
**Geração de imagens artificiais e extração de
características latentes aplicadas à classificação de
imagens**

Gabriela Salvador Thumé

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Gabriela Salvador Thumé

Geração de imagens artificiais e extração de características latentes aplicadas à classificação de imagens

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestra em Ciências – Ciências de Computação e Matemática Computacional. *EXEMPLAR DE DEFESA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientador: Prof. Dr. Moacir Antonelli Ponti

USP – São Carlos
Janeiro de 2016

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados fornecidos pelo(a) autor(a)

T634g Thumé, Gabriela Salvador
Geração de imagens artificiais e extração de
características latentes aplicadas à classificação
de imagens / Gabriela Salvador Thumé; orientador
Moacir Antonelli Ponti. - São Carlos - SP, 2016.
82 p.

Dissertação (Mestrado - Programa de Pós-Graduação
em Ciências de Computação e Matemática Computacional)
- Instituto de Ciências Matemáticas e de Computação,
Universidade de São Paulo, 2016.

1. Modelo. 2. Monografia de qualificação.
3. Dissertação. 4. Tese. 5. Latex. I. Ponti, Moacir Antonelli, orient. II. Título.

Gabriela Salvador Thumé

**Artificial images generation and latent feature extraction
applied to image classification task**

Master dissertation submitted to the Instituto de Ciências Matemáticas e de Computação – ICMC-USP, in partial fulfillment of the requirements for the degree of the Master Program in Computer Science and Computational Mathematics. *EXAMINATION BOARD PRESENTATION COPY*

Concentration Area: Computer Science and Computational Mathematics

Advisor: Prof. Dr. Moacir Antonelli Ponti

USP – São Carlos
January 2016

AGRADECIMENTOS

“”

()

RESUMO

THUMÉ, G. S.. **Geração de imagens artificiais e extração de características latentes aplicadas à classificação de imagens.** 2016. 82 f. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação (ICMC/USP), São Carlos – SP.

Cada imagem pode ser representada como uma combinação de diversas características, como por exemplo o histograma de intensidades de cor ou propriedades de textura da imagem. Essas características compõe um vetor multidimensional que representa a imagem. É comum esse vetor ser dado como entrada para um método de classificação de padrões que, após aprender através de diversos exemplos, pode indicar o quanto bem separadas as classes de imagens estão. Estudos sugerem evidências de que a preparação das imagens – por meio da especificação cuidadosa da aquisição, pré-processamento e segmentação – pode impactar significativamente a classificação. Além da falta de tratamento das imagens antes da extração de características, o desbalanceamento de classes também se apresenta como um obstáculo para que a classificação seja satisfatória. Imagens possuem características que podem ser exploradas para melhorar a descrição dos objetos de interesse. Portanto, é proposta a melhoria da classificação utilizando métodos de pré-processamento para a geração de imagens artificiais. Busca-se assim promover o balanceamento de bases de dados cujo número de exemplos de cada classe é desbalanceado. Esse estudo analisa a visualização do espaço de características após os métodos de geração artificial de imagens e de interpolação das características extraídas das imagens originais (SMOTE), contracenando com o espaço original. A ênfase se dá na facilidade de observação da importância do rebalanceamento das classes quando comparado com valores de métricas estatísticas, como a acurácia da classificação.

Porque deep learning? Mais descritores de forma e textura (dependentes da base) Explicar melhor o comportamento do SMOTE x ACC Bases maiores Algoritmos genéticos para geração de imagens artificiais?

Palavras-chave: Modelo, Monografia de qualificação, Dissertação, Tese, Latex.

ABSTRACT

THUMÉ, G. S.. **Geração de imagens artificiais e extração de características latentes aplicadas à classificação de imagens.** 2016. 82 f. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação (ICMC/USP), São Carlos – SP.

melhorar muito o abstract

The canonical steps of digital image processing can be summarized as acquisition, preprocessing, segmentation, feature extraction, transformation and recognition. Although the large number of methods available for image classification, only a subset of those steps are explored, specially related to feature extraction and recognition. The remaining steps are generally neglected. Meanwhile, some studies found evidence that image preparation — by means of an accurate specification of the acquisition, preprocessing and segmentation steps — can improve the classification results. Images captured from different kinds of sources like personal cameras, microscope, telescope and tomography, have features that can be explored to extend the description of objects. Therefore, we propose to improve image classification by using preprocessing methods. Among the possibilities for improvements we can highlight: using both image enhancement and restoration to extract latent features — non-visible characteristics in the original image; latent features learning by means of convolutional neural networks and restricted Boltzmann machines; and synthesis of random artificial images, to be used for database balancing when the number of samples from each class is originally unbalanced. This research has preliminary results which points to its potential. It aims to contribute to the research area by investigating methods to improve image classification while obtaining better feature spaces.

Key-words: Template, Qualification monograph, Dissertation, Thesis, Latex.

LISTA DE ILUSTRAÇÕES

Figura 1 – Etapas canônicas do reconhecimento de padrões desde a aquisição da imagem até sua posterior classificação.	31
Figura 2 – Visualização pixelizada de uma imagem da base COREL-1000.	36
Figura 3 – Sobre a imagem RGB de entrada foram realizadas operações de borramento, realce e de equalização de histograma. A imagem à direita é resultante dessas operações.	37
Figura 4 – Exemplo de filtragem gaussiana como operação de pré-processamento. . . .	39
Figura 5 – A imagem original, já em escala de cinza, foi realçada utilizando o método <i>unsharp masking</i>	40
Figura 6 – Conversão para a escala de cinza com os métodos de Intensidade e MSB. . .	42
Figura 7 – Características latentes de algas verdes.	45
Figura 8 – Componentes típicos de uma camada de uma rede neural de convolução. . .	47
Figura 9 – SMOTE: interpolação entre dois exemplos vizinhos no espaço de características. .	52
Figura 10 – Geração artificial a classe minoritária para rebalancear as classes.	58
Figura 11 – Geração artificial utilizando borramento com filtro bilateral.	59
Figura 12 – Geração artificial utilizando unsharp masking.	60
Figura 13 – Geração artificial utilizando adição de ruído de Poisson.	61
Figura 14 – Geração artificial utilizando o método SMOTE no espaço visual.	63
Figura 15 – Geração artificial utilizando uma mistura ponderada de duas imagens. . .	64
Figura 16 – Geração artificial utilizando uma mistura limiarizada de duas imagens. .	64
Figura 17 – Geração artificial utilizando uma mistura de duas imagens a partir da saliência da primeira imagem.	66
Figura 18 – Geração artificial utilizando uma composição de imagens.	67
Figura 19 – Fluxo de operações do experimento realizado.	69
Figura 20 – Base de imagens COREL-1000.	70
Figura 21 – Remoção de 50% das imagens de treino da classe Cavalo.	71
Figura 22 – Comparação dos exemplos de treinamento da geração com SMOTE e no campo visual. Em laranja estão representados os novos exemplos, projetados no plano da base original balanceada.	72
Figura 23 – Resultado do teste da classificação com 1-NN após o treinamento realizado com as bases rebalanceadas. A cor no interior dos marcadores quadrados representa a classe real dos exemplos e a borda representa a classe predita pelo classificador.	73

Figura 24 – Região de decisão com K-NN (K = 1)	74
Figura 25 – Melhores subespaços encontrados após a geração de novos exemplos para o SMOTE e para a geração artificial de imagens, e após a remoção de imagens para a projeção dos dados desbalanceados.	75
Figura 26 – Visualização do impacto do descritor de características.	76

LISTA DE QUADROS

LISTA DE ALGORITMOS

Algoritmo 1 – Algoritmo do SMOTE	51
Algoritmo 2 – Algoritmo de borramento com filtro bilateral	59
Algoritmo 3 – Algoritmo de aguçamento	60
Algoritmo 4 – Algoritmo da geração com ruído de Poisson	61
Algoritmo 5 – Algoritmo da geração com SMOTE visual	62
Algoritmo 6 – Algoritmo de mistura ponderada	63
Algoritmo 7 – Algoritmo de mistura limiarizada	65
Algoritmo 8 – Algoritmo de mistura saliente	66
Algoritmo 9 – Algoritmo de composição	68

LISTA DE CÓDIGOS-FONTE

LISTA DE TABELAS

LISTA DE SÍMBOLOS

SUMÁRIO

1	INTRODUÇÃO	29
1.1	Contextualização	31
1.2	Relevância e hipóteses	32
1.3	Objetivos	33
1.4	Estrutura do documento	34
2	FUNDAMENTOS	35
2.1	Considerações iniciais	35
2.2	Pré-processamento de imagens	35
2.2.1	<i>Filtragem espacial e convolução</i>	38
2.2.2	<i>Deconvolução</i>	38
2.2.3	<i>Realce de imagens</i>	40
2.2.4	<i>Restauração</i>	41
2.2.5	<i>Quantização</i>	42
2.3	Extração de características	43
2.4	Características latentes	44
2.5	Deep learning	45
2.5.1	<i>Redes de convolução</i>	47
2.6	Máquinas de Boltzmann restritas	48
2.7	Desbalanceamento de classes	50
2.7.1	<i>Sobreamostragem</i>	50
2.7.2	<i>Subamostragem</i>	51
2.8	Classificadores de padrões	53
2.8.1	<i>Algoritmo k-Vizinhos Mais Próximos</i>	53
2.9	Redução de dimensionalidade	53
2.9.1	<i>Análise de componentes principais</i>	54
2.10	Considerações finais	55
3	GERAÇÃO ARTIFICIAL DE IMAGENS	57
3.1	Considerações iniciais	57
3.2	Borramento	58
3.3	Aguçamento	60
3.4	Adição de ruído	61

3.5	SMOTE visual	62
3.6	Mistura ponderada	63
3.7	Mistura limiarizada	64
3.8	Mistura saliente	65
3.9	Composição	67
3.10	Considerações finais	68
4	RESULTADOS	69
4.0.1	<i>Base de Imagens</i>	70
4.0.2	<i>Protocolo</i>	70
4.0.3	<i>Resultados e Discussão</i>	71
5	CONCLUSÕES	77
5.1	Publicações	78
5.2	Trabalhos Futuros	78
	Referências	79



INTRODUÇÃO

A tarefa de classificação de imagens consiste em predizer corretamente uma imagem como pertencente a uma classe previamente determinada. Um exemplo prático é a classificação da imagem de um *oceano* como parte de uma classe denominada *praia*. Uma das formas de definir que certa imagem pertence à uma classe é especificar todas as regras que a caracterizam. Porém, para a maioria dos casos isso é impossível. Considere imagens coloridas, com três canais de cores e de tamanho 256×256 pixels. Dessa forma, cada um desses 65536 pixels pode ser representado por 256^3 combinações discretas de cores. Essa complexidade pode ser reduzida ao utilizar métodos de extração de características, os quais visam representar uma imagem com um número significativamente menor de valores vetoriais. Utilizando-se tal representação, pode-se desenvolver métodos computacionais que consigam definir e identificar a qual classe pertence uma imagem – sem a necessidade de se codificar todas as regras possíveis – por meio de algoritmos de Aprendizado de Máquina. Esses algoritmos possuem capacidade de generalização, crucial para classificar novos exemplos não contidos na base de imagens originalmente utilizada para o seu treinamento. Assim, “aprendem” a determinar a classe correta para imagens de entrada. Em uma etapa posterior pode-se validar esse aprendizado, aplicando o algoritmo a novos exemplos não contidos no treinamento.

O reconhecimento de padrões em imagens possui aspectos particulares para cada aplicação. Apesar da grande variedade de extratores de características disponíveis, nem sempre é possível representar as imagens de maneira satisfatória. Isso porque existem conjuntos de características que dificultam a diferenciação entre as classes. Um dos objetivos da área de Aprendizado de Máquina é encontrar quais são essas características que melhor discriminam as classes e, dessa forma, obtêm melhores resultados na etapa de reconhecimento. É comum concentrar o maior esforço dessa tarefa ao operar no espaço de características já extraídas. Para lidar com a deficiência da extração dessas características, podem ser utilizadas transformações do espaço ou sistemas de classificação complexos. No entanto, imagens obtidas de diferentes fontes, como imagens naturais, de microscopia, telescopia e tomografia, possuem características que

podem ser exploradas além dos métodos clássicos. Por isso é importante investigar métodos de processamento e preparação de imagens antes da extração dessas características, ao invés de lidar com a má representação das imagens. O uso desses métodos pode revelar *características latentes*, não visíveis nas imagens originais. Tais características podem melhor descrever certas classes, pois melhoram o conjunto de representações de imagens fornecidas à etapa de classificação.

Considerando que é comum realizar a extração de características a partir da imagem original, sem preocupação com a preparação da imagem, o enfoque desse estudo é na etapa de pré-processamento, destacada na Figura 19. Esta ilustra as etapas canônicas do reconhecimento de padrões desde a aquisição da imagem até sua posterior classificação. As etapas de pré-processamento e segmentação — apresentadas em destaque — são normalmente pouco exploradas, quando comparadas com as etapas posteriores.

Atualmente, o estado da arte de extração e classificação de imagens corresponde ao uso de redes neurais de convolução, conhecidas por CNN (SCHMIDHUBER, 2014). Essas redes são compostas por camadas de neurônios que têm por objetivo aprender quais são as melhores características que diferenciam as classes de imagens. O aprendizado, nesse caso, corresponde ao ajuste dos parâmetros para reduzir a diferença entre a saída esperada – classe verdadeira – e a produzida. Dessa forma, tais redes aprendem quais são as características latentes nas imagens de entrada. As redes neurais são discutidas na Seção 2.5, onde as máquinas de Boltzmann restritas (RBM) também são apresentadas. A representação das imagens de entrada, aprendida pela etapa de treinamento da RBM, será utilizada para definir quais imagens são relevantes para o aprendizado ou não. De certa forma essas técnicas produzem versões processadas das imagens de entrada, indicando que os filtros aprendidos são os que melhor diferenciam as classes de imagens.

O desbalanceamento de classes também se apresenta como um obstáculo para que a classificação de imagens seja satisfatória. Esse problema é caracterizado pela diferença entre o número de exemplos disponíveis para cada classe da base de imagens. Muitos métodos de transformação do espaço de características e de classificação assumem que as classes da base estão平衡adas, o que nem sempre é verdade. Portanto, é proposto a geração de imagens artificiais a partir do processamento das imagens originais, com o objetivo de rebalancear a base de imagens e consequentemente o modelo criado para a classificação. De maneira sumária, esta pesquisa busca melhorar a classificação de imagens, utilizando métodos de processamento com foco na extração de características latentes e no rebalanceamento de classes. Esse método foi comparado com o SMOTE, técnica de sobreamostragem dos vetores de características ao interpolar os exemplos mais próximos.

Os resultados obtidos, posteriormente apresentados na Seção 4, demonstram o potencial da geração de imagens artificiais. A visualização do espaço de características após o rebalanceamento das classes é crucial para analisar se as novas características extraídas são relevantes,

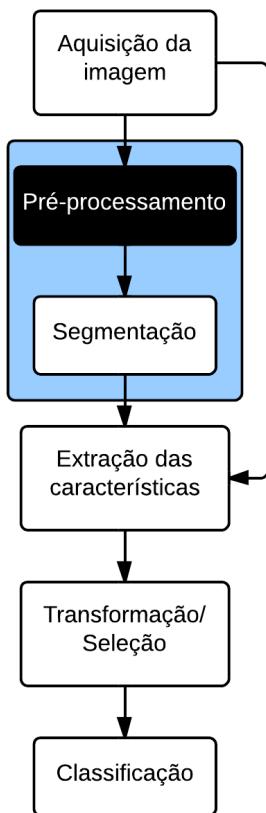


Figura 1 – Etapas canônicas do reconhecimento de padrões desde a aquisição da imagem até sua posterior classificação. As etapas de pré-processamento e segmentação — apresentadas em destaque — são normalmente pouco exploradas, quando comparadas com as etapas posteriores. O enfoque desse estudo é dar maior atenção à etapa de pré-processamento. *Fonte: Elaborado pela autora.*

ou seja, se adicionaram informações que estavam latentes ao aprendizado. De maneira sumária, esta pesquisa busca melhorar a classificação de imagens, utilizando métodos de processamento com foco na extração de características latentes e no rebalanceamento de classes. Os resultados preliminares obtidos, posteriormente apresentados na Seção ??, demonstram o potencial deste trabalho.

