

# Rede Neural Convolutacional (CNN) aplicada em identificação de embarcações que navegam nos rios da Amazônia

Gustavo Do Nascimento Tocantins<sup>1</sup>, André Guilherme Gouvêa Dos Anjos<sup>2</sup>

Brenda Do Socorro Da Silva Corrêa<sup>3</sup>

Fábio Ferreira Barros<sup>4</sup>

Joyce Luana Reis Coelho<sup>5</sup>

Kadu Naoki Ishikawa<sup>6</sup>

André Vinícius da Costa Araújo<sup>7</sup>

Marcus Pinto da Costa da Rocha<sup>8</sup>

Valcir João da Cunha Farias<sup>9</sup>

Laboratório de Análise Numérica e Otimização, UFPA, Belém, PA

A Inteligência Artificial (IA) está muito presente no cotidiano da sociedade nos últimos anos, visto que houve uma significativa evolução e uma crescente necessidade de criar ferramentas que possam facilitar ou gerar novas possibilidades para futuros projetos e desenvolvimentos, nesse viés, surgiu as redes neurais. As redes neurais são excelentes ferramentas que possibilitam que as máquinas aprendam, gradativamente, determinado assunto, categorias, classificações de imagens e diversas outras áreas que podem ser exploradas, tornando possível ter como objetivo uma ampliação de mecanismos que facilitem problemas cotidianos ou situações complexas de engenharia. O objetivo desse trabalho, é elaborar uma rede neural convolutacional para classificar até 10 tipos de embarcações que circulam nos rios da Amazônia.

Entre as várias arquiteturas de aprendizagem profunda, as Redes Neurais Convolutacionais (CNNs) vêm alcançando um notável destaque, tendo grandes avanços no processamento de áudio, vídeo e imagens. As CNNs foram inspiradas na organização hierárquica do córtex visual e nas noções clássicas de células simples e complexas. De acordo com os experimentos com gatos e macacos [3], foram observados que células simples foram ativadas quando apresentados padrões simples ao animal, como linhas; e células complexas foram ativadas quando padrões mais elaborados, compostos de combinações de padrões simples, foram apresentados ao animal.

As CNNs são redes neurais de multicamadas que utilizam matrizes responsáveis pela filtragem, extraíndo características para aprender padrões e detalhes da imagem [2]. Este tipo de rede é muito utilizado para classificação de imagens, embora essa também de forma incomum possa ser aplicada para reconhecimento de sons, o qual copia a voz de uma pessoa a partir de um áudio de poucos segundos e ao submeter um texto para a rede treinada, a mesma reproduz a leitura com a voz copiada [4].

---

<sup>1</sup>gustavo.tocantins@itec.ufpa.br

<sup>2</sup>andre.anjos@itec.ufpa.br

<sup>3</sup>brenda.correa@itec.ufpa.br

<sup>4</sup>Fabiobarros@gmail.com

<sup>5</sup>joyceluana121314@gmail.com

<sup>6</sup>kadu.ishikawa@itec.ufpa.br

<sup>7</sup>andrecosta@ufpa.br

<sup>8</sup>marcus.rocha@gmail.com

<sup>9</sup>vjcfarias@gmail.com

Uma das características mais importantes da CNN é o grande número de arquiteturas que podem ser configuradas, visto que as redes possuem camadas diferentes podendo ser combinadas, incluindo a variação dos parâmetros.

A arquitetura típica de uma Rede Neural Convolutiva é estruturada por uma série de estágios, sendo as primeiras etapas compostas pelas camadas convolucionais e de subamostragem (subsampling ou pooling) e as etapas seguintes por camadas totalmente conectadas. As unidades em uma camada convolutiva são 22 organizadas em mapas de características, nos quais, cada unidade é conectada a fragmentos locais nos mapas de características da camada anterior (ou camada de entrada) através de um conjunto de pesos chamados banco de filtros, que são compartilhados entre todas as unidades.

A rede CNN desenvolvida para a realização deste trabalho foi escrita em linguagem Python e dividida em sete camadas. A primeira camada é uma camada convolutiva, com 64 neurônios e kernel  $3 \times 3$ , usada para extrair recursos de imagens bidimensionais e também é responsável pela entrada das imagens. A segunda camada é uma MaxPooling2D de tamanho  $2 \times 2$ , usada para reduzir a dimensionalidade da saída convolutiva. A terceira e a quarta camada são Conv2D, com 128 neurônios e kernel  $3 \times 3$  e MaxPooling2D novamente. A quinta camada é do tipo Flatten, que converte uma matriz de entrada multidimensional em um vetor unidimensional. Em seguida, vem a sexta camada, que é uma camada densa com 256 neurônios. Por fim, a camada de saída é do tipo densa, com 10 neurônios que se referem às 10 classes de embarcação. Em seguida, o programa compila e otimiza a rede com 200 épocas e plota as imagens com a classificação dada pela rede e a porcentagem de acerto do total de imagens analisadas.

Por fim, a rede projetada foi treinada a partir de 746 imagens, de tamanho  $45 \times 45$ , para treino e teve cem por cento de acertos para 50 imagens de teste, demonstrando, dessa forma, uma boa capacidade de aprendizagem e excelentes resultados.[1]

## Referências

- [1] C. ACOSTA A. M.e SILVA. “Aplicação de análise envoltória de dados (DEA) para medir eficiência em portos brasileiros”. Em: **Nome do Congresso ou Periódico** 5 (dez de 2010), pp. 88–122.
- [2] Y. Guo e Y. Liu. “Deep learning for visual understanding: A review”. Em: **Neurocomputing** 308 (2018), pp. 154–173.
- [3] T. N. Hubel D. H. Wiesel. “Receptive fields and functional architecture of monkey striate cortex”. Em: **The Journal of physiology** 195.1 (1968), pp. 215–243.
- [4] X. JIA. “Transfer learning from speaker verification to multispeaker text-to-speech synthesis”. Em: **Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing**. 2019, pp. 6965–6969.