

Análise de Fotografias de Pílulas por Redes Neurais Convolucionais

Allan C. Carneiro¹, José G. F. Lopes¹, Flávio H. D. Araújo^{1 2}, Romuere R. V. Silva^{1 2},
Cornélia J. P. Passarinho¹, Jeová F. S. Rocha Neto³, Fátima N. S. Medeiros¹

¹Universidade Federal do Ceará, ²Universidade Federal do Piauí, ³Brown University, USA
e-mail: allanccarneiro@gmail.com

Abstract - The automatic recognition of pills through photography may reduce administration medication error and also enable forensic intelligence to establish links among drugs, laboratories, local and international trafficking routes. Considering the remarkable progress and performance of the convolutional neural networks (CNNs) to extract image characteristics, we propose the use of CNNs for automatic recognition of pill images. The results of classification and content-based image retrieval (CBIR) are evaluated quantitatively by the Accuracy and mean average precision (MAP) measures, and qualitatively by cluster visualization of the U-Matrix. The results demonstrated that the LeNet outperformed the pre-trained Inception Resnet on classification and CBIR tests for legal pills, while Inception Resnet outperformed LeNet on illegal pills.

Palavras-chave: CNN, CBIR, pílulas.

Introdução

Com o crescente acesso aos medicamentos por pacientes, as chances de ocorrência de erros na sua administração aumentam [1]. Isso se deve, muitas vezes, à dificuldade de identificação visual das pílulas por pacientes e profissionais da saúde. Nove em cada dez pessoas com idade acima dos 65 nos Estados Unidos tomam mais de um medicamento em forma de pílula [1]. Portanto, a identificação incorreta desses medicamentos é uma questão preocupante, visto que o reconhecimento correto poderia prevenir entre 6 e 8 mil mortes por ano [2].

Essa dificuldade de identificação pode também surgir durante a apreensão de pílulas ilícitas por autoridades competentes em que são demandadas tarefas de indexação e catalogação das mesmas. Os serviços de catalogação e indexação destas pílulas são realizados pelos órgãos de repressão ao tráfico e utilizam, em geral, testes laboratoriais de química forense. É possível ainda utilizar a identificação automática de pílulas ilícitas que circulam no mercado internacional para fins de rastreamento e apreensão pelas autoridades competentes [3]. Diante do exposto, surge a

necessidade de identificação automática de pílulas lícitas e ilícitas e do desenvolvimento de algoritmos de reconhecimento de padrões baseados em visão computacional e aprendizado de máquina [4, 5].

Existem vários serviços de indexação de pílulas tanto públicos (*pillreport.com*, *ecstasydata.org*, *erowid.org*, *PharmInfoNet*), quanto privados (DPD, Epocrates, Ident-A-Drug) que proporcionam uma ferramenta de busca simples [6]. No entanto, estes serviços não apresentam ferramentas para recuperação de imagens baseada em conteúdo.

Considerando o crescente progresso e desempenho das redes neurais convolucionais (CNNs) na caracterização de padrões de imagens [7, 8], a principal contribuição desse trabalho consiste em utilizar CNNs na extração de características de imagens de pílulas lícitas e ilícitas. Essas características constituem as entradas dos algoritmos de classificação e recuperação de imagens baseada em conteúdo (CBIR). Pretende-se disponibilizar uma abordagem de reconhecimento de pílulas por imagens que possa auxiliar tanto na caracterização como na identificação de pílulas lícitas e ilícitas.

Materiais e métodos

A abordagem proposta segue o fluxograma da Figura 1, consistindo nas seguintes etapas:

Base de imagens: bases de imagens contendo pílulas lícitas e ilícitas são utilizadas nos testes. A base pública de pílulas lícitas é oriunda da Biblioteca Nacional de Medicina (NLM) dos Estados Unidos da América [1]. Essa base contém 3 classes bem definidas com 454, 166 e 336 imagens, respectivamente, de acordo com o formato da pílula. A segunda base é composta de pílulas ilícitas, está disponível em *ecstasydata.org* e contém 284 imagens divididas em 31 classes não balanceadas. As 5 classes com maior quantidade de amostras, a saber, 16, 18, 20, 26 e 37 imagens foram utilizadas nos testes.

