational Intelligence, Idea Group Publishers*, 2006.



Carlos Silva é analista de sistemas da Prefeitura Municipal de São Francisco de Assis, Rio Grande dos Sul, Brasil. Graduado em informática pela Universidade da Região da Campanha (URCAMP) em 2010 e especialista em engenharia de sistemas pela (ESAB) em 2012. Atualmente é mestrando em Engenharia Elétrica pela Universidade Federal do Pampa (UNIPAMPA). Suas pesquisas concentram-se em análise e processamento de imagens, aprendizado de máquina e visão computacional. Membro ativo da SBC e IEEE.



Daniel Welfer é professor da Universidade Federal de Santa Maria (UFSM) lotado no Departamento de Computação Aplicada (DCOM). Concluiu seu doutorado em Ciência da Computação pela Universidade Federal do Rio Grande do Sul (UFRGS) em 2011. Suas áreas de interesse são processamento e análise de imagens médicas, morfologia matemática e sistemas de informação hospitalar. Atualmente é membro permanente do Programa de Pós-Graduação em Informática (PPGI) da UFSM.



Francisco Paulo Gioda possui graduação em medicina veterinária pela Universidade Federal de Santa Maria (UFSM) em 1978. É o médico veterinário responsável do Sindicato Rural de São Francisco de Assis, Rio Grande do Sul, Brasil. Atualmente é secretário de administração e planejamento da cidade de São Francisco de Assis e coordenador do setor de registro de marcas de gado do município.



Claudia Dornelles possui graduação em administração pela Universidade Norte do Paraná (UNOPAR) em 2014. É auxiliar administrativo da Prefeitura de São Francisco de Assis, Rio Grande do Sul, Brasil e responsável pelo registro e controle de marcas de gado do município.

APÊNDICE G - ARTIGO SUBMETIDO NA REVISTA IBEROAMERICANA DE INTELIGÊNCIA ARTIFICIAL

Inteligencia Artificial 20(59) (2017), 123-127
doi: 10.4114/intartf.vol20iss59pp123-127



Pattern Recognition in Cattle Brand using Bag of Visual Words and Support Vector Machines Multi-Class

Please, if you are submitting for review do not include author's names
Please, if you are submitting for review do not include author's affiliations
The journal uses double blind review
please@if.you.are.submitting.for.review.do.not.include.email

Abstract The recognition images of cattle brand in an automatic way is a necessity to governmental organs responsible for this activity. To help this process, this work presents a method that consists in using Bag of Visual Words for extracting of characteristics from images of cattle brand and Support Vector Machines Multi-Class for classification. This method consists of six stages: a) select database of images; b) extract points of interest (SURF); c) create vocabulary (K-means); d) create vector of image characteristics (visual words); e) train and sort images (SVM); f) evaluate the classification results. The accuracy of the method was tested on database of municipal city hall, where it achieved satisfactory results, reporting 86.02% of accuracy and 56.705 seconds of processing time, respectively.

Resumen Las imágenes de reconocimiento de la marca de ganado de manera automática es una necesidad para los órganos gubernamentales responsables de esta actividad. Para ayudar a este proceso, este trabajo presenta un método que consiste en el uso de Bolsa de Palabras Visuales para la extracción de características a partir de imágenes de marca de ganado y Máquinas de Vectores de Soporte Multiclase para clasificación. Este método consta de seis etapas: a) seleccionar base de datos de imágenes; b) extraer puntos de interés (SURF); c) crear vocabulario (*K-means*); d) crear un vector de las características de la imagen (palabras visuales); e) entrenar y ordenar imágenes (SVM); f) evaluar los resultados de la clasificación. La exactitud del método fue probada en la base de datos del ayuntamiento municipal, donde logró resultados satisfactorios, reportando un 86,02% de precisión y 56,705 segundos de tiempo de procesamiento, respectivamente.

Keywords: Computer vision, Pattern recognition, Machine learning, Bag of Visual Words, Support Vector Machines Multi-Class.

Palabras clave: Visión por computadora, Reconocimiento de patrones, Aprendizaje de máquinas, Bolsa de palabras visuales, Máquinas de vectores de soporte multiclase.

1 Introduction

The automatic recognition of cattle brands is a strategic need, since it encompasses a productive industry of great socioeconomic relevance to Brazil. According to the Food and Agriculture Organization – FAO, among the producing countries, Brazil and India have the largest herds, Brazil being the 1st, with an average of 209,215,666 cattle heads [1]. Thus, livestock has a relevant role in social formation, and is still today an activity of great importance in cultural expressions associated to it, as it is integrated to the culture and the countryside way of life, and it also has a role in the affirmation and building of individual and group identities [2].

The use of brands or symbols on cattle presupposes the public recognition of its property by an individual or group. Used since the beginnings of the Iberian colonization in America, its institutionalization began to occur with its recording in official agencies, recognized holders of public legitimacy [2]. After these records followed

regulations that seek to legitimize branding, as well as regulate the manner and timing to do it, discriminate how the records are made, assign fees to the records, regulate the craft of the irons, and government taxes. Generally, cattle brand records involve books with the drawings of the brands and the identification of their owner. In Brazil attempts and investments to upgrade the cattle branding recording system were always subject of controversy, due to the opposition from agriculturists. A major part of their concern is associated to a fear of losing family brands and the meaning they acquired through time. Currently, brand recording in Brazil is performed by town offices, generally without a more effective systematization and without instituted renewals.

In face of the context presented, this work intends to present and assess a tool that performs automatic cattle brand recognition, with the goal of replacing the manual control of cattle branding performed today, in order to potentially decrease the possibility of duplicate records, reducing waiting times for the recording of new brands, improving governmental administration regarding the brand archive under its care and aid security officials in preventing cattle raiding crimes.

This research presents the application of an automatic computational method through a software tool for cattle brand recognition. This research was supported by the "X" City Hall, Rio Grande do Sul, Brazil. Therefore, the employees in the Cattle Branding Record Section and in the Data Processing Center of this township validated the suggested tool.

2 Related work

In general, we could not find any works in the literature review that report the use of a set of visual words for the recognition of cattle branding images.

