



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE TELEINFORMÁTICA
MESTRADO ACADÊMICO EM ENGENHARIA DE TELEINFORMÁTICA

JOSÉ GERARDO FONTELES LOPES

REDES NEURAS CONVOLUCIONAIS E ENERGIA DE DOBRAMENTO
MULTIESCALA NA DESCRIÇÃO E ANÁLISE DE FORMAS

FORTALEZA

2019

JOSÉ GERARDO FONTELES LOPES

REDES NEURAS CONVOLUCIONAIS E ENERGIA DE DOBRAMENTO MULTIESCALA
NA DESCRIÇÃO E ANÁLISE DE FORMAS

Dissertação apresentada ao Curso de Mestrado Acadêmico em Engenharia de Teleinformática do Programa de Pós-Graduação em Engenharia de Teleinformática do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Engenharia de Teleinformática. Área de Concentração: Sinais e Sistemas

Orientadora: Prof^a. Dr^a. Fátima Nelsi-zeuma Sombra de Medeiros

FORTALEZA

2019

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- L853r Lopes, José Gerardo Fonteles.
 Redes Neurais Convolucionais e Energia de Dobramento Multiescala na Descrição e Análise de Formas / José Gerardo Fonteles Lopes. – 2019.
 46 f. : il. color.
- Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia de Teleinformática, Fortaleza, 2019.
 Orientação: Profa. Dra. Fátima Nelsizeuma Sombra de Medeiros.
1. Análise de formas. Redes Neurais Convolucionais. Energia de Dobramento Multiescala Normalizada. Descritor de Formas. Recuperação de Imagens pelo Conteúdo. I. Título.
- CDD 621.38
-

JOSÉ GERARDO FONTELES LOPES

REDES NEURAS CONVOLUCIONAIS E ENERGIA DE DOBRAMENTO MULTIESCALA
NA DESCRIÇÃO E ANÁLISE DE FORMAS

Dissertação apresentada ao Curso de Mestrado Acadêmico em Engenharia de Teleinformática do Programa de Pós-Graduação em Engenharia de Teleinformática do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Engenharia de Teleinformática. Área de Concentração: Sinais e Sistemas

Aprovada em: 27 de Junho de 2019

BANCA EXAMINADORA

Prof^a. Dr^a. Fátima Nelsizeuma Sombra de
Medeiros (Orientadora)
Universidade Federal do Ceará (UFC)

Prof. Dr. José Marques Soares
Universidade Federal do Ceará (UFC)

Prof. Dr. Antônio Oseas de Carvalho Filho
Universidade Federal do Piauí (UFPI)

À minha mãe Raquel Natalina Fonteles

AGRADECIMENTOS

Agradeço primeiramente a Deus, por ser meu recanto e por me proporcionar tantas alegrias ao longo dos anos.

À minha mãe e meu irmão por sempre me apoiarem em todos os desafios que enfrentei em minha vida.

À minha orientadora Profa. Fátima Sombra pela paciência na orientação e incentivo que tornaram possível a conclusão deste trabalho.

Aos colegas do Grupo de Processamento de Imagens pela amizade adquirida após todo esse tempo.

E por fim, aos amigos e colegas, pelo incentivo e pelo apoio constante.

Por fim, meu agradecimento a todos os cidadãos brasileiros, que por meio da FUN-CAP, apoiaram financeiramente este trabalho.

“Ideas and only ideas can light the darkness.”

(Ludwig Von Mises)

RESUMO

O reconhecimento automático de pílulas por meio de imagens digitais pode reduzir erros de administração de medicação por pacientes idosos e também possibilitar a polícia forense no rastreamento de rotas comerciais de drogas ilegais. Considerando o crescente uso de redes neurais convolucionais (CNNs) em problemas de reconhecimento de padrões, neste trabalho é proposta a aplicação das redes convolucionais *LeNet* e *Inception ResNet*, assim como do descritor multiescala de formas Energia de Dobramento Multiescala (NMBE) na extração de atributos de imagens de bases públicas de Pílulas Lícitas e Ilícitas para fins de reconhecimento automático das mesmas. Realizamos experimentos para avaliar a generalização dos algoritmos de descrição de objetos em uma base de imagens do alfabeto internacional da linguagem de sinais (*Hands*). Avaliamos quantitativamente os resultados de classificação e recuperação de imagem baseada em conteúdo (CBIR) por meio das medidas de Acurácia e *mean average precision* (MAP). Na análise qualitativa utilizamos a Matriz U para a visualização do arranjo dos agrupamentos de formas. Os experimentos de classificação e CBIR mostraram que a *LeNet* superou a *Inception Resnet* e a NMBE com resultados acima de 90% nas bases públicas que possuem quantidades de imagens suficientes para o seu treinamento, como as bases de Pílulas Lícitas e *Hands*.

Palavras-chave: Análise de formas. Redes Neurais Convolucionais. Energia de Dobramento Multiescala Normalizada. Descritor de Formas. Recuperação de Imagens pelo Conteúdo.

ABSTRACT

The automatic recognition of pills through image can reduce wrong drug administration in elderly patients and also enable forensic intelligence to establish links among illegal drug marketing. Considering the remarkable progress and performance of the convolutional neural networks (CNNs) in pattern recognition problems, this study proposes the application of neural network, namely *LeNet* and *Inception ResNet* as well as the normalize normalize multiscale bending energy (NMBE) for feature extraction of licit and illicit pills in order to develop algorithms for automatic recognition of pills. We assess the generalization of the algorithms in an image database from the international alphabet of sign language (*Hands*). We conduct classification and content-based image retrieval (CBIR) experiments and evaluate the results quantitatively using the Accuracy and mean average precision (MAP) measures. We perform the qualitative analysis through the U matrix for visualization of the cluster arrangement. The results showed that *LeNet* outperformed the *Inception Resnet* and NMBE for image databases with a great amount of images for training as the Licit Pills and *Hands* databases.

Keywords: Shape Analysis. Convolutional Neural Networks. Normalized Multiscale Bend Energy. Shape Descriptor. Content-Based Image Retrieval.

LISTA DE FIGURAS

Figura 1 – Etapas de processamento de uma imagem digital com extração de características e reconhecimento de objeto.	16
Figura 2 – Suavização de uma forma.	18
Figura 3 – Ilustração da arquitetura de uma <i>LeNet</i> para extração de características das imagens de entradas e suas camadas convolucionais, de <i>pooling</i> e totalmente conectadas.	22
Figura 4 – Arquitetura do módulo <i>Inception</i>	23
Figura 5 – Modelo de recuperação de imagem baseada em conteúdo (CBIR).	25
Figura 6 – Rede SOM.	26
Figura 7 – Exemplo de uma Matriz U.	28
Figura 8 – Metodologia proposta para descrição e análise de formas usando CNNs e NMBE.	29
Figura 9 – Amostras de imagens e suas respectivas máscaras de segmentação da base pública de Pílulas Lícitas.	30
Figura 10 – Amostras de imagens e suas respectivas máscaras de segmentação da base pública de Pílulas Ilícitas.	31
Figura 11 – Amostras de imagens de cada classe da base <i>Hands</i>	31
Figura 12 – Resultados de recuperação com a <i>LeNet</i> para amostras de Pílulas Lícitas e a <i>Inception Resnet</i> para as amostras da base de Pílulas Ilícitas. As bordas verdes indicam imagens corretamente retornadas e as vermelhas incorretamente retornadas.	35
Figura 13 – Matriz U para Pílulas Lícitas descritas pela rede (a) <i>LeNet</i> e (b) <i>Inception Resnet</i>	36
Figura 14 – Matriz U para Pílulas Ilícitas descritas pela (a) <i>LeNet</i> e (b) <i>Inception Resnet</i>	37
Figura 15 – Matrizes U para Pílulas Lícitas.	40
Figura 16 – Matrizes U para Pílulas Ilícitas.	40
Figura 17 – Matrizes U para a base <i>Hands</i>	41

LISTA DE TABELAS

Tabela 1 – Escalas do descritor NMBE.	21
Tabela 2 – Porcentagem de amostras de Pílulas Lícitas.	30
Tabela 3 – Porcentagem de amostras de base Pílulas Ilícitas	31
Tabela 4 – Acurácia da classificação da base de Pílulas Lícitas.	33
Tabela 5 – Acurácia da classificação da base de Pílulas Ilícitas.	33
Tabela 6 – Valores da Medida MAP para recuperação de imagens de base de Pílulas Lícitas.	34
Tabela 7 – Valores da Medida MAP para recuperação de imagens de base de Pílulas Ilícitas.	34
Tabela 8 – Resultados da Acurácia.	38
Tabela 9 – Resultados para CBIR (MAP).	39

SUMÁRIO

1	INTRODUÇÃO	12
1.1	Objetivos	13
1.2	Produção Científica	14
1.3	Organização da Dissertação	15
2	FUNDAMENTAÇÃO TEÓRICA	16
2.1	Energia de Dobramento Multiescala Normalizada (NMBE)	17
2.2	Redes Neurais Convolucionais para Extração de Atributos	21
2.3	Algoritmo de Classificação knn	23
2.3.1	<i>Avaliação da Classificação: Acurácia</i>	24
2.4	Sistemas de Recuperação de Imagens Baseada em Conteúdo	24
2.4.1	<i>Avaliação de Sistemas de Recuperação: Mean Average Precision (MAP)</i>	25
2.5	Self Organized Map (Rede SOM)	26
3	METODOLOGIA	29
3.1	Base de Imagens	30
3.1.1	<i>Pílulas Lícitas</i>	30
3.1.2	<i>Pílulas Ilícitas</i>	31
3.1.3	<i>Hands</i>	31
4	RESULTADOS	32
4.1	Experimentos com Imagens sem Pré-processamento	32
4.1.1	<i>Resultados dos Experimentos de Classificação</i>	33
4.1.2	<i>Experimentos de Recuperação</i>	34
4.2	Experimentos com as Imagens com Pré-processamento	37
4.2.1	<i>Experimentos de Classificação com Pré-processamento</i>	38
4.2.2	<i>Experimentos de Recuperação com Pré-processamento</i>	38
5	CONCLUSÕES E TRABALHOS FUTUROS	42
5.1	Trabalhos Futuros	43
	REFERÊNCIAS	44

1 INTRODUÇÃO

Com o crescente acesso da população a medicamentos, aumentam as chances de ocorrência de erros em sua administração. Esse problema deve-se, principalmente, à dificuldade de identificação visual dessas pílulas por pacientes e profissionais de saúde. Nove em cada dez pessoas idosas, com idade acima dos 65 anos nos Estados Unidos, ingerem mais de um medicamento em formato de pílula (USHIZIMA *et al.*, 2015). Dessa forma, a identificação correta desses medicamentos é uma questão também de saúde.

Existem vários serviços de indexação de pílulas públicos ¹, e também privados ² que fornecem uma ferramenta de busca simples (LOPATKA; HOUTEN, 2013). Porém, nenhum desses serviços possui as ferramentas necessárias para executar a recuperação de imagens baseadas em conteúdo.

A dificuldade de identificação de pílulas também pode surgir durante a apreensão de pílulas ilegais por autoridades competentes em que são demandadas tarefas de indexação e catalogação dessas pílulas. Os serviços de catalogação e indexação de pílulas são realizados pelos órgãos de repressão ao tráfico e utilizam, em geral, testes laboratoriais de química forense. Além disso, estimulantes do tipo anfetaminas representam um problema de saúde pública de acordo com Relatório Mundial de Drogas em 2011 publicado pelo Departamento das Nações Unidas em Drogas e Crime (*United Nations Office on Drugs and Crime*) (UNODC, 2011) que afeta entre 0,3% e 1,3% da população mundial, com idades variando entre 15 e 64 anos, são afetados por esse tipo de problema (CAMARGO *et al.*, 2012).

O reconhecimento automático de pílulas lícitas para auxílio ao paciente e de ilícitas para auxílio a polícia forense requer o desenvolvimento de algoritmos de reconhecimento de padrões que têm como base a visão computacional (LEE *et al.*, 2012; WONG *et al.*, 2017). O desenvolvimento desses sistemas de reconhecimento é baseado, principalmente, em marcação impressa nas pílulas, descritores de cor, textura de formas (LEE *et al.*, 2012; CAMARGO *et al.*, 2012; LOPATKA; HOUTEN, 2013; MADDALA *et al.*, 2017), e também com auxílio de redes neurais (DELGADO *et al.*, 2019).

Além dos problemas descritos anteriormente existem estudos baseados em visão computacional que têm por objetivo o desenvolvimento de sistemas para assistir pessoas com alguma limitação auditiva (MITRA; ACHARYA, 2007; ISLAM *et al.*, 2017; SOMBANDITH *et*

¹ (pillreport.com, ecstasydata.org, erowid.org, PharmInfoNetpillreport.com, ecstasydata.org, erowid.org, PharmInfoNet)

² (DPD, Epocrates, Ident-A-Drug)

al., 2017). A linguagem de sinais é a principal forma de comunicação utilizada por pessoas que possuem algum grau de surdez e, o desenvolvimento de ferramentas assistidas por computador para a interpretação dessa linguagem é baseada em reconhecimento de gestos (ISLAM *et al.*, 2017).