Extração de Atributos: atributos de forma, cor e textura de objetos podem ser quantificados e utilizados como descritores em experimentos

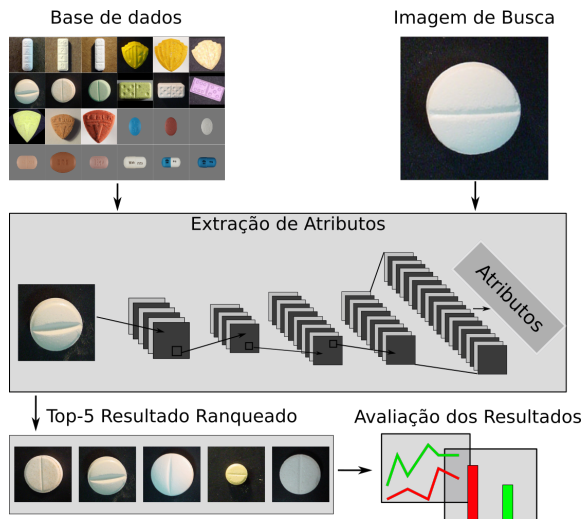


Figura 1 – Fluxograma do método proposto.

CBIR. Assim, cada objeto pode ser representado no espaço R^n em termos dos n atributos extraídos, e que sejam invariantes a transformações como rotação, escala e translação, de preferência.

Experimentos CBIR: sistemas CBIR empregam técnicas para realização de busca de imagens a partir de exemplares de consulta, utilizando características das próprias imagens [9]. Essas características podem ser extraídas em baixo nível (detecção de bordas [10]), alto nível (histograma de cor [11]) ou ambos. As imagens consultadas são comparadas às imagens da base por avaliação de similaridade. As N imagens da base mais similares à imagem consultada são retornadas e exibidas ao usuário, na ordem decrescente de similaridade.

Avaliação dos Resultados: na avaliação quantitativa dos resultados, os experimentos de classificação são avaliados pela medida Acurácia, que é definida pela razão entre a quantidade de elementos corretamente preditos e o número total dos elementos. Quanto aos experimentos CBIR, esses são aferidos quantitativamente pela medida *mean average precision* (MAP) [12] e na avaliação qualitativa aplica-se a Matriz U [9].

Método Proposto: a extração de características é realizada utilizando CNNs: *LeNet* [7] e *Inception Resnet* [8]. A *LeNet* é uma rede formada por apenas duas camadas convolucionais e duas totalmente conectadas e requer poucas imagens no treinamento. A *Inception Resnet* é uma rede convolucional profunda formada por dezenas de blocos de camadas, e requer milhares de imagens para treinamento e assim evitar o *overfitting*. Como as bases utilizadas nos testes não possuem amostras

suficientes para treinar a *Inception Resnet*, utilizou-se essa rede treinada com uma base de propósito geral denominada *ImageNet* [13].

A rede *LeNet* foi treinada de duas formas, devido à diferença na quantidade de amostras das bases. Na base de pílulas lícitas utilizamos 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 5 classes - aquelas com maior quantidade de imagens - para teste e avaliação dos resultados. Além disso, como as bases possuem classes desbalanceadas, foram necessárias operações de *data augmentation*. Assim, novas imagens foram geradas pela rotação e adição de ruído gaussiano aleatório nas 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 estas imagens foram utilizadas no 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, que é a camada de classificação. 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 *k-nearest neighbors* (kNN) para $k=5$ e foi avaliada pela Acurácia (A).

Resultados

A Tabela 1 apresenta os valores de Acurácia obtidos na classificação e da medida MAP na recuperação e ranqueamento das imagens utilizando as duas redes para as pílulas lícitas e ilícitas. Vale destacar que as imagens utilizadas na avaliação dos resultados não estavam no conjunto de treinamento da *LeNet*.

As medidas Acurácia e MAP mostram que mesmo utilizando a *Inception Resnet* treinada com uma base de imagens de propósito geral, a qual não possui pílulas, é possível caracterizar os padrões de entrada e identificá-los em diferentes classes para as duas bases de imagens. A *LeNet* obteve melhor desempenho na separação das classes das pílulas lícitas por haver mais imagens disponíveis para o treino. Em contrapartida, mesmo empregando *data augmentation* para o

Classe	<i>LeNet</i>		<i>Inception Resnet</i>	
	MAP	A	MAP	A
Pílulas Lícitas				
Redondos	84,88%	97,79%	51,33%	88,98%
Encapsulados	73,51%	92,85%	88,41%	99,39%
Ovais	88,58%	98,79%	83,55%	99,55%
Pílulas Ilícitas				
Alprazolam	44,67%	70,00%	66,29%	85,00%
Domino	45,20%	50,00%	70,98%	81,25%
Medium Cut	45,60%	70,00%	82,13%	89,18%
Tesla	49,46%	69,23%	90,59%	96,15%
Warner Bros	43,88%	66,67%	70,70%	83,33%

Tabela 1 – Medidas MAP e Acurácia (A) da *LeNet* e *Inception Resnet* para pílulas lícitas e ilícitas.

treinamento da *LeNet* com imagens de pílulas ilícitas, a quantidade de dados foi insuficiente para que as camadas convolucionais extraíssem características relevantes das imagens de entrada.

