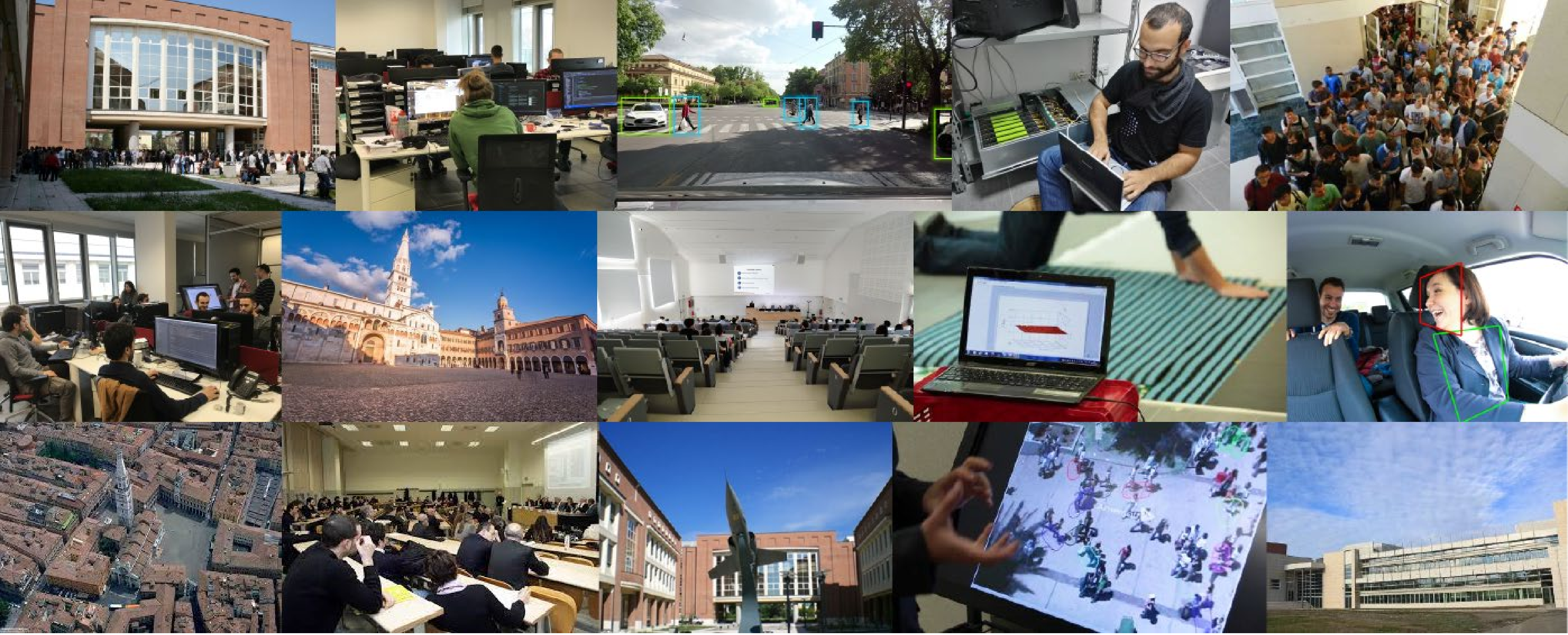


(QUASI) MODEL REPRO: THE DREYEVE PROJECT

Davide Abati

(davide.abati@unimore.it)

