CVPR WAD Video Segmentation Challenge

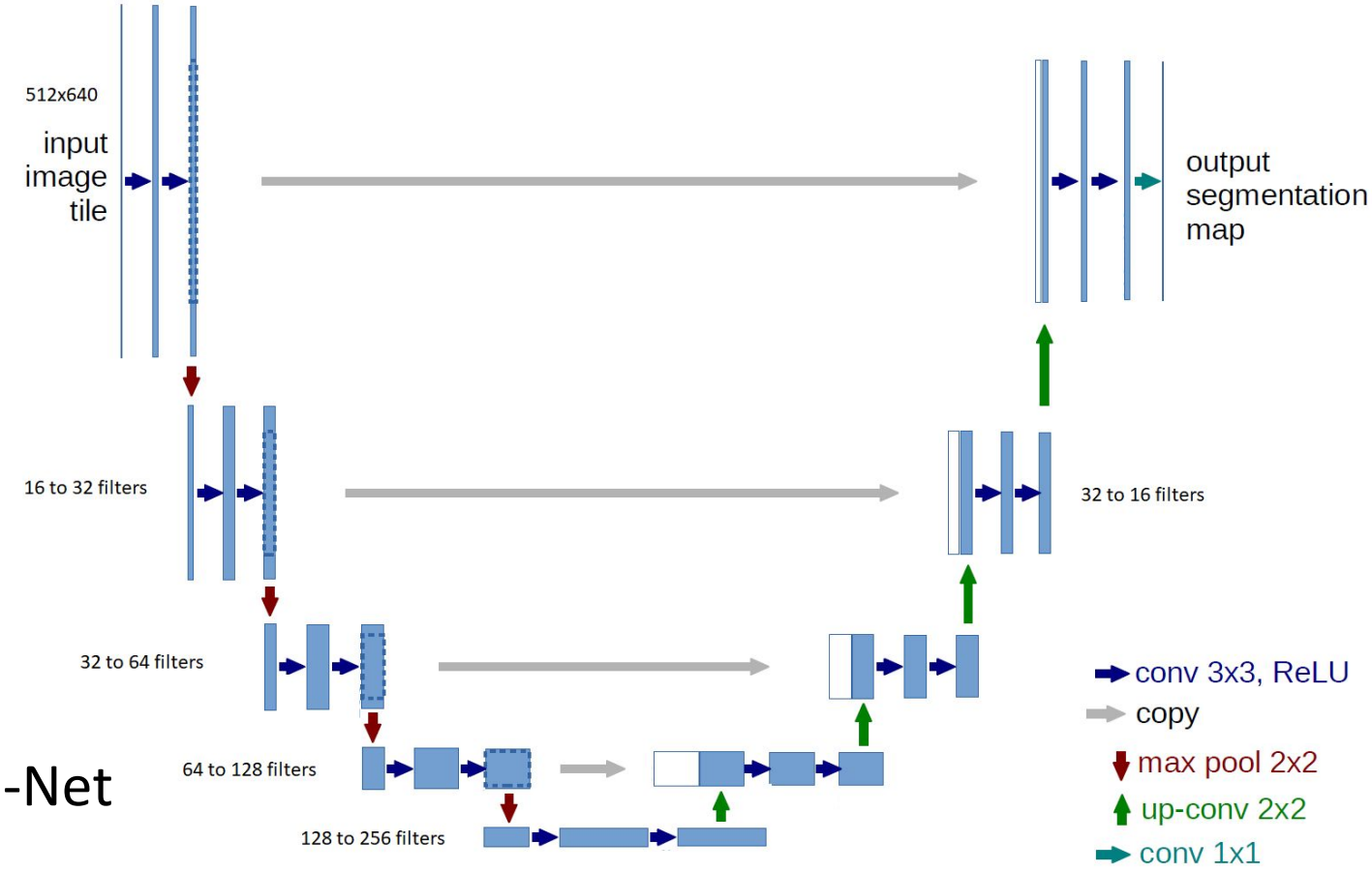
- Project group members
  - Sanchez, Connor
  - Cisewski, Trenton
- Team 9
- Demo: No
- Video: Yes
- Ok to post online: Yes
- Preferred presentation timeslot: Tgroup2

The Problem

- Object detection and semantic segmentation
- Dashcam footage from cars
- Classify and segment:
  - Car, motorcycle, bicycle, person, bus, tricycle, truck, other
- Add temporal coherence to model because data is sequential

The Solution

- U-Net architecture + Temporal



U-Net

Input images: $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$

Corresponding general representation: $\boldsymbol{z}_{\boldsymbol{\theta}}^{l}(\boldsymbol{x}_1)$ and $\boldsymbol{z}_{\boldsymbol{\theta}}^{l}(\boldsymbol{x}_2)$ in the l^{th} layer.