1.1 Contextualização

O grupo de pesquisa em Visualização, Imagens e Computação Gráfica (VICG), do Instituto de Ciências Matemáticas e de Computação (ICMC), tem atuado nas áreas de apoio para a classificação de coleções de imagens. Os trabalhos do grupo estão relacionados à visualização de informação com projeções multidimensionais e árvores (JOIA *et al.*, 2011), assim como à extração de características e classificação de coleções de imagens (PAIVA *et al.*, 2011). No que tange o processamento de imagens digitais, PICON, ROSSI e JR (2011) e PONTI (2013) focaram no pré-processamento para obter melhores resultados de classificação.

[Paiva et al. \(2011\)](#) mostraram que os espaços de características formados por cor e textura podem ser melhorados, porém há um limite até o qual as características podem ser transformadas, ou selecionadas, de forma a garantir a discriminação entre as classes. Tal projeto atua na investigação de métodos que permitam gerar espaços de características com maior discriminação entre as classes, facilitando a classificação.

Em outros dois trabalhos relacionados é possível ver a diferença na performance para problemas de classificação de imagens. No primeiro, os autores atingem acurácia acima de 98% na classificação de frutas após investigar alterações nos parâmetros de aquisição, realizar pré-processamento e obter a segmentação ([ROCHA et al., 2010](#)). No segundo, os autores indicam que o método utilizado para obter a imagem em escala de cinza (comumente utilizada por algoritmos de extração), pode impactar significativamente a classificação final de diversas bases de imagens ([KANAN; COTTRELL, 2012](#)).

Recentemente, [Ponti, Nazaré e Thumé \(2014\)](#) demonstraram que o uso de algoritmos de pré-processamento permite ao mesmo tempo obter vetores de características mais compactos e com maior capacidade de discriminação entre classes. Esta pesquisa pretende dar continuidade a esse trabalho, ao analisar técnicas de *deep learning*. Essas técnicas realizam múltiplas operações sobre imagens de entrada de forma a aprender quais operações permitem gerar características capazes de discriminar as classes ([BENGIO, 2009](#)).

1.2 Relevância e hipóteses

Conforme anteriormente mencionado, muitos aspectos influenciam a performance da classificação de coleções de imagens. É comum encontrar bases cuja extração de características é considerada difícil, onde algoritmos canônicos de extração não conseguem extrair características que diferenciem bem as classes, prejudicando sua posterior classificação. Normalmente, tenta-se lidar com as particularidades das características extraídas através de transformações no espaço de atributos ou mesmo projetando classificadores mais elaborados. Acredita-se que, ao invés disso, é importante investigar métodos de processamento e preparação de imagens antes da extração das características. Por isso, uma das hipóteses desse trabalho é que o uso desses métodos possa revelar características latentes – não visíveis nas imagens originais – que podem melhorar a acurácia da classificação.

Além disso, o desbalanceamento de classes é um obstáculo para uma classificação satisfatória, e por isso também será estudado. Em bases médicas, por exemplo, a quantidade de imagens relacionadas com uma doença rara é menor do que as imagens de pacientes sem a doença. Nessas situações, em que as imagens representam eventos importantes porém menos frequentes, o sistema de classificação pode ter problemas para lidar com a classe minoritária. A hipótese, nesse caso, é que a geração de imagens artificiais como preparação para a extração de características pode melhorar a acurácia da classificação, quando comparada à geração de

exemplos artificiais no espaço de atributos. Ou seja, gerar novas imagens artificiais — que serão posteriormente reduzidas a atributos — pode apresentar melhores resultados para a classificação do que o *bootstrap* de atributos artificiais.

1.3 Objetivos

Destacar as contribuições

O objetivo desta pesquisa é explorar as etapas de processamento de imagens com o intuito de melhorar a discriminação entre classes de uma coleção de imagens. As hipóteses são:

hipóteses repetidas?

- A geração de imagens artificiais pode contribuir com o balanceamento entre classes (em se tratando de problemas de classes desbalanceadas), melhorando a acurácia de algoritmos de classificação, quando comparada à geração de exemplos artificiais no espaço de atributos;
- O uso de métodos de pré-processamento permite a extração de características latentes que aumentem a variância entre as classes, sem aumentar, no entanto, a variância intra-classe. Melhorando, assim, a classificação.

Com o objetivo de confirmar tais hipóteses, uma das propostas é analisar as características aprendidas por uma rede neural de convolução (CNN). Essa rede permite encontrar as características mais relevantes da base de imagens, que os extratores de características canônicos não capturam. Isso porque ela possui uma hierarquia de camadas, desde a imagem original até uma etapa de classificação, com o objetivo de aprender qual o melhor processamento para as imagens de modo a melhor discriminá-las.

Após gerar as imagens artificiais, somente as imagens relevantes serão incluídas no treinamento. Para definir quais são as imagens que acrescentam informações na base, primeiramente será treinada uma máquina de Boltzmann restrita (RBM). A partir da matriz de características aprendida (memória associativa), é possível verificar se uma imagem acrescenta informações à base ou não. Por fim, conforme descrito na Seção ?? de resultados preliminares, será possível analisar operações simples e canônicas de pré-processamento de imagens.

Dados tais aspectos, pode-se então diferenciar os objetivos gerais e específicos:

Geral

Investigar os métodos de pré-processamento de imagens de forma a preparar uma coleção de imagens para a extração de características. Com isso, espera-se ao mesmo tempo obter características latentes e balancear o número de instâncias de diferentes classes.

Específicos

- Analisar o impacto da utilização de métodos canônicos de pré-processamento, como filtragem, adição de ruído e mistura, na classificação de bases de imagens. Investigar, também, o aprendizado de redes neurais de convolução, com o objetivo de observar quais são as características relevantes ao treiná-la com bases de imagens bem discriminadas em termos de cor, textura e forma;
- Modificar a base de imagens original de forma a tornar as características latentes visíveis. Com isso, pretende-se aumentar a variância entre as classes – antes da extração de características e classificação – com o auxílio dos métodos canônicos e CNN;
- Gerar imagens artificiais a partir das imagens pertencentes às classes minoritárias, compensando o desbalanceamento. Parte dessa tarefa já foi realizada e está descrita na Seção ???. O próximo passo é utilizar a matriz de características aprendida por máquinas de Boltzmann restritas para verificar se as imagens artificialmente geradas são relevantes para o aprendizado ou não, além de melhor escolher as imagens originais para a geração dessas imagens.

Considerando os objetivos aqui descritos, os resultados esperados desta pesquisa estão destacados na Seção 4.

1.4 Estrutura do documento

O conteúdo desta dissertação está estruturado como segue.

Capítulo ???: são conceituados os principais fundamentos necessários para o desenvolvimento desta pesquisa: pré-processamento de imagens, redes neurais de convolução, máquinas de Boltzmann restritas, extração de características e desbalanceamento de classes.

Capítulo ???: descreve-se a metodologia de pesquisa, assim como os resultados esperados. O cronograma, com suas respectivas atividades e o tempo previsto para conclusão desta pesquisa, é destacado em seguida.

Capítulo ???: apresenta e discute os resultados preliminares, ressaltando quais são os próximos passos.



FUNDAMENTOS

2.1 Considerações iniciais

Neste capítulo são destacados os tópicos mais relevantes para a compreensão do estado da arte de processamento de imagens, com foco na preparação das imagens para a extração de características. O problema do desbalanceamento de classes e seu efeito na classificação satisfatória também são abordados, assim como a extração de características para compreender as propriedades extraídas das imagens. A área de *deep learning* é definida e as redes de convolução e máquinas de Boltzmann restritas são contextualizadas no âmbito desta pesquisa. Alguns trabalhos relacionados são utilizados como exemplo, com o objetivo de elucidar tais tópicos.

Inicialmente, a Seção 2.2 apresenta alguns dos principais métodos utilizados para o pré-processamento de imagens (i.e. remoção de ruído, realce de imagens e convolução), relevantes para o desenvolvimento desta pesquisa. Tendo como proposta melhor compreender o que são as características latentes, a Seção 2.4 exemplifica as operações de pré-processamento utilizadas para visualizar tais características em uma aplicação específica. A seção seguinte (2.5) destaca o uso de *deep learning* para o aprendizado de características e a Seção 2.6 aborda as máquinas de Boltzmann restritas como método para aprendizado da distribuição das imagens de entrada. A extração de características é brevemente definida na Seção 2.3. Por fim, o problema do desbalanceamento de classes é apresentado na Seção 2.7. Nessa seção, além de caracterizar o problema, são apresentadas duas vertentes para solucioná-lo: sobreamostragem e subamostragem. Esses conceitos são necessários para a compreensão do capítulo de proposta desta dissertação.

2.2 Pré-processamento de imagens

Uma imagem digital I pode ser definida como uma função $f(x,y)$, onde x e y são as coordenadas cartesianas de um determinado ponto e f a intensidade (brilho) da imagem naquele ponto. Essa imagem é composta por finitos elementos chamados de pixels que podem ser

diretamente acessados através de sua posição x e y . Digitalmente, uma imagem é representada por uma matriz de valores com M linhas e N colunas onde cada elemento representa a sua intensidade. Uma imagem colorida I do sistema RGB possui três canais de cores representantes das cores vermelha, verde e azul. Cada elemento $I(x, y)$ corresponde a uma tripla (r, g, b) de números, com $0 \leq r \leq 255$, $0 \leq g \leq 255$ e $0 \leq b \leq 255$, onde 0 é a intensidade de cor mais escura e 255 a mais clara. A combinação dessas intensidades resulta na cor do pixel (GONZALEZ; WOODS, 2007). A Figura 2 mostra uma visualização ampliada dos pixels de uma imagem digital.



Figura 2 – Visualização pixelizada de uma imagem da base COREL-1000¹.
Fonte: Elaborado pela autora.

O processo de aquisição por um sistema de imageamento pode causar diversas imperfeições nas imagens, como pixels ruidosos, brilho inadequado e outras degradações. O pré-processamento de imagens é caracterizado por receber uma imagem de entrada e fornecer uma imagem de saída. Nessa etapa, efeitos indesejáveis podem ser eliminados e certas características realçadas (Figura 3). Considera-se que um determinado critério utilizado para uma imagem pode não ser o mais eficiente para outra, dependendo assim da área de aplicação.

Assim, técnicas de pré-processamento tornam os dados mais adequados para posterior análise, ao eliminar ou reduzir problemas como ruídos e imperfeições. Em Ponti (2012), o autor relata que a utilização de métodos de restauração na etapa de pré-processamento da imagem, antes da segmentação, resultou em uma qualidade superior para todos os testes, com valores

¹ Disponível em <http://wang.ist.psu.edu/docs/related/>

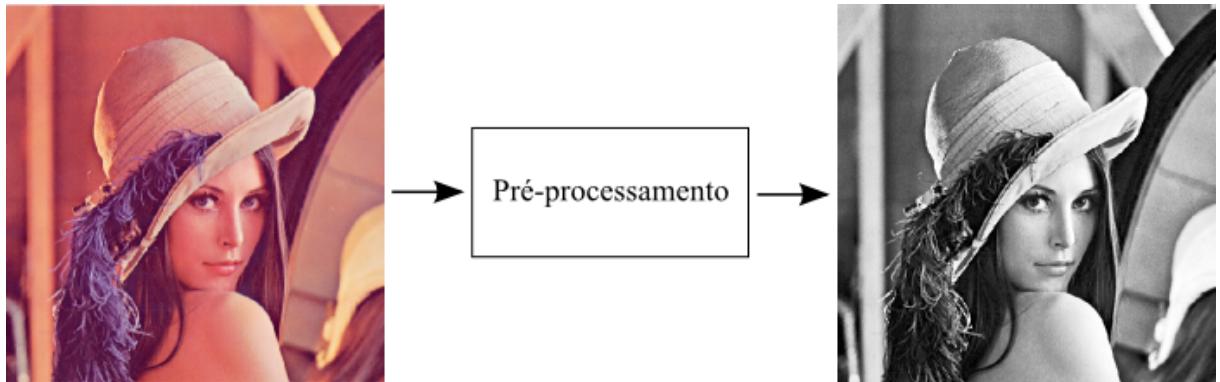


Figura 3 – O pré-processamento de imagens é caracterizado por receber uma imagem como entrada e fornecer uma imagem de saída. Sobre a imagem RGB de entrada (à esquerda) foram realizadas operações de borramento, realce e de equalização de histograma. A imagem à direita é resultante dessas operações. *Fonte: Elaborado pela autora.*

de erro e desvio padrão menores. No referido estudo, métodos de realce causaram perda de informação e por isso não são indicados para uso em imagens obtidas por microscópio. O método indicado para evitar a amplificação de ruído nessas imagens é o algoritmo iterativo Richardson-Lucy, que será apresentado na Seção 2.2.2.

Em contrapartida, [Ahonen, Hadid e Pietikäinen \(2006\)](#) propuseram uma representação para imagens faciais baseada em características de textura, sem utilizar pré-processamento. Este aparece somente como sugestão de trabalho futuro, como possível correção de problemas do sistema de captura (i.e. suavização causada pela captura fora de foco). O que implica que, apesar dos bons resultados, a melhoria com a utilização de pré-processamento não foi investigada. Pode-se imaginar, portanto, que o uso de pré-processamento pode melhorar os resultados já obtidos, através do realce de textura e eliminação de imperfeições nas imagens.

Como exemplo do uso de métodos de pré-processamento, considere imagens de algas verdes obtidas por um microscópio de alta resolução. Essas algas são mergulhadas em um líquido que normalmente causa problemas de ruído e pouco contraste. Para a preparação dessas imagens, antes da extração de características, [Borges et al. \(2013\)](#) cita algumas etapas comuns em processamento de imagens digitais:

- As imagens – originalmente em RGB – são convertidas para uma escala de cinza;
- A dimensão da imagem é reduzida para diminuir o tempo de execução dos passos subsequentes de processamento;
- O contraste é “ajustado”, para aumentar a diferença das intensidades dos pixels da imagem e corrigir o brilho;
- A imagem é filtrada, removendo ruídos causados pelo processo de captura;

- O contorno é realçado, pois a forma é uma das características mais importantes para discriminar algas (e outros objetos);
- Por fim, o histograma é equalizado.

2.2.1 *Filtragem espacial e convolução*

Um filtro espacial, também conhecido como *kernel*, máscara ou janela, consiste em uma matriz de vizinhanças e uma operação a ser realizada nos pixels de uma imagem. A filtragem cria um novo pixel com as mesmas coordenadas do centro da vizinhança, contendo o valor resultante da filtragem. Dessa forma, a imagem filtrada contém os pixels resultantes da passagem do centro do filtro espacial por todos os pixels da imagem original (GONZALEZ; WOODS, 2007).

O processo de percorrer a imagem com um filtro espacial é chamado de correlação. A convolução, que pode ser definida como o operador $*$ na Equação (2.1), é o mesmo processo, mas com o filtro previamente rotacionado em 180° (GONZALEZ; WOODS, 2007).

$$\text{Mapa de características} = \text{imagem de entrada} * \text{filtro} \quad (2.1)$$

Os métodos de filtragem possuem como objetivo aperfeiçoar certos aspectos da imagem de entrada. Essa filtragem pode ser realizada no domínio da frequência ou no domínio espacial. Um filtro de suavização típico no domínio espacial é o de Gaussiana, que resulta no borramento e redução de ruído, a fim de remover detalhes da imagem (Figura 4). Esse filtro utiliza uma função Gaussiana para calcular a transformação a ser realizada a cada pixel. A equação que representa a função Gaussiana em duas dimensões é definida por

$$G_\sigma(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}},$$

onde x, y são coordenadas de um determinado pixel da imagem e σ o desvio padrão que determina o raio da distribuição Gaussiana aplicada. Valores altos de variância fazem com que o resultado da função se aproxime da média.

2.2.2 *Deconvolução*

O processo de convolução, descrito na seção anterior, também pode ser definido como a passagem de uma imagem por um processo de aquisição que atue como um filtro de passa-baixas, resultando em uma imagem borrada. A deconvolução realiza a inversão desse borramento, o que pode facilitar a segmentação e detecção de características (PONTI, 2012). Para tal, é necessário conhecer previamente ou estimar a função que causou a degradação na imagem, geralmente denotada por $h(\mathbf{u})$, onde \mathbf{u} representa as coordenadas (x, y, z) para um sinal tridimensional.



Figura 4 – Exemplo de filtragem gaussiana como operação de pré-processamento. *Fonte: Elaborado pela autora.*

Em casos mais simples é possível utilizar o filtro pseudo-inverso, obtendo o filtro a partir da transformada de Fourier da função h , denotada por H em

$$W(\mathbf{u}) = \begin{cases} H(\mathbf{u}), & H(\mathbf{u}) > \gamma \\ \gamma, & \text{caso contrário,} \end{cases}$$

onde o limiar γ utilizado é em geral um valor entre 0,0001 e 0,1. O filtro W é utilizado para realizar a inversão, no domínio da frequências, da imagem g borrada, obtendo a imagem restaurada a partir de

$$\hat{F}(\mathbf{u}) = \frac{G(\mathbf{u})}{W(\mathbf{u})},$$

onde \hat{F} é a transformada de Fourier da imagem restaurada, G é a transformada de Fourier da imagem borrada, e W é o filtro pseudo-inverso, que realiza a deconvolução.