Neste trabalho foram estudadas técnicas de descrição de formas aplicadas a imagens de pílulas e de linguagem de sinais com ênfase em um descritor multiescala de forma, além de uma abordagem utilizando duas redes neurais convolucionais em uma base pública de Pílulas Lícitas, (YANIV *et al.*, 2016) e suas máscaras segmentadas disponível em (USHIZIMA *et al.*, 2015), uma base de Pílulas Ilícitas disponível em <<https://www.ecstasydata.org/>> e suas máscaras segmentadas encontradas em Carneiro *et al.* (2019). Além destas duas base, também foram realizados testes com uma base de linguagem de sinais (*Hands*) (BIRK *et al.*, 1997). O descritor de formas utilizado foi a Energia de Dobramento Multiescala (NMBE), do inglês *Normalized Multiscale Bending Energy* (COSTA; CESAR JR, 2009) e a razão dessa escolha se deve à sua invariância a escala, rotação, translação e robustez a ruído. As duas redes neurais convolucionais selecionadas são: a *LeNet* (LECUN *et al.*, 1998) e a *Inception Resnet* (SZEGEDY *et al.*, 2016).

O treinamento da *LeNet* foi realizado com uma parte das imagens da base, pois essa é uma rede com menos camadas e portanto, requer menos imagens para o seu treinamento. Quanto à *Inception ResNet*, utilizamos a rede pré-treinada dado que essa rede possui uma arquitetura com diversas camadas e requer milhares de imagens para seu treinamento (SZEGEDY *et al.*, 2016). A base de imagens utilizada para o pré-treinamento foi uma base de imagem de uso geral denominada *ImageNet* (DENG *et al.*, 2009).

A avaliação quantitativa para os três métodos de descrição foi realizada em termos da taxa de acertos do classificador para os experimentos de classificação supervisionada. A medida *mean average precision* (MAP) (HU *et al.*, 2008) foi utilizada para avaliar os resultados obtidos nos experimentos de recuperação de imagens baseada em conteúdo (CBIR), do inglês *Content-Based Image Retrieval* (WANG, 2001).

1.1 Objetivos

Este trabalho tem como objetivo geral a proposição de uma metodologia para descrição e categorização de formas de pílulas lícitas, ilícitas e uma base de linguagem de sinais. Tal metodologia engloba etapas de extração de características, classificação e/ou CBIR. Dentre os

objetivos específicos destacam-se:

- Descrição e análise de imagens digitais de pílulas e gestos da linguagem de sinais por meio de um descritor multiescala de formas e duas CNN's;
- Validação dos descritores selecionados em reconhecimento automático de pílulas e gestos da linguagem de sinais através de experimentos de classificação e CBIR;
- Criação de uma nova base de máscaras de Pílulas Ilícitas segmentadas e disponibilizadas publicamente.

1.2 Produção Científica

Artigos Publicados

1. CARNEIRO, A. C.; **LOPES, J. G. F.**; ARAÚJO, F. H. D.; SILVA, R. R. V.; PASSARINHO, C. J. P.; ROCHA NETO, J. F. S.; MEDEIROS, F. N. S.. *Análise de fotografias de pílulas por redes neurais convolucionais*. VIII Simpósio de Instrumentação e Imagens Médicas e VII Simpósio de Processamento de Sinais, São Bernardo do Campo, 2017.
2. CARNEIRO, A. C.; **LOPES, J. G. F.**; ROCHA NETO, J. F. S.; SOUZA, M. M. S.; MEDEIROS, F. N. S.; BEZERRA, F. N. *On the evaluation of the cost functions for parameter optimization of a multiscale shape descriptor*. In: IEEE. International Symposium on Signal Processing and Information Technology (ISSPIT). [S.l.], Bilbao, 18-20 Dec, p. 045-050, 2017.
3. SILVA, R. R. V.; **LOPES, J. G. F.**; ARAÚJO, F. H. D.; Medeiros, F. N. S.; USHIZIMA, D.. *Visão Computacional em Python Utilizando as Bibliotecas Scikit-image e Scikit-learn*. In: III Escola Regional de Informática do Piauí. (Org.). Livro Anais - Artigos e Minicursos. 1ed.: Sociedade Brasileira de Computação, v. 1, p. 407-428, 2017.
4. CARNEIRO, A. C.; **LOPES, J. G. F.**; SOUZA, M. M. S.; ROCHA NETO, J. F. S.; ARAÚJO, F. H. D.; SILVA, R. R. V. ; MEDEIROS, F. N. S.; BEZERRA, F. N. *Parameter optimisation of a multiscale descriptor for shape analysis on healthcare image datasets*. Pattern Recognition Letters, Elsevier, 2019.
5. Pinheiro, R. G.; **LOPES, J. G. F.**; CARNEIRO, A. C.; SOUZA, M. M. S.; MEDEIROS, F. N. S. *A multiscale shape descriptor based on differential entropy*. 14º Simpósio Brasileiro de Automação Inteligente, Ouro Preto, 2019.

1.3 Organização da Dissertação

A estrutura desta dissertação está organizada da seguinte forma:

- **Capítulo 2:** apresenta a fundamentação teórica, com os principais métodos de descrição de imagens, a saber: Energia de Dobramento Multiescala Normalizada e Redes Neurais Convolucionais. Além disso, mostra o algoritmo de classificação utilizado e o fluxo geral de um sistema CBIR e matriz U;
- **Capítulo 3:** introduz a metodologia utilizada para a descrição e reconhecimento de formas de bases públicas utilizadas nesse trabalho, além disso, apresenta as medidas de avaliação quantitativa e qualitativa de desempenho dos algoritmos.
- **Capítulo 4:** apresenta e discute os resultados obtidos para as bases de imagens de Pílulas Lícitas, Pílulas Ilícitas e *Hands* incluindo a avaliação quantitativa e qualitativa dos mesmos;
- **Capítulo 5:** destaca as principais conclusões alcançadas bem como as limitações das abordagens adotadas, possíveis melhorias e trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são apresentados os conceitos necessários para o entendimento deste trabalho, destacando-se o processo de extração de características das imagens.

As descrições de imagens podem ser realizadas por atributos de forma, cor e textura pois estas propriedades estão presentes em objetos registrados na maioria das imagens. Essas características podem ser quantificadas e usadas como descritores da imagem ou dos objetos numa cena (SINGH; SINGH, 2019). Os descritores compõem um conjunto de propriedades representadas em um vetor de escalares, denominado descritor da imagem, ou assinatura. Assim sendo, cada objeto pode ser representado por um ponto em um espaço definido pelo número de características. Em alguns casos, é desejável que os descritores apresentem a propriedade de invariância a transformações afins, ou seja, invariante a rotação, escala, e translação (HMIMID *et al.*, 2015).

Toda forma pode ser caracterizada por meio de atributos, sejam eles externos ou internos, os quais podem ser oriundos da bordas e pixels que compõem a forma, respectivamente (PEDRINI; SCHWARTZ, 2007). A escolha de um descritor para extrair características de imagens digitais pode ser o passo essencial para o reconhecimento de um objeto numa cena. O segundo passo é descrever o objeto ou região com base na representação escolhida. A Figura 1 representa uma sequência de etapas para extração de características de um objeto em uma imagem digital.



Figura 1 – Etapas de processamento de uma imagem digital com extração de características e reconhecimento de objeto.

A extração de características exibida na Figura 1 mostra que podem ser obtidos diferentes atributos de uma imagem digital na qual se deseja reconhecer um determinado objeto na mesma.

2.1 Energia de Dobramento Multiescala Normalizada (NMBE)

A energia de dobramento expressa a quantidade de energia necessária para transformar um contorno fechado em um círculo com o mesmo perímetro do contorno original (COSTA; CESAR JR, 2009). Essa energia é calculada por meio da curvatura do objeto $k(n)$ obtida do contorno parametrizado $u(n)$, em que n representa a quantidade de pontos do contorno. A curvatura pode ser obtida por meio da expressão:

$$k(n) = \frac{-Im\{\dot{u}(n) \ddot{u}(n)\}}{|\dot{u}(n)|^3}, \quad (2.1)$$

em que Im é a parte imaginária de um número complexo, $\dot{u}(n)$ e $\ddot{u}(n)$ representam a primeira e a segunda derivada do contorno da forma, respectivamente e o operador $|\cdot|$ representa o módulo.

A Equação 2.1 resulta em valores positivos ou negativos dependendo da concavidade local. Além disso, a curvatura é definida pela geometria do contorno. Por exemplo, a curvatura de uma reta é zero, de uma circunferência é constante e a de um objeto irregular repleto de cantos e junções apresenta grande variação. De acordo com a Equação 2.1 o sinal de curvatura em geral é contaminado pelo processo de aquisição da imagem e a quantização dos contornos dos objetos da imagem, o que poderia limitar sua utilização como descritor. Uma solução para isso é a suavização prévia do contorno $u(n)$, por meio de um filtro passa-baixas gaussiano dado por (COSTA; CESAR JR, 2009):

$$g_\sigma(n) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-n^2}{2\sigma^2}\right). \quad (2.2)$$

Portanto, o contorno suavizado é denotado por:

$$\hat{u}(n, \sigma) = u(n) * g_\sigma(n), \quad (2.3)$$

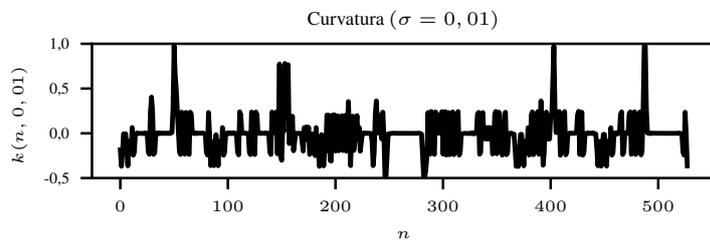
em que o símbolo $*$ representa o operador de convolução. O parâmetro σ da Equação 2.2 permite controlar o nível de detalhamento do contorno da forma. A Figura 2 exibe o processo de suavização do contorno de uma forma e sua curvatura. Vale ressaltar que o contorno perde detalhes ao ser suavizado e a curvatura também fica mais suave. Dessa forma, à medida que o contorno perde detalhes ao ser suavizado, a curvatura também fica mais suave, indicando a diminuição do ruído, mas também a perda de informações em relação à forma inicial.

A primeira e segunda derivadas de $\hat{u}(n, \sigma)$ em relação a n são definidas como:

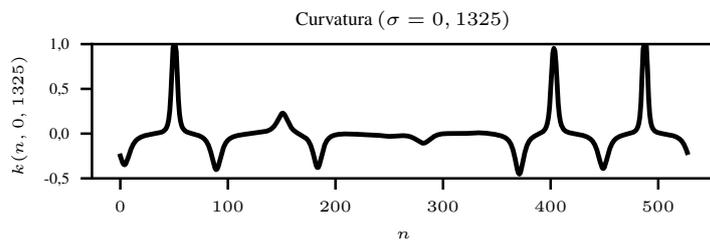
$$\hat{\dot{u}}(n, \sigma) = \dot{u}(n) * g_\sigma(n), \quad (2.4)$$

$$\hat{\ddot{u}}(n, \sigma) = \ddot{u}(n) * g_\sigma(n). \quad (2.5)$$

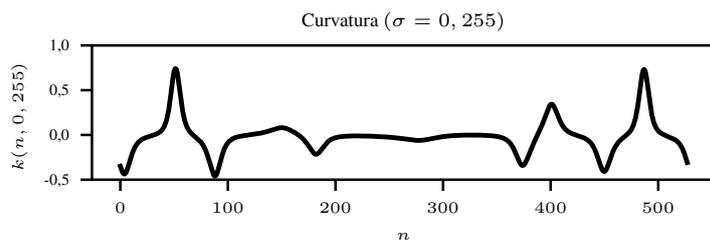
Contorno ($\sigma = 0, 01$)



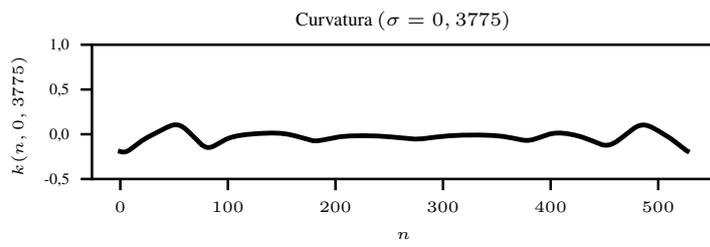
Contorno ($\sigma = 0, 1325$)



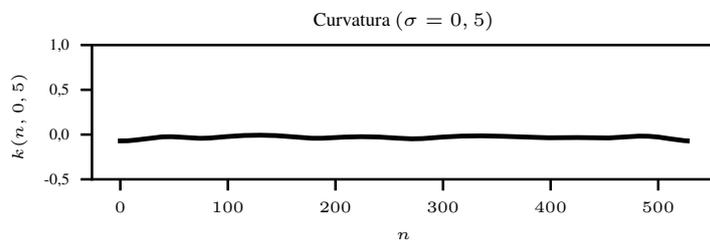
Contorno ($\sigma = 0, 255$)



Contorno ($\sigma = 0, 3775$)



Contorno ($\sigma = 0, 5$)



Fonte adaptado de CARNEIRO (2018)
 Figura 2 – Suavização de uma forma.

