

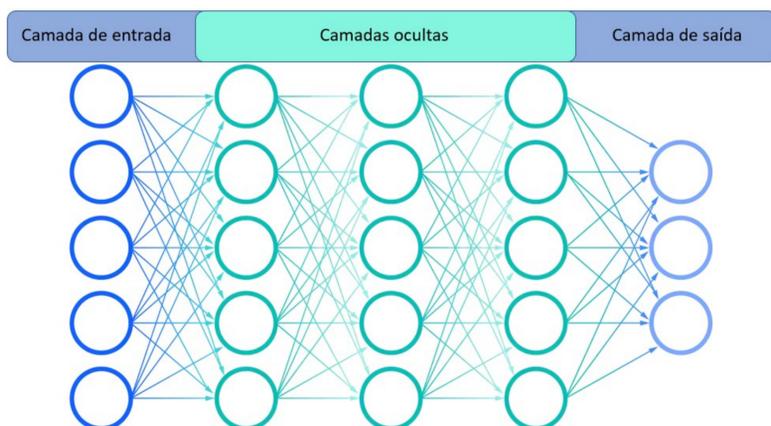
Reconhecimento de Ações Humanas com Deep Learning

Departamento de Ciência da Computação
Instituto de Matemática e Estatística - USP

Lucas Henrique Bahr Yau
Orientador: R. Hirata Jr.

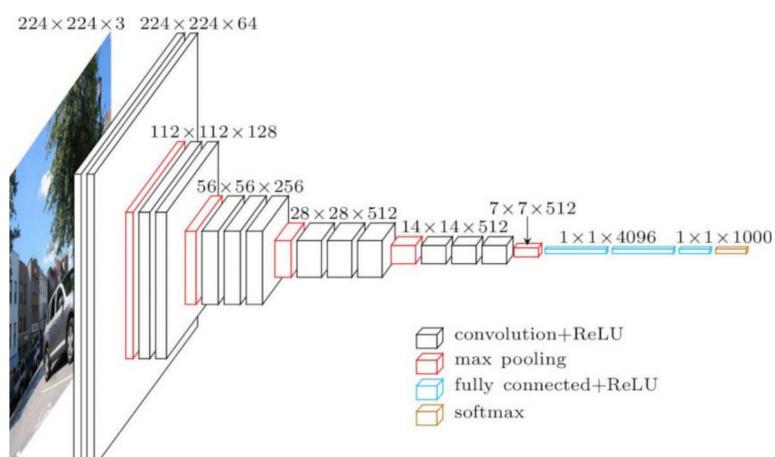
O reconhecimento de ações humanas não é um problema trivial. Existem diversos aspectos que devem ser levados em conta no momento de checar se uma imagem corresponde a uma certa ação humana. Exemplos de aspectos que devem ser medidos são o formato e a posição do corpo humano. Pensando nestes pontos, podemos fazer a seguinte pergunta: **é possível criar um modelo de Deep Learning para reconhecer certas ações humanas a partir de imagens?**

As redes neurais convolucionais são um dos pilares para a construção de modelos de Deep Learning para o reconhecimento de imagens. Para cada nova imagem de uma ação humana inserida na rede, ela “aprende” as características dessa ação, melhorando a capacidade de predição do modelo.



Forma de uma rede neural típica. Adaptado de: <https://www.ibm.com/cloud/learn/neural-networks>

Para o experimento, foi utilizado um *dataset* com 12600 imagens de treino, 5400 imagens de teste e 15 classes. O modelo de treino tem como base a rede neural convolucional VGG16, adaptada para o problema de 15 classes.



A arquitetura VGG16. Adaptado de: <https://neurohive.io/en/popular-networks/vgg16/>

O *dataset* utilizado possui algumas peculiaridades. As imagens de treino estavam rotuladas com apenas uma classe identificando-a. As imagens de teste originais do *dataset* não continham rótulos. Com a ajuda da ferramenta SLIL (<https://github.com/arturandre/SLIL>), foram rotuladas as 5400 imagens de teste, considerando o contexto no qual a ação é executada, e atribuindo todas as ações humanas contidas na imagem.



Um conjunto-exemplo de imagens de treino e de teste. Repare que as imagens de teste possuem mais de um rótulo atribuído.

Com o modelo treinado, foi executada uma avaliação sobre as imagens de teste, utilizando a métrica de acurácia para o cálculo de acertos, obtendo-se uma acurácia de 49%.



Dois exemplos de imagens avaliadas pelo modelo treinado.

Podemos concluir que o treinamento de imagens de ações humanas com Deep Learning é plausível. Além do mais, o reconhecimento de ações humanas é de fato um problema complexo, mas muito relevante para a atualidade.