

# Busca Reversa de Imagens: Técnicas de Hashing e CNNs

Sandy Hoffmann<sup>1</sup>

UFSC, Florianópolis, SC

Ricardo R. Ladeira<sup>2</sup>

IFC, Blumenau, SC

A busca reversa de imagens (RIS) é uma técnica que identifica imagens visualmente semelhantes, aplicada em plataformas como Google Lens e Bing Imagens. Este trabalho explora técnicas de *hashing* perceptivo, redes neurais convolucionais (CNN) e *hashing* sensível à localidade (LSH) para melhorar a acurácia e a eficiência da RIS.

O *hashing* perceptivo é uma técnica que gera valores de *hash* semelhantes para imagens visualmente parecidas, mantendo valores próximos mesmo quando a entrada é exposta a alterações sutis. A transformada discreta do cosseno (DCT) é um método comum de *hashing* perceptivo, que utiliza coeficientes de baixa frequência para gerar códigos *hash* [5]. Os códigos gerados por esse tipo de *hash* podem ser utilizados para medir a similaridade entre duas entradas distintas [2].

O LSH reduz o custo computacional ao particionar o espaço de busca em grupos (*buckets*), onde pontos próximos têm maior probabilidade de serem mapeados para o mesmo *bucket* [1]. A probabilidade  $P$  de dois vetores  $A$  e  $B$  colidirem<sup>3</sup>, ao serem processados pela função *hash*  $h$ , é dada pela Equação 1 [1], dependendo do ângulo  $\theta$  entre os vetores.

$$P[h(A) = h(B)] = 1 - \frac{\theta}{\pi}, \quad \theta = \arccos \left( \frac{|A \cap B|}{\sqrt{|A| \cdot |B|}} \right). \quad (1)$$

A CNN é amplamente utilizada no processamento de imagens devido à sua capacidade de extrair características específicas através de operações de convolução [3]. A arquitetura VGG16<sup>4</sup>, por exemplo, utiliza camadas de convolução, *pooling* e ativação (ReLU) para classificar e detectar imagens [4]. Durante o treinamento, a CNN é submetida a diversas transformações visuais, como rotação e corte, o que aumenta sua robustez frente a alterações nas imagens [2].

Nos experimentos<sup>5</sup>, foram comparadas as técnicas de DCT, CNN e LSH aplicado à CNN, considerando acurácia e tempo de execução. Foram avaliadas imagens com alterações visuais, sendo estas: desfoque gaussiano, dessaturação, alteração de resolução, realce de nitidez, rotação, espelhamento, corte e inserção de elementos gráficos. A taxa de acertos foi avaliada com base no retorno das nove variações de cada imagem consultada, incluindo as oito transformações visuais e a versão original. O espaço de busca consistiu em 4500 imagens (500 originais mais suas oito variações). Os resultados de acurácia e tempo de execução são apresentados no Gráfico 1 e na Tabela 1.

A CNN demonstrou maior robustez frente a alterações como corte, rotação e espelhamento, com taxas de acerto significativamente maiores que o DCT. No entanto, o DCT apresentou tempos de execução mais rápidos, especialmente em grandes volumes de dados. A combinação da CNN com o LSH mostrou-se eficaz, aumentando a acurácia e reduzindo o tempo de execução. O LSH

<sup>1</sup>sandy.hoffmann@posgrad.ufsc.br

<sup>2</sup>ricardo.ladeira@ifc.edu.br

<sup>3</sup>Uma colisão ocorre quando duas entradas distintas geram o mesmo valor de hash.

<sup>4</sup>Convolutional Network for Classification and Detection.

<sup>5</sup>Códigos disponíveis no Github: [https://github.com/SandyHoffmann/BuscaReversaImagens\\_CNN-DCT-LSH](https://github.com/SandyHoffmann/BuscaReversaImagens_CNN-DCT-LSH).

Tabela 1: Média de tempo de execução obtida com as técnicas DCT, CNN e LSH.

| Técnica | Tempo de execução (em segundos) |              |              |
|---------|---------------------------------|--------------|--------------|
|         | 900 Imagens                     | 2700 Imagens | 4500 Imagens |
| DCT     | 34,06                           | 38,85        | 66,14        |
| CNN     | 63,55                           | 165,07       | 307,60       |
| LSH     | 51,00                           | 123,59       | 216,15       |

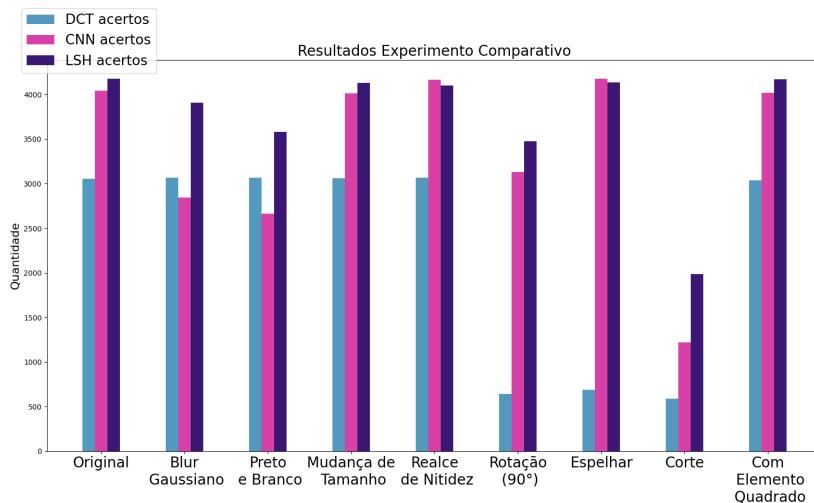


Figura 1: Porcentagem de acertos obtidos com as técnicas DCT, CNN e LSH. Fonte: Autora (2024).

obteve melhores resultados em sete das nove transformações testadas, destacando-se em operações como rotação e corte.

Este trabalho comparou técnicas de *hashing* perceptivo, CNN e LSH para busca reversa de imagens, demonstrando que a CNN é mais robusta frente a alterações visuais, enquanto o LSH aplicado à CNN melhora a eficiência da busca em grandes volumes de dados.

## Referências

- [1] M. S. Charikar. “Similarity estimation techniques from rounding algorithms”. Em: **Proceedings of the Thiry-Fourth Annual ACM Symposium on Theory of Computing**. STOC ’02. Montreal, Quebec, Canada: Association for Computing Machinery, 2002, pp. 380–388. ISBN: 1581134959. DOI: 10.1145/509907.509965.
- [2] M. Gaillard. **Perceptual Hashing using Convolutional Neural Networks for Large Scale Reverse Image Search**. INSA, 2017.
- [3] I. Goodfellow, Y. Bengio e A. Courville. **Deep Learning**. MIT Press, 2016.
- [4] K. Simonyan e A. Zisserman. “Very Deep Convolutional Networks for Large-Scale Image Recognition”. Em: **arXiv preprint arXiv:1409.1556** (2015). arXiv: 1409.1556 [cs.CV].
- [5] C. Zauner. **Implementation and Benchmarking of Perceptual Image Hash Functions**. Rel. técn. Upper Austria University of Applied Sciences, Hagenberg Campus, 2010.