A primeira e a segunda derivadas podem ser calculadas no domínio da frequência (CESAR JR; COSTA, 1997) e aplicando a propriedade da derivada da transformada de Fourier segundo:

$$\hat{u}(n, \sigma) = \mathcal{F}^{-1}\{U(f)G_{\frac{1}{\sigma}}(f)\}, \quad (2.6)$$

$$\hat{\dot{u}}(n, \sigma) = \mathcal{F}^{-1}\{j2\pi fU(f)G_{\frac{1}{\sigma}}(f)\}, \quad (2.7)$$

$$\hat{\ddot{u}}(n, \sigma) = \mathcal{F}^{-1}\{-(2\pi f)^2U(f)G_{\frac{1}{\sigma}}(f)\}, \quad (2.8)$$

onde $U(f)$ e $G_{\frac{1}{\sigma}}(f)$ são definidas por:

$$U(f) = \mathcal{F}\{u(n)\}, \quad (2.9)$$

$$G_{\frac{1}{\sigma}}(f) = \mathcal{F}\{g_{\sigma}(n)\} = e^{-2\pi^2 f^2 \sigma^2}. \quad (2.10)$$

$\mathcal{F}\{\cdot\}$ e $\mathcal{F}^{-1}\{\cdot\}$ são as transformadas de Fourier e sua inversa, respectivamente, ou seja, $U(f)$ e $G_{\frac{1}{\sigma}}(f)$ são as transformadas de $u(n)$ e $g_{\sigma}(n)$.

O efeito de suavização ou filtragem gaussiana modifica a amplitude do contorno causando o efeito do encolhimento dos contornos filtrados. Uma técnica para contornar esse problema consiste em empregar uma normalização dos contornos suavizados para que o sinal não seja afetado (COSTA; CESAR JR, 2009). Uma solução possível é a admissão de uma constante de normalização (P_{σ}) que seria a razão entre o perímetro do contorno original (L) e o perímetro do contorno suavizado (L_{σ}), de acordo com a seguinte equação:

$$P_{\sigma} = \frac{L}{L_{\sigma}}, \quad (2.11)$$

onde o cálculo de L e L_{σ} é dado por:

$$L = \frac{2\pi}{N} \sum_{n=0}^{N-1} |\dot{u}(n)|, \quad (2.12)$$

$$L_{\sigma} = \frac{2\pi}{N} \sum_{n=0}^{N-1} |\hat{\dot{u}}(n, \sigma)|. \quad (2.13)$$

O contorno e suas derivadas normalizadas são dados, respectivamente, por:

$$u(n, \sigma) = \hat{u}(n, \sigma)P_{\sigma}, \quad (2.14)$$

$$\dot{u}(n, \sigma) = \hat{\dot{u}}(n, \sigma)P_{\sigma}, \quad (2.15)$$

$$\ddot{u}(n, \sigma) = \hat{\ddot{u}}(n, \sigma)P_{\sigma}. \quad (2.16)$$

Com as Equações 2.14, 2.15 e 2.16 a curvatura multiescala pode ser definida como:

$$k(n, \sigma) = \frac{-\text{Im}\{\dot{u}(n, \sigma) \ddot{u}^*(n, \sigma)\}}{|\dot{u}(n, \sigma)|^3}, \quad (2.17)$$

em que cada valor σ corresponde a uma escala que permite a extração de diferentes detalhes da forma.

A energia de dobramento discreta, representada por E é definida por:

$$E = \frac{1}{N} \sum_{n=0}^{N-1} k(n)^2. \quad (2.18)$$

em que N é o número de pontos do contorno.

Para que a energia de dobramento seja invariante à escala, faz-se necessário sua normalização acrescentando um termo de normalização de L^2 à equação 2.18, então a versão da energia de dobramento normalizada é definida como:

$$E_n = \frac{L^2}{N} \sum_{n=0}^{N-1} k(n)^2. \quad (2.19)$$

Utilizando a definição de curvatura multiescala na Equação 2.17, a versão multiescala da energia de dobramento é obtida, ou seja, a energia de dobramento multiescala normalizada, do inglês *Normalized Multiscale Bending Energy* (NMBE) (COSTA; CESAR JR, 2009). A NMBE é definida por:

$$E_n(\sigma) = \frac{L^2}{N} \sum_{n=0}^{N-1} k(n, \sigma)^2, \quad (2.20)$$

onde $E_n(\sigma)$ é a energia de dobramento normalizada na escala σ .

Como a unidade de $E_n(\sigma)$ na Equação 2.20 é energia e este descritor apresenta grande variação ao longo das escalas, é de Souza (2016) aplicou a operação $\log(\cdot)$ nos diferentes valores de escalas, redefinindo assim o descritor NMBE segundo a expressão:

$$NMBE = [\log E_n(\sigma_1), \log E_n(\sigma_2), \dots, \log E_n(\sigma_m)], \quad (2.21)$$

em que m indica o número de escalas e σ_i é o i -ésimo parâmetro de escala.

Em (CESAR JR; COSTA, 1997) foi proposto o ajuste de escalas para o descritor NMBE de acordo com as seguintes equações:

$$\sigma_l = \left[\left(\frac{\sigma_{min}^{-1} - \sigma_{max}^{-1}}{oct_{max} - \sqrt{2}} \right) (oct_l - \sqrt{2}) + \sigma_{max}^{-1} \right], \quad (2.22)$$

sendo que

$$oct_l = (\sqrt{2})^l; l = [1, 2, \dots, m]. \quad (2.23)$$

Nas equações 2.22 e 2.23 o índice l indica a escala, e o número de escalas é definido como N_s . O termo oct indica o incremento em oitavas, enquanto σ_{min} e σ_{max} correspondem aos valores mínimos e máximos que um dado parâmetro de escala σ pode assumir. Por fim, o termo oct_{max} é o valor máximo que oct pode assumir. As escalas obtidas em (CESAR JR; COSTA, 1997) são escalas que adotam os seguintes parâmetros: $m = 32$, $\sigma_{min} = \frac{1}{64}$ e $\sigma_{max} = 2$, $oct_{min} = \sqrt{2}$ e $oct_{max} = (\sqrt{2})^{32}$. As escalas utilizadas nesse trabalho estão na Tabela 1:

Tabela 1 – Escalas do descritor NMBE.

Parâmetros	σ_1	σ_2	σ_3	σ_4	σ_5
Valor	0,4	0,654	1,1905	2,826	100

2.2 Redes Neurais Convolucionais para Extração de Atributos

As redes neurais convolucionais (CNNs) têm se apresentado como um novo padrão em pesquisas na área de visão computacional pois estas podem ser treinadas em grande volume de imagens ou dados, quando existem rótulos para as mesmas (ARAUJO, 2018). Métodos baseados em CNNs estão sendo utilizados em uma variedade de problemas, tais como: carros autônomos, diagnóstico médico por imagens, processamento de linguagem natural, saúde preventiva, entre outras áreas (FERREIRA *et al.*, 2018). Essas redes neurais provaram ser eficientes em diversas tarefas de reconhecimento de imagens e vídeos, sistemas de recomendação e processamento de linguagem natural (ARAUJO, 2018), dentre outros. As CNNs processam dados de uma imagem por meio de filtros que são aplicados na imagem original, preservando a relação entre os pixels vizinhos da imagem original por meio de operações de convoluções (GOODFELLOW *et al.*, 2016).

A *LeNet* (LECUN *et al.*, 1998), foi um dos primeiros projetos de CNNs, tendo a mesma auxiliado a impulsionar o campo de *Deep Learning*. Inicialmente, a *LeNet* foi utilizada para reconhecimento de caracteres, tais como código postal e dígitos numéricos. Novas arquiteturas foram propostas nos últimos anos como forma de melhoria da *LeNet*, embora as versões de CNNs melhoradas compartilhem os mesmos conceitos fundamentais. O valor da taxa de aprendizagem é de 0,01, foram utilizadas 50 épocas, *batch size* de 16, dropout de 0,5.

As CNNs são formadas por seqüências de camadas e cada uma destas possui uma função específica na propagação do sinal de entrada. A Figura 3 ilustra a arquitetura de uma *LeNet* e suas três principais camadas: convolucionais, de *pooling* e totalmente conectadas. As camadas convolucionais são responsáveis por extrair atributos dos volumes de entradas. As camadas de *pooling* são responsáveis por reduzir a dimensionalidade do volume resultante após as camadas convolucionais e ajudam a tornar a representação invariante a pequenas translações na entrada. As camadas totalmente conectadas são responsáveis pela propagação do sinal por meio da multiplicação ponto a ponto e o uso de uma função de ativação (LECUN *et al.*, 1998).

Após a etapa de treino, as características que representam as imagens de entrada são obtidas removendo a última camada totalmente conectada da rede, que é a camada de classificação.

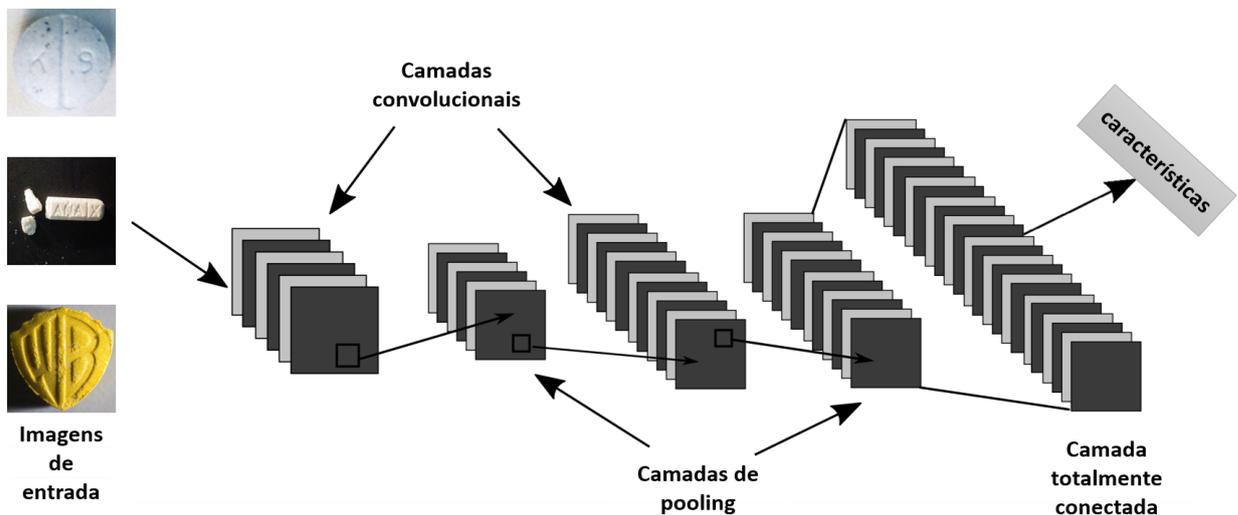
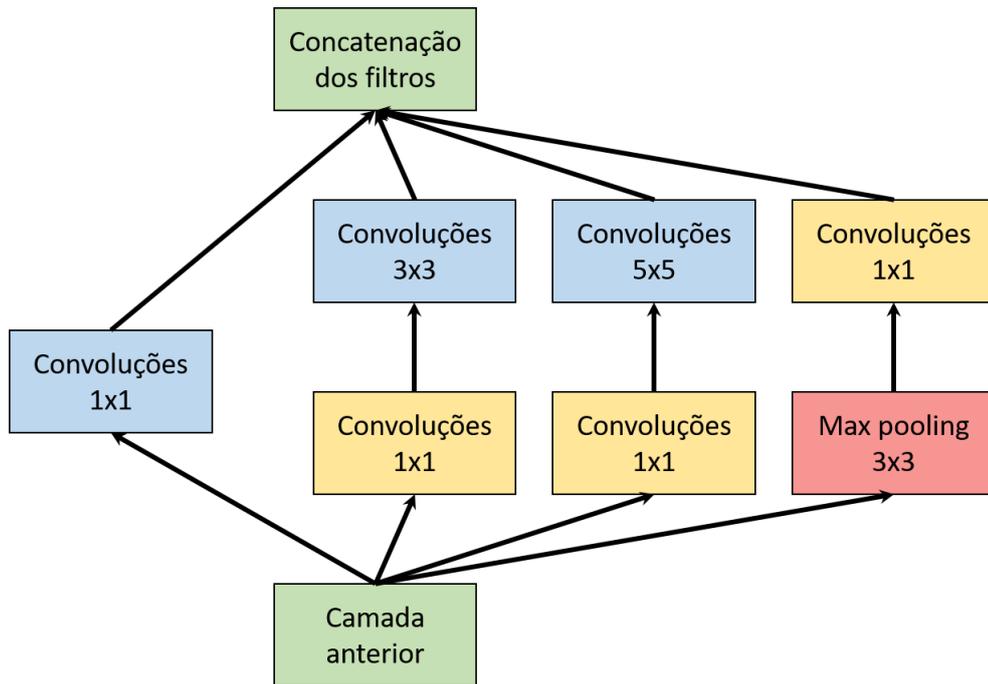


Figura 3 – Ilustração da arquitetura de uma *LeNet* para extração de características das imagens de entradas e suas camadas convolucionais, de *pooling* e totalmente conectadas.

As CNNs se tornaram extremamente úteis em categorizar conteúdos de imagens e de *frames* de vídeos. Dessa forma, diversas empresas como a *Google* estavam interessadas em melhorar a eficiência das arquiteturas existentes. Daí a rede *GoogLeNet* foi proposta por (SZEGEDY *et al.*, 2015) tendo alguns blocos de suas camadas executados em paralelo. Essas camadas ficaram conhecidas como *Inception* e consistiam de combinações paralelas de camadas com filtros convolucionais de tamanho 1x1, 3x3 e 5x5. A Figura 4 representa o módulo *Inception*. A vantagem desse módulo é a redução do número de características com o uso da convolução com filtros 1x1 para reduzir o número de características no bloco paralelo antes de realizar as convoluções com os filtros maiores.

