

O vídeo apresentado no módulo 17 retrata a computação visual e suas contribuições para a sociedade, iniciando com um exemplo de uma criança reconhecendo imagens, comparando com o período de treinamento de um modelo de aprendizado de máquina.

Em seguida a apresentadora discorre sobre as dificuldades de se treinar um modelo para reconhecer objetos, mostrando como um objeto não pode ser padronizado tão facilmente pois ele pode estar exposto em diferentes posições/ângulos, dificultando a interpretação do mesmo, porém para remediar parte deste problema, foi criado o imagenet, database com 15 milhões de imagens com 22 mil categorias diferentes, a fim de treinar modelos de computação visual com maior precisão.

É apresentado uma breve introdução à CNNs, pois este TED talk não é direcionado somente a pesquisadores da área, expondo um modelo que utiliza o imagenet, sendo um modelo razoavelmente grande, possuindo 24 milhões de nodos.

Por fim é exposto um projeto de computação visual, um modelo que recebe uma imagem e a descreve em linguagem natural, o mesmo possui uma precisão adequada, porém ainda comete erros, seja por receber um objeto não conhecido pelo mesmo ou por não conseguir diferenciar construtos dos objetos reais.