Sanchez et al [3] present a tool for recognition of cattle branding that uses Hu and Legendre moments for extracting features of images in a grey scale, and also a classifier of k-nearest neighbors (k-NN). The authors used Hu and Legendre moments to source features that were not prone to rotation, translation, and scale transformations. The peak percentage of correct classification presented by the authors was 99.3%, with a significant decrease in accuracy as the number of classified images increased, however. Another result they presented was the processing time for the classification. Since a k-NN classifier was used for each new object that was meant to be classified, training data were used to check which objects of the database resembled the most the new object that was meant to be classified. The object is classified in the most common class to which the objects that resemble it the most belong. Thus, classification occurs by analogy. No classification model is created. Instead, for each new object to be classified, training data are scanned, and the suggested classifier becomes computationally expensive.

Differently from the research presented by Sanchez et al [3], the work we propose here intends to show results that can be generalized or reproduced, deploying state-of-the-art techniques for feature extraction and statistical classification of digital images, such as Bag of Visual Words (BoW) and Support Vector Machines (SVM), in order to create a "model" responsible for classifying and retrieving cattle branding images by their content, but with efficient results when applied to large databases. The method described in [3] presents a significant loss of efficiency (accuracy and speed) when applied to large numbers of images.

The BoW method is also commonly referred to as 'visual word dictionary'. This method can be classified as a Content-Based Image Retrieval System (CBIR). Torres e Falcão [4] show some usual approaches in CBIR systems, where a vector is extracted from the images based on features such as shape, texture, and color distribution. In a new query, a vector with the same features is extracted from searched image and compared to the other vectors of existent features in the database through a distance function.

In order to develop the BoW method, feature descriptors and key points extracted from the images are used. The key points are saliences that contain local information of the image and are automatically obtained through key point detection methods [5], [6]. Once detected, the key points are represented by descriptors, such as Invariant Feature Transform (SIFT) [7], Speed Up Robust Features (SURF) [8], among others. Thus, in the visual word dictionary, each visual word is associated to a cluster of key point descriptors. Therefore, each visual word represents a specific local pattern shared by all descriptors of a given clustering. Once the visual word dictionary is defined, it is possible to associate each key point descriptor with the nearest visual word. Each image is represented by a histogram that indicates the frequency that each visual word from the dictionary occurs in the image.

According to the literature, we can find several researches related to the BoW technique. The work proposed by Sivic and Zisserman [9] presents the technique as an approach for recovering all occurrences of an object in frames from a given video. In order to achieve this, the objects are represented as a set of descriptors that do not vary according to scale, rotation, translation, illumination, and partial occlusion.

Csurka *et al* [10] apply this technique to find a generic process to deal with several types of objects, and, at the same time, to handle the variations in illumination, viewing, rotation and occlusion, typical of real-world scenarios. The BoW model has shown an outstanding performance in a wide range of tasks, such as action recognition [11], texture [12], gestures [13], image classification [14], etc. The model was used for nudity detection in videos in a work by Lopes *et al* [15]. On the other hand, in Batista *et al* [16], the BoW methodology was used for the automatic identification of images that contain façades and buildings in the digitized collection of the Minas Gerais Public Archive.

Li *et al* [17] used BoW based on blocks for face recognition. Wang *et al* [18] applied BoW weighing the visual words in medical image retrieval. Alternatively, in Wang *et al* [19], the authors have developed an algorithm based on BoW for the classification of breast tissue density images in mammographies. Barata *et al* [20] suggested two systems for melanoma detection in dermatological images, in which the first system used global methods to classify skin lesions, and the second one used local features, and the BoW method to classify the images. Li *et al* [21] employed a SURF descriptor and spatial pyramid in a BoW methodology to enhance image recognition and classification.

3 Materials

The images from cattle branding presented in this research were provided by the “X” City Hall, in Rio Grande do Sul. We used 12 cattle branding images, each one of them composed by 45 sub-images (samples), totaling 540 samples from original images, but with size and orientation variations. We intended to identify patterns with the greatest independence possible from these variable factors. The images were provided in high resolution in the Portable Network Graphics format at a size of 600 x 600 pixels. The brands used are displayed in figure 1.

			
802	803	804	805
			
811	812	813	814
			
815	821	822	1093

Figure 1. Images of cattle branding used in this article.

Brand codes, owners and number of samples of each brand are displayed in table 1.

Table 1: Brands owners and total samplings by brand.

Branding	Owner	Total samplings
802	Owner "A"	45
803	Owner "B"	45
804	Owner "C"	45
805	Owner "D"	45
811	Owner "E"	45
812	Owner "F"	45
813	Owner "G"	45
814	Owner "H"	45
815	Owner "I"	45
821	Owner "J"	45
822	Owner "K"	45
1093	Owner "L"	45

For the implementation of the proposed tool, as well as image database storage, algorithm processing and viewing of the results, we used a personal computer with an CPU Intel Core i5-3330 3 GHz, RAM of 8 GB DDR3 1600 MHz, and GPU NVIDIA GTX 750 Ti. Furthermore, we used the MATLAB software with the Parallel Computing and Statistics and Machine Learning libraries.

4 Methods

The proposed method consists of six steps, which are: image database selection; extraction of points of interest using the SURF algorithm; development of a visual word dictionary with K-means clustering; development of histograms and vectors for image features; training and classification of images by Support Vector Machines Multi-Class; and, finally, evaluation of the classification results. Figure 2 shows a summarized flowchart of the proposed method.

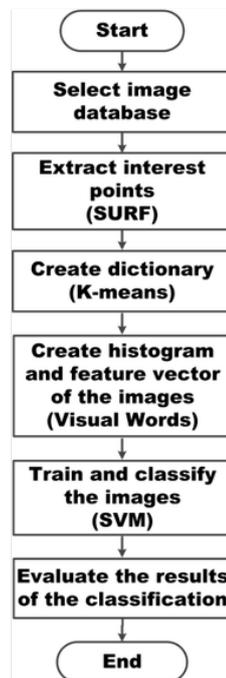


Figure 2. Summarized flowchart of the proposed method.

4.1 Image database

The image set used in the research is described on Section 3. With the application of the data augmentation technique, we generated 45 sub-images for each of the 12 brandings that were used in the experiments. Examples of sub-images generated with scale, translation and rotation variations originated from the brandings presented on figure 1 are illustrated on figure 3.