Department of Engineering, University of Modena and Reggio Emilia, Italy

AImage Lab

UNIMORE UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

TODAY'S AGENDA

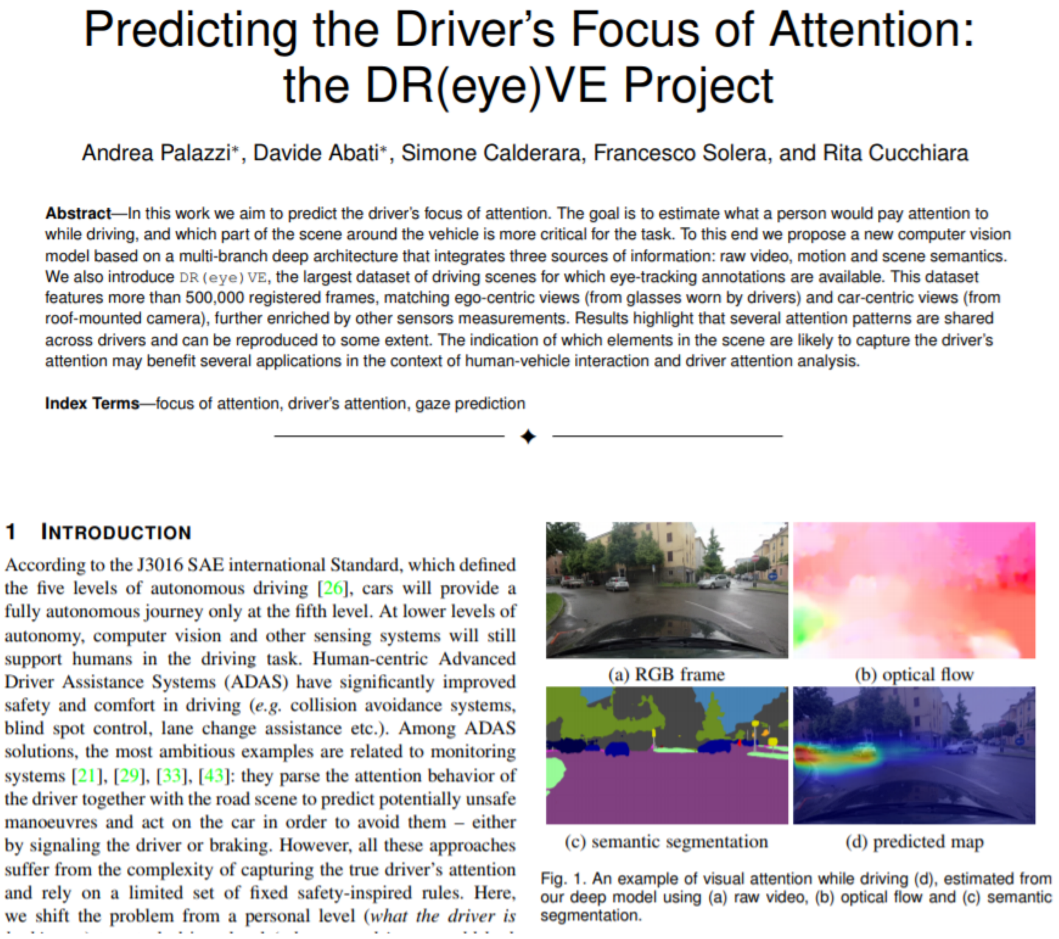
The goal of today's lab is to implement a network to predict the human attention while driving.

- In Pytorch
- We will refer to this paper

Why?

- Get your hands dirty
- Could be useful for the final project

But first…



Predicting the Driver's Focus of Attention: the DR(eye)VE Project

Andrea Palazzi*, Davide Abati*, Simone Calderara, Francesco Solera, and Rita Cucchiara

Abstract—In this work we aim to predict the driver's focus of attention. The goal is to estimate what a person would pay attention to while driving, and which part of the scene around the vehicle is more critical for the task. To this end we propose a new computer vision model based on a multi-branch deep architecture that integrates three sources of information: raw video, motion and scene semantics. We also introduce DR(eye)VE, the largest dataset of driving scenes for which eye-tracking annotations are available. This dataset features more than 500,000 registered frames, matching ego-centric views (from glasses worn by drivers) and car-centric views (from roof-mounted camera), further enriched by other sensors measurements. Results highlight that several attention patterns are shared across drivers and can be reproduced to some extent. The indication of which elements in the scene are likely to capture the driver's attention may benefit several applications in the context of human-vehicle interaction and driver attention analysis.

