

# DETECÇÃO DE MARCOS VISUAIS COM A UTILIZAÇÃO DE REDES NEURAS PROFUNDAS

Sylvio Rubens Juliani Neto<sup>1</sup>, Reinaldo Bianchi<sup>2</sup>

<sup>1,2</sup> Centro Universitário da FEI

[srjneto@hotmail.com](mailto:srjneto@hotmail.com), [reinaldo.bianchi@gmail.com](mailto:reinaldo.bianchi@gmail.com)

**Resumo:** O projeto consiste no desenvolvimento de um sistema de visão que utiliza técnicas de inteligência artificial, mais especificamente, as redes neurais profundas, para a detecção de marcos visuais em um campo de futebol. O projeto será desenvolvido com o auxílio de uma câmera monocular fixa ao robô humanoide. O objetivo do projeto, é auxiliar os sistemas de decisão e localização da equipe de futebol de robôs RoboFEI-Humanoide durante as partidas da competição RoboCup, detectando marcos como, robôs, bola e linhas.

## 1. Introdução

A visão computacional é um campo de estudo que não engloba apenas a si mesmo, mas, como é explicado em [1], um sistema de visão deve auxiliar na interação com o ambiente externo através da percepção de objetos e de variáveis como, iluminação e cores predominantes, para que assim, um sistema de decisão possa determinar a melhor ação a ser tomada em uma determinada situação. Essa percepção do agente pode ser feita a partir de diversos sensores, como ultrassom, sensores a laser ou como será apresentado neste projeto, uma câmera fixa ao próprio robô, a partir de técnicas de inteligência artificial.

Uma das constatações a partir das pesquisas realizadas, é que a maior parte dos pesquisadores não estão utilizando mais técnicas clássicas de visão, mas sim as famosas redes neurais artificiais, mais específico, as Redes Neurais Profundas (Deep Neural Networks) e Redes Neurais Convolucionais (Convolutional Neural Network). Isso se dá ao fato de que essas técnicas são capazes de realmente aprender e generalizar padrões a partir de uma grande quantidade de dados de treinamento.

A figura 1 mostra os passos realizados pela rede neural convolucional a partir de uma imagem de entrada e a figura 2 demonstra a estrutura de uma rede convolucional, apresentados em [2].

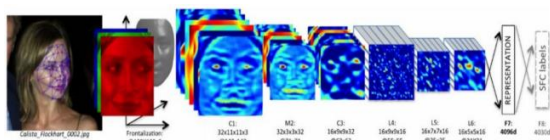


Figura 1 – Exemplo de execução das camadas de uma rede neural convolucional.

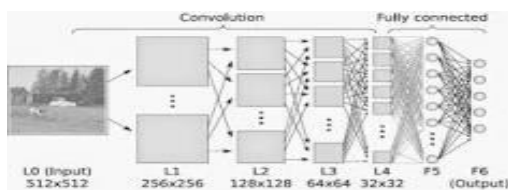


Figura 2 – Estrutura de uma rede convolucional.

As redes neurais convolucionais são uma classe especial das redes neurais artificiais. Este tipo de topologia foi criada especialmente para reconhecimento de padrões e características em duas dimensões, como no caso de objetos em imagens. Este tipo de técnica possui um alto grau de robustez quanto a distorções, como compressão, alargamento e variação de escala das características analisadas devido às camadas apresentadas nesta rede: Camada de convolução, onde são extraídos os mapas de características a partir de kernels, matrizes que se movimentam por toda a imagem; pooling, utilizada para reduzir o tamanho das matrizes resultantes da convolução, reduzindo a quantidade de parâmetros a serem aprendidos; e a camada completamente conectada, onde ocorre o aprendizado dessas características.

O trabalho [3] demonstra o funcionamento de todas as camadas e os cálculos que envolvem as redes convolucionais, assim como as aplicações dessas redes. Redes com esse tipo de estrutura (topologia), tem um papel de grande importância em aplicações como, tratamento de imagens, classificação, reconhecimento e segmentação a partir de características que a rede irá extrair para aprender determinados padrões.

As redes neurais convolucionais foram inicialmente propostas por [5], onde foi desenvolvido uma rede de múltiplas camadas capaz de classificar números escritos a mão, chamada LeNet-5, demonstrada na figura 3. Como outras redes convolucionais, esta pode ser treinada a partir do algoritmo de gradiente descendente estocástico mostrado na equação 1, minimizando o erro entre a saída desejada e a saída atual da rede a partir da ideia que em um gráfico de erro por valor de peso, no ponto mínimo da curva, a derivada é zero. Os valores dos kernels usados nas convoluções também são atualizados a partir do método de backpropagation.

$$\frac{\partial \text{erro}}{\partial w_1} = \frac{\partial \text{erro}}{\partial \text{saída}} * \frac{\partial \text{saída}}{\partial \text{camada2}} * \frac{\partial \text{camada2}}{\partial \text{camada1}} * \frac{\partial \text{camada1}}{\partial w_1} \quad (1)$$

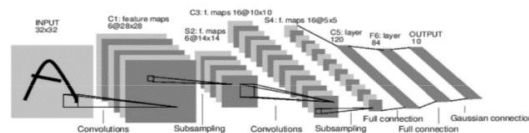


Figura 3– Estrutura da rede LeNet-5

## 2. Metodologia

Para o desenvolvimento deste trabalho, foi estudado o aprendizado de máquina, mais especificamente o aprendizado supervisionado, onde a partir de dados de entrada já classificados, será realizado um treinamento para que a rede calcule os melhores pesos possíveis para o dataset criado. Assim, sendo capaz de identificar características, a ponto de poder classificá-las corretamente a partir de novos dados de entrada não treinados.