A Figura 2 exibe resultados de experimentos CBIR utilizando a *LeNet* para amostras de pílulas lícitas e a *Inception Resnet* para as amostras ilícitas. Em cada linha da Figura 2, a primeira coluna a esquerda exibe a imagem de consulta e as colunas seguintes exibem as imagens mais semelhantes retornadas da consulta à base. As bordas verdes indicam que a imagem recuperada é da mesma classe da amostra de consulta, e as bordas vermelhas indicam que é de uma classe diferente. 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. Vale destacar que a base de pílulas ilícitas é desafiadora, pois apresenta imagens com baixa resolução, padrão de fundo variável e sombreamento.

As Figuras 3 e 4 exibem as Matrizes U obtidas com as redes *LeNet* e *Inception Resnet* para as duas bases. Para fins de visualização, hexágonos representam as pílulas, sendo as amostras de uma mesma classe marcadas com cores iguais. Quanto mais escura uma fronteira entre duas amostras, maior é a distância entre elas.

Para as pílulas lícitas, os grupos formados pelas características da *LeNet* estão melhor agrupados como exibe a Figura 3a. Na Figura 3b, os grupos formados com a *Inception Resnet* estão mais dispersos com as classes Redondos e Ovais se sobrepondo. Nota-se que, na Tabela 1, os valores de MAP e Acurácia para as classes Encapsulados e Ovais são semelhantes, no entanto os valores das duas medidas para a classe Redondos são bem superiores para a rede *LeNet*, o que está de acordo com as Matrizes U para as pílulas lícitas.

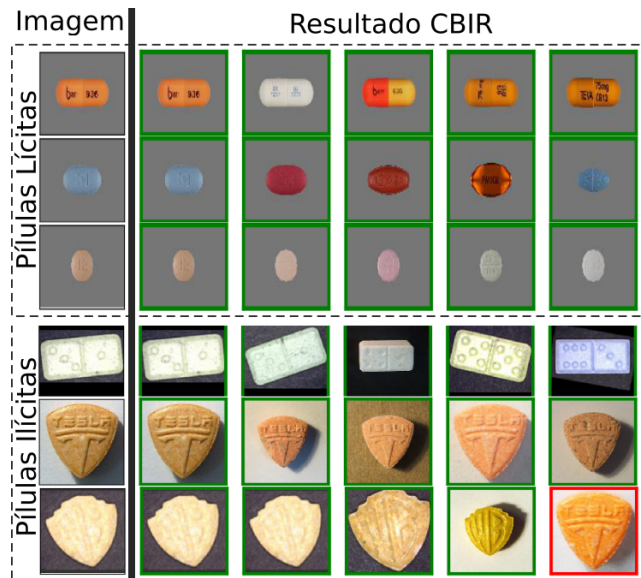


Figura 2 – Resultados de recuperação com a *LeNet* para amostras de pílulas lícitas e a *Inception Resnet* para as amostras ilícitas. As bordas verdes indicam imagens corretamente retornadas e as vermelhas incorretamente retornadas.

Para as pílulas ilícitas, a Matriz U da *LeNet* (Figura 4a) exibe maior dispersão e confusão entre as amostras, isso foi causado devido a pequena quantidade de imagens disponíveis para o treinamento. Os grupos da *Inception Resnet* são mais compactos, como mostra a Figura 4b, sendo estes constituídos em sua maioria por amostras das mesmas classes. As fronteiras entre as classes de pílulas são melhor definidas na Figura 4b, indicando assim maior distância entre amostras de classes diferentes. A Tabela 1 confirma os resultados das Matrizes U. Vale destacar que, na base de pílulas ilícitas, os pares Tesla e Warner Bros, Alprazolam e Domino se confundiram por terem cor e formato semelhantes.

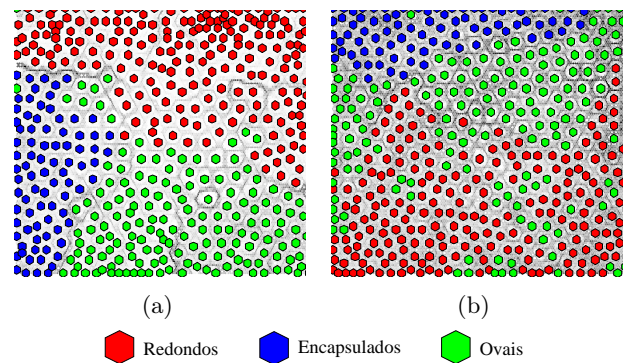


Figura 3 – Matriz U para pílulas lícitas com a rede (a) *LeNet* e (b) *Inception Resnet*.

