Detecção de Objetos Capturados por Drone

Augusto Berndt, Vinícius Zanandrea

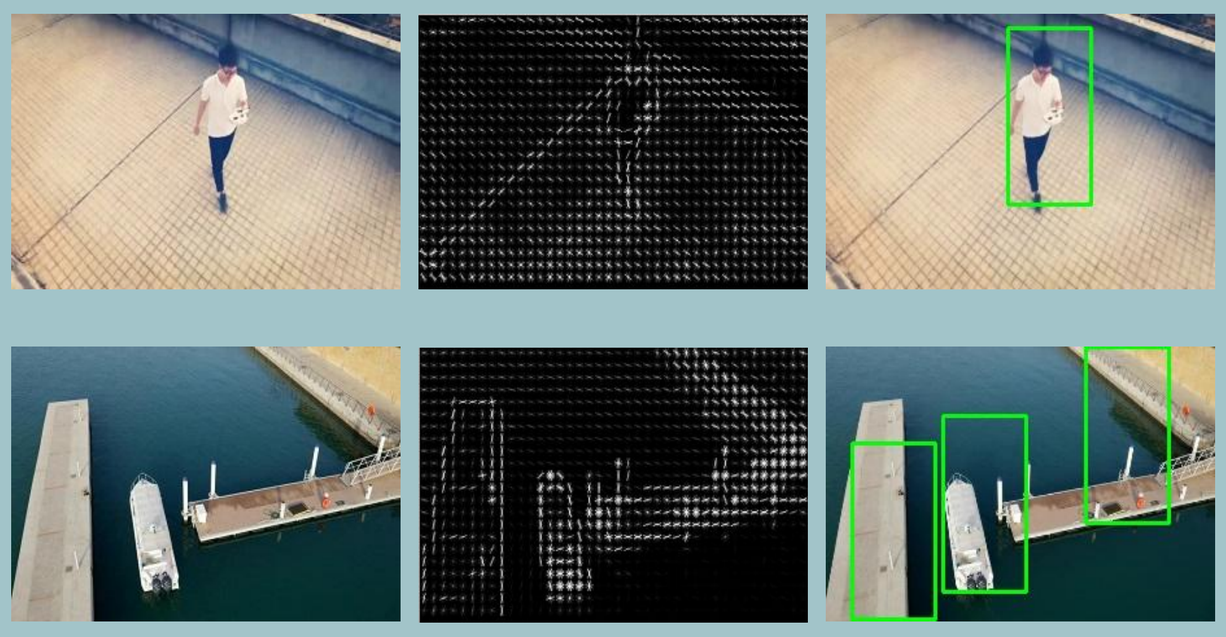
O Desafio: Realizar a detecção de diversos objetos filmados por drones. O desafio é proposto pelo congresso *Design Automation Conference* (DAC) que acontece anualmente [1]. O dataset de treinamento é disponibilizado pela empresa financiadora e contém filmagens de objetos em cenas com imagens em sequência. O modelo de solução do problema ainda deve ser embarcado em uma placa FPGA Ultra 96 v2, disponibilizada pelos organizadores.
O consumo de energia é crucial em um drone e uma detecção com baixo consumo energético é o principal objetivo. A agilidade do modelo em realizar o processo também é importante, refletida no processamento de frames por segundo (FPS). Devido a estes objetivos utilizamos a rede neural YOLO que realiza um método simples e rápido.

Introdução

O dataset [2] para treinamento possui um conjunto de imagens com tamanho total de 6GB (93520 arquivos). Existem diferentes classes de objetos para detecção: “car”, “building”, “person”, “boat”, “riding”, “wakeboard”, “drone”, “whale”, “paraglider”, “truck”. As imagens são disponibilizadas em formato JPG e os labels em formato XML, apresentando o bounding box do objeto a ser detectado.

