

Relatório 17 - Vídeo: Como Estamos Ensinando os Computadores a Compreender Fotos (III)

Thassiana Camilia Amorim Muller

How we teach computers to understand pictures | Fei Fei Li

Introdução

A ideia é mostrar a diferença entre coletar dados de imagens e de fato entender esses dados, em outras palavras, é ensinar o computador a realmente enxergar significado naquilo que lhe é mostrado.

Vídeo

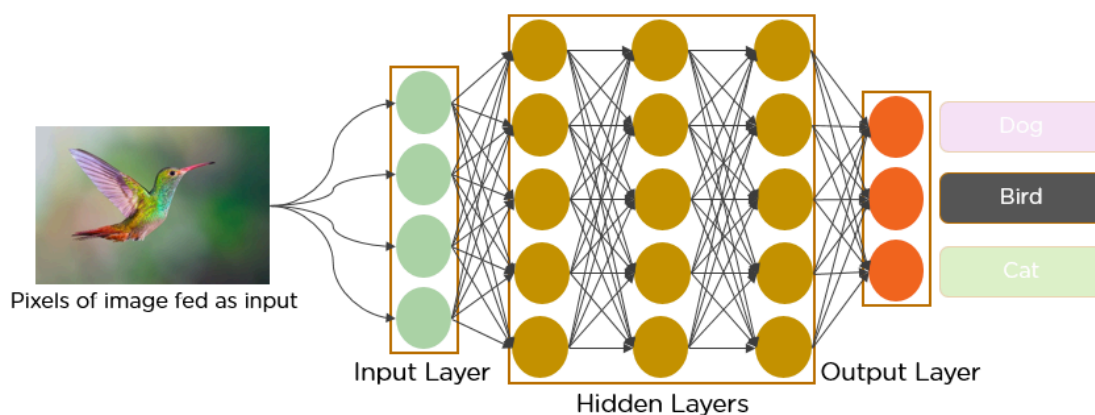
O vídeo introduz um comparativo de como uma criança de 3 anos já é capaz de descrever o conteúdo de fotos do mundo ao seu redor mostrando a importância de dar visão às máquinas, seja por questões de praticidade até questões de segurança. Se a visão inicia nos olhos ela ganha inteligência no cérebro, assim, a grande responsável por dar visão às máquinas não são somente as câmeras mas também as técnicas de machine learning aplicadas.

Dessa forma, podemos ensinar o computador o que é o mundo da mesma forma que ensinamos uma criança, mostrando a ele fotos e seus significados para que seja capaz de criar seus padrões, sua inteligência.

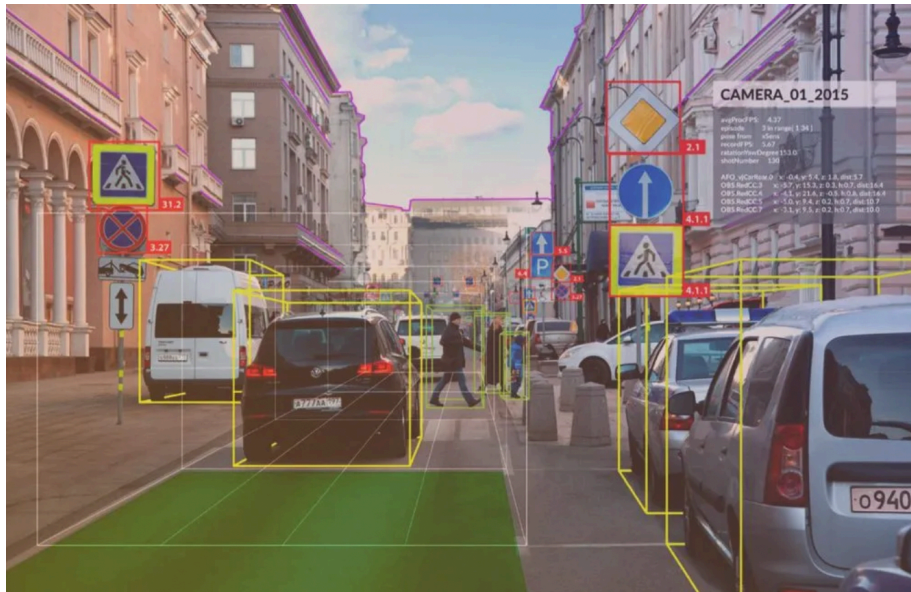
Para tornar esse processo real, é necessário que os humanos realizem a coleta, limpeza e organização de milhares de fotos relacionadas a um mesmo rótulo e exponham para que a máquina aprenda.