Um outro exemplo de algoritmo de deconvolução é o Richardson-Lucy ([PONTI et al., 2011](#)), que utiliza um processo iterativo para recuperar uma imagem degradada que foi borrada por algum processo conhecido. Utiliza uma metodologia probabilística, baseada em EM-ML (*Expectation-Maximization Maximum Likelihood*), para encontrar uma imagem que maximize a probabilidade de se visualizar a imagem original sem degradação, considerando um modelo de ruído de Poisson. O algoritmo é descrito na Equação 2.2, onde n é o número da iteração.

$$\hat{f}_{n+1}(\mathbf{u}) = \left[\left(\frac{g(\mathbf{u})}{\hat{f}_n(\mathbf{u}) * h(\mathbf{u})} \right) * h(\mathbf{u}) \right] \times \hat{f}_n(\mathbf{u}). \quad (2.2)$$

Algoritmos iterativos como o Richardson-Lucy tem a vantagem de permitir soluções parciais, evitando amplificação de ruído.

2.2.3 Realce de imagens

O realce de imagens é o processo de modificar uma imagem para que se torne mais adequada para uma aplicação específica do que na sua forma original. É subjetivo, porque depende do sujeito que está analisando a imagem dissennir a qualidade desse realce (GONZALEZ; WOODS, 2007).

Na Figura 5 está demonstrado o efeito do algoritmo de *unsharp masking*, utilizando como borramento um filtro de média. Com o objetivo de realçar imagens, os passos deste método são:

1. Borramento da imagem original;
2. Cálculo da diferença entre a imagem suavizada e a original;
3. Soma dessa diferença à imagem original.



(a) Original



(b) *Unsharp masking*

Figura 5 – A imagem original, já em escala de cinza, foi realçada utilizando o método *unsharp masking*. Fonte: Elaborado pela autora.

Um exemplo clássico de utilização de realce, é para compensar a variação de iluminação em diversas imagens. Em Gross e Brajovic (2003), os autores propuseram um algoritmo para o reconhecimento de faces invariante à iluminação. Eles ressaltam que, desconsiderando a variação da posição, a iluminação é o fator de maior impacto na aparência das faces. A luz varia durante o dia, entre um dia e outro e entre diferentes ambientes. Isso afeta o conjunto de imagens a ser analisado, que passa a conter imagens com diferentes contrastes, o que pode acentuar ou diminuir certas características faciais.

O contraste é a diferença de intensidade entre os níveis de maior e menor intensidade na imagem. Imagens com baixa resolução podem ser geradas a partir de uma iluminação pobre ou

outros problemas com a captura. Dessa forma, o processo de “esticar” o contraste expande os níveis de intensidade da imagem (GONZALEZ; WOODS, 2007).

Explicar melhor histograma e k

É possível aumentar o contraste de uma imagem ao manipular o seu histograma h , que pode ser definido como

$$h(i_k) = n_k,$$

onde n_k é o número de pixels de intensidade i_k . Ao observar os histogramas de diferentes imagens é possível notar que imagens com alto contraste possuem um histograma com níveis próximos a uma distribuição uniforme. Isso permite que certas operações, como a equalização de histograma, obtenham o melhor contraste de uma imagem. Essa operação é caracterizada pela máxima variância do histograma e pode ser definida como

$$s_k = T(i_k) = \frac{L-1}{MN} \sum_{j=0}^k n_j, \quad (2.3)$$

onde L é o número de intensidades e MN as dimensões da imagem. A imagem de saída é obtida ao mapear cada pixel de intensidade i_k em um nível s_k , com i entre $[0, L-1]$, sendo $i = 0$ um pixel preto e $i = L - 1$, branco (GONZALEZ; WOODS, 2007).

2.2.4 Restauração

Ao contrário do realce, a etapa da restauração é objetiva. Também procura melhorar os aspectos visuais de uma imagem, mas com base em modelos probabilísticos de degradação de imagens. Dessa forma, tendo-se um modelo de degradação, procura-se recuperar a imagem original (GONZALEZ; WOODS, 2007). Mas um problema comum em algoritmos de remoção de ruído é que alguns detalhes, como textura, irão sofrer alta suavização.

O objetivo dos algoritmos de remoção de ruído é recuperar a imagem original. Assim o modelo estatístico da imagem natural é crucial para a remoção de ruído. O tipo mais simples de ruído é o aditivo, que pode ser caracterizado como

$$p = p_0 + n,$$

onde p_0 é o valor real do pixel e n é a perturbação de ruído naquele pixel.

Um algoritmo alternativo para remoção de ruído é o de médias não locais. Isso porque pixels similares não necessariamente estão próximos em suas coordenadas cartesianas, assim, informações não locais podem ser utilizadas na redução de ruído. De acordo com Buades, Coll e Morel (2005), dada uma imagem ruidosa $v = v(i)|i \in I$, o valor estimado é calculado como uma média ponderada de todos os pixels em uma mesma imagem com

$$v(i) = \sum_{j \in I} w(i, j)v(j),$$

onde os pesos $w(i, j)$ dependem da similaridade entre pixels.



Figura 6 – Conversão para a escala de cinza com os métodos de Intensidade e MSB. *Fonte: Elaborado pela autora.*

2.2.5 Quantização

A maioria dos métodos de extração de características utiliza imagens de entrada em escala de cinza. Isso ocorre porque a complexidade de lidar com um pixel representado em três dimensões é muito maior do que em apenas uma dimensão. Assim, os métodos de quantização visam, de alguma forma, reduzir os canais de cores (24 bits) em apenas um (8 bits). [Kanan e Cottrell \(2012\)](#) demonstraram que os métodos para a quantização (conversão de uma imagem colorida para escala de cinza) influenciam a performance no reconhecimento de imagens. Eles também salientam que o método utilizado deveria estar claramente descrito nas publicações da área.

Em um trabalho recente ([PONTI; NAZARÉ; THUMÉ, 2014](#)), com participação da autora desta dissertação, quatro métodos de quantização foram investigados. Foi constatado que os melhores métodos de conversão para escala de cinza antes da utilização dos descritores de características a serem mencionados na Seção 2.3 são:

Intensidade: método mais simples, consiste em computar a média entre os canais RGB da imagem a partir de

$$Q_{\text{Intensidade}} = \frac{1}{3}(R + G + B)$$

e então realizar a correção por *gamma*.

Bits mais significativos (MSB): ao invés de realizar uma combinação linear dos canais de cores, ordena os bits dos canais coloridos em um único canal. Computa quantos bits de cada canal de cor contribuem para a imagem final e extrai os bits do código binário dos canais originais ([PONTI; ESCOBAR, 2013](#)).

A Figura 6 apresenta a conversão na escala de cinza obtida com o uso destes métodos.

2.3 Extração de características

O objetivo da extração de características é descrever as informações visuais relevantes em um vetor de características. Esse vetor pode ser utilizado como entrada para um algoritmo de classificação de padrões. Como por exemplo, em aplicações que envolvem a classificação de algas, uma informação muito importante para a discriminação entre classes é a forma (BORGES *et al.*, 2013). Essas características devem salientar as diferenças entre imagens de classes distintas e suavizar possíveis diferenças de imagens da mesma classe. Algumas características, segundo Gonzalez e Woods (2007), são:

Textura: na sua descrição estatística, possui propriedades como: suavidade, aspereza e uniformidade. Um exemplo de medida para descrever a textura é a entropia.

Forma: representa os objetos em termos de suas características externas, como, por exemplo, a medida da curvatura.

Cor: considera a distribuição espacial de cores na imagem. O histograma de uma imagem pode descrever essa configuração de forma global.

Exemplos de métodos conhecidos capazes de descrever forma e outras características são: histogramas de orientação de gradiente (WANG; HAN; YAN, 2009), curvatura, descritores de Fourier, métodos baseados na detecção de SUSAN (SMITH; BRADY, 1997), Harris-Affine (HARRIS; STEPHENS, 1998) e diferença de Gaussianas (LOWE, 2004). Os descritores utilizados para a obtenção dos resultados preliminares da Seção ?? e que vão ser utilizados no desenvolvimento desta pesquisa estão abaixo descritos.

Histograma global de cor (GCH): calcula o histograma global dos níveis de intensidade da imagem. É a alternativa mais simples para representar as informações de uma imagem (GONZALEZ; WOODS, 2007). Produz um vetor de N dimensões, sendo N o número de intensidades.

Vetor de coerência de cor (CCV): captura informações sobre como as cores são organizadas em regiões conectadas, de acordo com um *threshold*. Classifica os pixels da imagem em coerentes e incoerentes, computa os respectivos histogramas e os concatena (HUANG *et al.*, 1997). Dessa forma, o vetor de características produzido possui $2N$ dimensões.

Classificação de pixels de borda e interior (BIC): computa dois histogramas, um para pixels definidos como borda e outro como interior. Se um pixel possuir a mesma cor que seus vizinhos, é pixel de interior; caso contrário, será pixel de borda. Os histogramas são concatenados, gerando um vetor de $2N$ dimensões (STEHLING; NASCIMENTO; aO, 2002).

Auto-correlograma de cor (ACC): captura a correlação espacial entre cores idênticas. Computa a probabilidade de encontrar dois pixels com a mesma cor, a uma distância d um do outro (PASS; ZABIH; MILLER, 1996). Neste estudo, são consideradas quatro distâncias: 1, 3, 5 e 7; resultando em um vetor com $4N$ características.

Haralick-6: extrai seis características a partir uma matriz de coocorrência: entropia, homogeneidade, contraste, correlação, probabilidade máxima e uniformidade. O vetor resultante possui 6 dimensões (HARALICK; SHANMUGAM; DINSTEIN, 1973).

2.4 Características latentes

Assim como o método de quantização se mostrou relevante na classificação, nesta dissertação são estudados diversos métodos de pré-processamento de imagens, como filtros de borramento e deconvolução, com o objetivo de obter imagens processadas que sejam mais bem caracterizadas para a etapa de classificação (ou seja, aumentando a variância entre as classes, sem aumentar a variância intra-classe). O enfoque está em como realçar determinadas características, como por exemplo cor, textura e forma. A esses atributos pode-se dar o nome de características latentes, não visíveis na imagem original. Identificar e realçar tais características é objetivo deste estudo. Essa abordagem se diferencia das técnicas multi-resolução, pois não pretende encontrar apenas características em versões de diferentes resoluções (convoluídas com filtros passa-baixa), e sim também em versões deconvoluídas ou transformadas por outros operadores.

Se uma das principais características que diferenciam classes de uma certa base utiliza a sua forma, é possível utilizar o método de diferença de Gaussianas para ressaltá-la. O método DoG (*Difference of Gaussians*) se baseia na diferença entre duas imagens filtradas. A operação comumente é feita usando $\sigma = \sqrt{2}$ e sua definição para dois níveis de filtragem é definida por

$$\begin{aligned} f_1(x,y) &= G_\sigma(f(x,y)) \\ f_2(x,y) &= G_\sigma(f_1(x,y)) \\ DoG_1(x,y,\sigma) &= f_2 - f_1 \end{aligned}$$

A Figura 7 demonstra uma possível execução de filtragem e restauração seguida pela detecção de bordas por DoG, aplicadas em uma base de imagens de algas verdes.

A imagem 7a é uma imagem original segmentada de alga. As próximas colunas são imagens resultantes da deconvolução da imagem (coluna 2), filtragem Gaussiana (coluna 3) e filtragem Gaussiana seletiva (coluna 4). As modificações apresentadas na primeira linha geram informações em diferentes planos axiais da imagem de microscopia. Imagens na segunda coluna representam uma versão deconvoluída das imagens originais, ressaltando a textura na superfície das algas, resultantes da diferença de Gaussianas. Já a terceira linha demonstra imagens binárias obtidas por limiarização das imagens da segunda linha, realçando as características da base da célula.

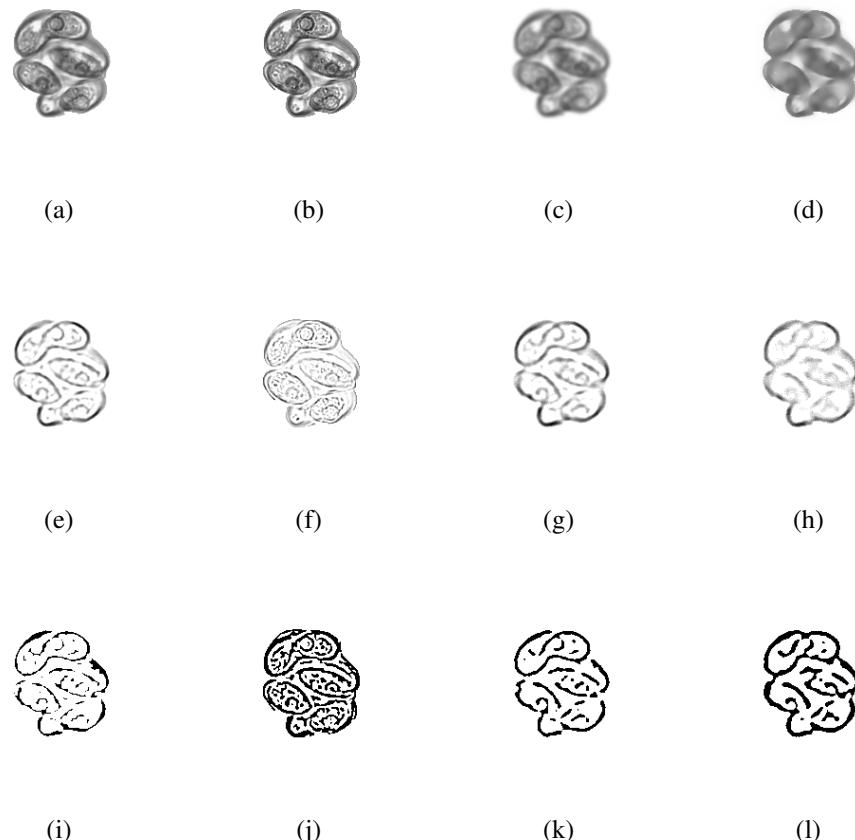


Figura 7 – Características latentes de algas verdes. A primeira imagem (a) é uma imagem original segmentada de alga. As próximas colunas são imagens resultantes da deconvolução da imagem (coluna 2), filtragem Gaussiana (coluna 3) e filtragem Gaussiana seletiva (coluna 4). A primeira linha mostra versões diferentes de imagens de algas, a segunda linha exibe imagens resultantes da diferença de Gaussianas, e a terceira linha demonstra imagens binárias obtidas por limiarização das imagens contidas na segunda linha).

Fonte: Elaborado pela autora.

Acredita-se que ao manipular as imagens através de técnicas de pré-processamento, pode-se gerar imagens artificiais com características latentes realçadas, melhorando a extração de características. As redes neurais, que aprendem automaticamente essas características, estão descritas nas próximas seções.

2.5 Deep learning

A habilidade humana de reconhecimento de padrões em imagens é surpreendente. Os algoritmos de aprendizado de máquina e de reconhecimento de padrões tentam habilitar os computadores a aprender, baseando-se no modelo de aprendizado de seres humanos (GEORGE, 2008). Considerando que existem milhões de neurônios em cada córtex visual e que o cérebro humano é composto por cinco desses córtex que evoluíram durante milhões de anos, pode-se afirmar que a tarefa de algoritmos de reconhecimento de padrões em imagens não é simples

(NIELSEN, 2015).

A capacidade humana de reconhecer novos padrões está diretamente relacionada com a capacidade de generalização, que para tal, utiliza hierarquias (GEORGE, 2008). Nessa linha, o aprendizado profundo ou *deep learning* é caracterizado por explorar o aprendizado ao simular o funcionamento do cérebro humano. Essa área está diretamente relacionada com aprendizado de máquina e inteligência artificial e lida com o aprendizado através de múltiplas camadas.

Explicitar quais são as regras para aprender um determinado conceito pode se tornar impossível. No caso de um avião, por exemplo, “voar” pode ser uma regra decomposta em “possuir largura maior que altura” e “ser aerodinâmico”. Essas subdivisões geram múltiplos níveis que possuem maior complexidade e abstração, de maneira similar ao aprendizado por camadas. Uma rede neural artificial utiliza imagens como treinamento para automaticamente inferir quais são as regras para o reconhecimento. São chamadas de redes neurais profundas ou *deep neural networks* as redes que possuem uma estrutura de muitas camadas – duas ou mais ocultas (NIELSEN, 2015). Assim, dada uma imagem como entrada, o problema é subdividido em problemas mais simples de serem resolvidos, podendo chegar ao nível de pixels isolados.