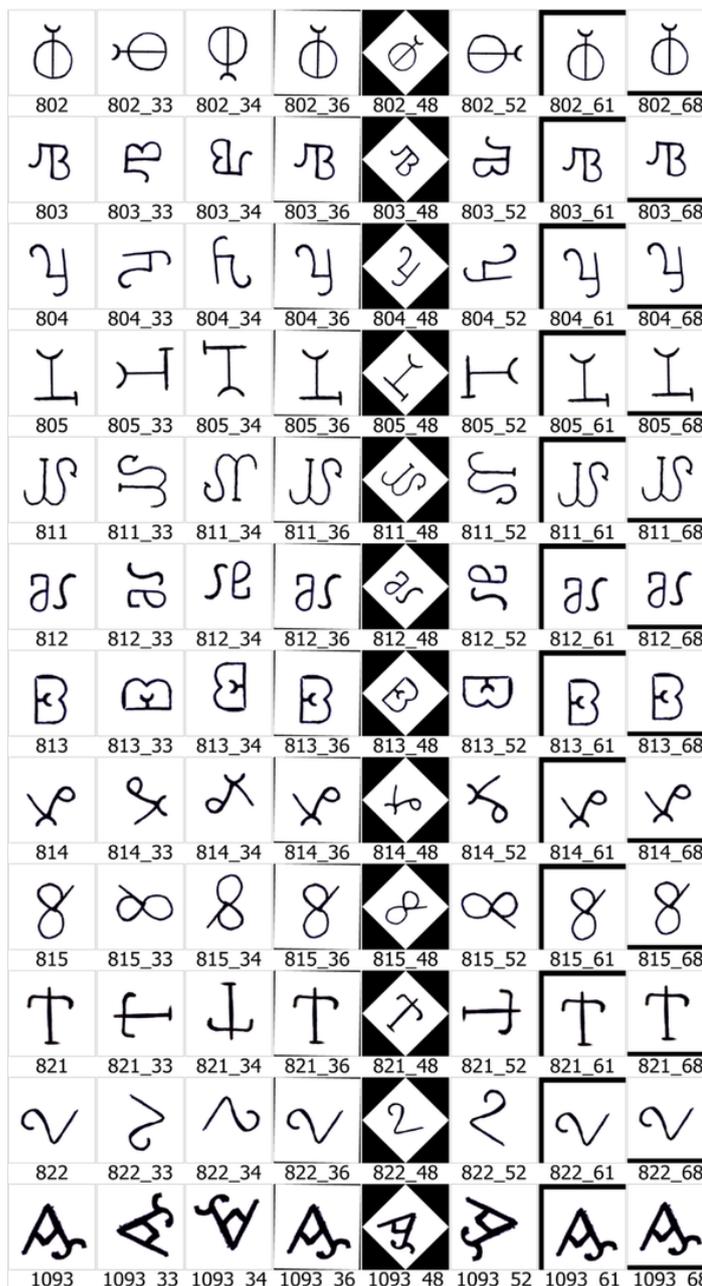


Figure 3. Examples of sub-images generated with scale, translation and rotation variations used in the experiments.

4.2 Extraction of points of interest from brands through the SURF algorithm

First, we locate an image dataset at the "X" City Hall FTP (File Transfer Protocol) server, in order to download the file, which contains a total of 12 brandings, and 540 sub-images. Next, this image database is instantiated using the MATLAB algorithm, thus the brandings were automatically sorted in categories. After sorting the brandings, the algorithm inspects each one of the created categories to check if the same number of images is available for each branding. In case the number is different, the branding category with the smallest amount of images is taken as a reference for the remaining ones. This process is conducted to balance the number of images in the training set for the next rounds of the method.

After instantiating the image database and sorting the brandings, we performed the extraction of points of interest from images through the SURF algorithm. This algorithm is based on the sum of the Haar 2-D wavelet answers and the usage of whole images to detect points of interest, and, for this reason, this algorithm is a robust local feature detector and descriptor. The SURF algorithm is used in several computer vision tasks, such as object recognition for 3D reconstruction. Although the SURF is inspired in the SIFT algorithm, SURF is quicker and more robust and also has the advantage of being an invariant region descriptor. In order to make the matching, SURF considers the Laplacian sign, or, in other words, the Hessian matrix trace. Figure 4 illustrates the extraction of 6,751 points of interest (key points) of brand "1093" using the SURF algorithm.

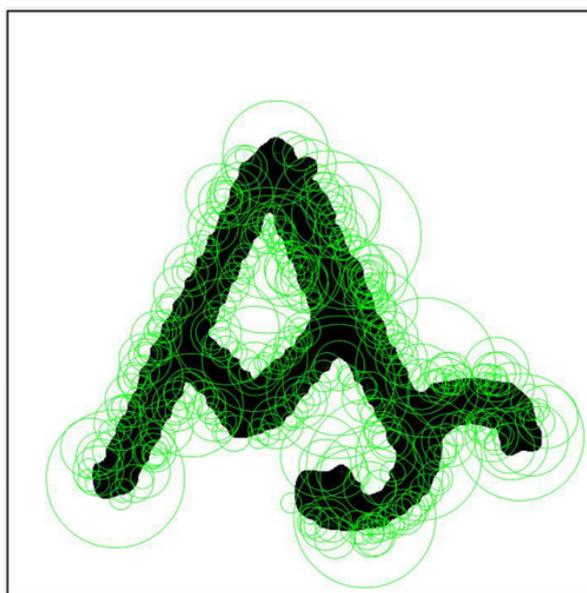


Figure 4. Extraction of 6,751 points of interest of brand "1093" by using the SURF algorithm.

The SURF algorithm was applied to the set of cattle brand images from the samples. Firstly, the algorithm detects the points of interest, and, later, it calculates the features in these points. First, the method finds features with prominent positions. The detection of those prominences is based on a Fast-Hessian multiscale and multi-orientation detector, with a descriptor based on the distribution of grey level change.

4.3 Development of a word dictionary using K-means

Clustering is a method used for grouping objects according to their similar features. The adopted method for clustering in the proposed work was K-means. The definition of the size of the word dictionary when applying the K-means method is one of the fundamental points in the development of the dictionary. The choice of dictionary size (number of clusters) is made empirically, usually after successive algorithm executions searching for the best sensitivity [22].

This choice is vital, for, besides its influence in the discriminatory ability of the dictionary, it also directly influences computational efficiency and memory usage during the processes of dictionary generation and image classification. Furthermore, the vectors of the features (visual words) that form this dictionary have high dimensionality, thus complicating the application of algorithm processing.