Para o treinamento das redes neurais, será proposto a utilização de uma rede neural convolucional chamada MobileNet [4], onde ela será treinada para reconhecer objetos em campo durante os jogos de futebol de robôs como, a bola, outros robôs e marcos no campo, tais como, intersecção de linhas e a marca do pênalti, como mostrado na figura 4. Com a rede já treinada, será possível utilizar o treinamento para captar estes marcos visuais durante o jogo, podendo assim, tomar as decisões necessárias durante a partida e até mesmo utilizar tais informações para análises de localização.

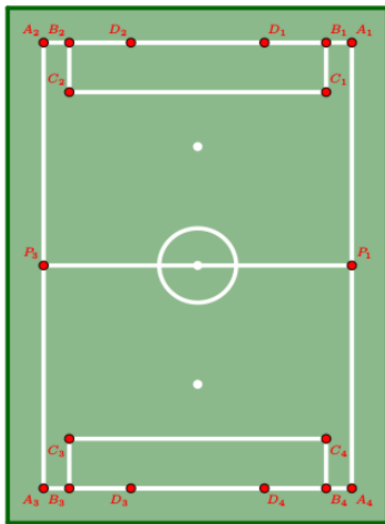


Figura 4 – Landmarks analisados nos treinamentos.

MobileNets é uma classe composta por diversos modelos (topologias) de classificação, construída para aplicações em dispositivos móveis e sistemas embarcados, devido ao fato de ser uma rede pequena e leve para treinamento e para utilização (classificação). Esses módulos são redes que demonstraram os melhores valores de acurácia a partir de treinamentos extensivos demonstrados em [4], tais modelos de rede incluem detecção de objetos, atribuições de faces e geo-localização em grande escala, como é ilustrado na figura 5.

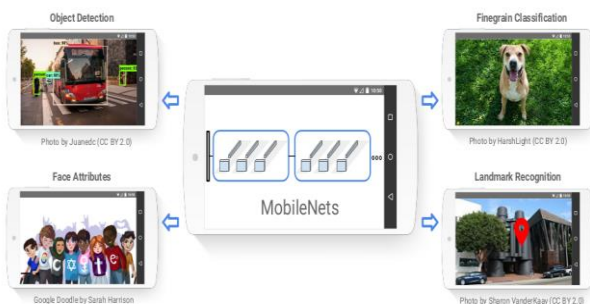


Figura 5 – Principais módulos de classificação da MobileNet.

A rede MobileNet foi desenvolvida como módulos de convoluções que permitem seu ajuste de tamanho, significando que esta rede é ajustada a partir de dois hiperparâmetros globais, que permitem ao desenvolvedor modificar sua estrutura (número de neurônios e resolução das imagens de entrada) em função das tarefas de

classificação desejadas. Estes parâmetros são chamados de "width multiplier" e "resolution multiplier".

Para a realização do treinamento das redes neurais, será utilizado um framework desenvolvido pela google, uma classe criada para operações com tensores, sendo assim, uma ótima ferramenta para tratamento de conjuntos de imagens, além de ser uma classe completa e de fácil utilização para o treinamento de redes neurais. O framework utilizado será o "tensorflow".

Para uma maior facilidade na questão de treinamento de redes, foram desenvolvidas classes prontas que automatizam o processo do próprio treinamento e teste, podendo ser acessadas facilmente por métodos e templates prontos para tensorflow.

### 3. Conclusões

A pesquisa em desenvolvimento é de grande importância para a área da robótica móvel, onde é necessário que o agente possa identificar objetos e padrões ao seu redor ou a partir de um banco de dados e gerar um aprendizado a partir de redes neurais. Isso possibilita um aprendizado próximo ao do cérebro humano, analisando e entendendo padrões nas imagens recebidas.

Como futuro da pesquisa e próximos passos, será feito o treinamento da rede, testando em campo, analisando o tempo de processamento e a acurácia obtida das classes treinadas a partir do tamanho do dataset de treinamento.

### 4. Referências

- [1] BIANCHI, Reinaldo AC; REALI-COSTA, Anna H. O Sistema de Visão Computacional do time FUTEPOLI de Futebol de Robôs. In: **Congresso Brasileiro de Automática**. 2000. p. 2156-2162.
- [2] ALBAWI, Saad; MOHAMMED, Tareq Abed; ALZAWI, Saad. Understanding of a convolutional neural network. In: **Engineering and Technology (ICET), 2017 International Conference on**. IEEE, 2017. p. 1-6.
- [3] GUO, Tianmei et al. Simple convolutional neural network on image classification. In: **Big Data Analysis (ICBDA), 2017 IEEE 2nd International Conference on**. IEEE, 2017. p. 721-724.
- [4] HOWARD, Andrew G. et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. **arXiv preprint arXiv:1704.04861**, 2017.
- [5] LECUN, Yann et al. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, v. 86, n. 11, p. 2278-2324, 1998.

### Agradecimentos

À instituição Centro Universitário FEI pela realização das medidas ou empréstimo de equipamentos.

<sup>1</sup> Aluno de IC do Centro Universitário FEI. Projeto com vigência de 02/18 a 12/18.