İTÜ

BLG561E Deep Learning
2010-2020 Final Project and Competition

Project Objectives and Data

The class project consists of developing/training a deep neural network for object detection in urban driving scenarios. You and/or your team will develop a network that takes a continuous video stream as the input and detects driving-related objects in each frame of the video, along with the bounding boxes of each object. The objective is to build a high performance detector that can run using a reasonable framerate, hence both detection accuracy and size of your network will be important in evaluating the performance of your network (See Project Instructions for details on scoring).

The object classes are

0. Person
1. Bicycle
2. Car
3. Motorcycle
4. Truck
5. Bus

Figures 1 - 5 show a sample frames from the test data, where each object is highlighted with the bounding boxes:

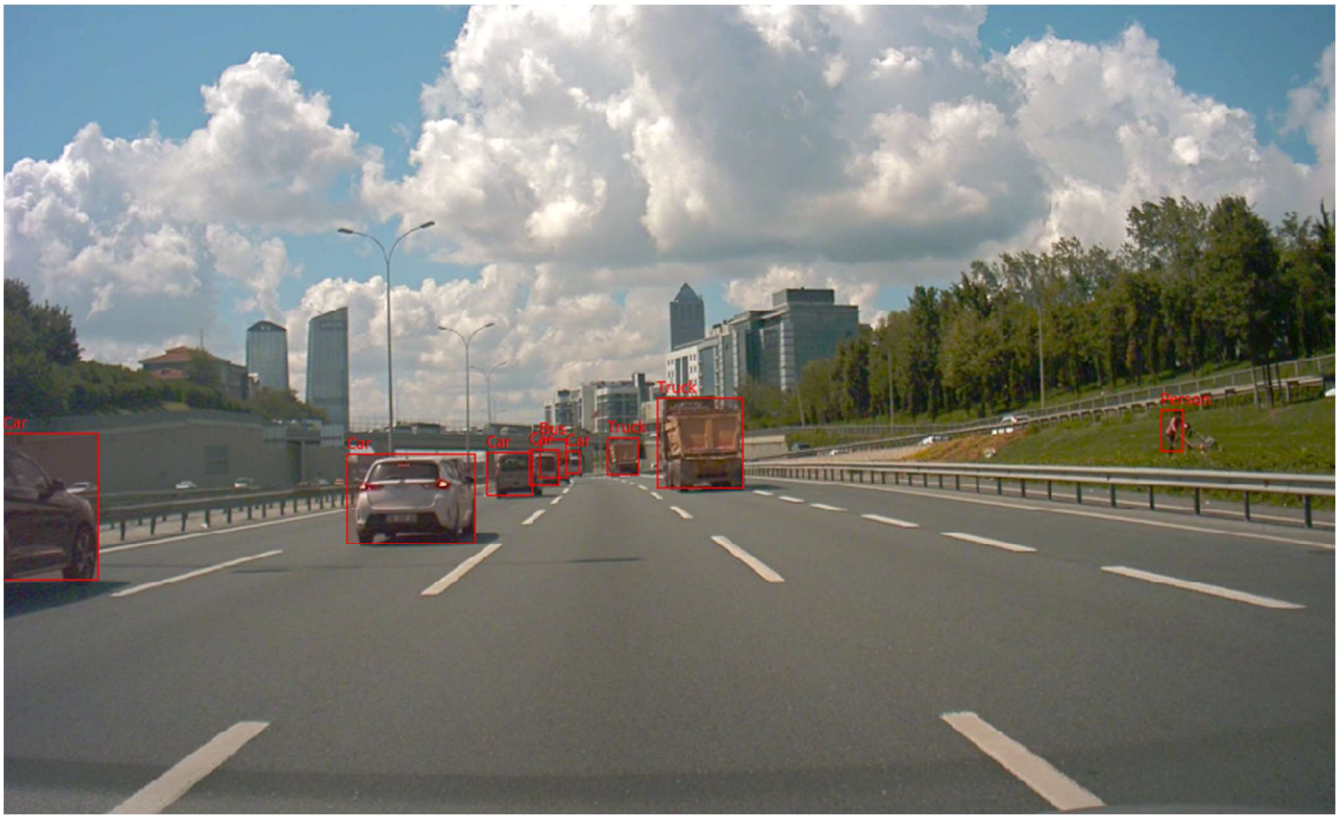

Figure 1: Sample Frame from Test Data

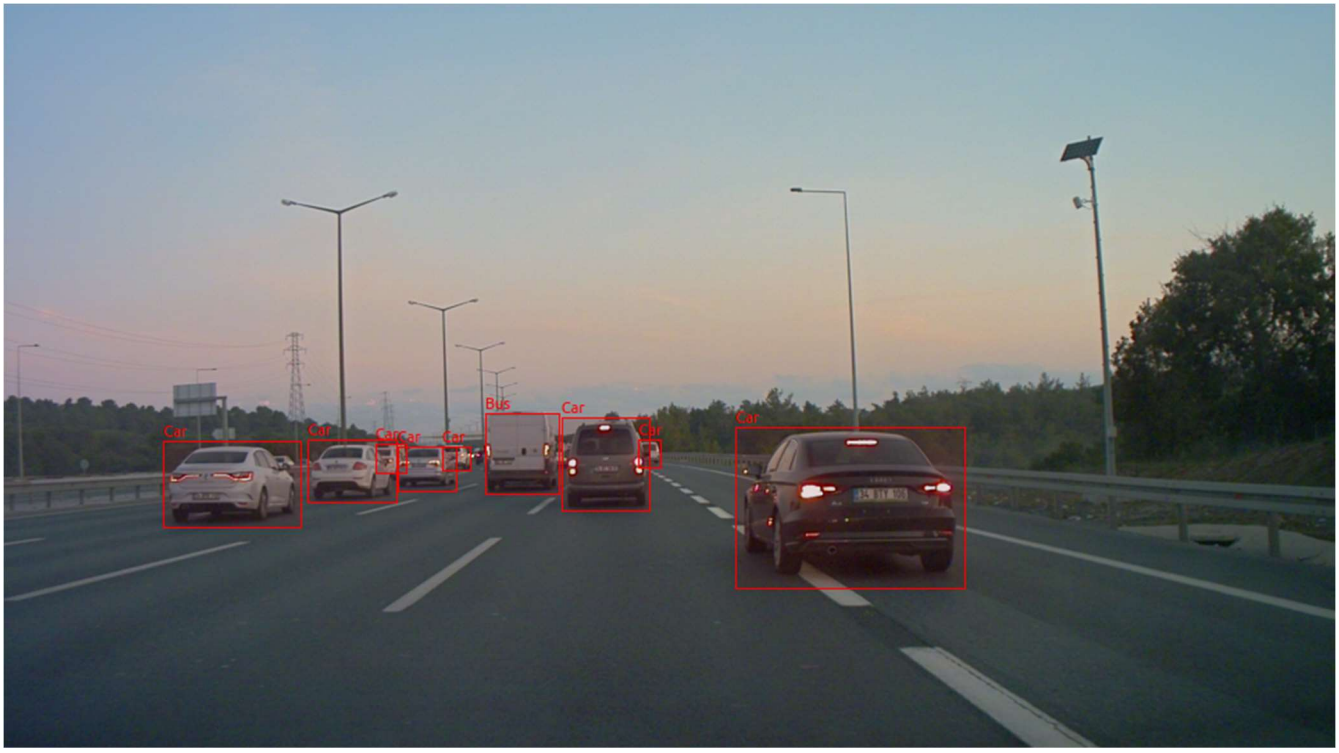

Figure 2: Sample Frame from Test Data

Figure 3: Sample Frame from Test Data

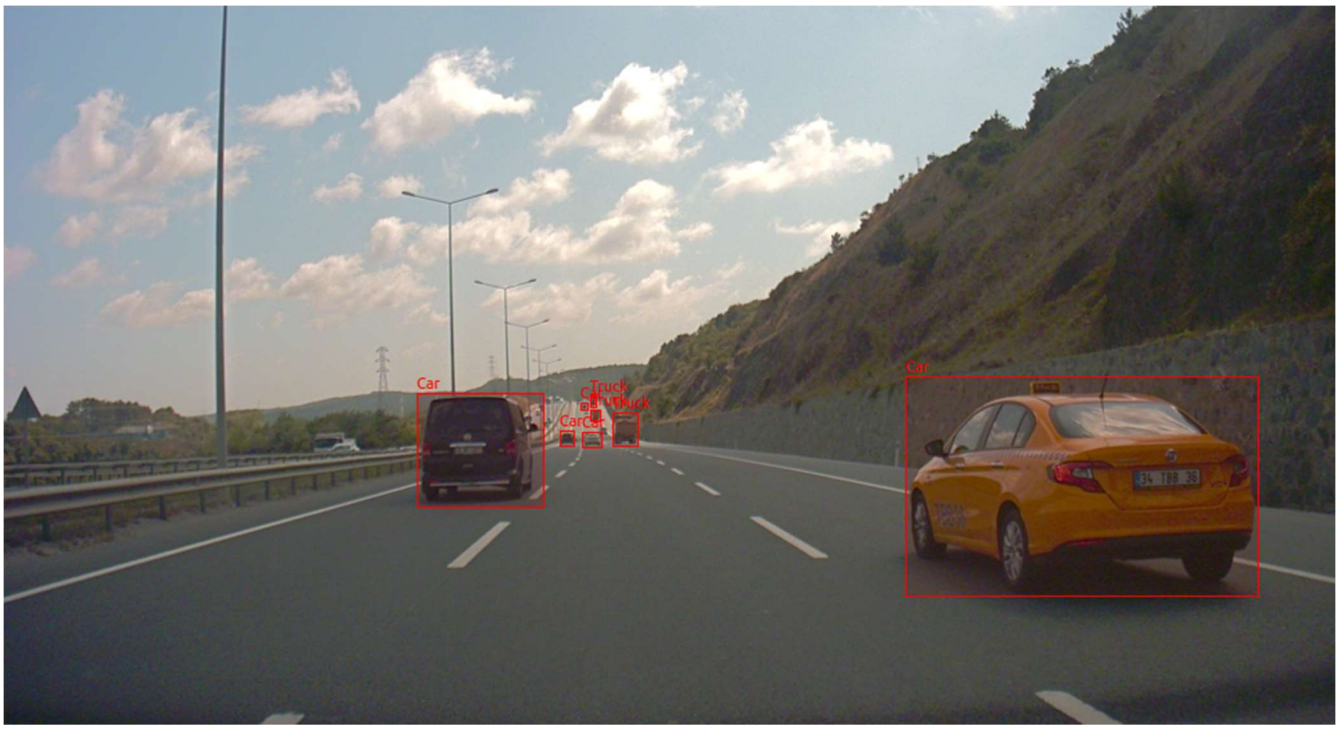

Figure 4: Sample Frame from Test Data

Figure5: Sample Frame from Test Data

Figure 6 shows a sample annotation file in XML format.

```
<?xml version="1.0" ?>
<frame000000>
   <object_1>
      <ID>2</ID>
      <bndbox>
         <x>1</x>
         <y>329</y>
         <w>207</w>
         <h>148</h>
      </bndbox>
   </object_1>
   <object_2>
      <ID>2</ID>
      <bndbox>
         <x>368</x>
         <y>342</y>
         <w>32</w>
         <h>23</h>
      </bndbox>
   </object_2>
   <object_3>
      <ID>2</ID>
      <bndbox>
         <x>430</x>
         <y>339</y>
         <w>19</w>
         <h>19</h>
      </bndbox>
   </object_3>
</frame000000>
```

Figure 6: Sample annotation File

The training data can be downloaded from the link below. The samples (~15000 frames and corresponding annotations) are collected from various driving scenes at Istanbul. The test data (which will be unavailable to the students and will be used by instructors for scoring your network) is also collected at Istanbul using a similar setup.