After conducting experiments to assess algorithm performance and recognition rate, the proposed clustering size of the presented work was 500. The experiments were conducted by attributing arbitrary values to the number of clusters used for creating the visual words dictionary. The definition of the ideal size of the dictionary for the proposed problem came from the observation of the tool accuracy rates. For dictionaries with less than 500 words, there was a considerable loss of General Accuracy. On the other hand, in case of dictionaries with larger clusters than what we propose in this research, the computational cost was much higher, which significantly elevated the processing time of the algorithm while performing the task of cattle branding recognition, without bringing any benefit to the general precision of the proposed tool.

Once the size of the visual word dictionary was defined, it was possible to associate each key point descriptor to the nearest visual word. Therefore, each image was represented by a histogram indicating how often each visual word occurs in the dictionary (an analogous process to what is done in textual information retrieval). We should mention that the K-means algorithm was applied only to the training image set, that is, a total of 14 images. The dictionary size can be verified in the X axis in figure 5.

4.4 Elaboration of a histogram and feature vector of visual words

After creating the visual word dictionary, the descriptors of all images were extracted from the training base, identifying to what visual word the descriptor belongs, and creating a histogram containing the amount of each visual word in the analyzed image. This histogram is a vector where each position corresponds to a visual word and its corresponding value to the amount of visual words from that kind of image. Processing was performed in all training images and the result is a histogram for each image. In the proposed method, the vectors with the image histograms are created by using the encode method of the `bagOfFeatures` class from MATLAB. Figure 5 shows the histogram with the number of visual word occurrences and the size of the dictionary.

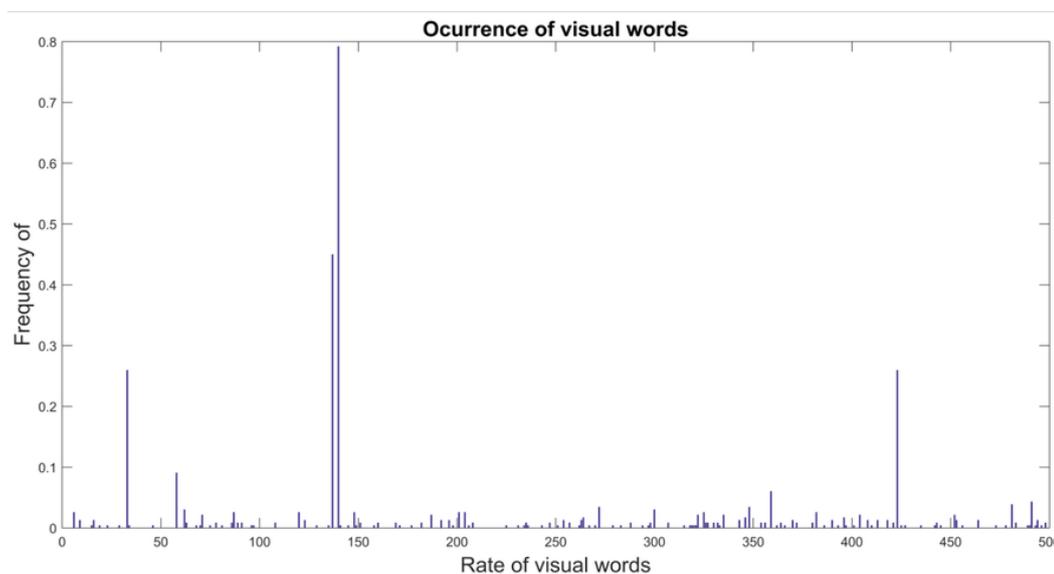


Figure 5. Histogram with the number of visual word occurrences and the size of the dictionary.

Figure 6 shows the performance of the Bag of Visual Words method applied in the proposed research.

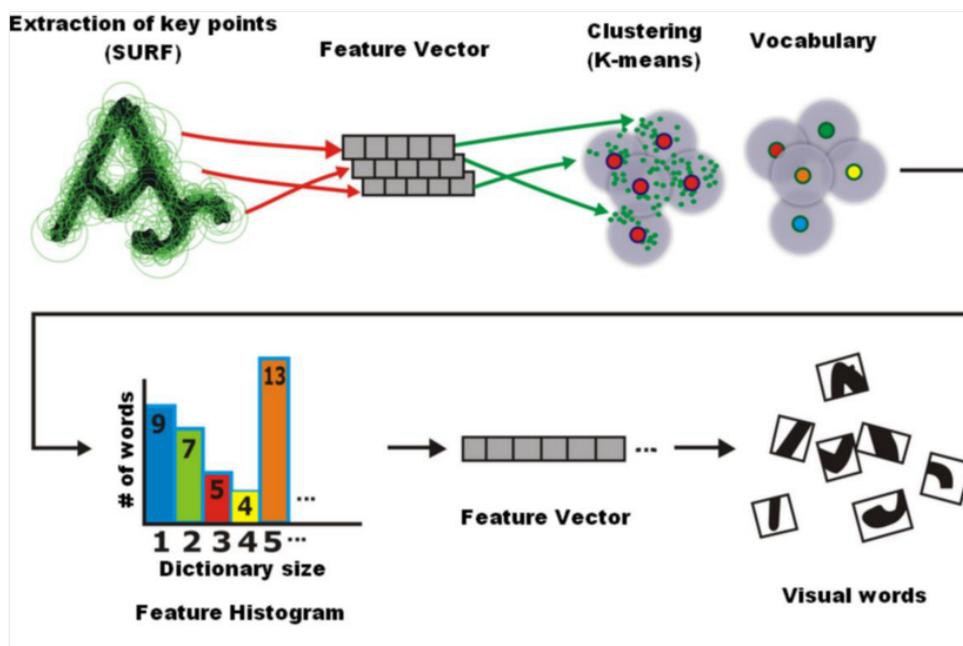


Figure 6. Bag of Visual Words method applied in the proposed research.

4.5 Training and classification of images with support vector machines

The fifth step of the method is the training of a linear classifier, to make it possible to determine the category to which the image belongs through the visual word histogram. The model for automatic learning adopted in the presented work was the Support Vector Machine (SVM) supervised classifier. Support Vector Machine is a classification algorithm known for its success in a wide range of applications. SVMs are one of the most popular approaches for data modeling and classification. Its advantages include their outstanding capacity for generalization, concerning the ability to correctly sort the samples that are not in the feature space used for training [23]. Considering two classes and a set of points attributed to these classes, the SVM determines the hyperplane that separates the points so that the higher number of points from the same class is placed in the same side, maximizing the distance from each class to that hyperplane, consequently being denominated as a maximum margin classifier [24]. Indeed, a great margin between the values corresponding to points from two data sub-sets entails a further minimized generalization risk of the classifier.