As redes neurais artificiais consistem em um método para solucionar problemas de forma a simular o comportamento do cérebro humano. Ao tentar uma determinada solução e errar, essas redes aprendem e podem tentar novamente. Ou seja, adquirem conhecimento através da experiência. Elas contém neurônios de entrada, ocultos e de saída. Aprender nesse contexto significa encontrar os pesos que fazem com que a rede neural exiba o comportamento esperado (SCHMIDHUBER, 2014).

A simulação desse modelo inspirado biologicamente utilizada hoje é de LeCun e Bottou (1998) e chama-se Rede Neural de Convolução (CNN ou ConvNet). Eles simplificaram tal arquitetura para utilizar um algoritmo de retropropagação para o treino. Desde então, essas redes vêm sendo utilizadas para detecção, reconhecimento, restauração, remoção de ruído e segmentação de imagens e vídeos, além de sua aplicação em áudio. Um exemplo de utilização dessa rede inspirada no modelo de LeCun foi desenvolvida pela empresa Google, com o objetivo de detectar faces e placas de carros para proteger a privacidade nas imagens de StreetView (FROME *et al.*, 2009).

Apesar de redes profundas representarem o estado da arte em visão computacional, um bom entendimento das suas propriedades ainda está faltando. Alguns artigos recentes como Mahendran e Vedaldi (2014), Zeiler, Taylor e Fergus (2011) Zeiler e Fergus (2013) e Simonyan, Vedaldi e Zisserman (2013) introduzem técnicas utilizadas para analisar o comportamento e operações internas dessas redes ao visualizá-las. Esta pesquisa situa-se nesse viés, ao analisar as características latentes extraídas por uma rede neural de convolução.

Para o uso em bases desbalanceadas, as imagens utilizadas para o rebalanceamento de forma visual podem ser geradas após o estudo das características latentes, encontradas

no treinamento da classe minoritária em uma CNN. Essas características também podem ser realçadas de forma a melhorar a classificação de imagens.

2.5.1 Redes de convolução

São um tipo de rede neural que utiliza uma operação chamada de convolução, previamente descrita na Seção 2.2.1. Pode ser entendida como sendo várias multiplicações de um filtro espacial pela imagem de entrada, resultando em um mapa de características ativadas. Esse filtro é um vetor de parâmetros capaz de aprender. Cada imagem possui seu conjunto de mapas de características, mas os filtros são comuns a todas as imagens. Dependendo dos valores do filtro, as características ativadas são diferentes. Nesse contexto, a camada de convolução consiste em muitas aplicações de convolução em paralelo, dado que um filtro é capaz de extrair apenas um tipo de característica (BENGIO; GOODFELLOW; COURVILLE, 2014).

Basicamente, uma CNN é uma arquitetura que pode aprender características invariantes. Com múltiplas camadas, ela pode aprender multiníveis de características (LECUN; KAVUKCUOGLU; FARABET, 2010). Ou seja, permite ao computador criar conceitos complexos a partir de conceitos simples. Como o conceito de *deep learning* enfatiza o uso de variáveis latentes, sugere que algoritmo de treinamento pode inventar os conceitos que precisa para representar determinada base de dados (BENGIO; GOODFELLOW; COURVILLE, 2014).

Cada camada é composta de três estágios LeCun, Kavukcuoglu e Farabet (2010), representados na Figura 8:

Estágio de convolução: várias convoluções em paralelo para produzir um conjunto de ativações pré-sinápticas;

Estágio de detecção: função de ativação, como a unidade linear retificada ou a sigmoide;

Estágio de pooling: retorna uma representação invariante a pequenas translações da entrada.

Um exemplo é utilizar o valor máximo entre vizinhos (conhecido como *max-pooling*).

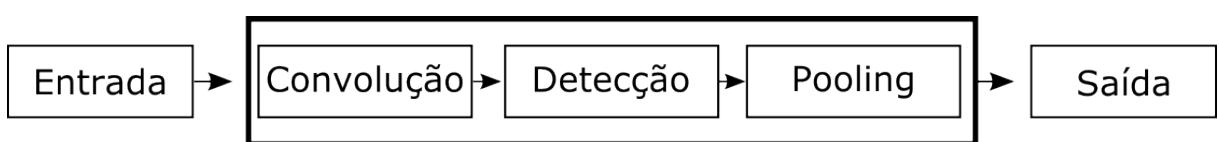


Figura 8 – Componentes típicos de uma camada de uma rede neural de convolução. Nessa terminologia a rede possui camadas complexas que possuem estágios (BENGIO; GOODFELLOW; COURVILLE, 2014).
Fonte: Elaborado pela autora.

De acordo com Bengio, Goodfellow e Courville (2014), durante a propagação são realizadas as convoluções e propagadas pelo resto da rede para então computar uma função de perda, que mede quão bem o sistema de aprendizado funcionou para cada exemplo. Na

retropropagação, recebe-se um gradiente para atualizar os pesos. O gradiente descendente estocástico atualiza o sistema de aprendizagem com base no erro medido para um único exemplo. Apesar de a convergência ser mais ruidosa do que calcular a média sobre todos os erros, diminui-se um custo constante do cálculo de custo.

Normalmente, o classificador é a última camada totalmente conectada, que computa o produto de um vetor de características (BENGIO; GOODFELLOW; COURVILLE, 2014). Mahendran e Vedaldi (2014) reportam que as camadas inteiramente conectadas podem ser invertidas a uma imagem resultante da composição de partes similares às encontradas na imagem original, mas não idênticas. Enfatizam também que todas as camadas de convolução mantêm uma representação fiel da imagem. Essa característica permite a análise de quais atributos são relevantes para o aprendizado da base de imagens, tarefa crucial para esta pesquisa.

Zeiler, Taylor e Fergus (2011) propuseram uma rede de deconvolução para a visualização das camadas, que pode ser vista como o caminho inverso de uma rede de convolução. Dessa forma, ao invés de mapear pixels em características, faz-se o contrário, permitindo a construção não supervisionada de representação de imagens hierárquicas, que podem, por exemplo, ser usadas para redução de ruído (ZEILER; FERGUS, 2013). Tenta gerar o sinal de entrada dada a soma das convoluções dos mapas de características com os filtros aprendidos. Apesar de destacar que, dado o mapa de características latentes, é possível sintetizar uma imagem, o artigo de Zeiler, Taylor e Fergus (2011) não discorre sobre essa síntese.

2.6 Máquinas de Boltzmann restritas

Uma máquina de Boltzmann restrita (RBM) é uma rede neural estocástica que treina um modelo a partir dos vetores de entrada. Ela aprende os parâmetros que definem a distribuição sobre um conjunto de observações sem retropropagação e baseando-se em energia (FISCHER; IGEL, 2014).

Diferente de uma rede de Hopfield, ela agrupa probabilidades e encontra a mínima energia global. Faz isso através do arrefecimento simulado (*simulated annealing*), ou seja: em um primeiro momento, quando possui energia alta, aceita configurações piores; e conforme a energia é reduzida, detém-se a uma região e a explora melhor, até o equilíbrio térmico (HINTON; OSINDERO; TEH, 2006).

Possuindo uma arquitetura muito mais simples que a CNN, é composta por duas camadas de neurônios: uma visível e outra oculta. Os pixels correspondem às unidades visíveis e os detectores de características às unidades ocultas. A sua restritividade se deve ao fato da falta de conectividade entre os neurônios de uma mesma camada, mas cada unidade visível é conectada a todas as unidades ocultas. Assim, a camada oculta, de detectores de características, modela a correlação entre os pixels.

A aprendizagem de uma RBM começa em um estado aleatório e atualiza seus pesos até encontrar uma distribuição que esteja em equilíbrio. Essa convergência ocorre quando o erro entre os exemplos treinados – fase positiva – e suas reconstruções – fase negativa – é menor que um certo limiar. Dessa forma, a cada iteração, um novo exemplo é treinado ao repetir iterativamente esses dois estágios e atualizar seus pesos (HINTON; OSINDERO; TEH, 2006).

Fase positiva: fixa a camada visível e computa o valor dos neurônios ocultos. A probabilidade de um neurônio oculto ser ativado (se torne igual a 1) é dada pela função

$$p(h_j = 1) = \frac{1}{1 + e^{-(b_j + \sum_{i \in \text{vis}} v_i w_{ij})}}, \quad (2.4)$$

onde $p(h_j)$ é a probabilidade do neurônio h_j ser ativado, utilizando uma função de ativação sigmóide, e w_{ij} o peso entre a entrada i e o neurônio j .

A contribuição da fase positiva é computada para cada par de unidades a_{ij} , ao conferir se os dois neurônios (x_i e x_j) estão ativados por meio de

$$\text{Fase}_+(a_{ij}) = x_i \cdot x_j$$

Fase negativa: a partir dos valores computados na fase positiva, os neurônios visíveis são “reconstruídos” ($p(v_i)$) e os ocultos novamente computados ($p(h_j)$) a partir da função de probabilidade da Equação (2.4).

A contribuição da fase negativa é calculada da mesma forma que a contribuição positiva, com

$$\text{Fase}_-(a_{ij}) = x_i \cdot x_j$$

Atualização dos pesos: os pesos w_{ij} são ajustados considerando a diferença entre os neurônios visíveis reconstruídos e os valores fornecidos de entrada inicialmente (etapa conhecida como *contrastive divergence*):

$$w_{ij} = w_{ij} + L(\text{Fase}_+(a_{ij}) - \text{Fase}_-(a_{ij})),$$

onde L é a taxa de aprendizado. Quanto menor esse valor, melhor o treinamento e maior tempo até a convergência.

Essas redes também podem ser usadas como classificadores, ao treiná-las para modelar a relação entre a distribuição de entrada e seus respectivos rótulos (FISCHER; IGEL, 2014). Devido à capacidade das máquinas de Boltzmann restritas de aprender representações das imagens, pode-se utilizá-las para identificar se uma imagem é relevante para o aprendizado. Pode-se, por exemplo, utilizar apenas um neurônio, com a matriz de pesos representando uma memória associativa, para entender quais imagens são mais relevantes dentro de uma classe arbitrária. Dessa forma, também é possível avaliar a relevância da informação contida em uma imagem artificialmente gerada.

2.7 Desbalanceamento de classes

Nesta seção é definido o problema do desbalanceamento de classes e apresentados os trabalhos relacionados que possuem duas diferentes abordagens: sobreamostragem (*over-sampling*) e subamostragem (*under-sampling*).

Em conjuntos de dados desbalanceados, determinadas classes possuem um número muito maior de instâncias do que outras. As classes com mais elementos são chamadas de classes majoritárias, e as com menos elementos, de minoritárias. O desempenho de algoritmos de Aprendizado de Máquina é prejudicado quando tratam de bases de dados desbalanceados. Esses algoritmos tendem a favorecer a classificação de um novo objeto à classe majoritária, pois esta fica muito melhor representada após o treinamento do que a minoritária. Considera-se, então, que esse problema é um obstáculo para a classificação satisfatória. Porém, como apontado por (BATISTA; PRATI; MONARD, 2004), o desbalanceamento não é o único responsável por reduzir o desempenho de algoritmos de aprendizagem. Eles sugerem que é possível haver uma ótima classificação mesmo contendo alto desbalanceamento na base de dados. Assim, a motivação do estudo de vários algoritmos para rebalanceamento não é apenas balancear os dados de treinamento, mas obter uma melhor diferenciação entre as classes. Isso porque o desbalanceamento por si só não é um problema, mas em conjunto com a sobreposição de classes pode diminuir significativamente a acurácia da classificação da classe minoritária. Os resultados reportados também mostram que a poda de árvores de decisão raramente levou a alguma melhora na classificação.

(CASTRO; BRAGA, 2011) destacam que duas abordagens têm sido utilizadas para solucionar esse problema: pré-processar os dados de forma a rebalancear as distribuições das classes ao reamostrar os dados; ou então modificar métodos de aprendizado – como através da adição de melhores funções de custo na classificação. Em geral, são reportados melhores resultados obtidos por algoritmos de *over-sampling*, os quais consistem em reamostrar os dados aumentando o número de elementos da classe minoritária (BATISTA; PRATI; MONARD, 2004). Esta pesquisa tem como enfoque o **pré-processamento dos dados**, ao rebalancear as classes através da **geração de imagens artificiais**.

2.7.1 Sobreamostragem

Realizar uma sobreamostragem (*over-sampling*) em um determinado conjunto de dados significa aumentar – utilizando alguma estratégia – o número de elementos desse conjunto. Em (CHAWLA; HALL; BOWYER, 2002), a simples replicação de exemplos pertencentes à classe minoritária não melhorou a classificação. Isso se deve ao reconhecimento de regiões muito específicas, causando *overfitting*.

O SMOTE (*Synthetic Minority Over-sampling Technique*) é um método desenvolvido por (CHAWLA; HALL; BOWYER, 2002) para rebalancear classes ao gerar artificialmente novos

elementos, ao invés de apenas replicá-los. É aplicado sobre os vetores de características previamente extraídos, com operações de perturbação dos dados de treino no espaço de características, e não no espaço dos dados. A diferença entre o vetor de características de um elemento e do seu vizinho mais próximo é multiplicada por um número $0 \leq x \leq 1$. Esse valor é adicionado ao vetor original, criando um novo elemento. Como pode ser visualizado na Figura 9, essa abordagem provoca a geração de um elemento resultante da interpolação dos dois vetores originais. Os exemplos sintéticos forçam uma região de decisão maior e mais geral para serem aprendidas como exemplos da classe minoritária.

Algoritmo 1: Algoritmo do SMOTE

Entrada: Imagem colorida I em formato BGR

Saída: Exemplos sintéticos

```

1  $N \leftarrow$  vizinhos(minoritária);
2 para cada exemplo  $sample$  faz
3    $nn \leftarrow$  vizinho aleatório de  $N$  para cada característica  $(x,y)$  faz
4     diferença  $\leftarrow nn(x,y) - sample(x,y);$ 
5     gap  $\leftarrow$  número aleatório entre 0 e 1;
6      $G(x,y) \leftarrow sample(x,y) + gap * diferença;$ 
7   fim
8    $S \leftarrow SUG;$ 
9 fim
```

Dessa forma, o SMOTE provê mais elementos para o classificador aprender, ao contrário da replicação de dados. Como trabalhos futuros, os autores Chawla, Hall e Bowyer (2002) apontam que diferentes estratégias para criar esses exemplos sintéticos podem melhorar a performance da classificação. Inclusive salientando exemplos que foram errôneamente classificados.

Uma variação desse algoritmo, denominada Borderline-SMOTE1 (HAN; WANG; MAO, 2005), considera que elementos fora da linha de borda de cada classe pouco contribuem para a classificação. Por isso, propõe a geração de elementos sintéticos utilizando apenas elementos de borda. Considera que se os vizinhos mais próximos são da classe majoritária, o exemplo é ruído, e se há mais vizinhos da classe majoritária do que da minoritária, considera esse elemento como sendo de borda. Como trabalho futuro, destacam a necessidade de considerar diferentes estratégias para definir em quais elementos realizar o over-sampling.

2.7.2 Subamostragem

Ao contrário da sobreamostragem, a subamostragem visa diminuir o número de elementos de um determinado conjunto. A ideia é eliminar elementos da classe majoritária que estão distantes da fronteira de decisão, isso porque eles são considerados menos relevantes para a aprendizagem.

Métodos para remoção de exemplos da classe majoritária normalmente apresentam resultados piores do que métodos de sobreamostragem, conforme relatado por [Batista, Prati e Monard \(2004\)](#) e [\(JAPKOWICZ; STEPHEN, 2002\)](#). Um dos motivos pela preferência natural à sobreamostragem é o fato de que ao realizar subamostragem pode-se remover informações essenciais dos dados originais. Mas não há uma estratégia única que funcione melhor para todos os cenários.

2.8 Classificadores de padrões

2.8.1 Algoritmo *k*-Vizinhos Mais Próximos

K-NN é um classificador supervisionado que considera a proximidade entre os dados para realizar previsões. Baseia-se na premissa de que os objetos do mesmo conceito são semelhantes. Na fase de treinamento, apenas armazena os exemplos rotulados do conjunto de dados de treinamento. Quando um novo exemplo deve ser classificado, calcula a distância entre os vetores de características do novo exemplo e aqueles já rotulados. O novo exemplo é então classificado como sendo da classe do exemplo de treinamento com menor distância (??).