A arquitetura da *Inception-ResNet-v2* é também formada pelos módulos *Inception* (SZEGEDY *et al.*, 2016). Sua arquitetura é formada por múltiplas sub-redes de várias camadas, obtendo padrões de classificação mais elaborados (SILVA, 2018). Dentre as diversas opções de CNNs, exploramos nesta dissertação duas diferentes arquiteturas, a saber: *LeNet* e *Inception-ResNet-v2*. Nesse trabalho somente a última camada é descartada e os valores resultantes após a camada totalmente conectada são utilizados como características que descrevem a imagem de entrada e a *Inception-ResNet-v2* será endereçada como *Inception Resnet*.



Fonte adaptado de (DESHPANDE, 2016)
 Figura 4 – Arquitetura do módulo *Inception*.

2.3 Algoritmo de Classificação knn

O algoritmo k vizinhos mais próximos, do inglês *k-nearest neighbors* knn (ARTHUR; VASSILVITSKII, 2007), classifica uma amostra de teste z tendo como base a classe dos vizinhos mais próximos do conjunto de treinamento C_T . Sendo C_T dado por:

$$C_T = ((x_1, y_1), (x_2, y_2), \dots, (x_Q, y_Q)), \quad (2.24)$$

onde x_i a i -ésima amostra de treinamento de dimensão M , y_i é o rótulo da classe de x_i e Q é o número total de amostras do conjunto de treinamento C_T . O algoritmo (knn) atribui a z a classe de maior frequência nas k amostras que pertencem a C_T com a menor distância a z .

Portanto, para se utilizar o knn é necessário que amostras sejam rotuladas formando um conjunto de treinamento, que se escolha uma função de distância e se definam o número de vizinhos k (MURTY; DEVI, 2011). Nesse trabalho os valores de k usados foram $k = 3$ e $k = 5$, e a função de distância foi a euclidiana.

2.3.1 Avaliação da Classificação: Acurácia

A medida de Acurácia de um classificador é definida pela razão entre a quantidade de elementos de um determinada base de dados que tiveram a classe corretamente predita e o número total dos elementos. Ela pode ser calculada pela seguinte equação:

$$Acurcia = \frac{VP + VN}{N_t} \quad (2.25)$$

em que VP representa o verdadeiro positivo, VN é o verdadeiro negativo e N_t é o número total de amostras. O valor da Acurácia está no intervalo de 0 a 100%, em que o último representa 100% da taxa de acerto.

2.4 Sistemas de Recuperação de Imagens Baseada em Conteúdo

Sistemas de recuperação de imagens baseada em conteúdo, do inglês *Content-Based Image Retrieval* (CBIR) realizam buscas utilizando propriedades pictoriais de imagens. Tais sistemas visam recuperar imagens que sejam do interesse do usuário de acordo com um padrão de consulta específico por meio de uma amostra (WANG, 2001).

A Figura 5 exibe um experimento CBIR clássico em que a partir de uma base de imagens são extraídas as características das imagens por meio de descritores, obtendo todos os vetores de assinaturas dessa base. A recuperação de imagens é feita utilizando uma imagem de consulta, e para essa imagem se extraem atributos que a caracterizam. A partir do vetor de atributos é realizada uma busca na base de assinaturas que irá retornar aquelas mais similares à que foi calculada para a imagem de consulta. O cálculo de similaridade se dá através da utilização de funções de distância, como a euclidiana. Por fim, o sistema terá como saída as imagens da base de dados ordenadas na ordem decrescente de similaridade dos atributos extraídos (SILVA, 2018).

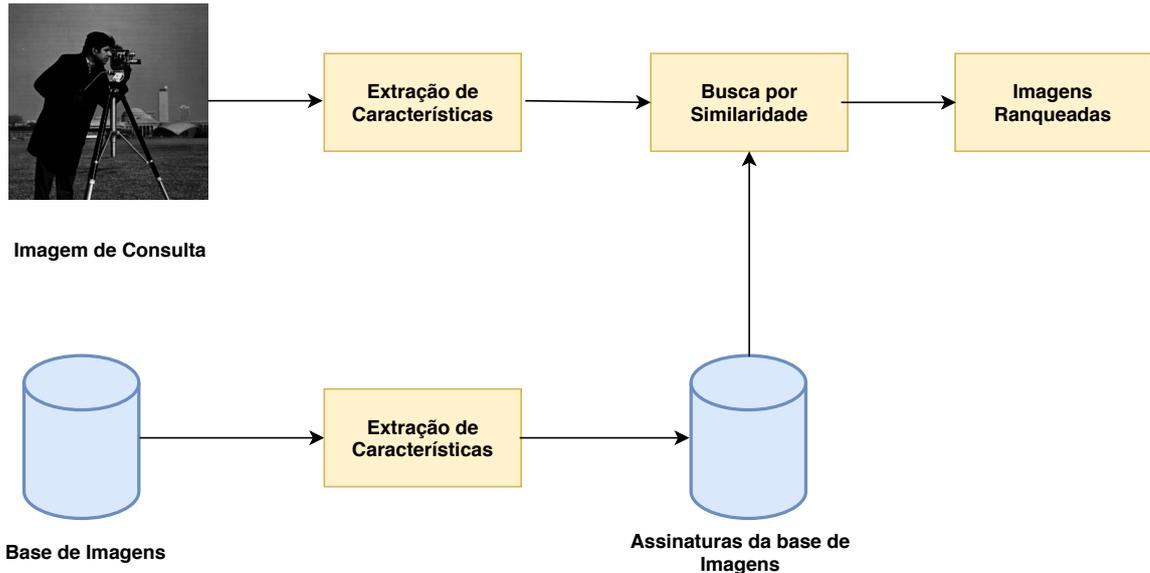


Figura 5 – Modelo de recuperação de imagem baseada em conteúdo (CBIR).

2.4.1 Avaliação de Sistemas de Recuperação: Mean Average Precision (MAP)

Nesta dissertação se avaliam os resultados dos experimentos de recuperação através da medida *mean average precision* (WANG *et al.*, 2015). Para se calcular o MAP, a precisão média $AP(Q)$ é obtida para cada imagem de consulta Q segundo a expressão:

$$AP(Q) = \frac{\sum_{n=1}^M (P(n) \times f(n))}{N}, \quad (2.26)$$

e,

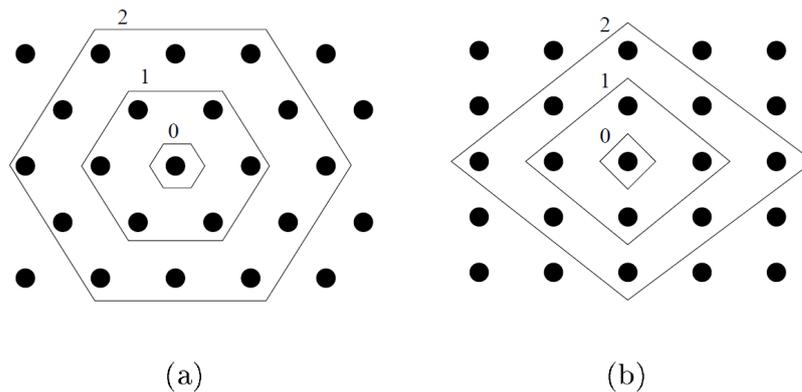
$$P(n) = \frac{VP}{VP + FN}, \quad (2.27)$$

em que o símbolo \times representa o produto escalar, $P(n)$ é a precisão até a posição n do ranqueamento, $f(n)$ é igual a 1 se a imagem n do ranqueamento pertence à mesma classe da imagem de referência, e 0 caso contrário. M corresponde ao número de imagens no ranqueamento e N é o número de imagens da mesma classe obtidas no ranqueamento. VP representa o verdadeiro positivo e FN o falso negativo. O valor de MAP é obtido pela média da AP em todas as imagens do ranqueamento. Quanto mais próximo de 1 o valor do MAP melhor o desempenho do sistema de recuperação. É possível calcular o MAP para a base completa bem como para cada classe, e assim observar o desempenho do sistema em diferentes classes de uma mesma base de imagens.

2.5 Self Organized Map (Rede SOM)

O mapa auto-organizável que foi proposto por Kohonen (KOHONEN, 1982), do inglês *Self Organized Map* (SOM), é uma rede neural não supervisionada, onde os seus neurônios são organizados de forma regular e dependem de um estímulo externo.

Esse mapa representa elementos multidimensionais em um mapa de dimensão reduzida, preservando os relacionamentos topológicos desses elementos. Portanto, a rede SOM pode ser usada para projetar os vetores apresentados em sua entrada de um espaço N-dimensional para um espaço bidimensional (SOUZA, 2016). A Figura 6 ilustra duas grades bidimensionais da rede SOM, os neurônios da rede são representados por círculos pretos. Na Figura 6a e Figura 6b foram adotadas grades hexagonal e retangular, respectivamente e os números indicam a vizinhança a um raio r do neurônio central (VESANTO *et al.*, 1999).



Fonte adaptado de (VESANTO *et al.*, 1999).

Figura 6 – Rede SOM.

A rede proposta por Kohonen é organizada como uma rede bidimensional, onde cada neurônio terá um vetor peso w_{ij} associado (KOHONEN, 2001). No momento em que um novo vetor x é apresentado à rede, o neurônio que possui um vetor peso que corresponde à menor distância para uma amostra x selecionada é escolhido, sendo denominado como neurônio ativo. O neurônio ativo é selecionado de acordo com a equação:

$$(i, j) = \arg \min_{p, q} \| x - w_{pq} \|, \quad (2.28)$$

em que $\| \cdot \|$ representa a norma euclidiana, o par (i, j) são as coordenadas do neurônio associados ao vetor peso w_{pq} , e $\arg \min$ é o argumento mínimo, que é o conjunto de valores dos argumentos de determinada função para os quais ela atinge o valor mínimo.

O treinamento do mapa de Kohonen acontece por meio de uma vizinhança de raio r , sendo esse parâmetro denominado raio da vizinhança. Os padrões da rede devem ser sorteados aleatoriamente e apresentados à rede. Sempre que um vetor é apresentado, a rede calcula o neurônio ativo por meio da Equação 2.28. Após o cálculo das células ativas deve-se calcular o novo vetor de pesos para a célula ativa e seus vizinhos de raio, atraindo-os para o vetor x . Na etapa de treinamento da rede de Kohonen se utiliza a equação:

$$\Delta w_{pq} = \alpha(x - w_{pq}), \quad (2.29)$$

em que p e q são coordenadas e α é o ganho de adaptação. O raio r e o ganho de adaptação α variam de acordo com uma lei prefixada, geralmente os valores de α e r são inicializados com valores que decrescem ao longo das iterações.

As etapas de treinamento do mapa de Kohonen são dadas a seguir (MARQUES, 2005):

1. Todos os vetores de pesos são inicializados de forma aleatória. O raio de vizinhança, r , e um ganho de adaptação α são inicializados.
2. Para o cálculo da célula ativa utiliza-se a equação 2.28, selecionando-se um vetor aleatório $x \in X$.
3. Os pesos são modificados de acordo com as seguintes equações:

$$w_{pq} = w_{pq} + \Delta w_{pq}, \forall (p, q) \in V_r(i, j),$$

$$\Delta w_{pq} = \alpha(x - w_{pq}).$$

4. Retornar ao passo 2 até atingir a condição de parada.

O vetor de pesos que mais se aproxima da entrada apresentada é escolhido conforme a Equação 2.28, sendo que a célula correspondente, chamada de neurônio ativo, irá representar a entrada topologicamente. Os vetores de peso do neurônio ativo e de sua vizinhança, denotada por $V_r(i, j)$, são atualizados pelo passo 3 do algoritmo. Para o treino da rede SOM foi utilizado o SOM Toolbox (VESANTO *et al.*, 1999). O valor de α obedece a seguinte equação:

$$\alpha = \frac{\alpha_0}{1 + \frac{100m}{T_{ir}}} \quad (2.30)$$

sendo α_0 o ganho de adaptação inicial, o qual depende do tamanho da rede (q_n) (quantidade de neurônios), T_{ir} é a duração do treinamento e m representa uma dada iteração do treinamento (VESANTO *et al.*, 2000). O critério de parada adotado foi a duração do treinamento que é dada por:

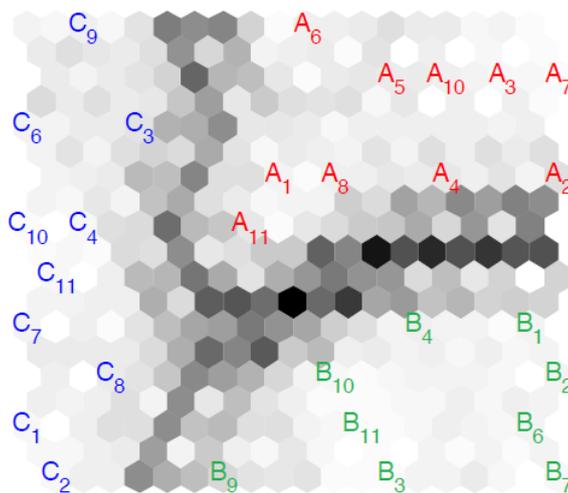
$$T_{tr} = \max\left(\frac{20q_n}{\eta}, 1\right), \quad (2.31)$$

onde η é o número de objetos da base de dados, os quais servem de entrada no treinamento da rede SOM, e $\max\{\cdot\}$ resulta no valor máximo dos operandos.