SVMs are used to classify and recognize patterns in several types of data; they are also employed in a wide range of applications, such as face recognition, clinical diagnosis, industrial process monitoring, and image processing and analysis [25]. In the proposed research, we randomly separated the set of images in two parts, where one of these parts was used for the training stage and the other for the validation stage, thus eliminating polarization in the results. The final result is the average of the results obtained in the validation stage. The percentage division used here was 30% for training and 70% for validation, in order to reduce the risk of overfitting, in which the sorting mechanism excessively adjusts to the peculiarities of the training set. Furthermore, this division also intends to evaluate the adaptability and robustness of the proposed tool regarding to recognition of patterns and the correct classification of large volumes of images, even in situations where the training set has a smaller number of samples when compared to the validation set.

4.6 Assessment of classification results – confusion matrix

The confusion matrix of a classifier indicates the number of correct classifications versus the predictions made in each case based on a group of examples. In this matrix, the lines depict the actual cases and the columns depict the predictions made by the model. Through the confusion matrix, it is possible to find information related to the number of correctly and incorrectly classified images for each group of samples. This is an $A \times A$ matrix, where A

is the amount of categories to which we apply the classifier. In our case, the experiment conducted included 12 brands, so we have a 12x12 confusion matrix in this situation.

5 Results and Discussion

Through the results found we were able to assess the proposed method. The assessment of experiment results was performed based on the recognition rate obtained in the confusion matrix rendered from the classification attained in the validation stage. Furthermore, the total processing time of the proposed method was also checked.

Figure 7 presents the confusion matrix for the best result obtained in the experiments, with a recognition rate of up to 86.02%. Through the analysis of the main diagonal we can observe that the correct classification rate is emphasized in two brandings, "802" and "803", in which the percentage of correctness reached 93.55%. We can also observe that the brandings with the lowest correct classification rates were "812", "814", and "822", with a percentage of 74.19%, 77.42% and 77.42%, in this order. The remaining brandings, "804"; "805"; "811"; "813" "815", "821" and "1093", had a correct classification rate of 83.87%; 90.32%; 87.09%; 83.87%, 90.32%, 90.32%, and 90.32%, respectively.

The hypothesis of wrong classification of cattle branding as shown in the confusion matrix may be associated to the complexity of the samples and to the size of the dictionary adopted for clustering. However, it is important to remark that a larger dictionary directly affects the algorithm performance.

In general, the image samples of brandings with a better descriptive power and better quality correctly classified more images, since they include more characteristics when compared to brandings with worse sample quality, and, consequently, less features extracted. The capacity of recognizing patterns in an image over a set of images depends on the amount of a priori information available about the object in question.

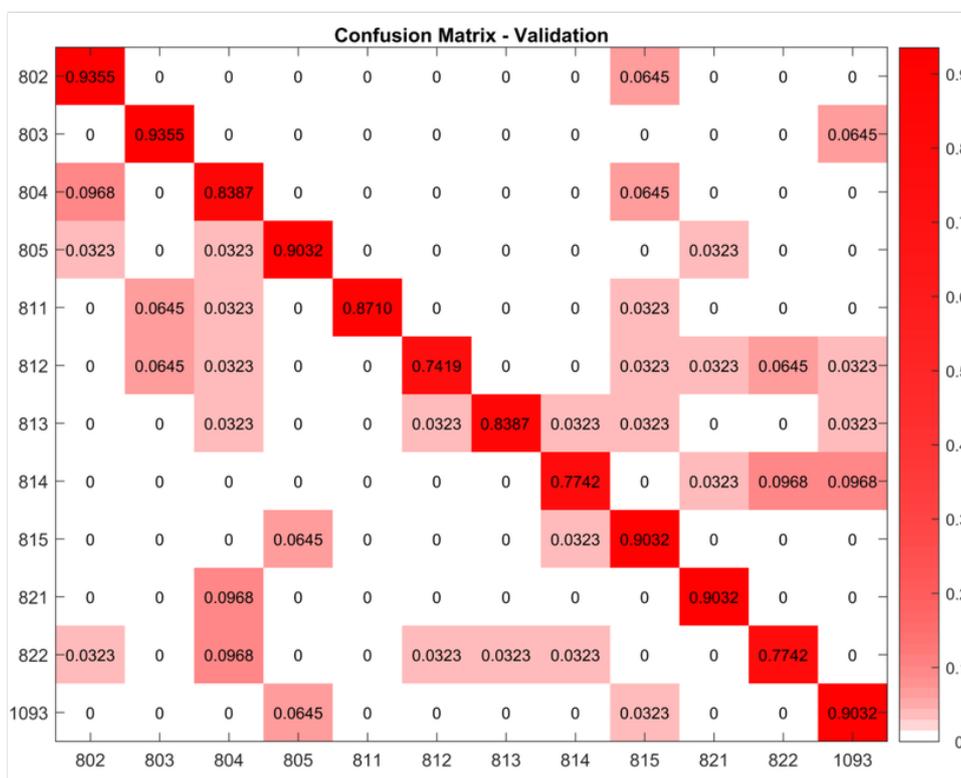


Figure 7. Confusion matrix obtained in the validation stage.

Figure 8 presents an illustration with the accuracy percentage of the brandings that were correctly classified in the experiment.