Esse classificador foi utilizado por ser suscetível à diferenciação entre as classes. Com $K = 1$, a previsão da classe corresponde ao exemplo mais próximo.

2.9 Redução de dimensionalidade

A visualização do espaço de características obtido após a geração artificial de imagens pode ajudar a verificar se as novas imagens melhoraram a definição da classe minoritária em relação ao espaço original (inclusive antes de imagens serem removidas para provocar o desbalanceamento). Ou seja, se o método utilizado revelou características latentes. Dessa forma, ao projetar os novos vetores no espaço das imagens originais, é possível analisar qual método (entre SMOTE ou geração de imagens no campo visual) mais se assemelha à distribuição original dos dados.

Considerando que um vetor de características extraído com extratores comuns pode possuir entre 6 (e.g. Haralick) e 512 (e.g. BIC) características, a visualização de um exemplo requer que seja realizado o mapeamento desses valores em apenas duas dimensões. Para isso, uma redução de dimensionalidade mapeia os dados de N dimensões para um espaço 1D, 2D ou 3D. A partir desses novos dados, pode então ser criada alguma representação visual que tente manter a relação de distância entre os novos e os originais (??).

2.9.1 Análise de componentes principais

O PCA (Análise de Componentes Principais) é uma técnica não supervisionada que pode ser utilizada para reduzir a dimensionalidade dos dados com a máxima variância possível. Cada imagem, originalmente representada por um vetor com N características, pode então ser representada por apenas um ou mais valores. Essa redução permite a projeção dessas imagens no espaço de características. O objetivo é extrair as informações mais importantes dos dados e representá-las como um conjunto de variáveis ortogonais chamadas de componentes principais. Para isso encontra-se uma outra base: uma combinação linear da base original, que melhor representa os dados ao assumir que as direções das maiores variâncias são as mais importantes. Ou seja, a variância associada com cada direção quantifica o quanto principal é aquela direção (??). Pode-se, portanto, enumerar os passos necessários para o PCA como:

1. Centraliza todos os atributos em zero ao subtrair a média de cada dimensão;
2. Calcula a matriz de covariância C_x dada por

$$C_x = XX^T, \quad (2.5)$$

onde X é a matriz de dados original e X^T sua transposta.

3. Encontra os autovalores e autovetores de C_x . Um autovetor u de uma matriz A pode ser definido por $Au = \lambda u$, onde λ é um autovalor escalar associado ao autovetor. Um vetor u é um autovetor da matriz A se o tamanho do vetor – e não sua direção – é modificado quando multiplicado por A . Os autovalores podem ser representados na diagonal de uma matriz λ (com outros valores como zero) e o conjunto dos autovetores de A em uma matriz U . Assim,

$$A = U\lambda U^T. \quad (2.6)$$

4. Então, os autovetores são ordenados de forma decrescente de acordo com seus autovalores correspondentes e escolhe-se os k principais autovetores (i.e. maiores autovalores) para formar uma matriz P de dimensão $n \times k$, onde cada coluna representa um autovetor. Esse valor k será a quantidade de dimensões do novo espaço de atributos.
5. O novo subespaço pode ser encontrado multiplicando essa matriz P pela matriz original, de acordo com a equação $Y = PX$, onde X representa o conjunto de dados original, Y é uma nova representação desses dados e P a matriz ortonormal que transforma X em Y . As linhas de P são os componentes principais de X .

2.10 Considerações finais

Deu-se destaque à discussão das etapas de pré-processamento e realce de características latentes, ambas foco deste estudo, assim como a geração de imagens artificiais para o balançamento de classes. Este capítulo apresentou diversos métodos para exemplificação, além de trabalhos similares.

A extração de características foi abordada, apresentando os principais descritores utilizados nesta pesquisa. A lacuna destacada é que existem características não passíveis de extração por descritores convencionais. Para isso, as redes de convolução são apresentadas, pois possuem capacidade de aprender as características relevantes das imagens de entrada. Justificando, assim, seu uso para análise das propriedades das bases de imagens. Podem também indicar possíveis operações para auxiliar na geração de imagens artificiais. Ainda, a memória associativa aprendida por máquinas de Boltzmann restritas pode ser indicadora de quais imagens geradas adicionam informações ao aprendizado. A geração dessas imagens visa rebalancear classes que diferem em número de imagens, e detalhes sobre esse problema também foram fundamentados nesse capítulo.

Esses fundamentos permitem compreender em que contexto esta dissertação de mestrado está inserida. O próximo capítulo abrangerá a proposta deste trabalho.



GERAÇÃO ARTIFICIAL DE IMAGENS

3.1 Considerações iniciais

O processo de manipular imagens para que elas se tornem mais satisfatórias para um determinado objetivo depende do domínio de aplicação. Ou seja, não existe uma teoria geral para melhorar qualquer tipo de imagem (GONZALEZ; WOODS, 2007): um método que processa melhor uma imagem bem definida pelas suas cores difere do processamento de imagens texturizadas, às quais um processamento sobre a intensidade dos pixels da imagem – como uma operação de borramento – pode ocasionar perda da textura. Assim, justifica-se a exploração de um vasto número de métodos de processamento de imagens e bases.

Nesta pesquisa oito métodos de processamento de imagens são aplicados nas imagens minoritárias originais, gerando imagens artificiais. Isso é realizado a fim de permitir a extração de informações latentes com o objetivo de melhorar a classificação com alguma técnica de Aprendizado de Máquina, o que reflete a melhora da diferenciação entre as classes. Dada a quantidade de imagens necessárias para rebalancear a base original, são geradas imagens utilizando cada um dos métodos, além de uma versão combinando todos eles (ou seja, compondo um conjunto com algumas imagens processadas por cada método) e outra apenas replicando as imagens como *baseline*. Como demonstrado na Figura 10, dado o conjunto de treinamento da classe (ou classes) com menor número total de imagens, é realizado o rebalanceamento ao aplicar os métodos descritos neste capítulo e posteriormente essas imagens resultantes são utilizadas como treinamento.

Neste capítulo os métodos de geração artificial para o rebalanceamento de classes de imagens são descritos. Os experimentos posteriormente destacados no Capítulo 4 foram realizados utilizando: borramento; mistura ponderada; *unsharp masking*; composição; combinação de thresholds; combinação com saliência; visual SMOTE; e adição de ruído.

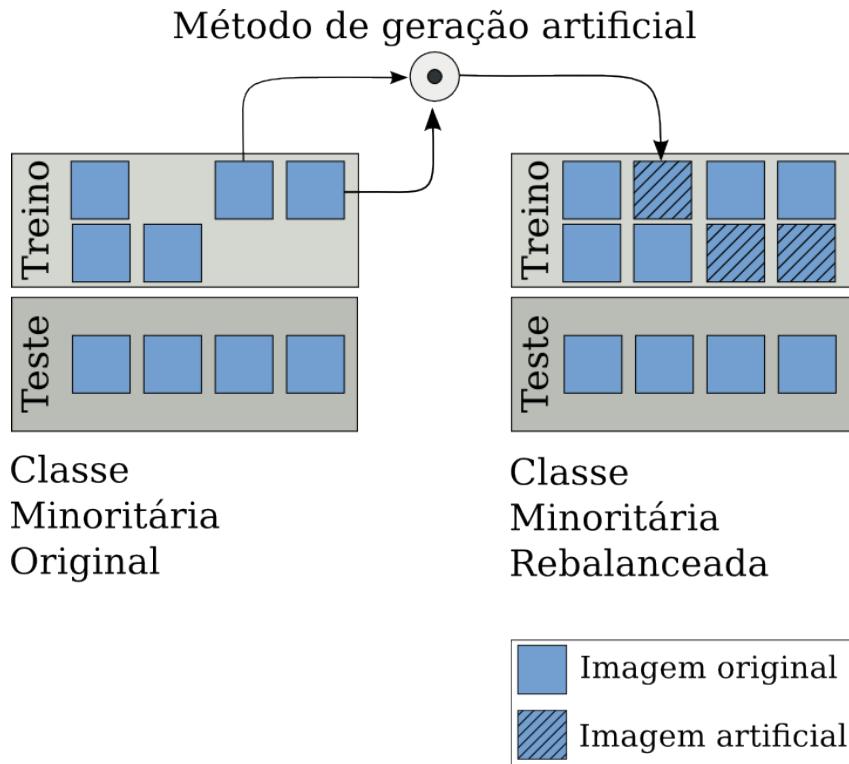


Figura 9 – Geração artificial a classe minoritária para rebalancear as classes. *Fonte:* Elaborado pela autora.

3.2 Borramento

Também conhecido como filtro de suavização, o borramento é uma operação de processamento comumente utilizada com o objetivo de filtrar as baixas frequências de uma imagem, removendo ruídos e detalhes não relevantes. Normalmente esse tipo de filtro provoca também um certo borramento das bordas, como pode ser observado na Figura 4.

Esse comportamento não é esperado quando devemos gerar novas imagens, pois informações relevantes podem ser removidas. Dessa forma, a operação de borramento utilizada é a de filtro bilateral. Ela substitui o valor do pixel (x,y) por uma média dos pixels de intensidade similar na imagem e dos pixels vizinhos (TOMASI; MANDUCHI, 1998). Ou seja, é uma média ponderada das intensidades. O Algoritmo 2 descreve os passos desse filtro e a Figura 11 exemplifica o seu funcionamento: à esquerda está demonstrada a imagem original e à direita a imagem borrada.

Algoritmo 2: Algoritmo de borramento com filtro bilateral

Entrada: Imagem colorida I em formato BGR

Entrada: Diâmetro d de pixinhança de pixels

Entrada: Sigma do espaço de cor

Entrada: Sigma do espaço de coordenadas

Saída: Imagem gerada G

1 **Algoritmo de filtro bilateral**



(a) Original



(b) Imagem artificial

Figura 10 – Geração artificial utilizando borramento com filtro bilateral. *Fonte: Elaborado pela autora.*

Parâmetros e suas variações Conforme descrito no Algoritmo 2, os parâmetros para essa geração são: o diâmetro d de pixinhança de pixels, o σ do espaço de cor e o σ do espaço de coordenadas. Esses parâmetros dependem das propriedades das imagens e dos resultados pretendidos. Dessa forma, o tamanho do filtro é um valor escolhido arbitrariamente para cada aplicação em específico (TOMASI; MANDUCHI, 1998). Como o nosso objetivo com a geração das imagens não foi especializar no comportamento de uma classe de imagens específica, um valor foi escolhido aleatoriamente e a partir dele os parâmetros de entrada foram definidos.

Limitações Esse filtro tende a remover texturas e a criar novos contornos. Dependendo dos valores, pode gerar uma imagem “cartoonizada”.

Métodos relacionados São diversos os métodos de borramento descritos na literatura, como a filtragem Gaussiana, a de mediana e a de médias.

Visualização É interessante notar na Figura ?? o comportamento da adição de imagens borradas para o rebalanceamento de classes bem descritas pela propriedade da cor.

colocar aqui alguma figura do espaço...

3.3 Aguçamento

Diferentemente da suavização, o processamento de aguçamento procura enfatizar as transições de intensidade. Um processo bem conhecido para atingir tal objetivo é o *unsharp mask*. Ele borra a imagem, subtrai a imagem borrada da original e adiciona essa diferença na imagem original (Ver Algoritmo 3). A imagem resultante, ilustrada na Figura 12, é uma versão realçada da imagem original, dado que soma à imagem justamente o que é removido com um filtro de suavização.

Algoritmo 3: Algoritmo de aguçamento

Entrada: Imagem colorida I em formato BGR

Saída: Imagem gerada G

```

1 borrada  $\leftarrow$  filtro de suavização( $I$ ) para cada pixel  $(x,y)$  faça
2   |   diferença  $\leftarrow I(x,y) - \text{borrada}(x,y);$ 
3   |    $G(x,y) \leftarrow I(x,y) + k * \text{diferença};$ 
4 fim
```



(a) Original



(b) Imagem artificial

Figura 11 – Geração artificial utilizando unsharp masking. *Fonte: Elaborado pela autora.*

Parâmetros e suas variações Pode-se variar o parâmetro k de forma a ponderar a soma dessa diferença. Para a geração das imagens da classe minoritária, foi utilizado $k = 1$.

Limitações É possível que existam pixels com valor negativo no resultado final. Isso pode causar o aparecimento de uma áurea em volta das bordas, efeito não desejado (GONZALEZ; WOODS, 2007).

Métodos relacionados Outros algoritmos de aguçamento conhecidos são: utilizar primeira derivada (gradiente) ou a segunda derivada da imagem (Laplaciano).

Visualização

3.4 Adição de ruído

O ruído de Poisson ocorre na contagem de fótons de dispositivos ópticos. Ele segue a distribuição de Poisson, que expressa a probabilidade de um certo número de eventos ocorrerem em um intervalo fixo de tempo e/ou espaço se esses eventos ocorrem com uma taxa média conhecida. O efeito da adição de ruído pode ser visto na Figura 13.



(a) Original



(b) Imagem artificial

Figura 12 – Geração artificial utilizando adição de ruído de Poisson. *Fonte: Elaborado pela autora.*

descrever o algoritmo

A distribuição de Poisson segue a equação:

Uma possível implementação para encontrar os valores de Poisson foi desenvolvida por Knuth e pode ser vista no Algoritmo 4.

Algoritmo 4: Algoritmo da geração com ruído de Poisson

Entrada: Imagem colorida I em formato BGR

Saída: Imagem gerada G

```

1 para cada pixel  $(x,y)$  faça
2    $L \leftarrow \exp(-I(x,y));$ 
3    $p \leftarrow 1;$ 
4    $k \leftarrow 0;$ 
5   faça
6      $k \leftarrow k + 1;$ 
7      $p \leftarrow p * \text{número aleatório entre } 0 \text{ e } 1;$ 
8   enquanto  $p > L;$ 
9    $G(x,y) \leftarrow k - 1;$ 
10 fim
```

Limitações A adição de ruído é normalmente indesejável e a utilizamos para englobar mais um processamento de imagens. (??)

Métodos relacionados

Visualização

3.5 SMOTE visual

Conforme visto na Seção 2.7.1, o SMOTE é um método de rebalanceamento aplicado após a extração de características. A ideia dessa geração, chamada de SMOTE visual, é imitar esse funcionamento no nível de pixels. A diferença é que não é feito entre as imagens mais próximas, mas sim entre duas imagens escolhidas de forma aleatória do conjunto de treinamento da classe minoritária.

Para cada pixel é calculado a diferença entre as duas imagens. Essa diferença é então multiplicada por um número aleatório no intervalo $[0 - 1]$ e adicionado na imagem original (Ver Algoritmo 5). O efeito que esse processamento causa na imagem pode ser visualizado na Figura 14.

Algoritmo 5: Algoritmo da geração com SMOTE visual

Entrada: Imagem colorida I em formato BGR

Entrada: Imagem colorida I_2 em formato BGR

Saída: Imagem gerada G

```

1 para cada pixel  $(x, y)$  faça
2   |   diferença  $\leftarrow I(x, y) - I_2(x, y);$ 
3   |   gap  $\leftarrow$  número aleatório entre 0 e 1;
4   |    $G(x, y) \leftarrow I(x, y) + gap * \text{diferença};$ 
5 fim
6 mínimo  $\leftarrow$  menor valor de  $G;$ 
7 máximo  $\leftarrow$  maior valor de  $G;$ 
8 para cada pixel  $(x, y)$  faça
9   |    $G(x, y) \leftarrow G(x, y) - \text{mínimo};$ 
10 fim
11 para cada pixel  $(x, y)$  faça
12   |    $G(x, y) \leftarrow G(x, y) * (255 / (\text{máximo} - \text{mínimo}));$ 
13 fim

```

Limitações Esse método adiciona texturas e bordas que não estavam originalmente nas imagens.



Figura 13 – Geração artificial utilizando o método SMOTE no espaço visual. *Fonte: Elaborado pela autora.*

Métodos relacionados Esse método é visualmente parecido com o de mistura ponderada, apresentado na próxima seção.

Visualização

3.6 Mistura ponderada

Essa geração calcula a soma ponderada de duas imagens, de acordo com o Algoritmo 6. O efeito dessa mistura pode ser visto na Figura 15, onde dadas duas imagens como entrada, a imagem da direita corresponde a soma delas.