Index Terms—focus of attention, driver's attention, gaze prediction

1 INTRODUCTION

According to the J3016 SAE international Standard, which defined the five levels of autonomous driving [26], cars will provide a fully autonomous journey only at the fifth level. At lower levels of autonomy, computer vision and other sensing systems will still support humans in the driving task. Human-centric Advanced Driver Assistance Systems (ADAS) have significantly improved safety and comfort in driving (*e.g.* collision avoidance systems, blind spot control, lane change assistance etc.). Among ADAS solutions, the most ambitious examples are related to monitoring systems [21], [29], [33], [43]: they parse the attention behavior of the driver together with the road scene to predict potentially unsafe manoeuvres and act on the car in order to avoid them – either by signaling the driver or braking. However, all these approaches suffer from the complexity of capturing the true driver's attention and rely on a limited set of fixed safety-inspired rules. Here, we shift the problem from a personal level (*what the driver is*

(a) RGB frame (b) optical flow

(c) semantic segmentation (d) predicted map

Fig. 1. An example of visual attention while driving (d), estimated from our deep model using (a) raw video, (b) optical flow and (c) semantic segmentation.

LET'S SET UP THE ENVIRONMENT

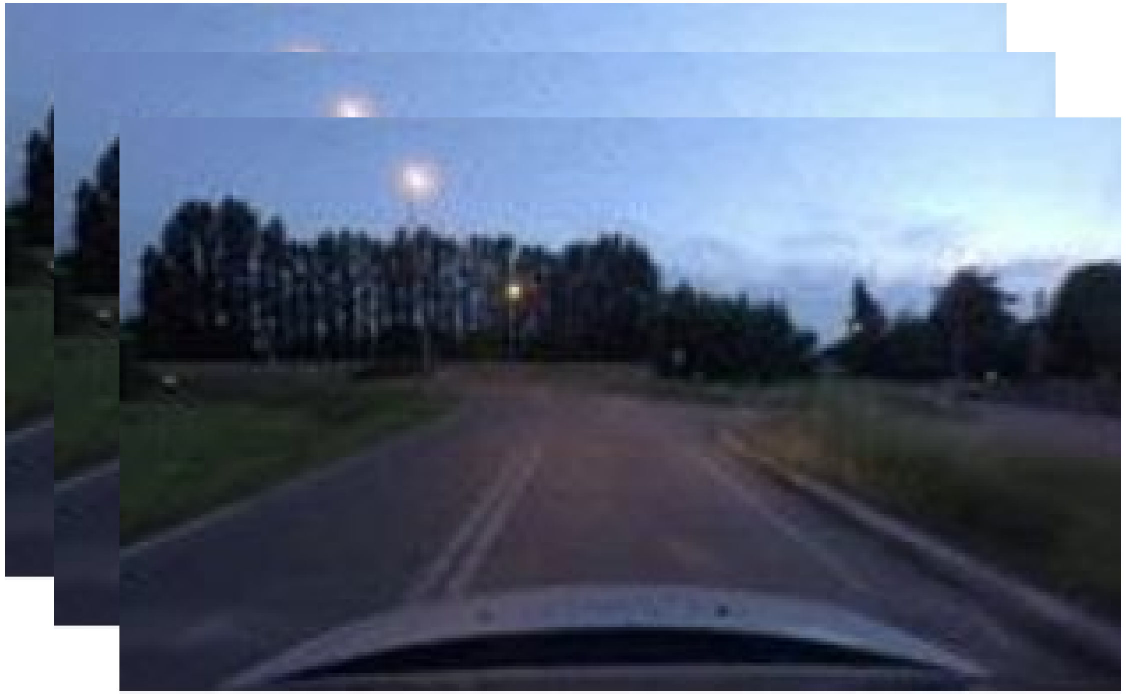
- Clone the github repository
- Set up Pycharm with the remote interpreter
- Open a ssh session to aimagelab-srv-00, with X server forwarding.