Link for training data (~10 GB):
https://drive.google.com/drive/folders/18NSm2OmZ0N2xc2phWkV_0TE15BHmBIKE

Note that by downloading this data you agree that it can only be used for personal/academic research. Any commercial use for any commercial gain is prohibited.

It is recommended to download the data as soon as possible since it might be removed in upcoming days.

Project Instructions

- First step is forming your team. Each team can consist of up to 4 students. Forming larger teams is encouraged. However, note that if your team wins the competition then the prize money will be split among the team members (please see the Competition section for details)

- After forming your team, you can start developing and training your object detection networks. You are free to use any public training data in addition to the data we provided you. Here are some recommended datasets:

http://cocodataset.org/#download
http://host.robots.ox.ac.uk/pascal/VOC/voc2012/
http://apolloscape.auto/scene.html
https://bdd-data.berkeley.edu/
https://www.nuscenes.org/

- You are allowed to use pre-trained networks and perform transfer learning as a part of your model. **However**, if you are using a pre-existing architecture (such as YOLO) you need to perform at least one non-trivial modification (such as adding a new layer/operation, modifying the loss function in a novel way, fuse object detection with Kalman Based filtering etc.). If you just take an existing network and re-train it on the data we provided **your project will be disqualified**. If you are not sure what constitutes as a non-trivial modification, consult us and/or TAs before your submission.

- The official programming language for the project is PyTorch, please do not use any other framework. Your code will be tested on a PyTorch script.

- You need to submit your code and a technical report (max. 10 pages long) by the deadline **7th of January 2020. (deadline might change).** We will score the technical report based on the novelty of your approach and the clarity of your writing.

- After the deadline, instructors will score the performance of your network on the test data based on the following formula:

$$score(network) = 0.5 * accuracy + 0.5 * speed$$

The accuracy will be measured by computing MAP (mean average precision) formula of your network on the unreleased test data. Details on how MAP is calculated can be found in the following link:

https://medium.com/@jonathan_hui/map-mean-average-precision-for-object-detection-45c121a31173

Accuracy score of all networks will be normalized according to the best performing network. Hence the network with the best performance will get score 1.0 and the rest of the networks will be scored accordingly.

The speed of the network will be calculated by computing the FPS (frame per second) performance of the network on a GPU (GeForce GTX 1080 Ti with Intel® Core i9-7980XE Extreme Edition CPU). Similar to the accuracy score, speed will be normalized according to the best network.

- An average of your network score and technical report score will determine your project score. A number of selected teams with the top performance will be invited to the competition track. The details of the competition are given below.

Competition Info (Sponsored by EATRON Technologies)

- After the evaluation of the network, teams ranked among the Top 8 will be invited to present their work to a jury. The competition will take place at ITU Magnet **around 10th of January (Exact date TBA).** BLG 561E professors, technical staff from Eatron and a number of AI experts from academia and industry will be present during the competition.

- Each team will give a short presentation (2-3 minutes) on how they used data, how they developed the network structure and how they trained the network.

- After each presentation, the team's algorithm will be ran online on a test stream, where the audience will see the performance of the object detection algorithm in real time. The score of the team will also be displayed in a leaderboard, which will be updated after every presentation.

- After the presentations and demonstrations end, Top 3 teams in the leaderboard will be announced on the site along with the following prizes:

    - Best Team: 1500 Turkish Lira
    - Second Team: 1000 Turkish Lira
    - Third Team: 500 Turkish Lira

- The money prize will be divided equally among team members

- Note that entering or winning the competition has no impact on your letter grade.

Important Note: Project regulations are subject to change during the term (such as the details on score and deadline)