Matriz U

Uma forma de representação da rede SOM é a Matriz U que ilustra a distância entre os vetores de pesos por níveis de cinza ou de cores (ULTSCH; SIEMON, 1990). Caso a distância de uma vizinhança m_i seja pequena, uma intensidade de cinza clara será empregada e caso contrário uma intensidade mais escura será utilizada (KOHONEN, 1982).

A Figura 7 exhibe como a Matriz U é construída. O nível de cinza das células na matriz representa a distância dos vetores pesos. De acordo com a Figura 7, nota-se que os elementos A_k estão mais próximos, ou seja, a uma distância menor entre eles caracterizando-os como um grupo, assim como os elementos B_k e C_k que são próximos entre si. Os elementos de classes distintas estão, portanto, distantes de elementos de outras classes e isso é exibido pela coloração escura de fronteiras.



Fonte adaptado de (CARNEIRO, 2018)
Figura 7 – Exemplo de uma Matriz U.

3 METODOLOGIA

A Figura 8 mostra as etapas da metodologia adotada neste trabalho. Foram realizados dois experimentos, o primeiro deles com as imagens das bases de Pílulas Lícitas e Pílulas Ilícitas sem processamento prévio. O segundo experimento utilizou apenas imagens segmentadas ou pré-processadas. Para a primeira etapa, de metodologia as CNNs foram usadas somente para a extração de características e seus resultados avaliados em experimentos de classificação e recuperação.

A segunda abordagem utiliza imagens segmentadas de bases públicas disponíveis de pílulas e do alfabeto internacional da linguagem de sinais para descrição e análise de suas formas. Nessa abordagem, o descritor NMBE foi usado como extrator de características juntamente com as duas CNNs. Os resultados dos experimentos de classificação e recuperação foram avaliados quantitativa e qualitativamente.

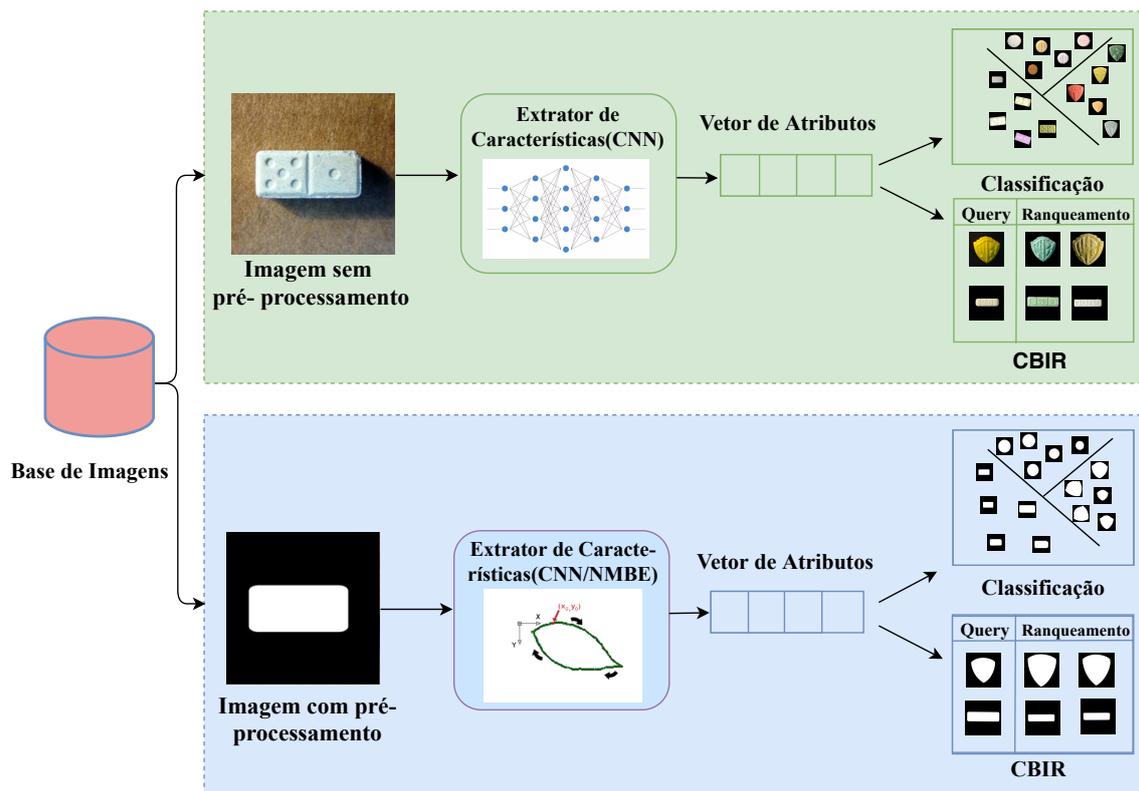


Figura 8 – Metodologia proposta para descrição e análise de formas usando CNNs e NMBE.

3.1 Base de Imagens

As bases de imagens utilizadas nesse trabalho são voltadas para o reconhecimento de pílulas e para o alfabeto internacional da linguagem de sinais. Duas dessas bases são de imagens de pílulas, sendo uma delas de pílulas lícitas, ou seja, de medicamentos comercializados nos Estados Unidos da América e uma segunda de pílulas ilícitas que tem amostras de drogas sintéticas. Essas bases de pílulas possuem suas respectivas máscaras segmentadas. Além dessas bases, utilizamos uma terceira base composta de imagens da linguagem de sinais que também possui máscaras de segmentação.

3.1.1 Pílulas Lícitas

A base pública de imagens de pílulas da Biblioteca Nacional de Medicina dos Estados Unidos da América (NLM) foi disponibilizada por Yaniv *et al.* (2016) e suas máscaras segmentadas em (USHIZIMA *et al.*, 2015). Esta base está nomeada como Pílulas Lícitas. A Figura 9 exibe uma amostra de cada classe da base e sua respectiva máscara.

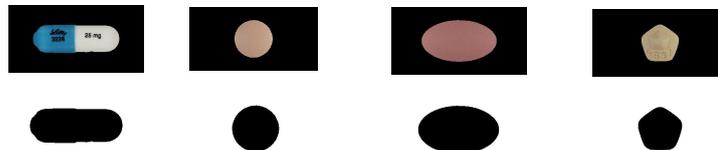


Figura 9 – Amostras de imagens e suas respectivas máscaras de segmentação da base pública de Pílulas Lícitas.

De acordo com (USHIZIMA *et al.*, 2015), as pílulas que são aprovadas para consumo recebem um código (*National Drug Code*) e um rótulo (*Structured Product Labeling*) e a Tabela 2 exibe a distribuição de pílulas em suas classes. Esta base é composta de 3 classes bem definidas de acordo com o formato da pílula, as quais contêm 454, 166 e 336 imagens, respectivamente. A classe *Diversos* não foi utilizada neste trabalho devido à grande variabilidade intraclasse.

Tabela 2 – Porcentagem de amostras de Pílulas Lícitas.

Classes	<i>Redondos</i>	<i>Ovais</i>	<i>Encapsulados</i>	<i>Diversos</i>
Amostras	454	336	166	44

3.1.2 Pílulas Ilícitas

A segunda base é composta de Pílulas Ilícitas ¹ e contém 284 imagens divididas em 31 classes não balanceadas. Neste trabalho foram utilizadas nos testes somente as 5 classes com maior quantidade de amostras, a saber, 16, 18, 20, 26 e 37 de imagens. As amostras das classes são exibidas na Tabela 3 e são as únicas classes que possuem segmentação (CARNEIRO *et al.*, 2019). A Figura 10 mostra um exemplar de cada classe e sua máscara.

Tabela 3 – Porcentagem de amostras de base Pílulas Ilícitas

Classes	<i>Medium Cut</i>	<i>Tesla</i>	<i>Alprazolam</i>	<i>Warner Bros</i>	<i>Domino</i>
Amostras	37	26	20	18	16

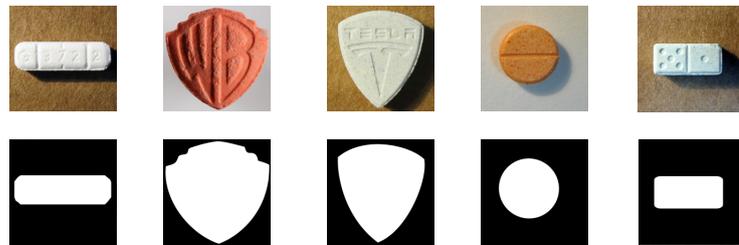


Figura 10 – Amostras de imagens e suas respectivas máscaras de segmentação da base pública de Pílulas Ilícitas.

3.1.3 Hands

A terceira base é composta por imagens da linguagem de sinais e consiste de 2038 amostras do alfabeto internacional de sinais distribuídas entre 24 classes. A Figura 11 ilustra amostras de cada classe da base.

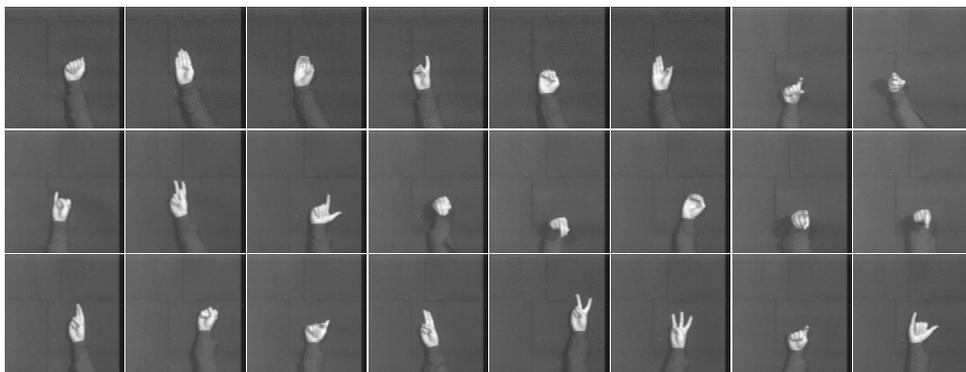


Figura 11 – Amostras de imagens de cada classe da base *Hands*.

¹ <<https://www.ecstasydata.org>>

4 RESULTADOS

Este capítulo apresenta os resultados obtidos da aplicação das duas abordagens propostas nessa dissertação para extração de características. Os métodos de extração foram utilizados em experimentos de classificação e CBIR com três bases públicas de imagens: Pílulas Lícitas, Pílulas Ilícitas e *Hands*.

4.1 Experimentos com Imagens sem Pré-processamento

Nestes experimentos a extração de características é realizada com as CNNs *LeNet* e *Inception Resnet* diretamente das imagens sem uso das máscaras de segmentação das bases Pílulas Lícitas e Pílulas Ilícitas. Para a *LeNet*, que é uma rede formada por apenas duas camadas convolucionais e duas totalmente conectadas, foram utilizadas as imagens das três bases para seu treinamento. No entanto, a *Inception Resnet*, que é uma rede convolucional profunda formada por dezenas de camadas, requer milhares de imagens para treinamento. Como as bases utilizadas nos testes não possuem amostras suficientes para treinamento da *Inception Resnet*, utilizamos essa rede treinada com a base de propósito geral denominada *ImageNet* esse processo é conhecido como transferência de aprendizado (LU *et al.*, 2015).

Por conta da diferença na quantidade de amostras das bases, a rede *LeNet* foi treinada de duas formas distintas. Na base de Pílulas Lícitas foi utilizada metade das imagens de cada classe para treino. Na base de Pílulas Ilícitas, que possui menos imagens por classe, foram utilizadas 26 classes para treino. E as 5 classes restantes para teste e a avaliação dos resultados foram empregadas as classes com maior quantidade de imagens. Como estas bases possuem classes desbalanceadas, foi necessário aplicar operações de *data augmentation* (MIKOŁAJCZYK; GROCHOWSKI, 2018) às mesmas durante a fase de treino. Assim, novas imagens foram geradas pela rotação e adição de ruído gaussiano aleatório às imagens originais. Após as operações de *data augmentation*, cada uma das 3 classes da base de Pílulas Lícitas totalizou 664 imagens e cada uma das 26 classes da base de Pílulas Ilícitas totalizou 85 imagens, sendo que somente estas imagens foram utilizadas na fase de treinamento da rede *LeNet*.

Após a etapa de treino, as características que representam as imagens de entrada são obtidas removendo a última camada totalmente conectada da rede. Com isso, cada imagem da base de teste passa a ser representada por um vetor de características. Em seguida, a distância euclidiana é utilizada para identificar as imagens mais similares da base a uma dada amostra de

consulta. A classificação foi feita com o algoritmo knn para $k = 5$ e foi avaliada pela Acurácia.

Nas subseções seguintes os resultados e a avaliação quantitativa e qualitativa dos experimentos de classificação e recuperação utilizando as imagens das máscaras de segmentação das bases de Pílulas Lícitas e Pílulas Ilícitas.

4.1.1 Resultados dos Experimentos de Classificação

As Tabelas 4 e 5 apresentam os valores de Acurácia obtidos na classificação utilizando as duas redes como extratores de características para ambas as bases de Pílulas Lícitas e Pílulas Ilícitas. Vale destacar que para o treinamento da *LeNet* foram utilizadas imagens diferentes daquelas usadas para o teste e a avaliação dos resultados. A *LeNet* extraiu 192 características de cada base, enquanto a *Inception Resnet* obteve 1536 características.