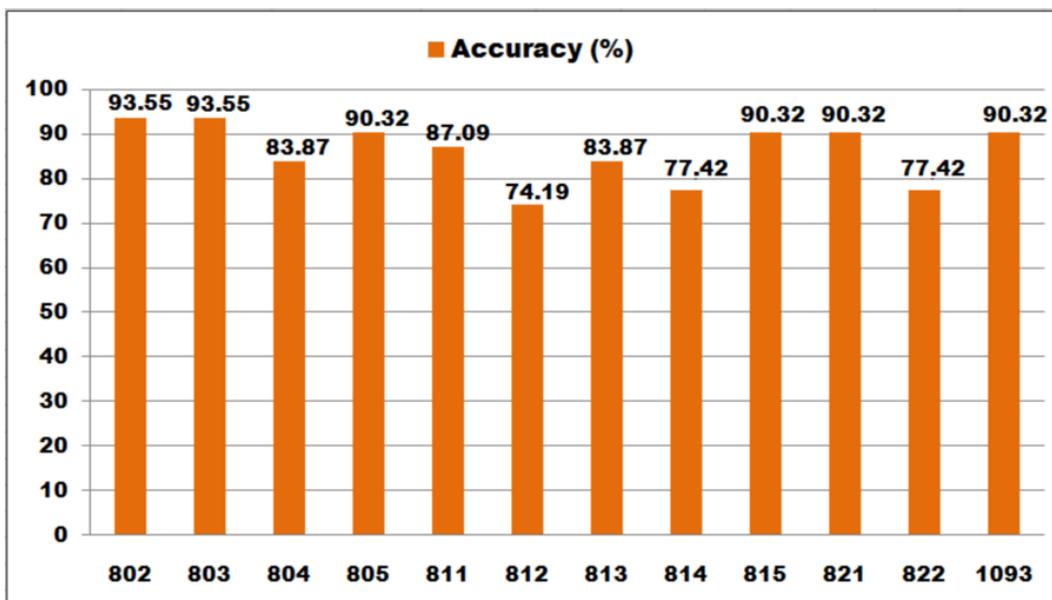


Figure 8. Percentage of correctly classified brandings.

The orange bars represent the percentage of correctly classified cattle brandings (accuracy). Amongst the 12 brandings we analyzed, 2 presented 93.55% of correct classifications, namely, "802" and "803". The cattle brandings with the lowest correct percentage were "812", "814", and "822", with 74.19%, 77.42% and 77.42%, respectively.

For the evaluation of the proposed tool, we also used the Precision, Recall and F1-score measures, commonly used in the recovery of information and applications computational vision applications. These metrics are calculated based on the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), from the confusion matrix generated by the experiment.

Precision and Recall are measurements originated from Information Recovery and used in Classification when working with non-balanced classes. Precision is the percentage of instances that were correctly classified as positive among all of the data that were classified as positive, while Recall is the percentage of instances that were correctly classified as positive among the ones that really were positive, and F1-score is the harmonic mean between precision and recall [26]. The advantage of the F1-score is that it offers only one quality metric, facilitating a better understanding for end users.

Precision metrics are calculated according to Equation 1, the obtained Recall through Equation 2, and the F1-score defined by Equation 3.

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \tag{3}$$

Table 2 demonstrates the results of Precision, Recall, and F1-score obtained in the experiment, during the validation step. The results suggest that the proposed tool have achieved satisfactory results regarding the recognition of cattle branding images.

Table 2: Precision, Recall, and F1-score results obtained during the validation step.

Branding	Precision (%)	Recall (%)	F1
802	93.55	85.29	0.8923
803	93.55	87.88	0.9062
804	83.87	72.22	0.7761
805	90.32	87.50	0.8889
811	87.10	100.00	0.9310
812	74.19	92.00	0.8214
813	83.87	96.30	0.8966
814	77.42	88.89	0.8276
815	90.32	77.78	0.8358
821	90.32	90.32	0.9032
822	77.42	82.76	0.8000
1093	90.32	80.00	0.8485

The processing time of the algorithm based on the number of cattle branding samples is shown in figure 9. The processing times of the proposed method were measured for the classification of five sample groups, respectively. Each group contained 108; 216; 324; 432; and 540 images. The processing times for the classification of the images in each group was 14.341s; 28.605s; 39.123s; 47.039s, and 56.705s, in this order. When analyzing the graphic in figure 9, we notice that the processing time observed in the Y-axis varies in direct proportion of the increase of samples of cattle branding classified in the X-axis, i. e., a pattern of linear growth of the function is observed, even if the growth rate is not exactly a constant number.

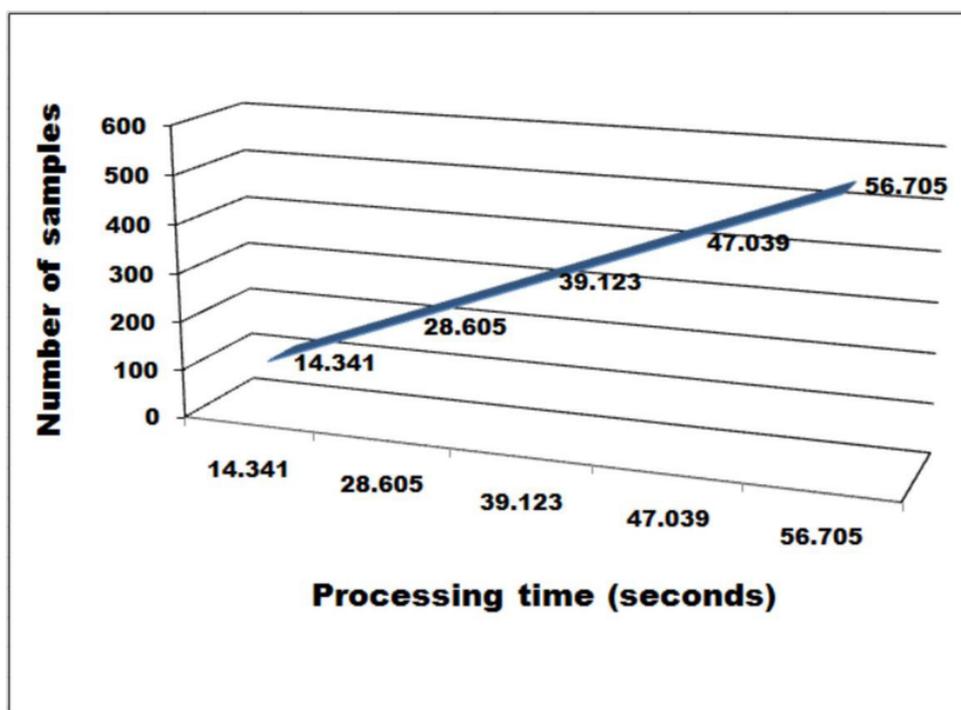


Figure 9. Processing time of the algorithm according to number of samples.

The results obtained with the experiments performed with 12 cattle brandings and 540 image samples obtained a recognition rate of 86.02%, an error rate of 13.98%, and a total processing time of 56.705 seconds.

The recognition rate was obtained through the calculation of the arithmetic average of brands properly classified from the confusion matrix, and the total processing time was obtained with the use of the MATLAB software which, at the end of code processing, performs the breakdown of the algorithm processing speed.