Algoritmo 6: Algoritmo de mistura ponderada

Entrada: Primeira imagem colorida I em formato BGR

Entrada: Segunda imagem colorida I_2 em formato BGR

Saída: Imagem gerada G

- 1 $\alpha \leftarrow$ número aleatório entre 10 e 80;
 - 2 $\beta \leftarrow 100 - \alpha;$
 - 3 **para cada** pixel (x, y) **faça**
 - 4 $| G(x, y) \leftarrow \beta.I(x, y) + \alpha.I_2(x, y);$
 - 5 **fim**
-

Parâmetros e suas variações Os parâmetros α e β são escolhidos de forma aleatória. Um valor entre 10% e 80% é escolhido para α ; e o β é o restante para completar 100%.

Limitações Assim como todas as gerações artificiais que envolvem a mistura de imagens, efeitos são adicionados às imagens originais. Dependendo da combinação dos métodos de descrição, quantização e classificação, isso pode piorar a acurácia da classificação.



Figura 14 – Geração artificial utilizando uma mistura ponderada de duas imagens. *Fonte: Elaborado pela autora.*

Métodos relacionados É um método de combinação de imagens primitivo. Algoritmos similares são muito mais complexos, como os de threshold e saliência descritos a seguir.

Visualização

3.7 Mistura limiarizada

A combinação de *thresholds* é uma composição do fundo (*background*) de uma imagem e do objeto da cena (*foreground*) de outra imagem. A Figura 16 mostra a mistura dos *thresholds* de duas imagens originais para compor uma nova imagem. O Algoritmo 7 descreve as operações necessárias para realizar tal processamento.



Figura 15 – Geração artificial utilizando uma mistura limiarizada de duas imagens. *Fonte: Elaborado pela autora.*

Algoritmo 7: Algoritmo de mistura limiarizada

Entrada: Imagem colorida I em formato BGR

Entrada: Imagem colorida I_2 em formato BGR

Saída: Imagem gerada G

- 1 $I_{cinza} \leftarrow$ escala de cinza(I);
 - 2 $I_{threshold} \leftarrow OTSU(I_{cinza})$;
 - 3 $I_{morfologica} \leftarrow$ abertura e dilatação($I_{threshold}$);
 - 4 $I_{foreground} \leftarrow$ aplica máscara($I_{morfologica}, I$);
 - 5 $I_{morfologica} \leftarrow$ oposto($I_{morfologica}$);
 - 6 $I_{background} \leftarrow$ aplica máscara($I_{morfologica}, I_2$);
 - 7 $G \leftarrow I_{background} + I_{foreground}$;
-

Parâmetros e suas variações No âmbito desta pesquisa, os parâmetros estão fixos, mas é possível modificar o tamanho dos elementos estruturantes que fazem as operações de abertura e dilatação para remover pequenas regiões.

Limitações Dependendo da quantidade de informações da imagem, o *threshold de OTSU* pode não conseguir extrair nenhuma informação relevante ou mesmo a imagem toda.

Métodos relacionados Essa geração está fortemente correlacionada com a mistura a partir da saliência da imagem, apresentada a seguir.

Visualização

3.8 Mistura saliente

A combinação de regiões salientes é muito similar com o método anterior de combinação de *thresholds*, porém, utiliza um algoritmo mais robusto que detecta a saliência da imagem a partir do método SLIC. A Figura 17 mostra a combinação da região saliente da imagem original à esquerda com a imagem central, resultando na imagem combinada à direita.

As operações aplicadas na imagem para extrair a região mais saliente são: SLIC; rotulação por conectividade; *threshold de OTSU*; e operações morfológicas. O Algoritmo 8 apresenta os passos para o cálculo do *background* e *foreground*.

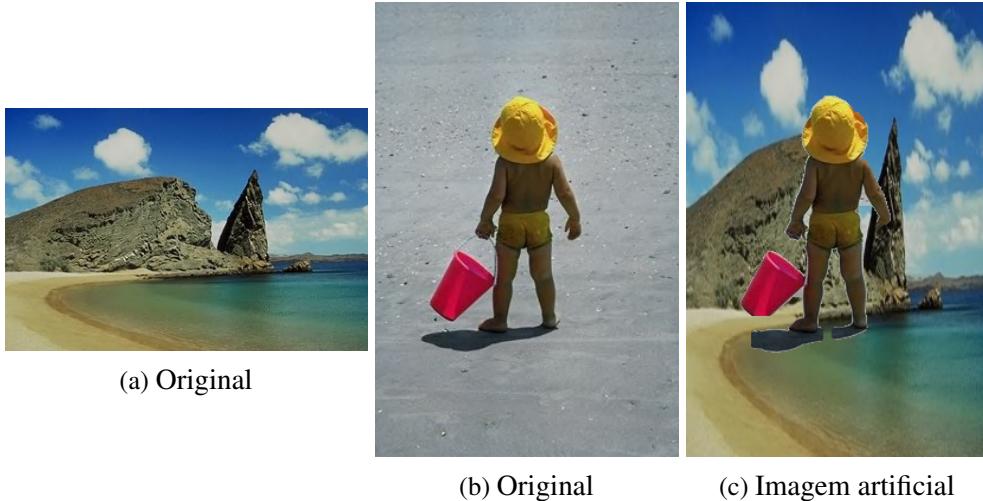


Figura 16 – Geração artificial utilizando uma mistura de duas imagens a partir da saliência da primeira imagem.
Fonte: Elaborado pela autora.

Algoritmo 8: Algoritmo de mistura saliente

Entrada: Imagem colorida I em formato BGR

Entrada: Imagem colorida I_2 em formato BGR

Saída: Imagem gerada G

- 1 $I_{\text{rotulada por segmento}} \leftarrow SLIC(I);$
 - 2 $I_{\text{mapa de saliência}} \leftarrow \text{rotulação por conectividade}(I_{\text{rotulada por segmento}});$
 - 3 $I_{\text{threshold}} \leftarrow OTSU(I_{\text{mapa de saliência}});$
 - 4 $I_{\text{morfologica}} \leftarrow \text{abertura e dilatação}(I_{\text{threshold}});$
 - 5 $I_{\text{foreground}} \leftarrow \text{aplica máscara}(I_{\text{morfologica}}, I);$
 - 6 $I_{\text{morfologica}} \leftarrow \text{oísto}(I_{\text{morfologica}});$
 - 7 $I_{\text{background}} \leftarrow \text{aplica máscara}(I_{\text{morfologica}}, I_2);$
 - 8 $G \leftarrow I_{\text{background}} + I_{\text{foreground}};$
-

Parâmetros e suas variações Assim como no método anterior, os parâmetros são relacionados ao tamanho do elemento estruturante para a abertura e dilatação e estão fixos.

Limitações Não é garantido que o algoritmo de saliência consiga extrair a melhor região, ou mesmo que sempre haja uma região.

Métodos relacionados Similar à mistura por *thresholds*.

Visualização

3.9 Composição

Essa geração pretende compor informações de diversas imagens em uma única imagem. Assim é feito um mosaico com várias imagens, conforme pode ser visto na Figura 18. Se as imagens possuem um elemento centralizado, essa geração pode resultar em uma imagem de um objeto que parece uma mutação dos objetos centrais, conforme pode ser visualizado na Figura ??.



Figura 17 – Geração artificial utilizando uma composição de imagens. *Fonte: Elaborado pela autora.*

Para cada quadrado a ser preenchido sorteia uma imagem do conjunto de treinamento, realiza uma operação de borramento, aguçamento, mistura ponderada ou visual SMOTE e adiciona essa imagem no quadrado respectivo. Os passos para tal composição estão descritos no Algoritmo 9.

Algoritmo 9: Algoritmo de composição

Saída: Imagem gerada G

```

1 enquanto  $total < \text{número de quadrados } q$  faça
2    $I \leftarrow$  imagem aleatória do conjunto de treinamento;
3    $operação \leftarrow 1 + (\text{rand}() \% 3);$ 
4   selecione  $operação$  faça
5     caso 1
6        $I \leftarrow \text{borramento}(I);$ 
7     caso 2
8        $I \leftarrow \text{mistura ponderada}(I);$ 
9     caso 3
10     $I \leftarrow \text{aguçamento}(I);$ 
11    caso 4
12     $I \leftarrow \text{visual SMOTE}(I);$ 
13  fim
14   $x \leftarrow \text{posição aleatória em } x \text{ de } I;$ 
15   $y \leftarrow \text{posição aleatória em } y \text{ de } I;$ 
16   $qx \leftarrow \text{posição atual para o quadrado em } x \text{ de } G;$ 
17   $qy \leftarrow \text{posição atual para o quadrado em } y \text{ de } G;$ 
18   $G(qx, qy) \leftarrow I(x, y);$ 
19   $total ++;$ 
20 fim

```

Parâmetros e suas variações O parâmetro q controla quantos quadrados serão criados na nova imagem. Nesta pesquisa foram realizados testes com 4 e 16.

Limitações O término brusco de uma imagem para início da outra ao formar a grade de imagens tenha efeitos colaterais de inserção de textura que não excedam a vantagem de compor uma mesma imagem com várias cores, texturas e formas das imagens originais.

Métodos relacionados

Visualização

3.10 Considerações finais

CAPÍTULO
4

RESULTADOS

Essa seção descreve os resultados encontrados, com foco na análise do espaço de características utilizando uma projeção multidimensional. Na figura 19 é possível observar o fluxo de operações realizadas para analisar o impacto da geração de imagens no rebalanceamento de classes. O mesmo protocolo de conversão para escala de cinza, extração de características e classificação foi seguido para três experimentos: base desbalanceada; base rebalanceada com interpolação dos vetores de características; e base rebalanceada com a geração artificial de imagens. Portanto inicia-se a seção com a descrição destas bases de imagens, assim como a descrição do protocolo adotado e os resultados obtidos a partir de seu uso.

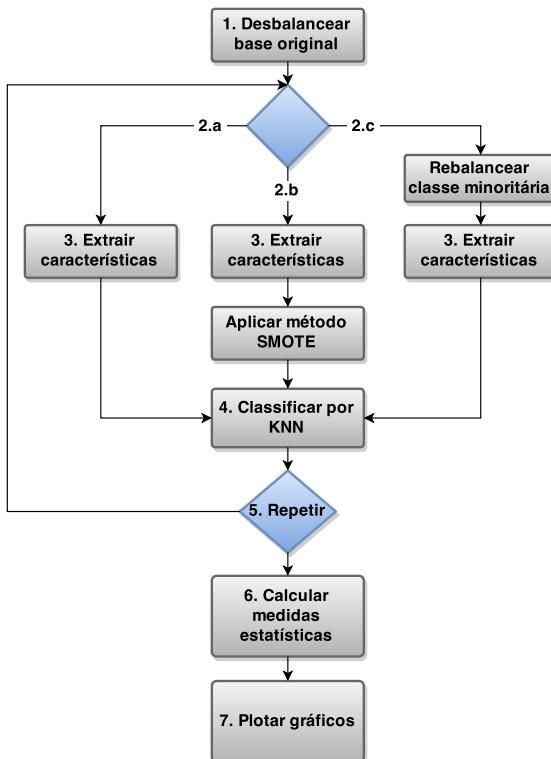


Figura 18 – Fluxo de operações do experimento realizado.

4.0.1 Base de Imagens

Os resultados foram obtidos utilizando a base de imagens COREL¹, composta por fotografias que representam classes variadas: tribos africanas, praia, construções, ônibus, dinossauros, elefantes, flores, cavalos, montanhas e tipos de comidas. São 10 classes balanceadas com 100 imagens cada. Para fins de exemplificação, são apresentadas amostras das imagens que representam essas classes na Figura 20.



Figura 19 – Base de imagens COREL-1000 utilizada. Estão representadas as 10 classes da base. *Fonte: Elaborado pela autora.*

4.0.2 Protocolo

O seguinte protocolo foi seguido para a obtenção dos resultados:

1. **Classes de imagens originais:** classes Elefante e Cavalo da COREL-1000;
2. **Amostragem dos dados:** 50% para treino e 50% para teste de cada classe;
3. **Desbalanceamento:** remoção de 50% do conjunto de treino da classe Cavalo;
4. **Método para geração artificial:** mistura de duas imagens originais, exemplificado na Figura ??;
5. **Quantização:** Intensidade. É o método mais simples de converter uma imagem colorida em tons de cinza. Consiste em computar a média entre os canais RGB da imagem a partir de

$$Q_{\text{Intensidade}} = \frac{1}{3}(R + G + B) \quad (4.1)$$

e então realizar uma correção por *gamma*;

6. **Extração de características:** classificação de pixels de borda e interior (BIC). Computa dois histogramas, um para pixels definidos como borda e outro como interior. Se um pixel possuir a mesma cor que seus vizinhos, é considerado pixel de interior; caso contrário, será

¹ Disponível em <http://wang.ist.psu.edu/docs/related/>

pixel de borda. Os histogramas são concatenados em um vetor de $2N$ dimensões, onde N é o número de intensidades de cor da imagem (STEHLING; NASCIMENTO; aO, 2002);

7. **Classificação:** classificador supervisionado K=NN com $K = 1$ (para mais detalhes ver Seção ??);
8. **Projeção multidimensional:** projetados os dois componentes principais encontrados ao aplicar PCA nos vetores de características para redução de dimensionalidade (Seção 2.9.1). A implementação desenvolvida para esta projeção pode ser encontrada em <<https://github.com/GabiThume/msc-src/blob/master/visualization/plot.py>>. Vale destacar que uma animação da adição de cada novo exemplo foi realizada para fim de visualizar o melhor subespaço gerado.

4.0.3 Resultados e Discussão

As classes Elefante e Cavalo possuem 100 imagens cada. O primeiro passo é remover imagens de uma das classes, tornando a base desbalanceada. Como o foco é na visualização do espaço de características, é relevante ter o modelo do espaço ideal das classes balanceadas, por isso esse experimento em específico não trata de uma base naturalmente desbalanceada. Na Figura 21 está ilustrada a remoção de 50% das imagens de treino da classe Cavalo, originalmente balanceada. Essa e as próximas projeções desta seção foram obtidas com a técnica para redução de dimensionalidade PCA, descrita na Seção 2.9.1 e são referentes aos dois componentes principais com maiores autovalores.

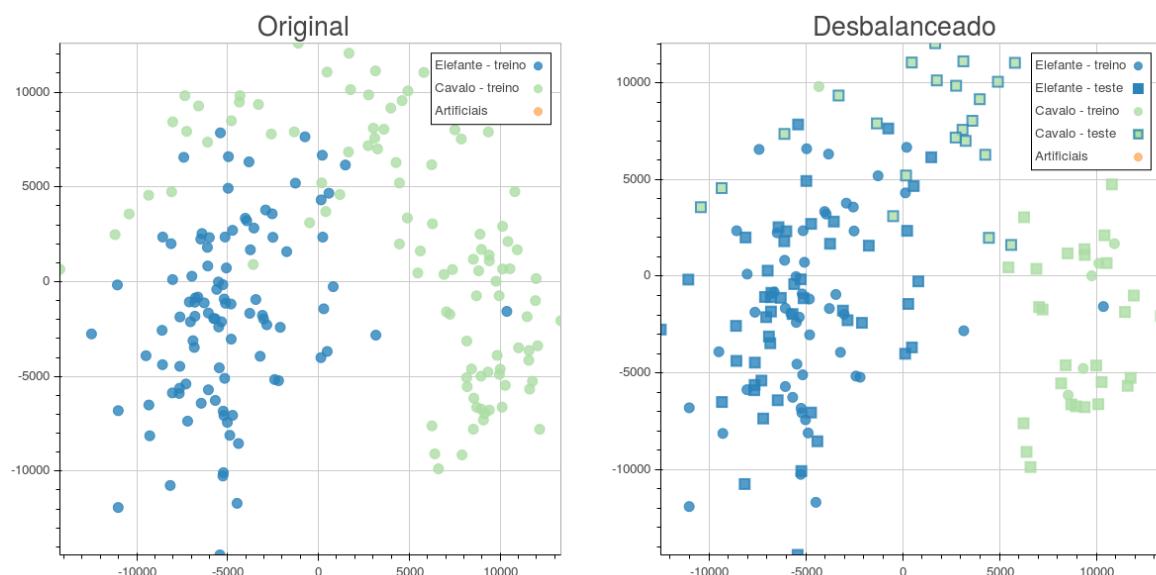


Figura 20 – Remoção de 50% das imagens de treino da classe Cavalo.

A classificação dos três experimentos utilizando KNN reportou que o *f-score* da geração artificial de imagens utilizando o método de mistura teve um ganho de mais de 10% em relação ao

rebalanceamento no espaço de características com o SMOTE. Para confirmar que a geração aqui proposta inseriu mais informação na classe minoritária do que apenas povoar os espaços entre os exemplos (i.e. SMOTE), a classe rebalanceada utilizando ambos métodos está demonstrada na Figura 22. Em laranja estão representados os novos exemplos de treinamento, projetados no plano da base original balanceada.