Tabela 4 – Acurácia da classificação da base de Pílulas Lícitas.

Classe	<i>LeNet</i>	<i>Inception Resnet</i>
<i>Redondos</i>	97,79%	88,98%
<i>Encapsulado</i>	92,85%	99,39%
<i>Ovais</i>	98,79%	99,55%

Tabela 5 – Acurácia da classificação da base de Pílulas Ilícitas.

Classe	<i>LeNet</i>	<i>Inception Resnet</i>
<i>Alprazolam</i>	70,00%	85,00%
<i>Domino</i>	50,00%	81,25%
<i>Medium Cut</i>	70,00%	89,18%
<i>Tesla</i>	69,23%	96,15%
<i>Warner Bros</i>	66,67%	83,33%

As medidas de Acurácia das Tabelas 4 e 5 mostram que mesmo utilizando a *Inception Resnet* que é uma base pré-treinada com uma base de imagens de propósito geral, foi possível caracterizar os padrões de entrada e identificá-los em diferentes classes para as duas bases de imagens.

A *LeNet* apresentou melhor desempenho nas classes das Pílulas Lícitas, possivelmente pelo fato desta base conter maior quantidade de imagens para o treinamento da rede.

Porém, os resultados da *LeNet* foram inferiores para a base de Pílulas Ilícitas, mesmo empregando *data augmentation* para o seu treinamento. Isso se deve provavelmente à quantidade de dados que foi insuficiente para que as camadas convolucionais extraíssem características relevantes das imagens, tais como: textura, forma e cor, além da qualidades das imagens da Pílulas Ilícitas, além da baixa qualidade das imagens que prejudicam a caracterização das mesmas.

4.1.2 Experimentos de Recuperação

As Tabelas 6 e 7 apresentam os resultados obtidos para os experimentos de recuperação. A *Inception Resnet* alcançou resultados acima de 80% para as classes *Encapsulados* e *Ovais* para a base de Pílulas Lícitas. Isso significa que a extração de atributos de imagens dessas classes assimilou nuances e detalhes de modo a recuperá-las corretamente em mais de 83%. A classe *Redondo* foi a mais desafiadora para a *Inception Resnet*, e possivelmente esse padrão não se assemelha com as imagens de treino da base *ImageNet*. No caso da base de Pílulas Ilícitas, esta rede alcançou somente para as classes *Tesla Medium Cut* valores da medida MAP acima de 80%. Os valores obtidos de MAP para todas as classes podem ser influenciados pelo cálculo da própria medida que tende a penalizar erros nas primeiras posições da recuperação.

Tabela 6 – Valores da Medida MAP para recuperação de imagens de base de Pílulas Lícitas.

Classe	<i>LeNet</i>	<i>Inception Resnet</i>
<i>Redondos</i>	84,88%	51,33%
<i>Encapsulado</i>	73,51%	88,41%
<i>Ovais</i>	88,58%	83,55%

Tabela 7 – Valores da Medida MAP para recuperação de imagens de base de Pílulas Ilícitas.

Classe	<i>LeNet</i>	<i>Inception Resnet</i>
<i>Alprazolam</i>	44,67%	66,29%
<i>Domino</i>	45,20%	70,98%
<i>Medium Cut</i>	45,60%	82,13%
<i>Tesla</i>	49,46%	90,59%
<i>Warner Bros</i>	43,88 %	70,70%

A *LeNet* apresentou melhor desempenho na recuperação das classes de *Pílulas Lícitas*, e isso se deve à maior quantidade de elementos dessa base. Além disso, a qualidade

da aquisição dessas imagens é melhor, as imagens possuem o mesmo *background*, apresentam resolução maior e ausência de sombreamento. Para a base de *Pílulas Ilícitas*, a extração de características realizada pela rede não alcançou bons resultados, não tendo sido capaz de extrair detalhes das formas que as diferenciavam umas das outras. Nessa base, a quantidade de imagens é menor, possuem *background* diferentes, apresentam baixa resolução, e sombreamento. Vale destacar que esta base de *Pílulas Ilícitas* se mostrou bastante desafiadora para as duas abordagens propostas por apresentar essas características.

A Figura 12 exibe resultados de experimentos CBIR utilizando a *LeNet* para amostras de *Pílulas Lícitas* e a *Inception Resnet* para as amostras de *Pílulas Ilícitas*. Em cada linha dessa imagem, a primeira coluna a esquerda exibe a imagem de consulta e as colunas seguintes exibem as imagens mais semelhantes retornadas da consulta à base.

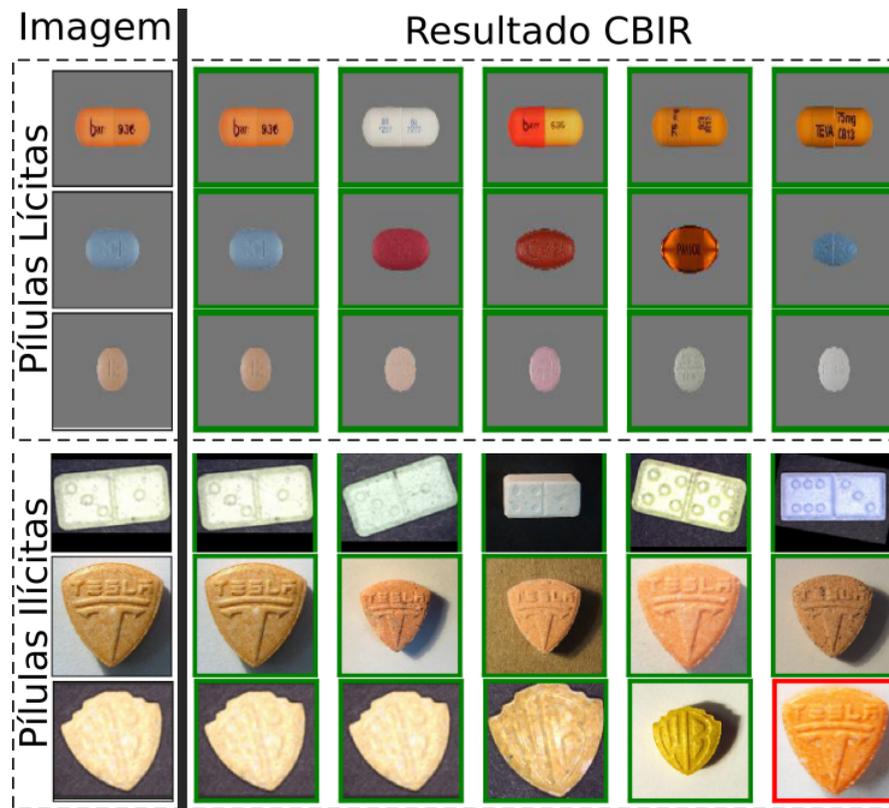


Figura 12 – Resultados de recuperação com a *LeNet* para amostras de *Pílulas Lícitas* e a *Inception Resnet* para as amostras da base de *Pílulas Ilícitas*. As bordas verdes indicam imagens corretamente retornadas e as vermelhas incorretamente retornadas.

As bordas verdes indicam que a imagem recuperada é da mesma classe da amostra de consulta enquanto que as bordas vermelhas indicam que a imagem recuperada é de classe

diferente da consulta. Uma vantagem da abordagem proposta é a utilização das imagens originais nos testes sem o uso de pré-processamento, o que implica em um menor tempo de processamento.

As Figuras 13 e 14 representam as Matrizes U obtidas com os atributos extraídos das redes convolucionais *LeNet* e *Inception Resnet*. Os hexágonos são representações de amostras da base e as amostras de uma mesma classe são marcadas com a mesma cor para cada classe.

De acordo com a Figura 13, os agrupamentos da base de Pílulas Lícitas estão melhor distribuídos, particularmente quando a caracterização das formas se realiza com a *LeNet*, quando comparados aos resultados da *Inception Resnet*. Na Figura 13b, as amostras das classes *Ovais* e *Redondos* se sobrepõem, confirmando que há dificuldades em agrupar essas duas classes, ou seja, discriminar suas formas a partir de seus atributos. Esse resultado também confirma os dados das Tabelas 4 e 6, principalmente na classe *Redondos* que obteve os menores valores de Acurácia e MAP para ambos os experimentos.

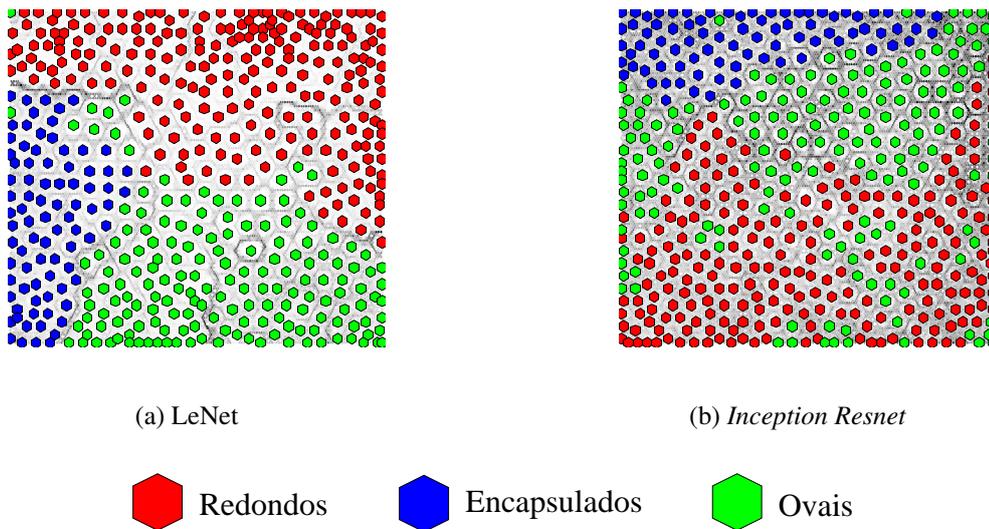
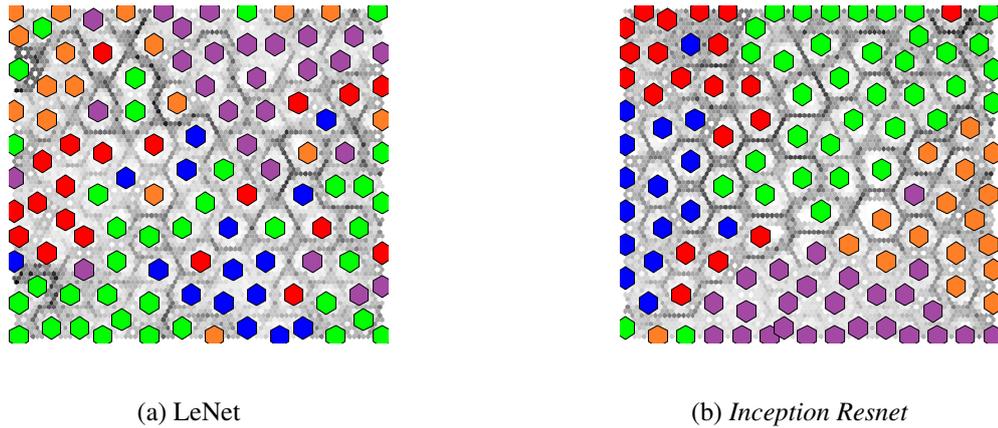


Figura 13 – Matriz U para Pílulas Lícitas descritas pela rede (a) *LeNet* e (b) *Inception Resnet*.

Para as Pílulas Ilícitas descritas pela *LeNet*, a Matriz U exibida na Figura 14a apresenta maior dispersão intra e entre os agrupamentos do que a base de Pílulas Lícitas, quando descritas pela mesma rede. Esse resultado se deve principalmente à dificuldade da rede em ser treinada em um base com menor quantidade de amostras.

Os agrupamentos gerados pela Matriz U ao utilizar em descritores oriundos da *Inception Resnet* são mais compactos e menos dispersos, como mostra a Figura 14b, sendo estes constituídos em sua maioria por amostras das mesmas classes. As fronteiras entre os agrupamentos ou classes de pílulas estão mais bem definidas o que indica numa maior distância

entre amostras de classes diferentes. As Tabelas 5 e 7 confirmam esses resultados das Matrizes U. Vale destacar que para a base de Pílulas Ilícitas, os descritores dos pares de pílulas Tesla e Warner Bros, Alprazolam e Domino não foram capazes de caracterizar adequadamente as mesmas, tendo gerado assim confusão, possivelmente, devido a semelhanças de cor e forma de objetos pertencentes a essas classes.



■ Alprazolam
 ■ Domino
 ■ Medium Cut
 ■ Warner Bros
 ■ Tesla
 Figura 14 – Matriz U para Pílulas Ilícitas descritas pela (a) *LeNet* e (b) *Inception Resnet*.

4.2 Experimentos com as Imagens com Pré-processamento

Nesses experimentos, a extração de características é realizada utilizando o descritor NMBE e as redes *LeNet* e *Inception Resnet*. Para extração de atributos das amostras das bases, somente as regiões resultantes da segmentação com as máscaras foram utilizadas. Nesses experimentos incluímos a base *Hands* para testes de generalização das abordagens propostas em experimentos de classificação e CBIR.