Abordagem Clássica

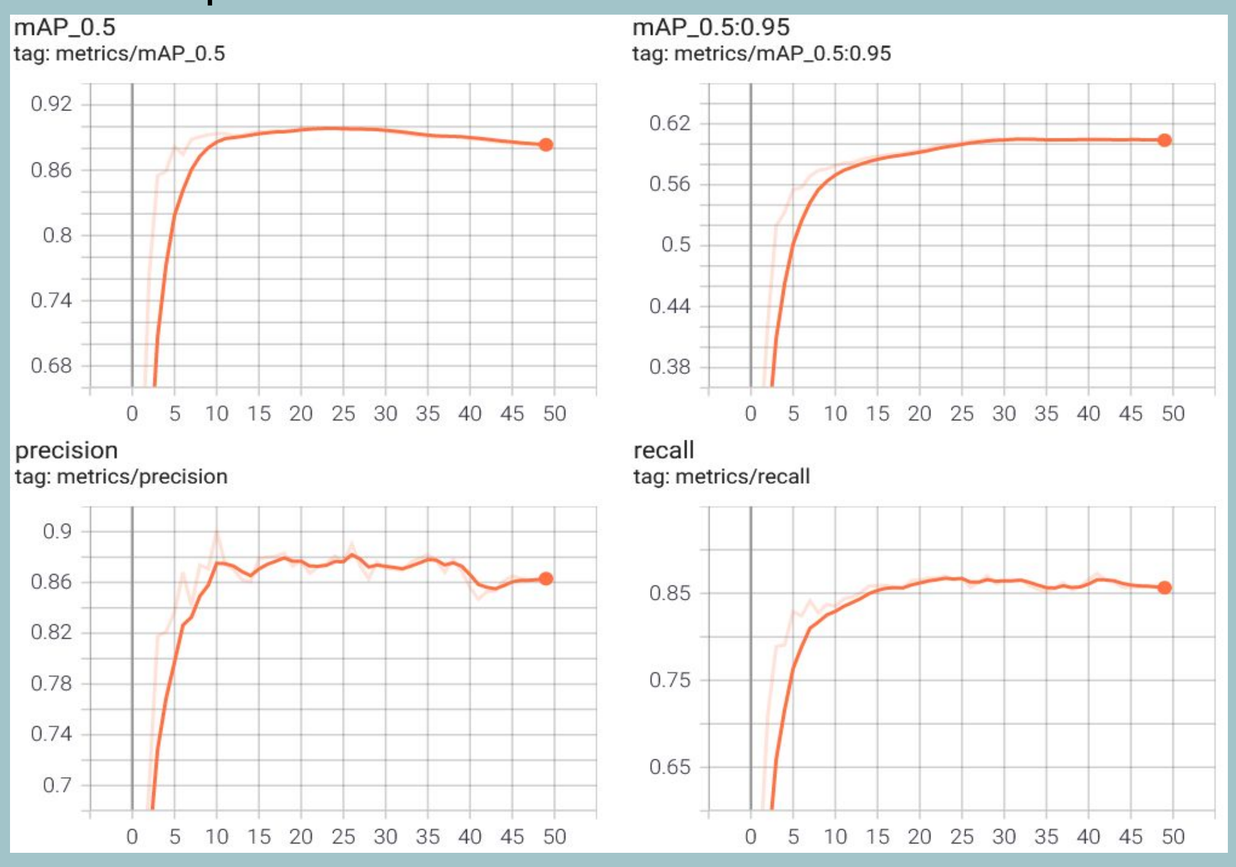
Inicialmente, verificou-se a utilização de métodos clássicos de visão computacional para detecção das imagens. Em específico, utilizou-se Histogram of Oriented Gradients (HOG) para extrair as features e Support Vector Machines (SVM) como classificador para treinar o modelo a reconhecer os objetos. A técnica de Non Maximum Suppression (NMS) foi aplicada para selecionar o melhor bounding box. Observamos que o modelo implementado consegue identificar relativamente bem o objeto a ser detectado em determinadas imagens, como mostrado para a classe “person”. Entretanto, para classes como “boat”, a detecção não apresenta uma acurácia aceitável. Uma possível explicação para este comportamento pode ser pelo fato do dataset conter mais imagens de determinada classe, o que facilita na extração de features.



Abordagem com Deep Learning

Seguindo a abordagem de Deep Learning, Redes Neurais Convolucionais (CNNs) foram utilizadas para realizar a detecção de imagens do desafio. A rede YOLOv5 (you only look once version 5) [3] foi utilizada para efetuar o reconhecimento. Esta rede é o estado da arte no reconhecimento de imagens com baixo consumo.

O modelo yolov5 obteve uma média de **0.878 IOU** e **41 fps** para todas as imagens da competição, apresentando ótimos resultados de detecção. Este fps foi obtido em uma NVIDIA Geforce GTX 1060. As figuras a seguir demonstram o *mean Average Precision,* para thresholds com confiança de 0.5 e 0.5:9.5, seguido dos valores de precision e recall. Todos resultados apresentados são de um treinamento com 50 épocas e uma divisão do dataset em 80-10-10% para treinamento, validação e teste respectivamente.



Conclusões

A utilização de uma rede neural convolucional trouxe um resultado significativamente melhor quando comparada com a abordagem clássica. O treinamento é um processo que exige tempo e a quantidade de épocas utilizadas demonstram que ainda há possibilidade de aprendizado.
Esta solução para a competição ainda está sob desenvolvimento e seu código será divulgado após sua submissão.

Referências
[1] DAC (Design Automation Conference) de 2021. Disponível em: https://dac-sdc-2021.groups.et.byu.net/doku.php
[2] Dataset disponível em: https://byu.app.box.com/s/hdgztcu12j7fij397jmd68h4og6ln1jw
[3] Interface YOLOv5. Disponível em: https://github.com/ultralytics/yolov5