DATA COLLECTION: ACQUISITION RIG

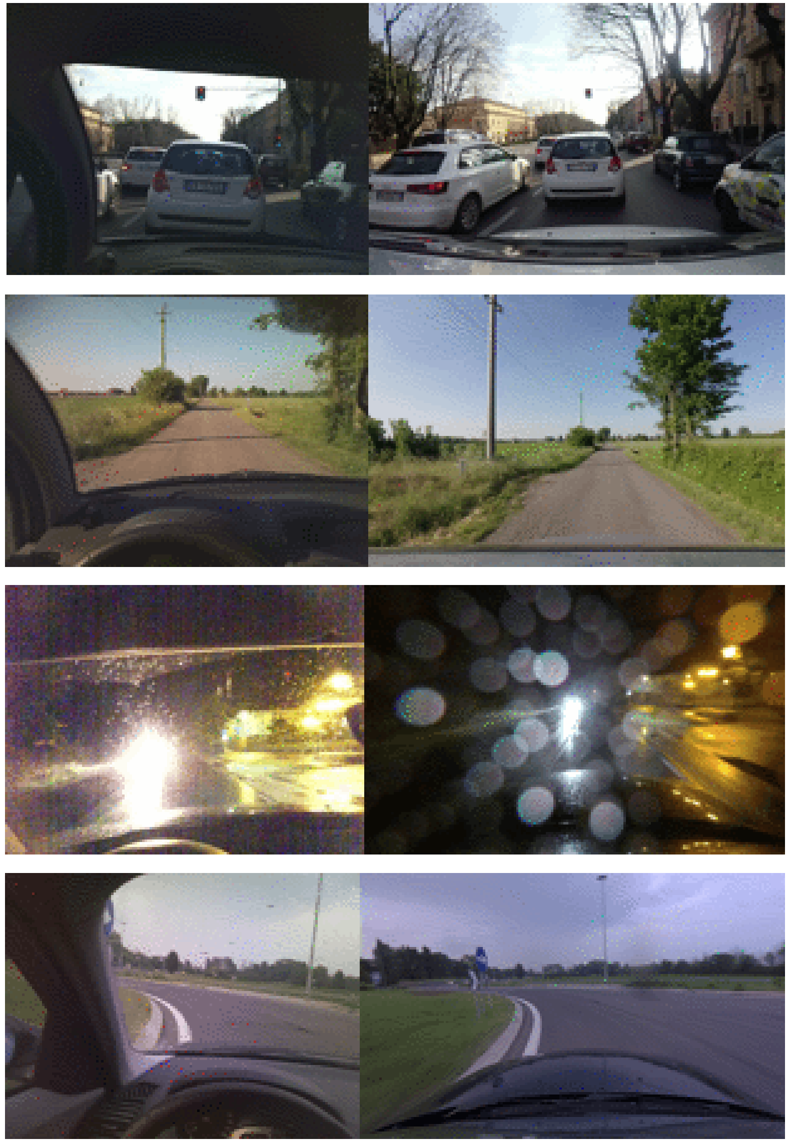
AImage Lab

First-person camera: eye tracker SMI ETG HD camera 720p/30fps

Dashboard camera: Garmin VirbX 1080p/25fps, embedded GPS



DATA COLLECTION: OVERVIEW



- 8 different drivers
- 3 different landscapes
  {Highway, Countryside, Downtown}
- 3 different weather conditions:
  {Sunny, Cloudy, Rainy}
- 3 different lighting conditions:
  {Morning, Evening, Night}