Em 2009 a equipe da Li reuniu em um projeto chamado “ImageNet” 15 milhões de fotos e as categorizou em 22 mil categorias e as disponibilizou para que todos pudessem ter acesso. Por exemplo, foram mais de 62 mil fotos coletadas de gatos das mais diferentes cores, raças, tamanhos e posições. Para que a máquina fosse capaz de aprender, foi utilizado a técnica de redes neurais convolucionais, o modelo treinado chegou a 24 milhões de nós, 140 milhões de parâmetros e 15 bilhões de conexões.

Exemplo de rede neural convolucional



Exemplo de como um computador pode interpretar o mundo

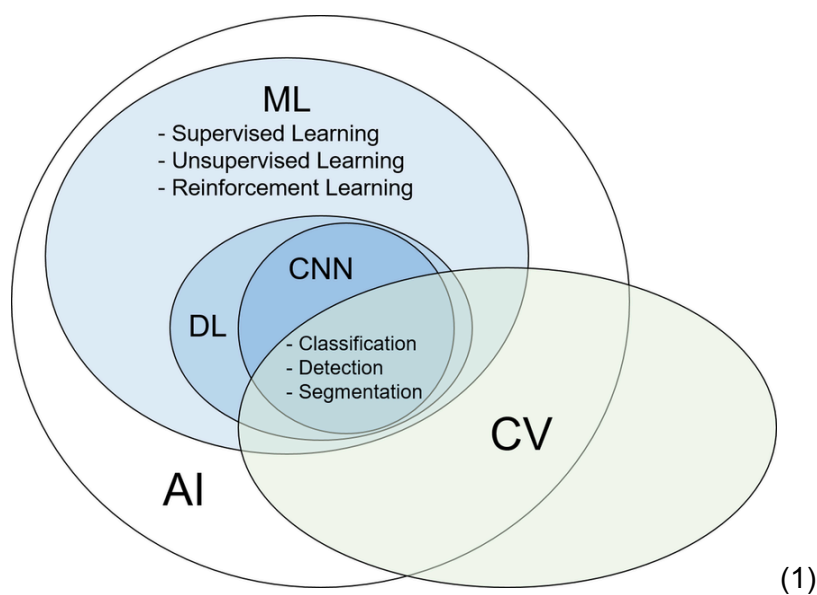


A palestra continua evidenciando a importância do Big Data para a visão computacional para que ela seja capaz de além de identificar apenas objetos isolados nas imagens poder entender o contexto dos objetos em uma foto. Li ainda demonstra de forma humorada alguns exemplos errados que podem ter vindo de um possível overfitting já que, parafraseando-a, “quando o modelo vê muitos gatos, ele acha que tudo é um gato”.

Ainda há a provocação de talvez um dia ensinar a máquina a enxergar não somente os objetos em seus contextos, mas também emoções, beleza e sentimentos.

Conclusão

É muito interessante observar como um trabalho que para muitos seria impossível 10 anos atrás se prova não só possível como também pioneiro em relação à visão computacional atual. Assim, cada bloco de inovações tecnológicas se junta para nos aproximarmos cada vez mais da verdadeira inteligência artificial.



Referências

- (1) Kitaguchi, Daichi & Takeshita, Nobuyoshi & Hasegawa, Hiro & Ito, Masaaki. (2021). Artificial intelligence-based computer vision in surgery: Recent advances and future perspectives. *Annals of Gastroenterological Surgery*. 6. 10.1002/ags3.12513.