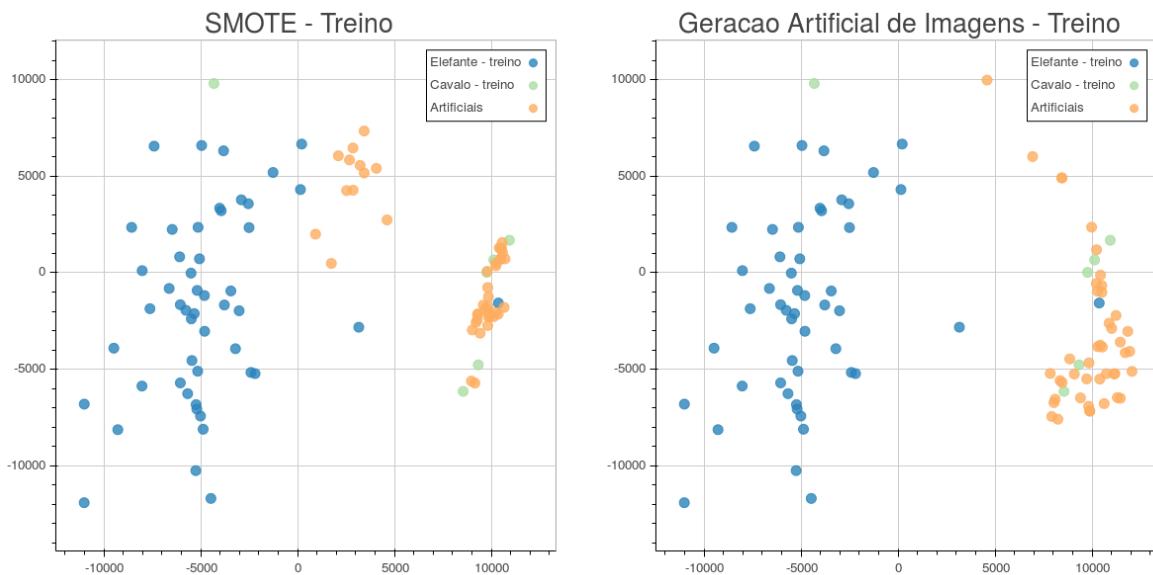


Figura 21 – Comparação dos exemplos de treinamento da geração com SMOTE e no campo visual. Em laranja estão representados os novos exemplos, projetados no plano da base original balanceada.

Após o treinamento realizado com as novas imagens e exemplos, o conjunto de teste foi fornecido ao classificador 1-NN e o resultado das predições está ilustrado na Figura 23. A cor no interior dos marcadores quadrados representa a classe real dos exemplos e a borda representa a classe predita pelo classificador. Nota-se que a melhoria na classificação com a geração de imagens fica visível e corresponde ao aumento do *f-score*.

De uma forma geral, pode-se dizer que a geração de imagens melhorou a definição da classe minoritária e foi o método que mais se assemelhou à distribuição dos dados originais. Além disso, um dos problemas do SMOTE pode ser verificado nessas projeções: ao realizar a interpolação dos vetores de características originais, pode-se criar exemplos em regiões do espaço que fazem parte da outra classe. Ficou claro também que o método SMOTE não possui capacidade de extrapolar a sua região, como pode ser observado no grupo de exemplos gerados à direita do espaço. O SMOTE gerou novos elementos em linha reta, enquanto a geração de imagens proporcionou uma abrangência maior em volta desse espaço, com maior dispersão.

É válido também visualizar a região de decisão, observando suas modificações frente aos métodos. Pode ser observado que em ambas técnicas a região da classe minoritária apresenta-se melhor representada. Além disso, é possível verificar que o SMOTE ocasionou uma certa “invasão” do espaço de características da classe majoritária.

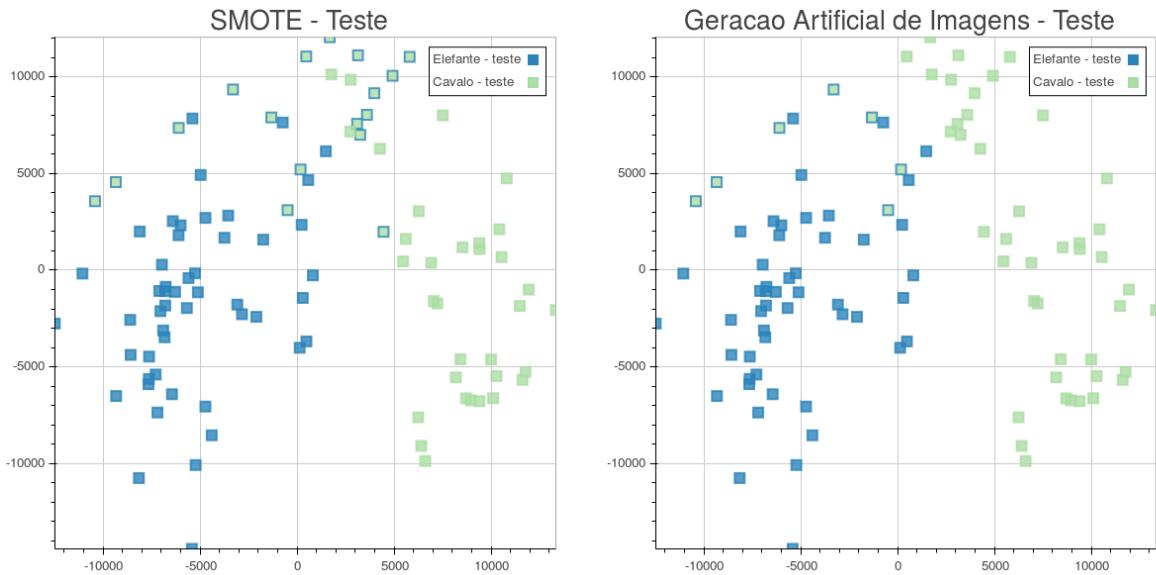


Figura 22 – Resultado do teste da classificação com 1-NN após o treinamento realizado com as bases rebalanceadas. A cor no interior dos marcadores quadrados representa a classe real dos exemplos e a borda representa a classe predita pelo classificador.

Nas figuras anteriores os exemplos foram projetados no plano criado pelas suas componentes principais com maior autovalores da base original balanceada. Se após a geração de novos exemplos essas componentes forem recalculadas (Figura 25), pode-se notar que a geração de imagens artificiais proporciona a criação de um subespaço que melhor discretiza as classes, quando comparado com SMOTE ou com a base desbalanceada.

Como relatado no início dessa seção, o extrator de características utilizado foi o BIC. Fundamentalmente ele captura informações de intensidade de cor das imagens. Na Figura 26 as próprias imagens foram utilizadas como marcadores na projeção do melhor subespaço após a geração artificial com o método de mistura. É nítido o impacto da etapa de extração de características na separação das classes e também no método de geração de imagens antes dessa extração.

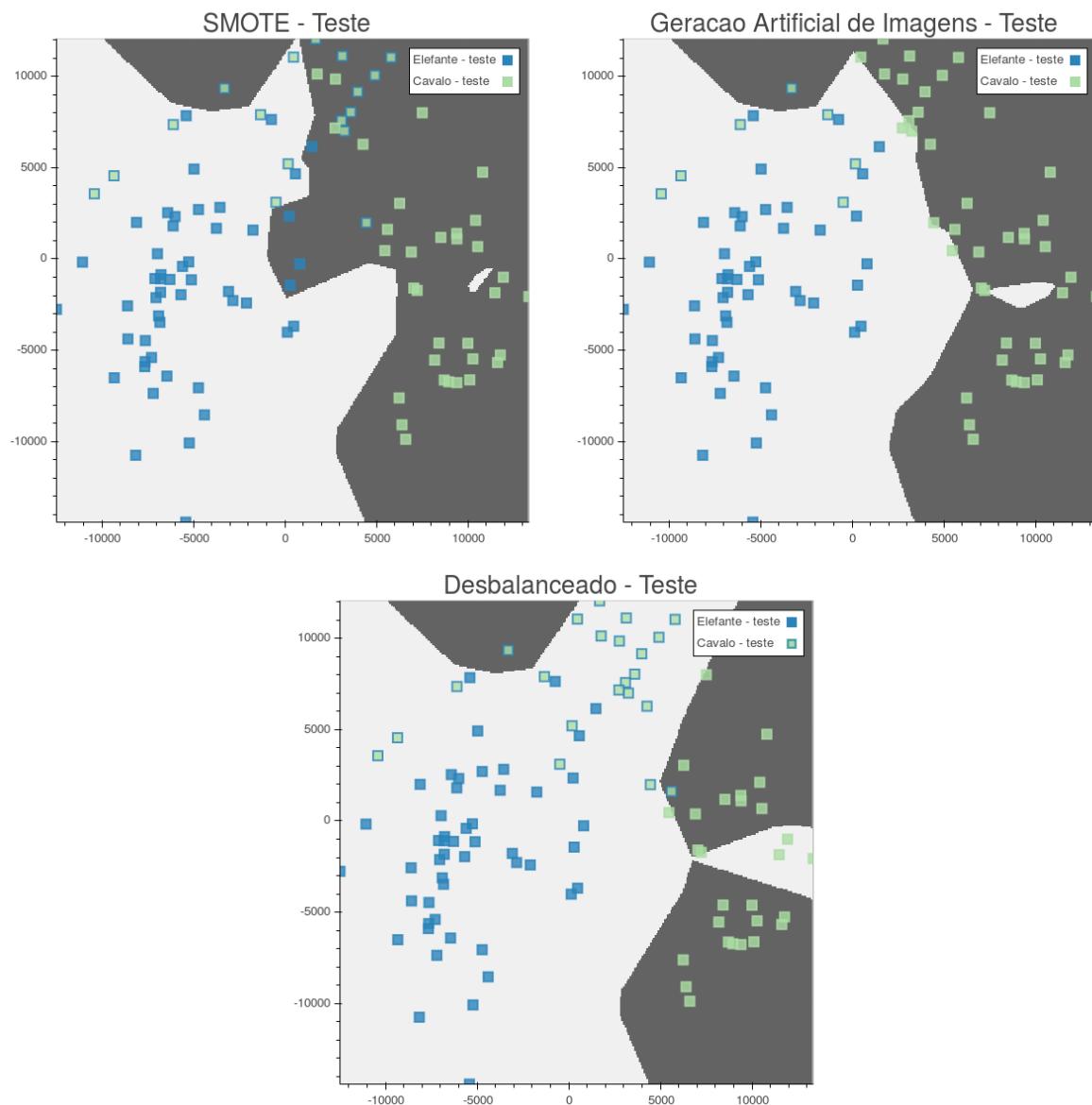


Figura 23 – Região de decisão com K-NN ($K = 1$)

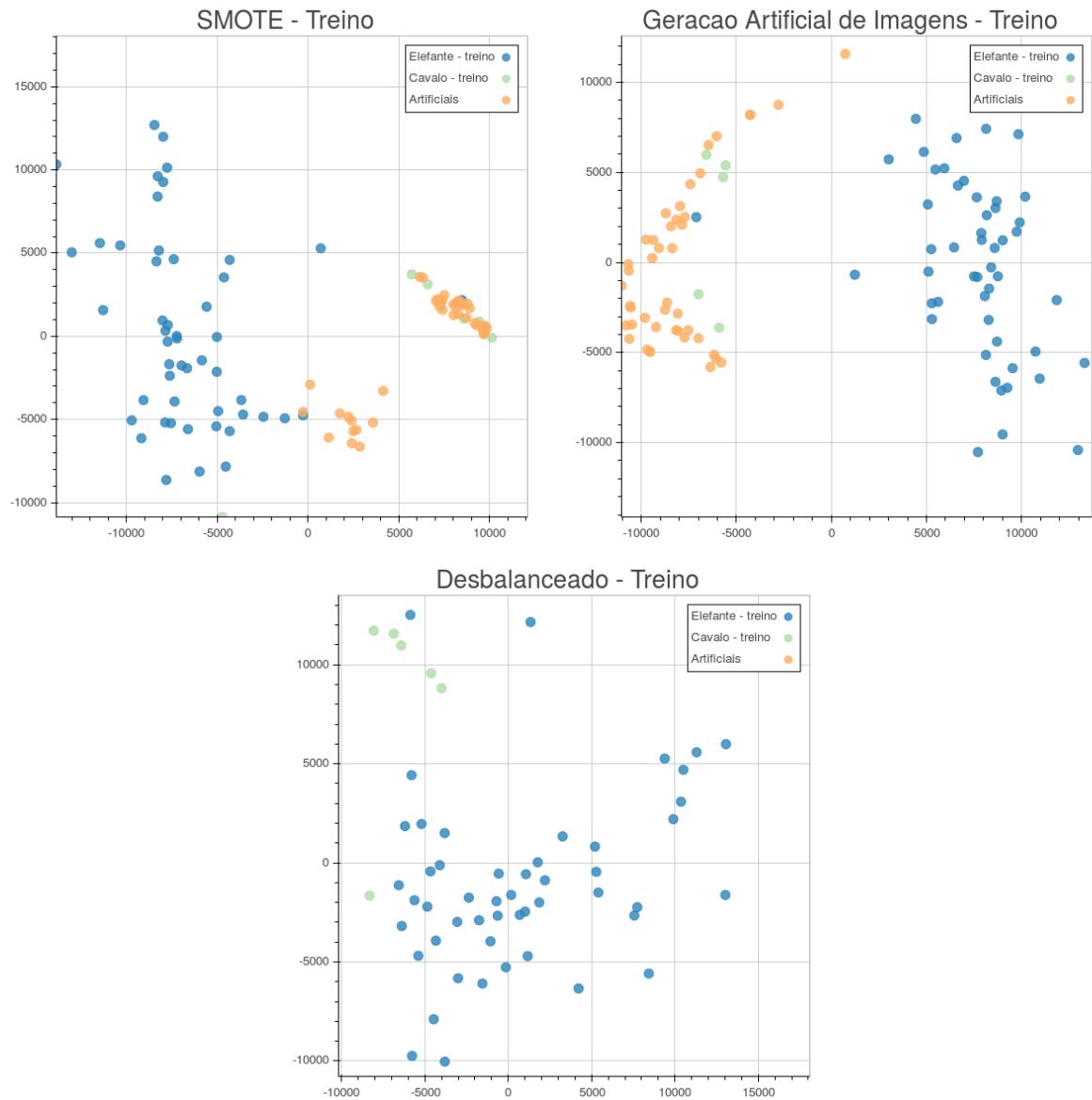


Figura 24 – Melhores subespaços encontrados após a geração de novos exemplos para o SMOTE e para a geração artificial de imagens, e após a remoção de imagens para a projeção dos dados desbalanceados.

