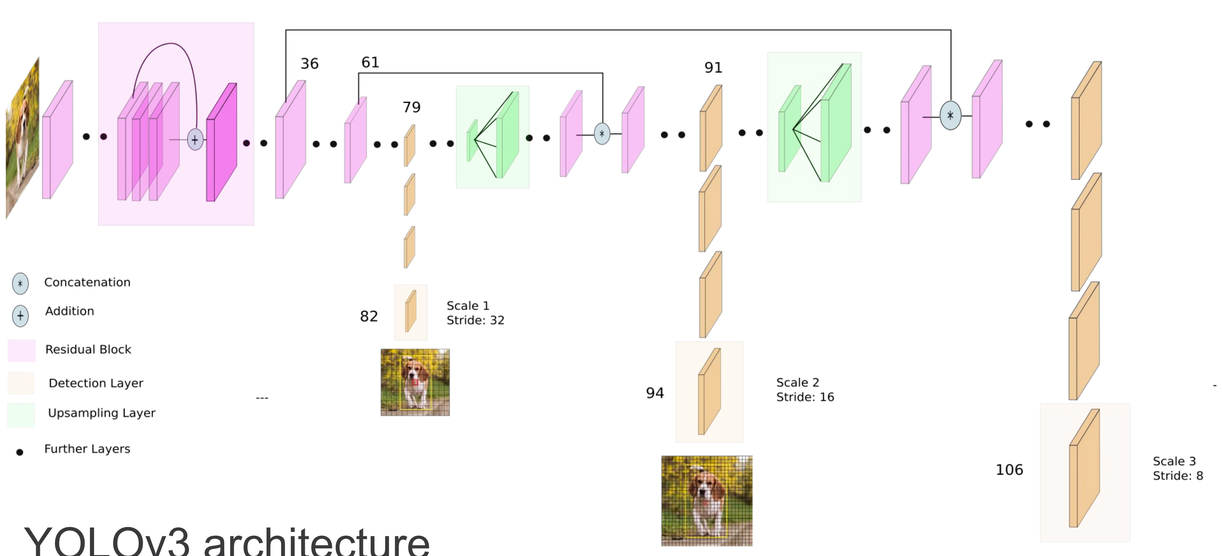
StyriaAI Team
Open Images Object Detection Task

Bartol Freškura, Leon Luttenberger, Tome Radman, Antonio Šajatović

Overview

Google AI has publicly released the Open Images dataset, which the Open Images Challenge is based on. The training set is the largest of its kind, with more varied and complex bounding-box annotations spanning 500 classes. Our goal was to build the best performing algorithm for automatic object detection, pushing the field of machine vision even further.



YOLOv3 architecture

Courtesy of Ayoosh Kathuria:
https://towardsdatascience.com/yolo-v3-object-detection-53fb7d3bfe6b

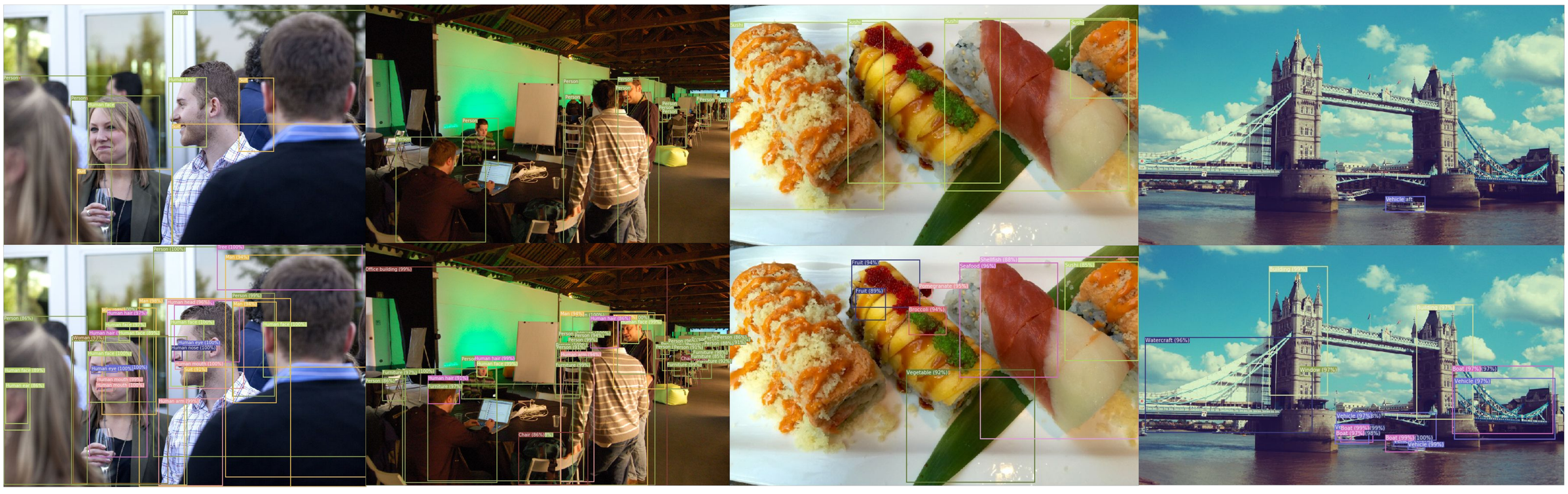
Approach

Our approach used the YOLOv3 [1] paper as a reference, which we slightly modified to fit our needs. We trained an ensemble of models with the pre-trained ResNet50 [2], ResNet18 [2], InceptionV3 [3], and DenseNet161 [4] backbones on the ImageNet [5] dataset.

The model input size was fixed to 465x465 which we found was an ideal balance of the training speed and accuracy due to our hardware limitations.

We also concatenated the X and Y coordinate grid with the input image for one of our models. This approach was inspired by CoordConv [6].

Class imbalance had a severe negative impact on the model accuracy and train time which was remedied by oversampling and undersampling the minority and majority classes accordingly. Images were also randomly augmented using horizontal and vertical flips, saturation, brightness, contrast, and scaling.



Ground truths (upper row) and our predictions (bottom row). Most of the ground truth labels are inconsistent as can be seen in the examples above. The image annotations are frequently missing objects which our model successfully detects. Human related labels are quite common in the dataset resulting in near-perfect results on the images alike. However, the model is still limited in detecting very small objects and infrequent objects such as particular types of food.

Results

Public leaderboard score: 0.20743 mAP

Public ranking: 74 / 440

Literature

[1] - Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement."

[2] - He, Kaiming, et al. "Deep residual learning for image recognition."

[3] - Szegedy, Christian, et al. "Rethinking the inception architecture for computer vision.

[4] - Iandola, Forrest, et al. "Densenet: Implementing efficient convnet descriptor pyramids."

[5] - Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database."

[6] - Novotny, David, et al. "Semi-convolutional Operators for Instance Segmentation."