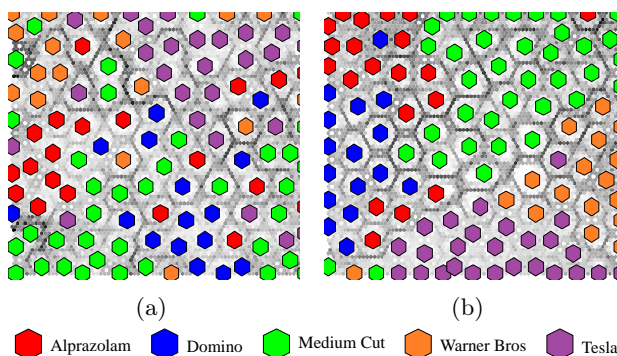


Figura 4 – Matriz U para pílulas ilícitas com a rede (a) *LeNet* e (b) *Inception Resnet*.

Conclusões

Esse trabalho emprega duas arquiteturas de CNNs em experimentos de classificação e CBIR com imagens de pílulas que podem auxiliar na identificação e no monitoramento das mesmas. Os testes foram realizados em imagens de pílulas sem pré-processamento, embora essas possuam diferentes padrões de fundo, sombreamento e resolução. Mesmo diante desses desafios, os resultados alcançados com a rede *Inception Resnet* foram satisfatórios nas imagens testadas. A *LeNet* obteve bom desempenho somente 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 ilícitas é decorrente da pequena quantidade de amostras de treinamento.

Agradecimentos

Agradecemos a CAPES e ao CNPq (444784/2014-4, 401442/2014-4, 306600/2016-1, 150288/2017-4) pelo apoio financeiro.

Referências

- [1] D. Ushizima, A. Carneiro, M. Souza, and F. Medeiros, *Investigating Pill Recognition Methods for a New National Library of Medicine Image Dataset*, vol. 9475, pp. 410–419. Springer, 2015.
- [2] D. Hale, “Pill meets poka to enhance patient safety,” in *Medicine 2.0 Congress*, 2009.
- [3] R. Chakravarti and X. Meng, “A study of color histogram based image retrieval,” in *6th International Conference on*

Information Technology : New Generations, pp. 1323–1328, IEEE, 2009.

- [4] Y.-B. Lee, U. Park, A. K. Jain, and S.-W. Lee, “Pill-id: Matching and retrieval of drug pill images,” *Pattern Recognition Letters*, vol. 33, no. 7, pp. 904–910, 2012.
- [5] Y. F. Wong, H. T. Ng, K. Y. Leung, K. Y. Chan, S. Y. Chan, and C. C. Loy, “Development of fine-grained pill identification algorithm using deep convolutional network,” *Journal of Biom. Informatics*, vol. 74, pp. 130–136, 2017.
- [6] M. Lopatka and W. Van Houten, “Automated shape annotation for illicit tablet preparations: A contour angle based classification from digital images,” *Science & Justice*, vol. 53, no. 1, pp. 60–66, 2013.
- [7] Y. Lecun *et al.*, “Gradient-based learning applied to document recognition,” in *Proceedings of the IEEE*, pp. 2278–2324, 1998.
- [8] C. Szegedy *et al.*, “Inception-v4, inception-resnet and the impact of residual connections on learning,” *Computing Research Repository*, 2016.
- [9] J. Z. Wang, *Integrated Region-Based Image Retrieval*, vol. 11. Springer Science & Business Media, 2001.
- [10] T. Kato, Kurita, *et al.*, “A sketch retrieval method for full color image database-query by visual example,” in *11th International Conference on Pattern Recognition (IAPR), Computer Vision and Applications, Vol.1.*, pp. 530–533, IEEE, 1992.
- [11] C. W. Niblack *et al.*, “Qbic project: querying images by content, using color, texture, and shape,” in *IS&T/SPIE’s Symposium on Electronic Imaging: Science and Technology*, pp. 173–187, International Society for Optics and Photonics, 1993.
- [12] Y. Hu *et al.*, “Multiple-instance ranking: Learning to rank images for image retrieval,” in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, IEEE, 2008.
- [13] O. Russakovsky *et al.*, “ImageNet Large Scale Visual Recognition Challenge,” *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.