6 Conclusions and Future Research

In this research, we presented an automated method for the recognition of cattle branding, using a set of visual words. The project was developed and conducted in two institutions: “X” City Hall and Federal University “Y”. In the experiments conducted in this research, the Bag of Visual Words method was used, in which was applied the SURF algorithm to extract the points of interest from the images, the K-means clustering to group the histograms, and an SVM supervised classifier to classify the cattle branding images. The experiments were all conducted using the cattle branding base provided by the “X” City Hall. The usage of the suggested method delivered a mean accuracy of 86.02% and an algorithm processing time of 56.705 seconds for the 12 assessed brands, in a total of 540 samples.

The deployed method has effectively and efficiently performed the recognition of different cattle brands, but its main limitation was the empirical definition of the size of the dictionary defined for the clusters. Several experiments were performed with different dictionary sizes in order to obtain an optimal value, which could maintain satisfactory recognition levels and reasonable computational cost and processing time. The next step of this research is related to the improvement of the applied method striving for more accurate results.

In general, there is no specific method in the Literature for determining the exact value of a dictionary size for the clusters, but it is possible to obtain a more precise estimation through some techniques, such as cross-validation, silhouette method, Gaussian-means (G-means) algorithm, among others. For that matter, in future works, we aim to conduct new experiments with methods that allow for a more appropriate choice of number of clusters of K-means algorithm, in order to obtain better results both for precision and processing time. Furthermore, we intend to test other techniques for extracting points of interest from cattle images, such as the BRISK, FREAK and ORB algorithms, in order to obtain faster results when compared to the SURF algorithm, that we used in this researched.

References

- [1] Secretaria do Planejamento e Desenvolvimento Regional – *Governo do Estado do Rio Grande do Sul*, Brasil. 2015. URL: <http://www.scp.rs.gov.br>. Accessed 19 July 2015.
- [2] Arnoni, R. Os Registros e Catálogos de Marcas de Gado da Região Platina. Pelotas: *Revista Memória em Rede da UFPEL*, 2013.
- [3] Sanchez, G.; Rodriguez, M. Cattle Marks Recognition by Hu and Legendre Invariant Moments. *ARPN Journal of Engineering and Applied Sciences*, Vol. 11, Nº 1, 2016.
- [4] Torres, R.; Falcão, A. Content-based image retrieval: Theory and applications. *Revista de Informática Teórica e Aplicada*, v. 13, p. 161-185, 2006.
- [5] Lindeberg, T. Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention. *International Journal of Computer Vision*, 11: 283-318, 1993. doi: 10.1007/BF01469346
- [6] Lazebnik, S.; Schmid, C.; Ponce, J. Affine-invariant local descriptors and neighborhood statistics for texture recognition. In *Proc. ICCV*, pages 649-655, 2003. doi: 10.1109/ICCV.2003.1238409
- [7] Lowe, D. Object recognition from local scale-invariant features. *Computer Vision, IEEE International Conference on*, 2:1150, 1999. doi: 10.1109/ICCV.1999.790410
- [8] Bay, H.; Tuytelaars, T.; Gool, L. Surf: Speeded up robust features. In *ECCV*, pages 404-417, 2006. doi: 10.1007/11744023_3
- [9] Sivic, J.; Zisserman, A. Video google: A text retrieval approach to object matching in videos. *Proceedings of the Ninth IEEE International Conference Computer Vision*, 2:1470, 2003. doi: 10.1109/ICCV.2003.1238663
- [10] Csurka, G.; Dance, C.; Fan, L.; Willamowski, J.; Bray, C. Visual categorization with bags of keypoints. In *Workshop on Statistical Learning in Computer Vision, ECCV*, pages 1-22, 2004.

- [11] Ullah, M.; Parizi, S.; Laptev, I. Improving bag-of-features action recognition with non-local cues. In: Labrosse, F.; Zwiggiellar, R.; Liu, Y.; Tiddeman, B. (Ed.). *Proceedings of the British Machine Vision Conference* [S.l.]: BMVA Press, 2010.
- [12] Zhang, J.; Marszalek, M.; Lazebnik, S.; Schmid, C. Local features and kernels for classification of texture and object categories: A comprehensive study. In. *J. Computer Vision*, 2007. v. 73, n. 2, p. 213-238, 2007. doi: 10.1007/s11263-006-9794-4
- [13] Dardas, N.; Chen, Q.; Georganas, N. Hand gesture recognition using bag-of-features and multi-class support vector machine. *Proceedings of IEEE International Symposium on Haptic Audio-Visual Environments and Games*, p. 1-5, 2010. doi: 10.1109/HAVE.2010.5623982
- [14] Nowak, E.; Jurie, F.; Triggs, B. Sampling strategies for bag-of-features image classification. In: *Computer Vision. [S.l.]: 9th European Conference on Computer Vision*, v. 3954, p. 490-503, 2006. doi: 10.1007/11744085_38
- [15] Lopes, A.; Avila, S.; Peixoto, A.; Oliveira, R. Coelho; Araújo, A. Nude detection in vídeo using bag-of-visual-features. In: *SIGGRAP '09 Proceedings of the 2009 XXII Brazilian Symposium on Computer Graphics and Image Processing*. [S.l.: s.n.], 2009. doi: 10.1109/SIBGRAP.2009.32
- [16] Batista, N.; Lopes, A.; Araújo, A. Vocabulários visuais aplicados à detecção de edifícios em fotografias históricas. In: *XXXV Conferência Latinoamericana de Informática (Latin-American Conference on Informatics)*, CLEI, Pelotas, RS, Brazil. [S.l.: s.n.], 2009.
- [17] Li, Z.; Imai, J.; Kaneko, M. Robust face recognition using block-based bag of words. In: *2010 International Conference on Pattern Recognition*. [S.l.: s.n.], 2010. doi: 10.1109/ICPR.2010.320
- [18] Wang, J.; Li, Y.; Zhang, Y.; Wang, C.; Xie, H.; Chen, G.; Gao, X. Bag-of-features based medical image retrieval via multiple assignment and visual words weighting. *IEEE Trans. Med. Imaging*, v. 30, n. 11, p. 1996-2011, 2011. doi: 10.1109/TMI.2011.2161673
- [19] Wang, J.; Li, Y.; Zhang, Y.; Xie, H.; Wang, C. Bag-of-features based classification of breast parenchymal tissue in the mammogram via jointly selecting and weighting visual words. In: *ICIG '11 Proceedings of the 2011 Sixth International Conference on Image and Graphics*. [S.l. s.n.], 2011. doi: 10.1109/ICIG.2011.192
- [20] Barata, C.; Marques, J.; Mendonça, T. Bag-of-features classification model for the diagnose of melanoma in dermoscopy images using color and texture descriptors. *10th International Conference, ICIAR 2013*, Póvoa do Varzim, Portugal, June 26-28, v. 7950, p. 547-555, 2013. doi: 10.1007/978-3-642-39094-4_62
- [21] Li, K.; Wang, F.; Zhang, L. A new algorithm for image recognition and classification based on improved bag of features algorithm. *Optik*, 2016. v. 127, p. 4736-4740, 2016. doi: 10.1016/j.ijleo.2015.08.219
- [22] Ferraz, C. Novos descritores de textura para localização e identificação de objetos em imagens usando Bag-of-Features. Tese de Doutorado – Programa de Pós-Graduação em Engenharia Elétrica – Escola de Engenharia de São Carlos da Universidade de São Paulo, 2016.
- [23] Teixeira, A. Desenvolvimento de uma Interface Gráfica para Classificadores de Imagem. 2016. URL: <https://repositorio.ipcb.pt/bitstream/10400.11/1155/1/disserta%C3%A7ao.pdf>. Accessed 15 July 2015.
- [24] Lu, H.; Huang, Y. Chen, Y.; Yang, D. Real-Time Facial Expression Recognition Based on Pixel Pattern-Based Texture Feature. In: *Proc. Electronic Letters*, pp. 916-918, 2007. doi: 10.1049/el:20070362
- [25] Tchangani, A. Support Vector Machines: A Tool for Pattern Recognition and Classification. *Studies in Informatics & Control Journal* 14: 2. 99-110, 2005.
- [26] Huang, H.; Xu, H.; Wang, X.; Silamu, W. Maximum F1-Score Discriminative Training Criterion for Automatic Mispronunciation Detection. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, v. 23, n. 4, p. 787-797, 2015. doi: 10.1109/TASLP.2015.2409733

