Object Detection in Videos with Modernized SSD

Jiajun Zhang, Shawn Pachgade, Aditya Huddedar

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science

Data, Models, and Code: github.com/shpach/ssd_keras
Video on YouTube: https://www.youtube.com/watch?v=kzWk6Gs8vw0

Modernized SSD Architecture

We build a Modernized Single Shot Object Detection model and then apply it to a video. We make two noteworthy contributions: 1) We build the SSD model based on VGG19 network from scratch and optimize it by adding two additional bounding boxes in each layer. 2) In the video application, we use Kernel Correlation Filters between the frames to smoothen the output, avoid missing and overlap detections, etc. In this poster, we present the main idea behind Modernized SSD Architecture and the results on both the single images and videos.

Bounding Boxes

Loss Calculation

Modernized SSD Architecture
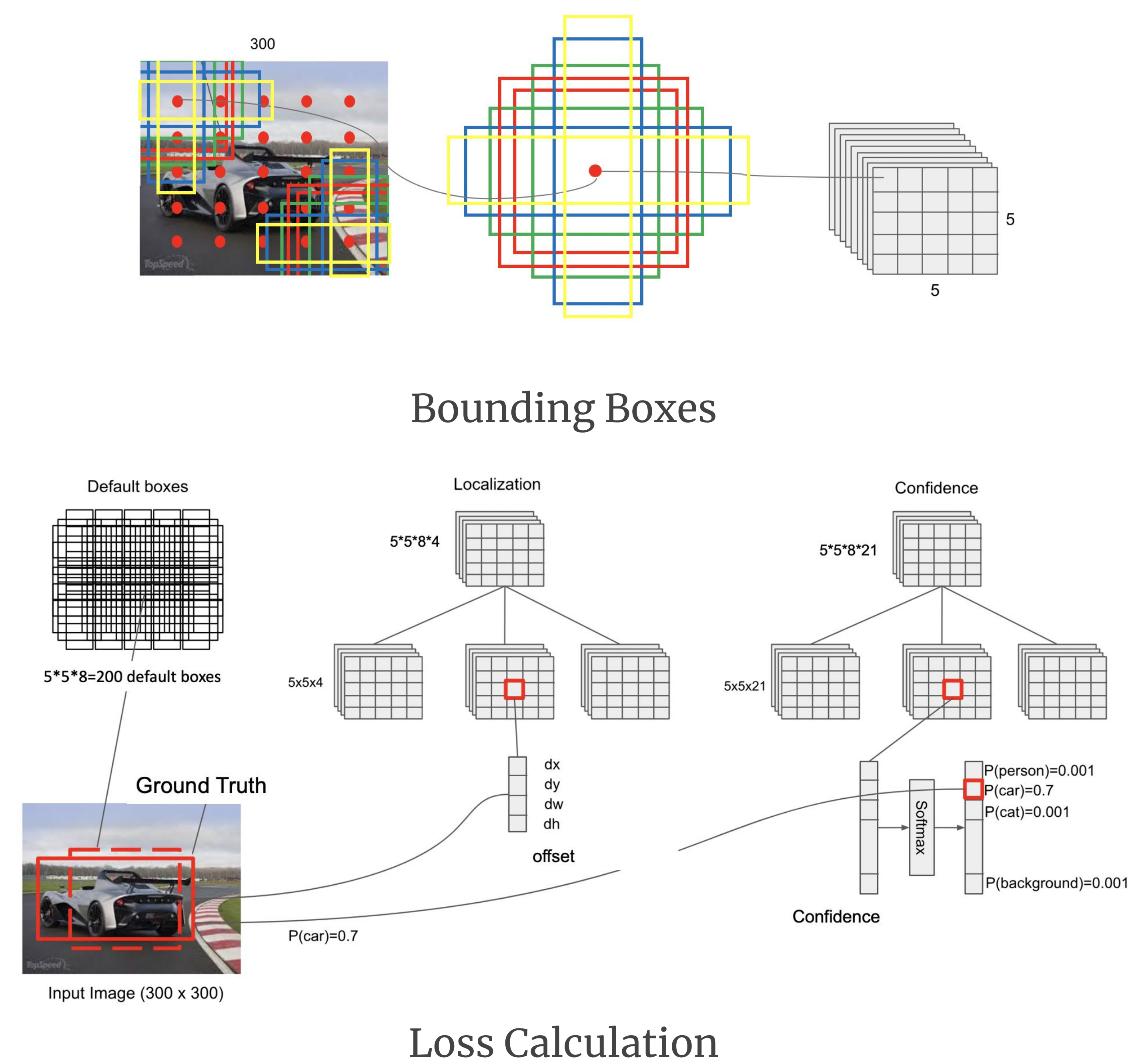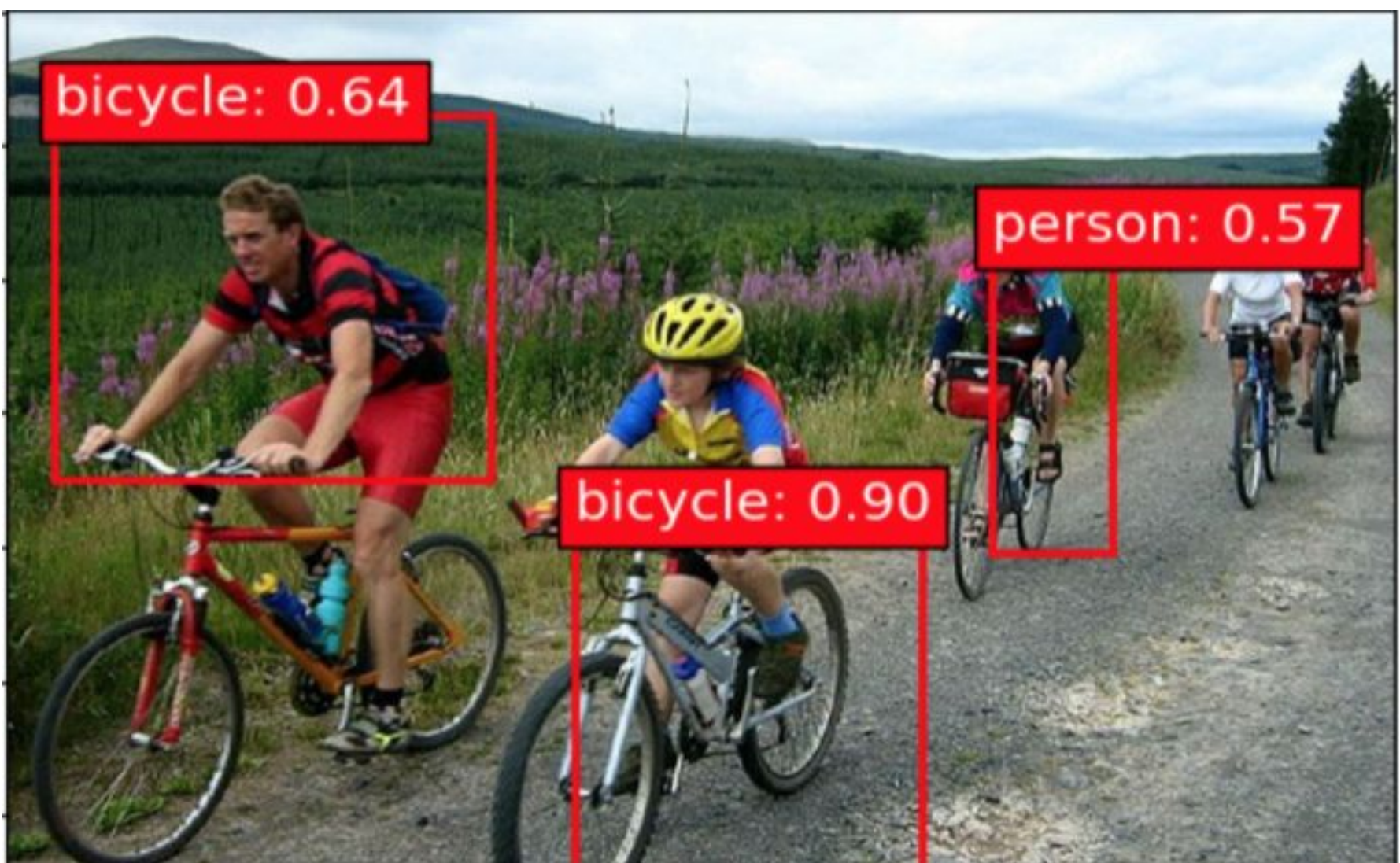
Results

