

## REDES NEURAIS DE CONVOLUÇÃO EM DIFERENTES FORMAS DE REPRESENTAÇÃO DE OBJETOS

PEDRO LEMOS BALLESTER<sup>1</sup>; RICARDO MATSUMURA ARAUJO<sup>2</sup>

<sup>1</sup>Universidade Federal de Pelotas – [pedballester@gmail.com](mailto:pedballester@gmail.com)

<sup>2</sup>Universidade Federal de Pelotas – [ricardo@inf.ufpel.edu.br](mailto:ricardo@inf.ufpel.edu.br)

### 1. INTRODUÇÃO

Redes neurais são reconhecidas pela sua grande capacidade de identificação de padrões não lineares em conjuntos de dados. Porém, com o desenvolvimento de técnicas para o treinamento de redes neurais, criou-se a possibilidade de análises de conjuntos de dados muito maiores e mais complexos. Dessa forma, as redes neurais passaram a desempenhar um papel importante em *Machine Learning*, criando um processo que se chama atualmente de *Deep Learning*.

*Deep Learning* é a utilização de redes neurais com diversas camadas, fornecendo uma grande capacidade de abstração de dados, buscando uma boa representação do conjunto de dados fornecido. Dessa forma, o *Deep Learning* passou a ser aplicado à outras áreas de conhecimento. Uma dessas redes utilizadas é a Rede Neural de Convolução.

Redes Neurais de Convolução tem sido de grande importância em diversas áreas, sendo uma delas a visão computacional. Criada em 1980 FUKUSHIMA (1980), a rede tenta imitar o comportamento das interações entre os neurônios responsáveis pelo mapeamento na detecção de objetos no cérebro humano.

Nota-se pelos trabalhos de LECUN (1998), KRIZHEVSKY (2012), SZEGEDY (2014), a grande capacidade da rede de identificar objetos, dado que seus dados de treinamento tenham alguma relação com os dados de teste. Porém, neste trabalho será abordada a interpretação de outra forma de representação do mesmo objeto, o desenho feito por um humano.

O conjunto de dados TU-Berlin, fornecido por EITZ et al. (2012), teve em sua criação um método de reconhecimento de desenhos feitos a mão, chegando a conquistar 56% de acertos. *O método envolvido no reconhecimento de sketches se baseou em técnicas tradicionais de visão computacional, utilizando uma variação do descritor SIFT LOWE (2004), porém somente guardando informação de orientação, em conjunto com um Bag of Features CSURKA (2004), algoritmo responsável pela criação de histogramas de features, utilizando-o como representação da imagem original.*

Além disso, para o conjunto de dados TU-BERLIN, o resultado de redes neurais de convolução, tem uma alta taxa de acerto em YU et al. (2015), ultrapassando o resultado dos humanos na mesma tarefa.

### 2. METODOLOGIA

A rede, treinada por um conjunto de dados utilizando imagens reais dos objetos, é fornecida em LECUN et al. (2014), e é amplamente reconhecida pela sua alta performance em detecção de objetos na mesma forma de representação do treinamento, nesse caso, fotos. Utiliza-se, então, o conjunto de dados de TU-Berlin para teste, em que milhares de imagens desenhadas a mão *são fornecidas*.

Pode-se realizar uma análise de como a rede se portaria enfrentando uma forma de representação diferente da que ela foi treinada para identificar. Dessa forma, a rede recebeu como entrada, diversas imagens desenhadas a mão, podendo ser realizada uma análise quantitativa de como ela se comportou.

Para a melhor interpretação dos dados, foi apresentada para a rede 80 imagens de cada grupo, ou seja, imagens que representam o mesmo objeto, selecionando os 10 objetos, indicados pela rede, com maior probabilidade de representar o objeto da imagem. Após a rede receber 80 imagens do mesmo grupo, fornecendo 10 objetos possíveis para cada uma delas, criou-se um gráfico contendo a frequência em que cada objeto é citado como uma das 10 possibilidades, retirando assim as classificações da rede sobre o que aquele grupo representa.

### 3. RESULTADOS E DISCUSSÃO

Como hipótese, acreditava-se que o resultado recebido deveria ser inferior ao que é testado com a mesma representação, porém, a rede acabou se provando incapaz de detectar o objeto representado nos desenhos que foram feitos. Nas figuras 1 e 2 fornecidas abaixo, têm-se uma análise quantitativa do resultado fornecido pela rede. O número de grupos foi reduzido devido ao espaço disponível.

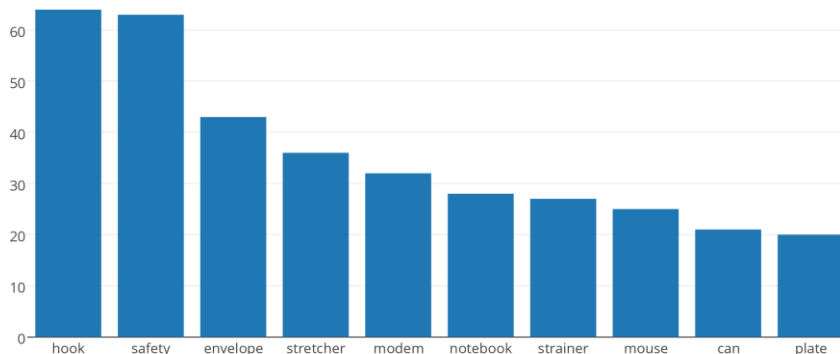


Fig. 1 – Gráfico de frequência do resultado do grupo “Caminhão”