Anexos

ANEXO A - CONJUNTO DE IMAGENS DE MARCAS DE GADO UTILIZADO NO TRABALHO

										
802	803	804	805	811	812	813	814	815	821	822
										
823	824	825	831	832	833	834	835	841	842	843
										
844	845	851	853	854	855	861	862	863	864	865
										
942	943	944	1081	1082	1083	1084	1085	1086	1087	1088
										
1089	1090	1091	1092	1093	1094	1095	1096	1097	1098	1099
										
1100	1101	1102	1103	1104	1105	1106	1107	1364	1365	1368
										
1369	1370	1371	1372	1373	1374	1375	1376	1377	1378	1379
										
1380	1381	1382	1383	1384	1385	1386	1387	1388	1389	1390
										
1391	1392	1393	1394	1395	1396	1397	1398	1399	1400	1401
										
1402	1404	1405	1406	1407	1408	1409	1450	1451	1452	1454
										
1455	1456	1457	1458	1459	1460	1461	1462	1463	1464	1465
										
1466	1467	1468	1469	1471	1472	1473	1474	1475	1476	1477
										
1478	1479	1480	1481	1482	1483	1484	1485	1486	1487	1488
										
1489	1770	1771	1772	1773	1774	1775	1776	2697	1777	1778
										
1779	1780	1782	1783	1784	1785	1786	1787	1788	1789	1790

Fonte: Setor de Registro de Marcas de Gado - São Francisco de Assis (2016).

										
1791	1793	1794	1795	1796	1797	1798	1799	1800	1802	1803
										
1804	2306	2307	2308	2310	2311	2312	2313	2314	2315	2316
										
2317	2318	2319	2320	2321	2322	2323	2324	2325	2326	2327
										
2329	2330	2331	2332	2333	2334	2384	2386	2387	2388	2390
										
2391	2392	2393	2394	2396	2397	2398	2399	2400	2404	2405
										
2406	2408	2410	2411	2412	2414	2415	2416	2417	2418	2420
										
2421	2422	2423	2424	2426	2427	2428	2429	2430	2432	2433
										
2434	2435	2436	2437	2438	2439	2440	2441	2443	2444	2445
										
2446	2447	2448	2449	2450	2451	2452	2453	2455	2456	2457
										
2458	2459	2460	2462	2463	2464	2465	2466	2467	2469	2470
										
2471	2472	2473	2474	2476	2477	2478	2479	2480	2481	2483
										
2484	2485	2486	2487	2488	2490	2491	2492	2494	2495	2496
										
2498	2499	2500	2502	2505	2506	2507	2509	2510	2511	2511A
										
2512	2513	2514	2515	2516	2517	2518	2519	2520	2521	2523
										
2524	2525	2526	2527	2528	2529	2530	2531	2532	2533	2534

Fonte: Setor de Registro de Marcas de Gado - São Francisco de Assis (2016).

2535	2536	2537	2538	2539	2540	2541	2542	2543	2544	2545
2546	2547	2548	2549	2550	2551	2552	2553	2554	2556	2557
2558	2559	2560	2561	2562	2563	2564	2565	2566	2567	2568
2569	2570	2571	2572	2573	2574	2575	2576	2577	2578	2579
2580	2581	2582	2585	2586	2587	2588	2590	2591	2592	2593
2594	2596	2597	2598	2599	2600	2602	2603	2604	2605	2606
2607	2608	2609	2610	2611	2612	2613	2614	2615	2616	2617
2618	2619	2620	2621	2622	2623	2624	2625	2626	2627	2628
2629	2630	2631	2632	2633	2634	2635	2636	2637	2638	2639
2640	2641	2642	2643	2644	2645	2646	2648	2649	2650	2651
2652	2653	2654	2665	2666	2667	2668	2669	2670	2671	2672
2673	2674	2675	2676	2677	2678	2679	2680	2681	2682	2683
2684	2685	2687	2690	2691	2692	2693	2694	2695	2696	2698
2699	2700	2701	2702	2703	2704	2705	2706	2707	2708	2709
2710	2711	2712	2713	2714	2715	2716	2717	2718	2719	2720
2721	2722	2723	2724	2725						

Fonte: Setor de Registro de Marcas de Gado - São Francisco de Assis (2016).