A rede *LeNet* foi treinada utilizando 70% das amostras de cada uma das três bases e a parte restante das amostras foi utilizada para avaliar os experimentos. A operação de *data augmentation* foi novamente utilizada considerando somente 10 rotações, pois as bases são binárias. A base *Hands* não pode ser rotacionada, pois a rotação de qualquer imagem de gesto descaracteriza o seu significado para o usuário da linguagem de sinais. Após a operação de *data augmentation*, a base de Pílulas Lícitas totalizou 3480 amostras e a base de Pílulas Ilícitas 800 amostras. Após a etapa de extração de características para os três métodos, a classificação foi feita com o classificador knn para $k=3$.

4.2.1 Experimentos de Classificação com Pré-processamento

A Tabela 8 mostra os resultados alcançados utilizando os três métodos de extração de atributos estudados: NMBE, *LeNet* e *Inception Resnet*. De acordo com os resultados de classificação a base de dados *Hands* apresentou o pior desempenho para o NMBE, pois o descritor não foi capaz de capturar detalhes globais e locais formas desses conjunto de dados. A base de Pílulas Lícitas mostrou ser uma base menos desafiadora, pois todos os métodos alcançaram resultados acima de 90 %. A base de Pílulas Ilícitas foi a que apresentou os piores resultados para a *LeNet*, e isso se deve principalmente à quantidade de imagens nesse experimento, uma vez que, apenas 117 imagens da base são segmentadas e por isso o treinamento foi realizado com menos imagens.

Tabela 8 – Resultados da Acurácia.

Base	NMBE	LeNet	<i>Inception Resnet</i>
Pílulas Lícitas	99,36	98,67	98,37
Pílulas Ilícitas	90,11	84,33	93,64
<i>Hands</i>	79,48	98,91	95,48

A Tabela 8 confirma ainda que os métodos baseados em redes neurais são os mais adequados na caracterização das Pílulas Lícitas e *Hands*, pois alcançaram os melhores resultados, em particular a *Inception Resnet*. Vale ressaltar que existe uma relação entre a quantidade de amostras da base e a Acurácia da classificação obtida pela *LeNet*, ou seja, bases com maior quantidade de amostras são melhor caracterizadas pelos descritores extraídos pelas redes.

4.2.2 Experimentos de Recuperação com Pré-processamento

A Tabela 9 apresenta os resultados obtidos para os experimentos de recuperação. A base de Pílulas Lícitas se mostrou bem caracterizada pelos métodos extratores de atributos utilizados, tanto na recuperação como na classificação. Isso demonstra que as formas de cada classe são distintas entre as classes facilitando os experimentos.

A base *Hands* se mostrou mais desafiadora para os métodos NMBE e *Inception Resnet*, pois ambos alcançaram resultados inferiores a 60% na recuperação. De fato, os vetores de características obtidos pelo descritor NMBE e pela *Inception Resnet*, não foram capazes de adquirir todos os detalhes das formas dessas bases. Para o NMBE a dificuldade está em

identificar todas as nuances do contorno após a suavização. A dificuldade para a *Inception Resnet* está na ausência de imagens com essas características na base de seu pré-treinamento *ImageNet*.

A *LeNet* apresentou melhor resultado em MAP para as bases de Pílulas Lícitas e *Hands*, porque ambas apresentam maior quantidade de amostras. Com isso a rede teve a capacidade de aprender os detalhes de cada forma para essas classes. No entanto, quando a base a ser analisada possui poucas amostras sua capacidade de descrição é reduzida e isso impactou nos resultados da base de Pílulas Ilícitas.

Tabela 9 – Resultados para CBIR (MAP).

Base	NMBE	LeNet	<i>Inception Resnet</i>
Pílulas Lícitas	90,38	91,57	85,33
Pílulas Ilícitas	72,58	80,69	83,87
<i>Hands</i>	43,12	93,53	56,18

As Figuras 15, 16 e 17 representam as Matrizes U para as bases Pílulas Lícitas, Pílulas Ilícitas e *Hands*, respectivamente. As Matrizes U para a *LeNet* apresentam menor quantidade de amostras que as demais matrizes, porque exibem apenas as amostras utilizadas no teste da rede, uma vez que a base foi treinada utilizando a operação de *data augmentation*.

De acordo com a Figura 15, os descritores de Pílulas Lícitas levaram a agrupamentos com pequena dispersão intraclasse e cujas fronteiras desses agrupamentos denotam uma maior separação entre as classes. Esse resultado confirmou o que havia sido encontrado nos outros dois experimentos cujos resultados são mostrados nas Tabelas 8 e 9. Com isso concluímos que os três métodos de extração de atributos se mostram adequados na caracterização de objetos dessa base.

De acordo com a Figura 16a o descritor NMBE foi responsável pelo agrupamento de objetos das classes *Tesla* e *Warner Bros* em um mesmo grupo, resultando portanto no pior desempenho de recuperação. Na Figura 16b a Matriz U formada pela caracterização de objetos pela *LeNet* mostra que esta foi capaz de distinguir, em particular, amostras de objetos da classe *Medium Cut* e *Tesla*. No entanto, para o restante da base essa matriz exhibe uma maior dispersão e confusão entre as classes.

Na Figura 16c oriunda da descrição de objetos pela *Inception Resnet*, se observa que essa rede foi capaz de extrair adequadamente características dos mesmos e assim separar objetos das classes *Medium Cut*, *Warner Bros* e *Tesla*. As classes *Alprazolam* e *Domino* formaram um único grupo, mostrando com isso que estes objetos constituem um desafio para o descritor, no

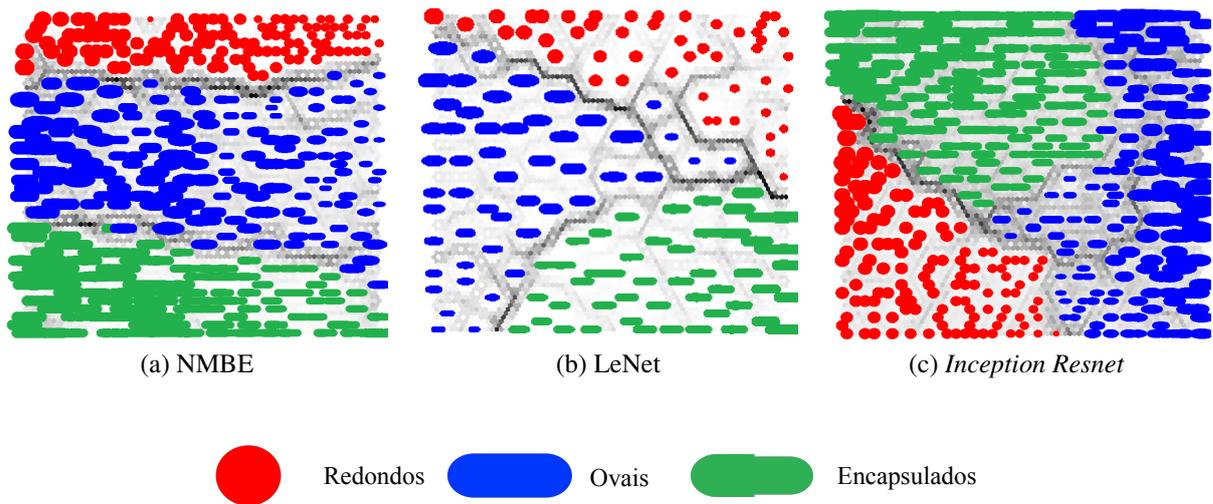


Figura 15 – Matrizes U para Pílulas Lícitas.

caso a *Inception Resnet*.

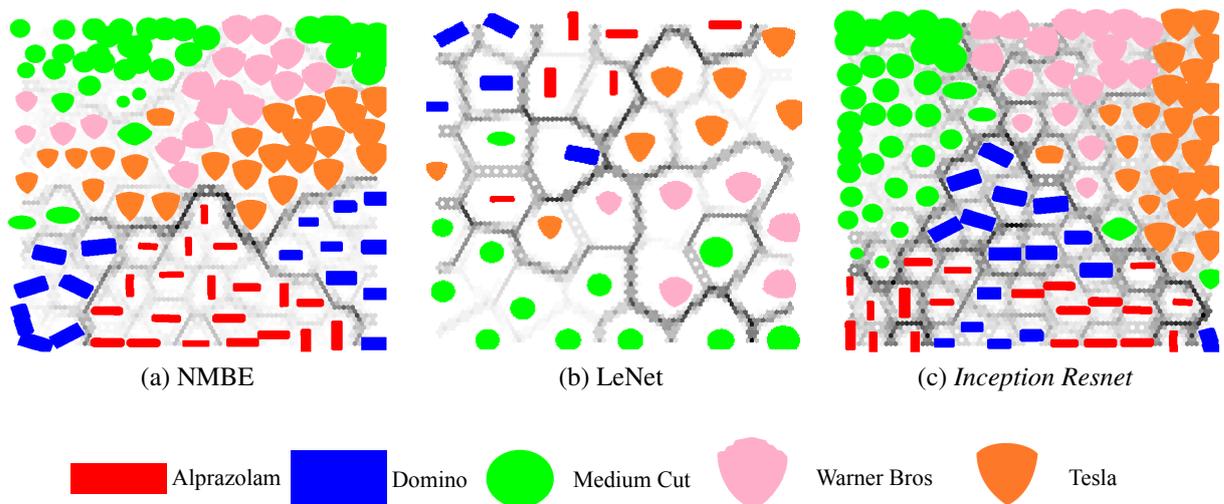


Figura 16 – Matrizes U para Pílulas Ilícitas.

As Matrizes U para a base *Hands* são mostradas na Figura 17. A Figura 17a confirma que o descritor NMBE não foi capaz caracterizar os objetos adequadamente, pois a A Matriz U exhibe agrupamentos dispersos e com a existência de subgrupos de uma mesma classe.

Na Figura 17b a Matriz U para a rede *Lenet* exhibe a melhor configuração de arranjo dos agrupamentos dessa base, contendo os elementos das classes agrupados e próximos e com fronteiras escuras entre as classes. Em síntese, o arranjo da matriz apresenta menor dispersão intra classes e maior dispersão entre as classes (fronteiras escuras). Esses resultados são confirmados pelos valores obtidos nas Tabelas 8 e 9.

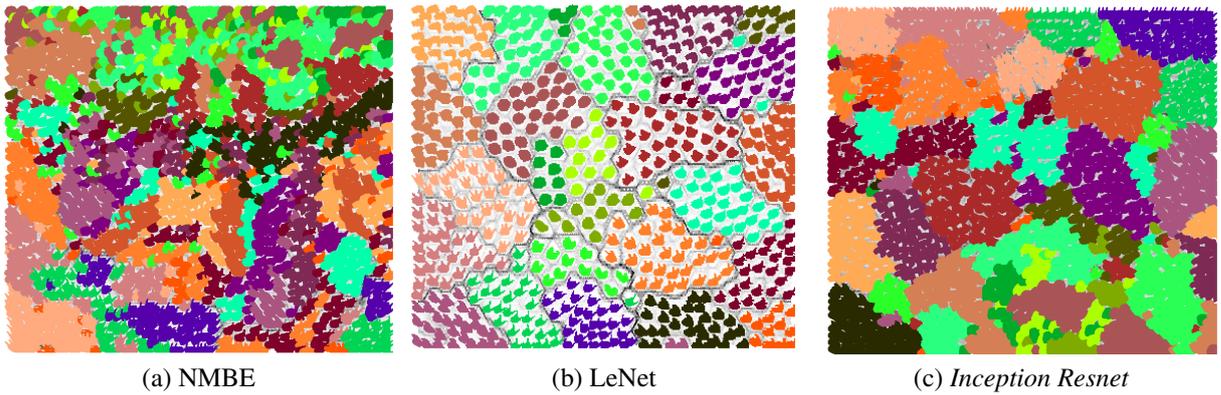


Figura 17 – Matrizes U para a base *Hands*.

De acordo com a Figura 17c a Matriz U da rede *Inception Resnet* apresenta subgrupos de uma mesma classe no arranjo da matriz, mostrando que esta rede não foi capaz de caracterizar bem estes objetos dispersos.

5 CONCLUSÕES E TRABALHOS FUTUROS

Nessa dissertação foram empregados três métodos distintos para descrição e análise de objetos, a saber, o descritor de formas NMBE e as redes *LeNet* e *Inception Resnet* que são duas arquiteturas de CNNs. Experimentos de classificação e CBIR foram realizados para testar a robustez dos descritores na caracterização dos objetos de três bases de imagens relacionadas a cuidados na saúde, e assim identificar o descritor mais adequado a uma determinada base no reconhecimento automático de objetos. Os testes foram realizados em imagens de pílulas sem pré-processamento, ou seja, sem segmentação prévia para as CNNs. Além disso, também foram realizados experimentos com imagens segmentadas com os três descritores no intuito de avaliar somente a capacidade de descrição deles com relação a atributos de forma do objeto. Mesmo diante dos desafios impostos pelas nuances próprias dos objetos de cada base, os resultados alcançados com a rede *Inception Resnet* foram satisfatórios nas imagens testadas. A *LeNet* alcançou bons resultados para a base de imagens lícitas, pois existiam imagens suficientes para o treinamento da rede. O desempenho inferior da *LeNet* para a base de imagens de Pílulas Ilícitas foi decorrente, principalmente, da menor quantidade de amostras de treinamento disponível.