Sample Detection Result based on our SSD model
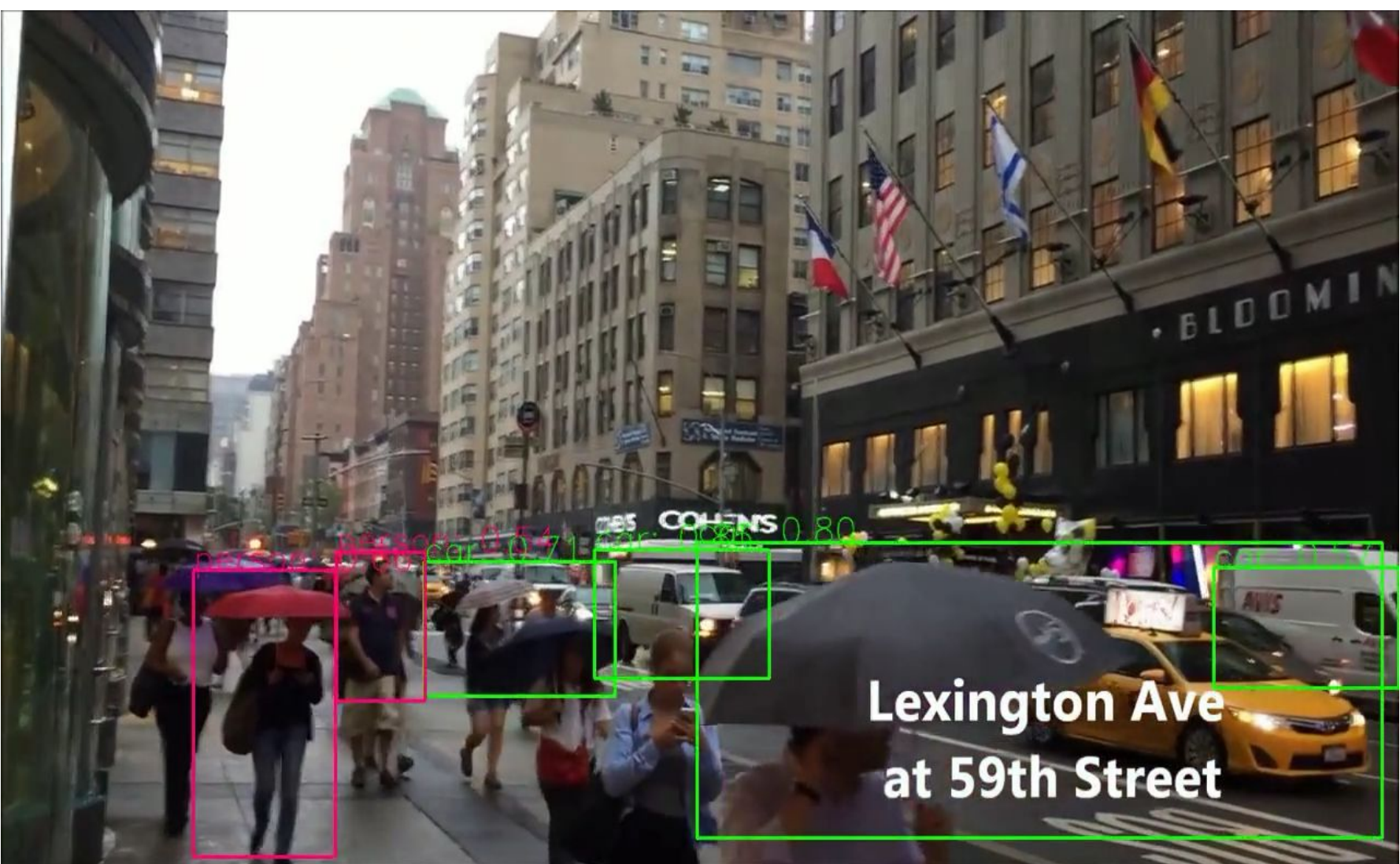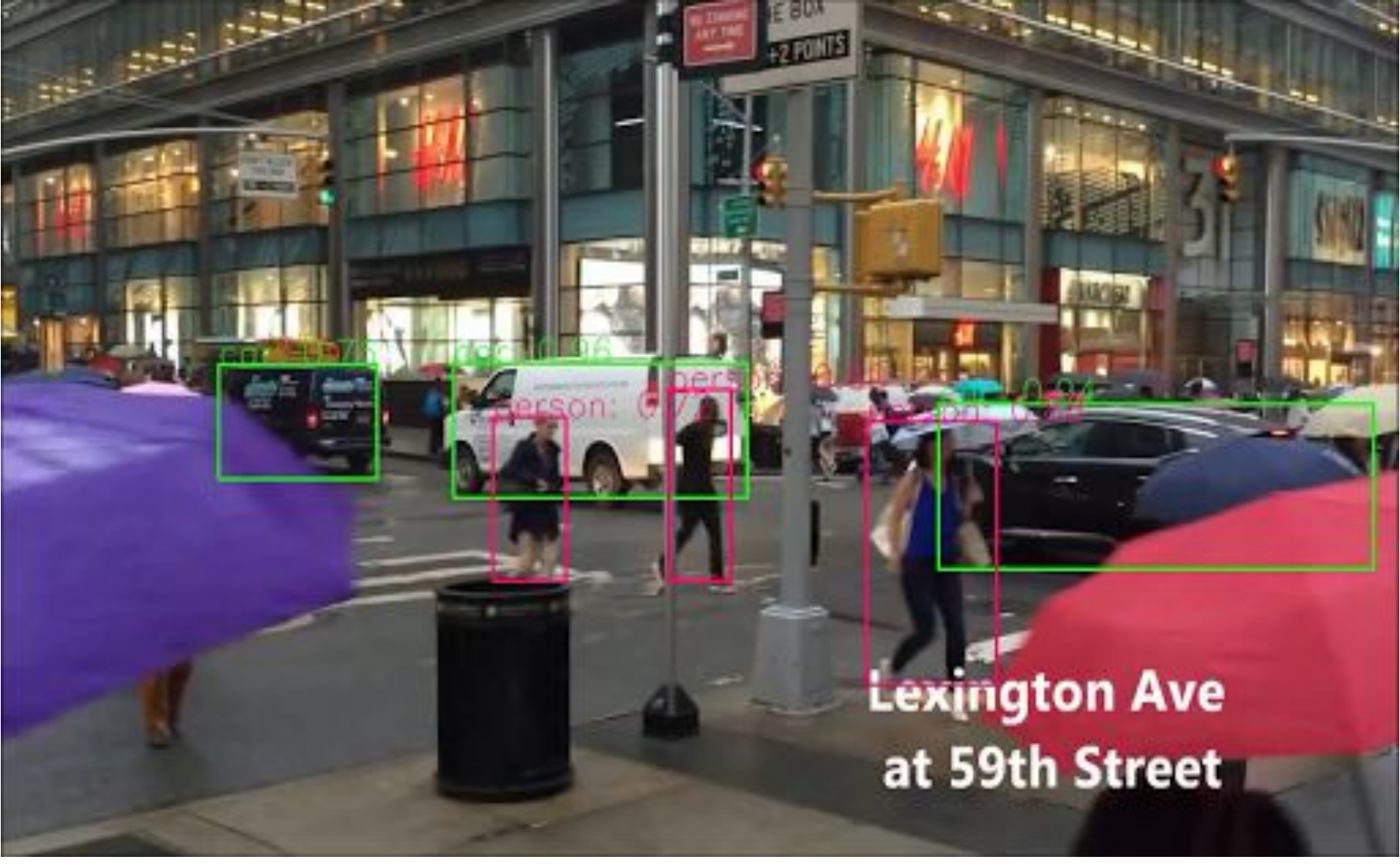
Screenshot of our SSD model with KCF applied on video

	Train Loss	Val Loss
Original SSD300	3.9	4.6
VGG19 SSD300	5.13	5.09
InceptionResNetV2 SSD300	5.07	83.36
VGG19 + more default boxes SSD300	4.72	4.76

Training Result Comparison

	Mean Average Precision
Original SSD300	0.738
VGG19 + more default boxes SSD300	0.52

mAP Result Comparison