- 74 videos of 5 minutes each

- 555 000 annotated frames

DATA COLLECTION: TEMPORAL INTEGRATION

DATASET ANALYSIS: DATA OVERVIEW

X: frames from first-person and dashboard camera

Y: driver's fixation maps on the scene

MIMICKING THE DRIVER: INTUITION

Exploiting multi-source information

Appearance

Motion

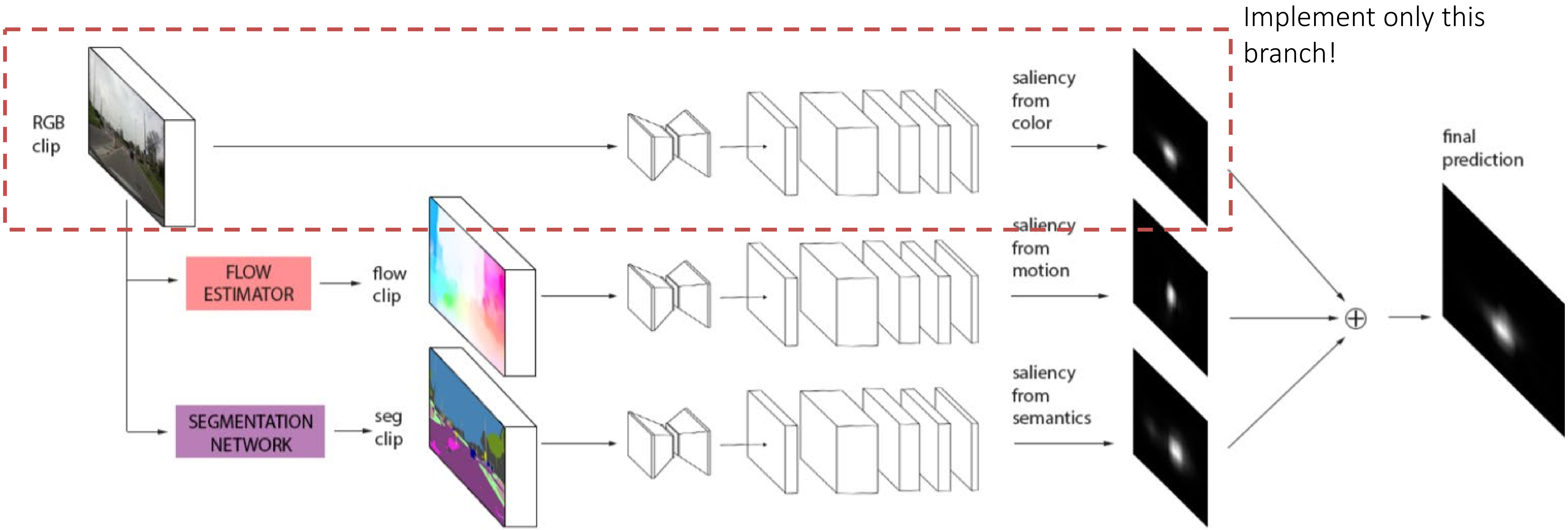
Semantics

Automatically predicted FoA

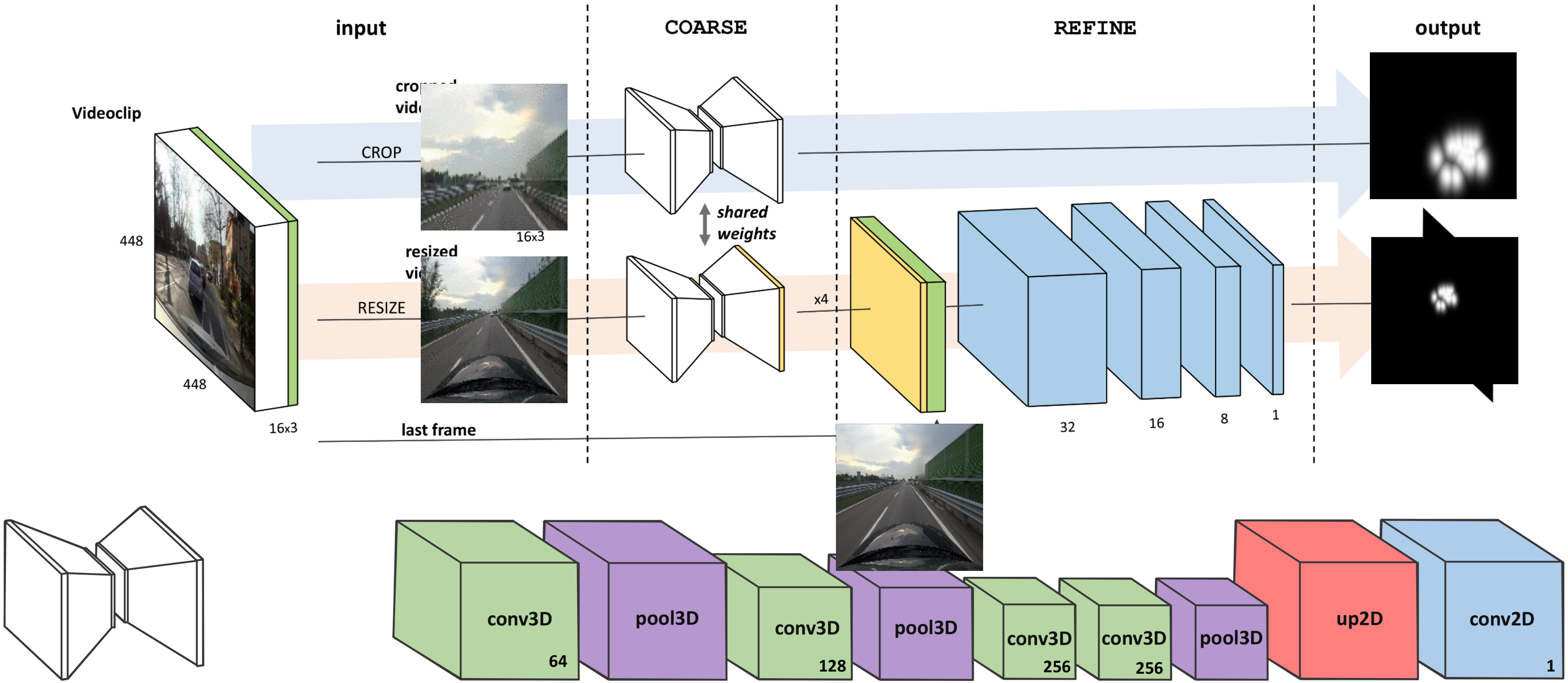
MULTI-BRANCH MODEL

Deep learning model (approximately 43 M of learnable parameters)