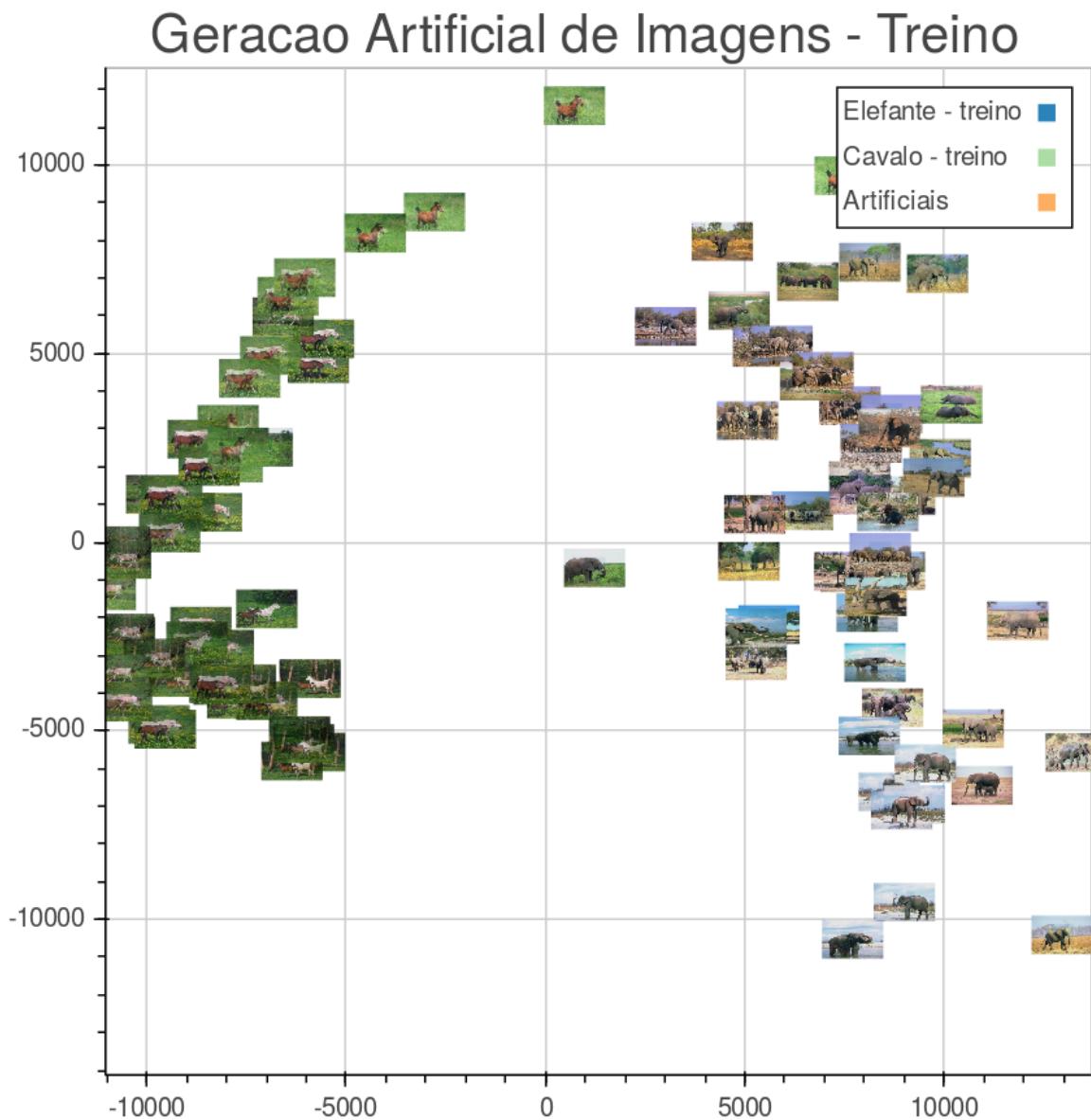


Figura 25 – Visualização do impacto do descritor de características.



CONCLUSÕES

Com os experimentos realizados foi possível notar que a geração de imagens artificiais pode gerar novas informações para a classificação das imagens. Assim a geração de elementos no espaço visual provou contribuir com o balanceamento entre classes (em se tratando de problemas de classes desbalanceadas), melhorando a acurácia de algoritmos de classificação, quando comparada à geração de exemplos artificiais no espaço de atributos (i.e. SMOTE).

A visualização do espaço de características utilizando a técnica de análise de componentes principais (PCA) se mostrou crucial para confirmar visualmente que a melhora da acurácia da classificação se deve à melhor definição da classe minoritária.

Para validar a ideia da geração artificial de imagens, as características das novas imagens – extraídas com o método BIC – e os exemplos resultantes da interpolação de vetores originais foram projetados no plano das imagens originais antes do desbalanceamento. Dessa forma foi possível visualizar que a geração de imagens artificiais proposta foi capaz de ocupar uma região do espaço mais abrangente do que o SMOTE. Este último, comprovadamente, possui o ponto negativo de não extrapolar os limites da classe minoritária. Ainda, está suscetível à criação de novos exemplos em regiões da classe majoritária, o que também prejudica a classificação.

A projeção forneceu uma análise intuitiva sobre o comportamento das técnicas que não é facilmente analisando quando olhado apenas os valores de acurácia ou *f-score*.

Como extensão dos experimentos reportados pode ser feita a análise dos espaços encontrados para os diferentes métodos de geração artificial de imagens. Além disso, o impacto de tais métodos em diferentes extractores de características pode sugerir quais são as características latentes percebidas com cada extrator.

5.1 Publicações

5.2 Trabalhos Futuros

REFERÊNCIAS

- AHONEN, T.; HADID, A.; PIETIKÄINEN, M. Face description with local binary patterns: application to face recognition. **IEEE transactions on pattern analysis and machine intelligence**, v. 28, n. 12, p. 2037–2041, dez. 2006. ISSN 0162-8828. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/17108377>>. Citado na página 37.
- BATISTA, G. E.; PRATI, R. C.; MONARD, M. C. A study of the behavior of several methods for balancing machine learning training data. **ACM Sigkdd Explorations Newsletter**, v. 6, n. 1, p. 20–29, 2004. Disponível em: <<http://dl.acm.org/citation.cfm?id=1007735>>. Citado 2 vezes nas páginas 50 e 53.
- BENGIO, Y. Learning Deep Architectures for AI. **Foundations and Trends in Machine Learning**, Now Publishers Inc., v. 2, n. 1, p. 1–127, jan. 2009. ISSN 1935-8237. Disponível em: <<http://dl.acm.org/citation.cfm?id=1658423.1658424>>. Citado na página 32.
- BENGIO, Y.; GOODFELLOW, I. J.; COURVILLE, A. Deep learning. Book in preparation for MIT Press. 2014. Disponível em: <<http://www.iro.umontreal.ca/~bengioy/dlbook>>. Citado 2 vezes nas páginas 47 e 48.
- BORGES, V.; OLIVEIRA, M. C. de; FERREIRA, T. G.; VIEIRA, A. A. Feature Extraction and Interactive Visualization to Assist Green Algae Taxonomic Classification. In: **XXVI SIBGRAPI - Conference on graphics, patterns and images**. Arequipa, Peru: [s.n.], 2013. p. 4. Disponível em: <<http://www.ucsp.edu.pe/sibgrapi2013/eproceedings/wip/115151.pdf>>. Citado 2 vezes nas páginas 37 e 43.
- BUADES, A.; COLL, B.; MOREL, J. M. A Review of Image Denoising Algorithms, with a New One. **Multiscale Modeling & Simulation**, Society for Industrial and Applied Mathematics, v. 4, n. 2, p. 490–530, jan. 2005. ISSN 1540-3459. Disponível em: <<http://pubs.siam.org/doi/abs/10.1137/040616024>>. Citado na página 41.
- CASTRO, C. L.; BRAGA, A. P. Aprendizado supervisionado com conjuntos de dados desbalanceados. **Sba Controle & Automação**, v. 22, n. 5, p. 441 – 446, 2011. Citado na página 50.
- CHAWLA, N. V.; HALL, L. O.; BOWYER, K. W. SMOTE: Synthetic Minority Over-sampling Technique. **Journal of Artificial Intelligence Research**, v. 16, p. 321–357, 2002. Disponível em: <<http://link.springer.com/chapter/10.1007/3-540-45428-4\5http://arxiv.org/abs/1106.1813>>. Citado 2 vezes nas páginas 50 e 51.
- FISCHER, A.; IGEL, C. Training restricted Boltzmann machines: An introduction. **Pattern Recognition**, v. 47, n. 1, p. 25–39, jan. 2014. ISSN 00313203. Citado 2 vezes nas páginas 48 e 49.
- FROME, A.; CHEUNG, G.; ABDULKADER, A.; ZENNARO, M.; BISSACCO, A.; ADAM, H.; NEVEN, H.; VINCENT, L. Large-scale privacy protection in Google Street View. In: **2009 IEEE 12th International Conference on Computer Vision**. IEEE, 2009. p. 2373–2380. ISBN

978-1-4244-4420-5. ISSN 1550-5499. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5459413>>. Citado na página 46.

GEORGE, D. **How the brain might work: a hierarchical and temporal model for learning and recognition**. Tese (Dissertation) — Stanford University, 2008. Citado 2 vezes nas páginas 45 e 46.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing**. 3. ed. [S.l.]: Prentice-Hall, 2007. ISBN 013168728X. Citado 7 vezes nas páginas 36, 38, 40, 41, 43, 57 e 60.

GROSS, R.; BRAJOVIC, V. An image preprocessing algorithm for illumination invariant face recognition. **Audio and Video-Based Biometric Person Authentication**, p. 10–18, 2003. Disponível em: <http://link.springer.com/chapter/10.1007/3-540-44887-X_2>. Citado na página 40.

HAN, H.; WANG, W.-Y.; MAO, B.-H. Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning. **Advances in intelligent computing**, Alvey Vision Club, v. 17, n. 12, p. 878–887, 2005. Disponível em: <http://link.springer.com/chapter/10.1007/11538059_91>. Citado na página 51.

HARALICK, R. M.; SHANMUGAM, K.; DINSTEIN, I. Textural Features for Image Classification. **IEEE Transactions on Systems, Man, and Cybernetics**, v. 3, n. 6, p. 610–621, nov. 1973. ISSN 0018-9472. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4309314>>. Citado na página 44.

HARRIS, C.; STEPHENS, M. A combined corner and edge detector. In: **Proceedings of the Alvey Vision Conference**. Alvey Vision Club, 1998. p. 147—152. Disponível em: <<http://www.bmva.org/bmvc/1988/avc-88-023.html>>. Citado na página 43.

HINTON, G.; OSINDERO, S.; TEH, Y. W. A fast learning algorithm for deep belief nets. **Neural computation**, 2006. Disponível em: <http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6796673>. Citado 2 vezes nas páginas 48 e 49.

HUANG, J.; KUMAR, S. R.; MITRA, M.; ZHU, W.-J.; ZABIH, R. Image indexing using color correlograms. In: **Computer Society Conference on Computer Vision and Pattern Recognition**. IEEE, 1997. p. 762–768. ISBN 0-8186-7822-4. ISSN 1063-6919. Disponível em: <<http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=609412>>. Citado na página 43.

JAPKOWICZ, N.; STEPHEN, S. The class imbalance problem : A systematic study. **Intelligent data analysis**, IOS Press, v. 6, n. 5, p. 429–449, 2002. Citado na página 53.

JOIA, P.; PAULOVICH, F. V.; COIMBRA, D.; CUMINATO, J. A.; NONATO, L. G. Local Affine Multidimensional Projection. **IEEE Transactions on Visualization and Computer Graphics**, v. 17, n. 12, p. 2563–2571, 2011. Citado na página 31.

KANAN, C.; COTTRELL, G. W. Color-to-grayscale: does the method matter in image recognition? **PloS one**, v. 7, n. 1, p. e29740, jan. 2012. ISSN 1932-6203. Disponível em: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3325461/>>. Citado 2 vezes nas páginas 32 e 42.

LECUN, Y.; BOTTOU, L. Gradient-based learning applied to document recognition. In: **Proceedings of the IEEE**. [s.n.], 1998. v. 86, p. 2278—2324. Disponível em: <http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=726791>. Citado na página 46.

- LECUN, Y.; KAVUKCUOGLU, K.; FARABET, C. Convolutional networks and applications in vision. In: **International Symposium on Circuits and Systems**. IEEE, 2010. p. 253–256. ISBN 978-1-4244-5308-5. Disponível em: <<http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=5537907>>. Citado na página 47.
- LOWE, D. G. Distinctive image features from scale-invariant keypoints. **International Journal of Computer Vision**, v. 60, n. 2, p. 91–110, 2004. Disponível em: <<http://link.springer.com/article/10.1023/B:VISI.0000029664.99615.94>>. Citado na página 43.
- MAHENDRAN, A.; VEDALDI, A. Understanding deep image representations by inverting them. ArXiv preprint arXiv:1412.0035. 2014. Disponível em: <<http://arxiv.org/abs/1412.0035>>. Citado 2 vezes nas páginas 46 e 48.
- NIELSEN, M. A. **Neural Networks and Deep Learning**. [S.l.]: Determination Press, 2015. Citado na página 46.
- PAIVA, J. G. S.; FLORIAN, L.; PEDRINI, H.; TELLES, G. P.; MINGHIM, R. Improved similarity trees and their application to visual data classification. **IEEE Transactions on Visualization and Computer Graphics**, IEEE, v. 17, n. 12, p. 2459–2468, 2011. Citado 2 vezes nas páginas 31 e 32.
- PASS, G.; ZABIH, R.; MILLER, J. Comparing images using color coherence vectors. In: **Proceedings of the fourth ACM international conference on Multimedia**. New York, USA: ACM Press, 1996. p. 65–73. ISBN 0897918711. Disponível em: <<http://dl.acm.org/citation.cfm?id=244130.244148>>. Citado na página 44.
- PICON, C. T.; ROSSI, I.; JR, M. P. P. Análise da classificação de imagens por descritores de cor utilizando várias resoluções. In: **Workshop of Undergraduate Works**. SIBGRAPI, 2011. Disponível em: <http://www.icmc.usp.br/~moacir/papers/Picon_WUW2011.pdf>. Citado na página 31.
- PONTI, M.; ESCOBAR, L. Compact color features with bitwise quantization and reduced resolution for mobile processing. In: **Global Conference on Signal and Information Processing**. Austin, TX, Estados Unidos: [s.n.], 2013. p. 751–754. Disponível em: <<http://link.springer.com/article/10.1007/s11760-011-0216-x>>. Citado na página 42.
- PONTI, M.; MASCARENHAS, N.; FERREIRA, P.; SUAZO, C. Three-dimensional noisy image restoration using filtered extrapolation and deconvolution. **Signal, Image and Video Processing**, v. 7, n. 1, p. 1–10, fev. 2011. ISSN 1863-1703. Disponível em: <<http://link.springer.com/10.1007/s11760-011-0216-x>> http://link.springer.com/article/10.1007/s11760-011-0216-x. Citado na página 39.
- PONTI, M.; NAZARÉ, T.; THUMÉ, G. Image quantization as a dimensionality reduction procedure in color and texture feature extraction. Submitted to Neurocomputing. 2014. Citado 2 vezes nas páginas 32 e 42.
- PONTI, M. P. Microscope volume segmentation improved through non-linear restoration. **Nature-Inspired Computing Design, Development, and Applications**, IGI Global, 2012. Disponível em: <<http://www.igi-global.com/article/microscope-segmentation-improved-through-non/52614>>. Citado 2 vezes nas páginas 36 e 38.

_____. Segmentation of Low-Cost Remote Sensing Images Combining Vegetation Indices and Mean Shift. **Geoscience and Remote Sensing Letters**, IEEE, v. 10, n. 1, p. 67–70, jan. 2013. ISSN 1545-598X. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6202674>>. Citado na página 31.

ROCHA, A.; HAUAGGE, D. C.; WAINER, J.; GOLDENSTEIN, S. Automatic fruit and vegetable classification from images. **Computers and Electronics in Agriculture**, v. 70, n. 1, p. 96–104, jan. 2010. ISSN 01681699. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S016816990900180X>>. Citado na página 32.

SCHMIDHUBER, J. Deep learning in neural networks: An overview. **Neural Networks**, p. 88, out. 2014. ISSN 08936080. Disponível em: <<http://arxiv.org/abs/1404.7828>>. Citado 2 vezes nas páginas 30 e 46.

SIMONYAN, K.; VEDALDI, A.; ZISSERMAN, A. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. In: **ICLR Workshop**. [s.n.], 2013. p. 1–8. Disponível em: <<http://arxiv.org/abs/1312.6034>>. Citado na página 46.

SMITH, S. M.; BRADY, J. M. SUSAN—a new approach to low level image processing. **International journal of computer vision**, v. 23, n. 1, p. 45–78, 1997. Disponível em: <<http://link.springer.com/article/10.1023/A:1007963824710>>. Citado na página 43.

STEHLING, R. O.; NASCIMENTO, M. A.; aO, A. X. F. A compact and efficient image retrieval approach based on border/interior pixel classification. In: **Proceedings of the eleventh international conference on Information and knowledge management**. New York, USA: ACM Press, 2002. p. 102–109. ISBN 1581134924. Disponível em: <<http://dl.acm.org/citation.cfm?id=584792.584812>>. Citado 2 vezes nas páginas 43 e 71.

TOMASI, C.; MANDUCHI, R. Bilateral filtering for gray and color images. In: **Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)**. Narosa Publishing House, 1998. p. 839–846. ISBN 81-7319-221-9. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=710815>>. Citado 2 vezes nas páginas 58 e 59.

WANG, X.; HAN, T. X.; YAN, S. An HOG-LBP human detector with partial occlusion handling. **IEEE 12th International Conference on Computer Vision**, IEEE, 2009. Citado na página 43.

ZEILER, M. D.; FERGUS, R. Visualizing and Understanding Convolutional Networks. **Computer Vision–ECCV 2014**, Springer, p. 818–833, nov. 2013. Disponível em: <<http://arxiv.org/abs/1311.2901>>. Citado 2 vezes nas páginas 46 e 48.

ZEILER, M. D.; TAYLOR, G. W.; FERGUS, R. Adaptive deconvolutional networks for mid and high level feature learning. In: **International Conference on Computer Vision**. IEEE, 2011. p. 2018–2025. ISBN 978-1-4577-1102-2. ISSN 1550-5499. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6126474>>. Citado 2 vezes nas páginas 46 e 48.