Para as imagens pré-processadas, o descritor NMBE alcançou o melhor resultado quando comparado à *LeNet* para a base de Pílulas Ilícitas, apesar desse descritor não depender do treinamento. A base *Hands* se apresentou como a mais desafiadora para o NMBE, principalmente, por conta das formas dos objetos em análise, que ao serem suavizadas no processamento deste descritor, perdem detalhes relevantes para a sua caracterização, quando comparada às outras duas bases de imagens. Concluimos ainda que quando a *LeNet* tem ao seu dispor quantidade maior de dados para treinamento, seu desempenho supera tanto a *Inception Resnet* pré-treinada quanto o descritor NMBE. Esse maior número de imagens de treino implica em mais informação disponível sobre a forma do objeto, nuances do contorno, textura e cor, o que implica em maior acerto na descrição e reconhecimento do objeto. A *Inception Resnet* apresentou resultados próximos aos da *LeNet*, exceto para a base *Hands*, considerando que neste trabalho não houve treinamento nas bases de testes, pois a usamos pré-treinada com a *ImageNet*. Apesar do descritor multiescala NMBE ser um método menos robusto quando comparado às redes neurais, ele apresentou um desempenho competitivo com as CNNs na base de Pílulas Lícitas. Como limitações das abordagens propostas, destacamos a utilização da descrição de objetos pela rede *LeNet* pois o desempenho da mesma é dependente da quantidade de imagens presente na base, tendo em vista o seu treinamento.

5.1 Trabalhos Futuros

Como perspectivas de trabalhos para a continuidade desse projeto, identificamos as seguintes possibilidades:

- Utilização de outras arquiteturas de redes neurais convolucionais na descrição de objetos;
- Desenvolvimento de um *framework* que associe esses métodos de descrição e experimentos CBIR com uma ferramenta de geolocalização das pílulas ilícitas para facilitar o rastreamento de drogas pela polícia em tarefas de buscas e apreensões.

REFERÊNCIAS

- ARAÚJO, F. H. D. **Convnets na caracterização, recuperação e ranqueamento de células** 2018. 107 f. Tese (Doutorado em Engenharia de Teleinformática) — Centro de Tecnologia, Universidade Federal do Ceará, Fortaleza, 2018. Disponível em: <<http://www.repositoriobib.ufc.br/000047/00004741.pdf>>. Acesso em: 14 ago. 2018.
- ARTHUR, D.; VASSILVITSKII, S. K-means++: The advantages of careful seeding. In: **Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms**. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2007. (SODA '07), p. 1027–1035. ISBN 978-0-898716-24-5. Disponível em: <<http://dl.acm.org/citation.cfm?id=1283383.1283494>>.
- BIRK, H.; MOESLUND, T. B.; MADSEN, C. B. Real-time recognition of hand alphabet gestures using principal component analysis. In: **10th SCANDINAVIAN CONFERENCE ON IMAGE ANALYSIS**. Lappeenranta, Finland: [s.n.], 1997. v. 1, p. 261–268.
- CAMARGO, J.; ESSEIVA, P.; GONZALEZ, F.; WIST, J.; PATINY, L. Monitoring of illicit pill distribution networks using an image collection exploration framework. **Forensic Science International**, v. 223, n. 1, p. 298 – 305, 2012. ISSN 0379-0738.
- CARNEIRO, A. C.; LOPES, J. G.; MEDEIROS, F. N. S. **Illicit Pills**. 2019. <<https://data.mendeley.com/datasets/bmsz34n3dk/2>>.
- CARNEIRO, C. A. **O papel da função custo na otimização de um descritor de formas multiescala** 2018. 57 f. Dissertação (Mestrado em Engenharia de Teleinformática) — Centro de Tecnologia, Universidade Federal do Ceará, Fortaleza, 2018. Disponível em: <<http://www.repositoriobib.ufc.br/000043/00004306.pdf>>. Acesso em: 23 abr. 2018.
- CESAR JR, R. M.; COSTA, L. F. Application and assessment of multiscale bending energy for morphometric characterization of neural cells. **Review of Scientific Instruments**, AIP Publishing, v. 68, n. 5, p. 2177–2186, 1997. ISSN 1089-7623.
- COSTA, L. F.; CESAR JR, R. M. **Shape Classification and Analysis: Theory and Practice**. 2. ed. Boca Raton, Florida, FL, USA: CRC Press Taylor & Francis Group, 2009. ISBN 978-0-8493-7929-1.
- DELGADO, N. L.; USUYAMA, N.; HALL, A. K.; HAZEN, R. J.; MA, M.; SAHU, S.; LUNDIN, J. Fast and accurate medication identification. **NPJ DIGITAL MEDICINE**, v. 2, n. 10, 2019. ISSN 2398-6352.
- DENG, J.; DONG, W.; SOCHER, R.; LI, L.-J.; LI, K.; FEI-FEI, L. Imagenet: A large-scale hierarchical image database. In: **2009 IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION**. Miami Beach, USA.: [s.n.], 2009. p. 248–255. ISSN 1063-6919.
- DESHPANDE, A. **The 9 Deep Learning Papers You Need To Know About (Understanding CNNs)**. 2016. <<https://adeshpande3.github.io/adeshpande3.github.io/The-9-Deep-Learning-Papers-You-Need-To-Know-About.html>>. Acessado em 02/06/2019.
- FERREIRA, M. V. dos S.; FILHO, A. O. de C.; SOUSA, A. D. de; SILVA, A. C.; GATTASS, M. Convolutional neural network and texture descriptor-based automatic detection and diagnosis of glaucoma. **Expert Systems with Applications**, Elsevier, v. 110, p. 250–263, 2018.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Cambridge, Massachusetts: MIT Press, 2016. <<http://www.deeplearningbook.org>>.

HMIMID, A.; SAYYOURI, M.; QJIDAA, H. Fast computation of separable two-dimensional discrete invariant moments for image classification. **Pattern Recognition**, Elsevier, v. 48, n. 2, p. 509–521, 2015.

HU, Y. *et al.* Multiple-instance ranking: Learning to rank images for image retrieval. In: **IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION**. Anchorage, USA: [s.n.], 2008. p. 1–8.

ISLAM, M. M.; SIDDIQUA, S.; AFNAN, J. Real time hand gesture recognition using different algorithms based on american sign language. In: **2017 IEEE INTERNATIONAL CONFERENCE ON IMAGING, VISION PATTERN RECOGNITION (ICIVPR)**. Himeji, Japan: [s.n.], 2017. p. 1–6.

KOHONEN, T. Analysis of a simple self-organizing process. **Biological cybernetics**, Springer, v. 44, n. 2, p. 135–140, 1982.

KOHONEN, T. **Self-organizing Maps**. 3. ed. Berlin, GER: Springer Series in Information Sciences, 2001. v. 30. ISBN 3-540-67921-9.

LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, v. 86, n. 11, p. 2278–2324, Nov 1998.

LEE, Y.-B.; PARK, U.; JAIN, A. K.; LEE, S.-W. Pill-id: Matching and retrieval of drug pill images. **Pattern Recognition Letters**, Elsevier, v. 33, n. 7, p. 904–910, 2012.

LOPATKA, M.; HOUTEN, W. V. Automated shape annotation for illicit tablet preparations: A contour angle based classification from digital images. **Science & Justice**, Elsevier, v. 53, n. 1, p. 60–66, 2013.

LU, J.; BEHBOOD, V.; HAO, P.; ZUO, H.; XUE, S.; ZHANG, G. Transfer learning using computational intelligence: A survey. **Knowledge-Based Systems**, v. 80, p. 14 – 23, 2015. ISSN 0950-7051. 25th anniversary of Knowledge-Based Systems. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0950705115000179>>.

MADDALA, K. T.; MOSS, R. H.; STOECKER, W. V.; HAGERTY, J. R.; COLE, J. G.; MISHRA, N. K.; STANLEY, R. J. Adaptable ring for vision-based measurements and shape analysis. **IEEE Transactions on Instrumentation and Measurement**, v. 66, n. 4, p. 746–756, April 2017.

MARQUES, J. S. **Reconhecimento de padrões: métodos estatísticos e neuronais**. 2. ed. Lisboa, POR: IST (Instituto Superior Técnico) PRESS, 2005. ISBN 9789728469085.

MIKOŁAJCZYK, A.; GROCHOWSKI, M. Data augmentation for improving deep learning in image classification problem. In: **2018 INTERNATIONAL INTERDISCIPLINARY PHD WORKSHOP (IIPHDW)**. Szczecin, Poland: [s.n.], 2018. p. 117–122.

MITRA, S.; ACHARYA, T. Gesture recognition: A survey. **IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)**, IEEE, v. 37, n. 3, p. 311–324, 2007.

MURTY, M. N.; DEVI, V. S. **Pattern Recognition: An Algorithmic Approach**. 1. ed. [S.l.]: Springer, 2011. (Undergraduate Topics in Computer Science). ISBN 978-0-85729-494-4.

PEDRINI, H.; SCHWARTZ, W. R. **Análise de Imagens Digitais: Princípios, Algoritmos e Aplicações**. 1. ed. São Paulo, São Paulo, SP, BRA: Editora Thomson Learning Edições Ltda, 2007. ISBN 978-85-221-0595-3.

SILVA, R. R. V. e. **Detecção Automática de Células Cervicais com Base em Atributos Radiais** 2018. 83 f. Tese (Doutorado em Engenharia de Teleinformática) — Universidade Federal do Ceará, Fortaleza, 2018. Disponível em: <<http://www.repositoriobib.ufc.br/000047/00004742.pdf>>. Acesso em: 14 ago. 2018.

SINGH, C.; SINGH, J. Geometrically invariant color, shape and texture features for object recognition using multiple kernel learning classification approach. **Information Sciences**, Elsevier, v. 484, p. 135–152, 2019.

SOMBANDITH, V.; WALAIRACHT, A.; WALAIRACHT, S. Hand gesture recognition for lao alphabet sign language using hog and correlation. In: **2017 14TH INTERNATIONAL CONFERENCE ON ELECTRICAL ENGINEERING/ELECTRONICS, COMPUTER, TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY (ECTI-CON)**. Phuket, Thailand: [s.n.], 2017. p. 649–651.

SOUZA, M. M. S. **Um método para otimização de descritores multiescala em análise de formas**. 2016. 77 f. Tese (Doutorado em Engenharia de Teleinformática) — Universidade Federal do Ceará, Fortaleza, 2016. Disponível em: <<http://www.repositoriobib.ufc.br/000030/00003057.pdf>>. Acesso em: 10 out. 2017.

SZEGEDY, C.; IOFFE, S.; VANHOUCHE, V. Inception-v4, inception-resnet and the impact of residual connections on learning. **Computing Research Repository**, abs/1602.07261, 2016.

SZEGEDY, C.; WEI LIU; YANGQING JIA; SERMANET, P.; REED, S.; ANGUELOV, D.; ERHAN, D.; VANHOUCHE, V.; RABINOVICH, A. Going deeper with convolutions. In: **2015 IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR)**. Boston, USA: [s.n.], 2015. p. 1–9. ISSN 1063-6919.

ULTSCH, A.; SIEMON, H. P. Kohonen's self organizing feature maps for exploratory data analysis. In: **PROCEEDINGS OF INTERNATIONAL NEURAL NETWORKS CONFERENCE**. Paris, France: Kluwer Academic Press, 1990. v. 1, p. 305–308.

UNODC. **United Nations Office on Drugs and Crime: World Drug Report 2010**. Vienna, Austria, 2011.

USHIZIMA, D.; CARNEIRO, A.; SOUZA, M.; MEDEIROS, F. Investigating pill recognition methods for a new national library of medicine image dataset. In: **SPRINGER. INTERNATIONAL SYMPOSIUM ON VISUAL COMPUTING**. Las Vegas, USA, 2015. p. 410–419.

VESANTO, J.; HIMBERG, J.; ALHONIEMI, E.; PARHANKANGAS, J. Self-organizing map in matlab: the som toolbox. In: **Proceedings of the Matlab DSP conference**. Espoo, Finland: [s.n.], 1999. v. 99, p. 16–17.

VESANTO, J.; HIMBERG, J.; ALHONIEMI, E.; PARHANKANGAS, J. **SOM Toolbox for Matlab 5**. Helsinki, Poland, 2000.

WANG, B.; BROWN, D.; GAO, Y.; SALLE, J. L. March: Multiscale-arch-height description for mobile retrieval of leaf images. **Information Sciences**, v. 302, p. 132 – 148, 2015. ISSN 0020-0255.

WANG, J. Z. **Integrated Region-Based Image Retrieval**. US: Springer Science & Business Media, 2001. v. 11. ISBN 978-1-4615-1641-5.

WONG, Y. F.; NG, H. T.; LEUNG, K. Y.; CHAN, K. Y.; CHAN, S. Y.; LOY, C. C. Development of fine-grained pill identification algorithm using deep convolutional network. **Journal of Biom. Informatics**, Elsevier, v. 74, p. 130–136, 2017.

YANIV, Z.; FARUQUE, J.; HOWE, S.; DUNN, K.; SHARLIP, D.; BOND, A.; PERILLAN, P.; BODENREIDER, O.; ACKERMAN, M. J.; YOO, T. S. The national library of medicine pill image recognition challenge: An initial report. In: **2016 IEEE APPLIED IMAGERY PATTERN RECOGNITION WORKSHOP (AIPR)**. Washington, USA: [s.n.], 2016. p. 1–9.