Trained to predict human FoA given a driving scene



MODEL: SINGLE BRANCH



EXAMPLE INPUT

X_res

B,3,16,112,112

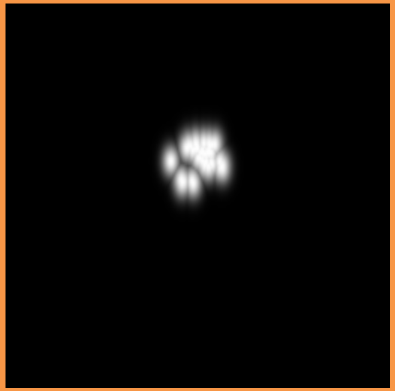
X_crp

B,3,16,112,112

X_ff

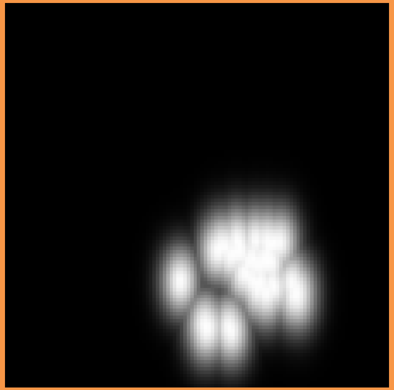
B,3,448,448

Y



B,1,448,448

Y_crop



B,1,112,112

EXAMPLE INPUT

X_res

B,3,16,112,112

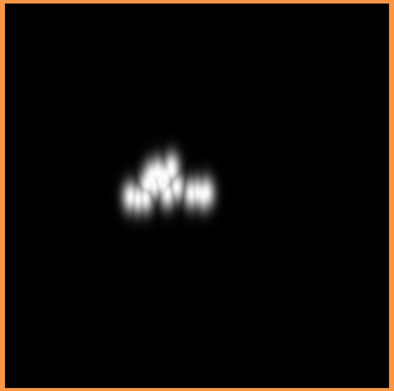
X_crp

B,3,16,112,112

X_ff

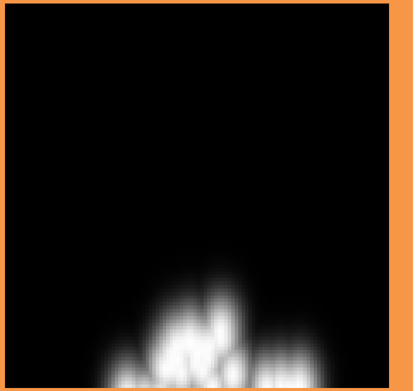
B,3,448,448

Y



B,1,448,448

Y_crop



B,1,112,112

MIMICKING THE DRIVER: RESULTS