

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

Willian Augusto Soder de Souza

O objetivo deste relatório é apresentar os principais conhecimentos adquiridos ao assistir ao vídeo “Fei-Fei Li: Como estamos ensinando os computadores a compreender fotos”, em que a cientista da computação Fei-Fei Li discute os desafios enfrentados em sua carreira e suas motivações no campo da visão computacional e da inteligência artificial.

No início de sua palestra, Fei-Fei Li mostra como uma criança de 3 anos descreve imagens e propõe a ideia de como um computador poderia desenvolver essa capacidade de identificar objetos em uma imagem, considerando que esta é formada por pixels (valores numéricos). Inicialmente, ela menciona como um gato em uma foto poderia ser visto por meio de formas geométricas. No entanto, isso apresenta um problema: se o gato estiver em uma posição diferente ou escondido atrás de um objeto, sua forma mudaria. Então, ela comenta que algo mudou sua perspectiva, que foi perceber que ninguém ensina um ser humano a enxergar; ele aprende naturalmente ao ver aquele objeto milhões de vezes. Ou seja, para um computador aprender a identificar um objeto, o caminho mais eficaz seria expô-lo a milhões de imagens para que ele compreenda o que caracteriza um gato ou um cachorro. Em 2007, foi lançado o ImageNet, um banco de dados com 15 milhões de imagens de 22 mil categorias. Com esse banco de dados e o auxílio de uma rede neural convolucional, foi possível criar algoritmos capazes de identificar diversos objetos em fotos, mesmo em diferentes poses.

Fei-Fei Li também introduz outro desafio na visão computacional: a descrição de fotos utilizando linguagem natural, de forma semelhante ao que um ser humano faria. Além de identificar os objetos na foto, o computador precisaria formular frases que descrevam o que foi identificado. Fei-Fei Li menciona que conseguiu desenvolver um algoritmo capaz de realizar essa tarefa. Entretanto, surgem dois problemas: primeiro, o computador não é capaz de expressar a emoção humana ao descrever uma foto ou momento; segundo, ao se deparar com um objeto desconhecido, o computador tende a confundi-lo com outros objetos que ele já conhece.

Ela encerra a palestra ressaltando a importância de investir no campo da visão computacional e da inteligência artificial, pois essas áreas podem ser úteis em diversos setores, como na medicina, em carros autônomos, na busca por pessoas em desastres, ou até mesmo na exploração de áreas e locais inacessíveis para o ser humano.

CONCLUSÃO:

Após assistir a esse vídeo, é possível concluir que o campo da visão computacional já revolucionou o mundo e ainda vai revolucionar mais. Essa tecnologia pode ser aplicada em tarefas cotidianas, como identificar uma criança em risco dentro de casa e gerar um aviso para os pais, ou proporcionar conforto, como é o caso dos carros autônomos. Além disso, a visão computacional pode ser utilizada em áreas mais complexas, como na medicina, em diagnósticos por imagem para detecção precoce de doenças, ou na realização de cirurgias assistidas por robôs. Com o avanço desse campo, a humanidade passará por uma verdadeira revolução, pois, com o desenvolvimento de tecnologias como a visão computacional e a inteligência artificial, será possível salvar vidas e resolver problemas cotidianos de maneira eficiente e assertiva.