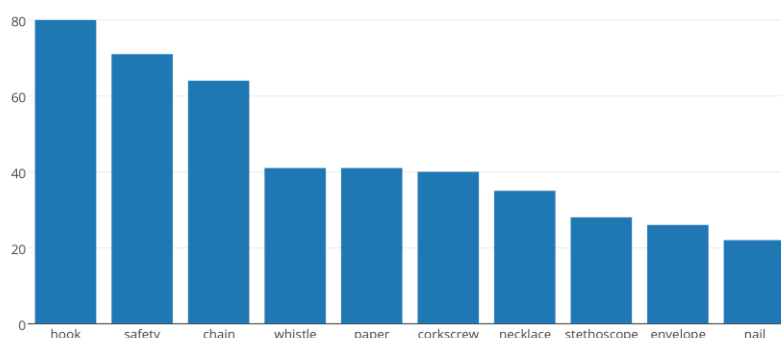


Fig. 2 – Gráfico de frequência do resultado para o grupo “Cavalo”.

Nota-se pelas Figuras 1 e 2, que a persistência da rede em citar objetos com bordas bem definidas demonstra que ela não consegue compreender a diferença entre objetos e um desenho. Isso se dá pela maneira que o ser humano desenha objetos quando não procura fidelidade.

Porém, pode-se demonstrar através da Figura 3, que a medida com que o desenho toma maior similaridade com o objeto real, a rede acaba por conquistar resultados mais condizentes com a realidade.

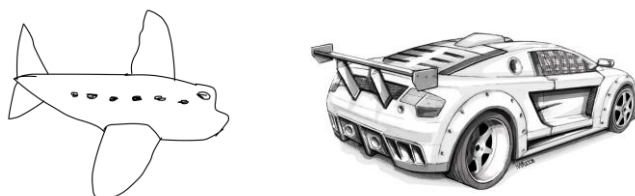


Fig. 3 – Avião e carro desenhados a mão.

Como esperado, o resultado da imagem do avião, assemelha-se com o resultado obtido para caminhão e cavalo, tendo como classe mais provável para o desenho em questão, um gancho. Já o carro, teve como classes mais prováveis, um carro de corrida, uma roda de um carro, e um carro esportivo.

#### 4. CONCLUSÕES

Inicialmente os resultados parecem inconclusivos, porém, estes fazem surgir algumas dúvidas sobre o funcionamento das redes neurais computacionais e do reconhecimento de objetos do cérebro humano.

Primeiramente, tendo como base a facilidade com que um ser humano detecta objetos desenhados a mão sem treinamento prévio, provavelmente as redes neurais de convolução ainda não são modelos suficientemente fiéis de como o cérebro reconhece objetos, tendo novas funcionalidades internas que devem ser descobertas no futuro.

Em segundo, se as redes neurais de convolução são, de fato, modelos de como o ser humano detecta os objetos, este provavelmente deve conter redes especializadas em detecção em desenhos, o que parece improvável, já que crianças desenhavam objetos que conhecem no mundo real sem observação de desenhos prévios.

Como trabalhos futuros, tem-se a intenção de estudar os motivos da rede não interpretar como o ser humano em tarefas de diferentes formas de representação, tentando responder as perguntas anteriores, e propor se há, ou não, a necessidade de existirem redes paralelas responsáveis por tipos de detecções diferentes.

## 5. REFERÊNCIAS BIBLIOGRÁFICAS

FUKUSHIMA, K. "Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position". **Biological Cybernetics** 36 (4): 193–202, 1980.

EITZ, M.; HAYS, J.; ALEXA, M. "How do Humans Sketch Objects?". **ACM Trans. Graph. (Proc. SIGGRAPH)**: 44:1-44:10, 2012.

YU, Q.; YANG, Y.; SONG, Y. -Z.; XIANG, T.; HOSPEDALES, T. "Sketch-a-Net that Beats Humans". **ArXiv e-prints**: 2015.

LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. "Gradient-based learning applied to document recognition". **Proceedings of the IEEE**, 1998.

KRIZHEVSKY, A.; SUTSKEVEER, I.; HINTON, G. E. "ImageNet Classification with Deep Convolutional Neural Networks". **Advances in Neural Information Processing Systems** 25 (NIPS 2012): 1097—1105, 2012.

SZEGEDY, C.; LIU, W.; JIA, Y.; SERMANET, P.; REED, S.; ANGUELOV, D.; ERHAN, D.; VANHOUCHE, V.; RABINOVICH, A. "Going Deeper with Convolutions". **ArXiv e-prints**: 2014

LOWE, D. G. "Distinctive image features from scale-invariant keypoints". *Int'l. Journal of Computer Vision* 60": 91-110, 2004. --ARRUMAR POIS É JOURNAL--

CSURKA, G.; DANCE, C. R.; FAN, L.; WILLIAMOWSKI, J.; BRAY, C. "Visual categorization with bags of keypoints". **Workshop on Statistical Learning in Computer Vision, ECCV**: 2004.

SERMANET, P.; EIGEN, D.; ZHANG, X.; MATHIEU, M.; FERGUS, R.; LECUN, Y.; "OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks". **International Conference on Learning Representations (ICLR 2014)**. 2014