$$L_{coh}(\boldsymbol{\theta}, \boldsymbol{x}_1, \boldsymbol{x}_2) = \quad (4)$$

$$\begin{cases} ||\boldsymbol{z}_{\boldsymbol{\theta}}^{l}(\boldsymbol{x}_1) - \boldsymbol{z}_{\boldsymbol{\theta}}^{l}(\boldsymbol{x}_2)||_1, & \text{if } \boldsymbol{x}_1, \boldsymbol{x}_2 \text{ consecutive} \\ \max(0, \delta - ||\boldsymbol{z}_{\boldsymbol{\theta}}^{l}(\boldsymbol{x}_1) - \boldsymbol{z}_{\boldsymbol{\theta}}^{l}(\boldsymbol{x}_2)||_1), & \text{otherwise} \end{cases}$$

Temporal Coherence

No coherence > with coherence: small object detection

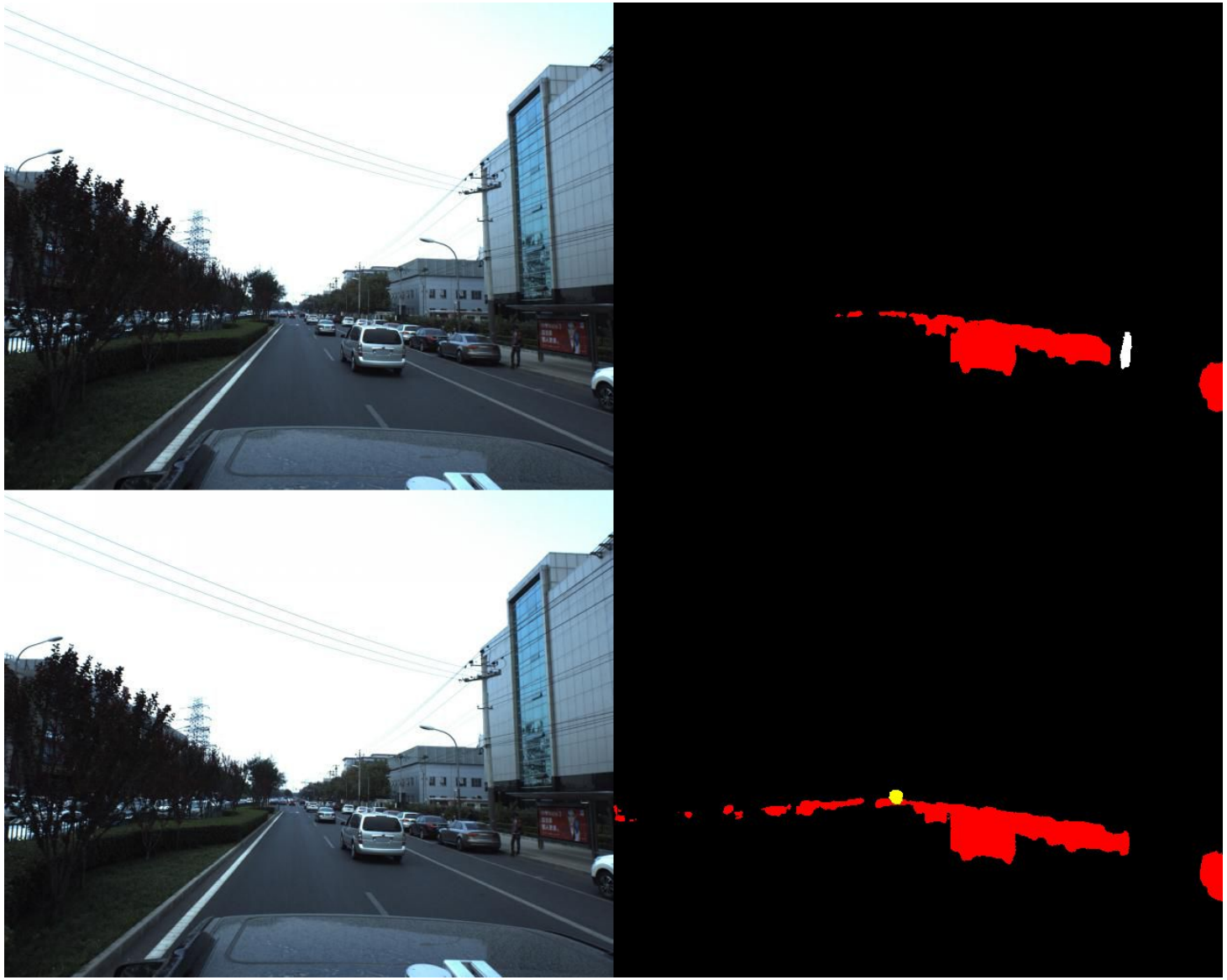
= Car

= Bus

= Person

Top Picture: No coherence
Bottom Picture: With coherence

No coherence > with coherence: small object detection



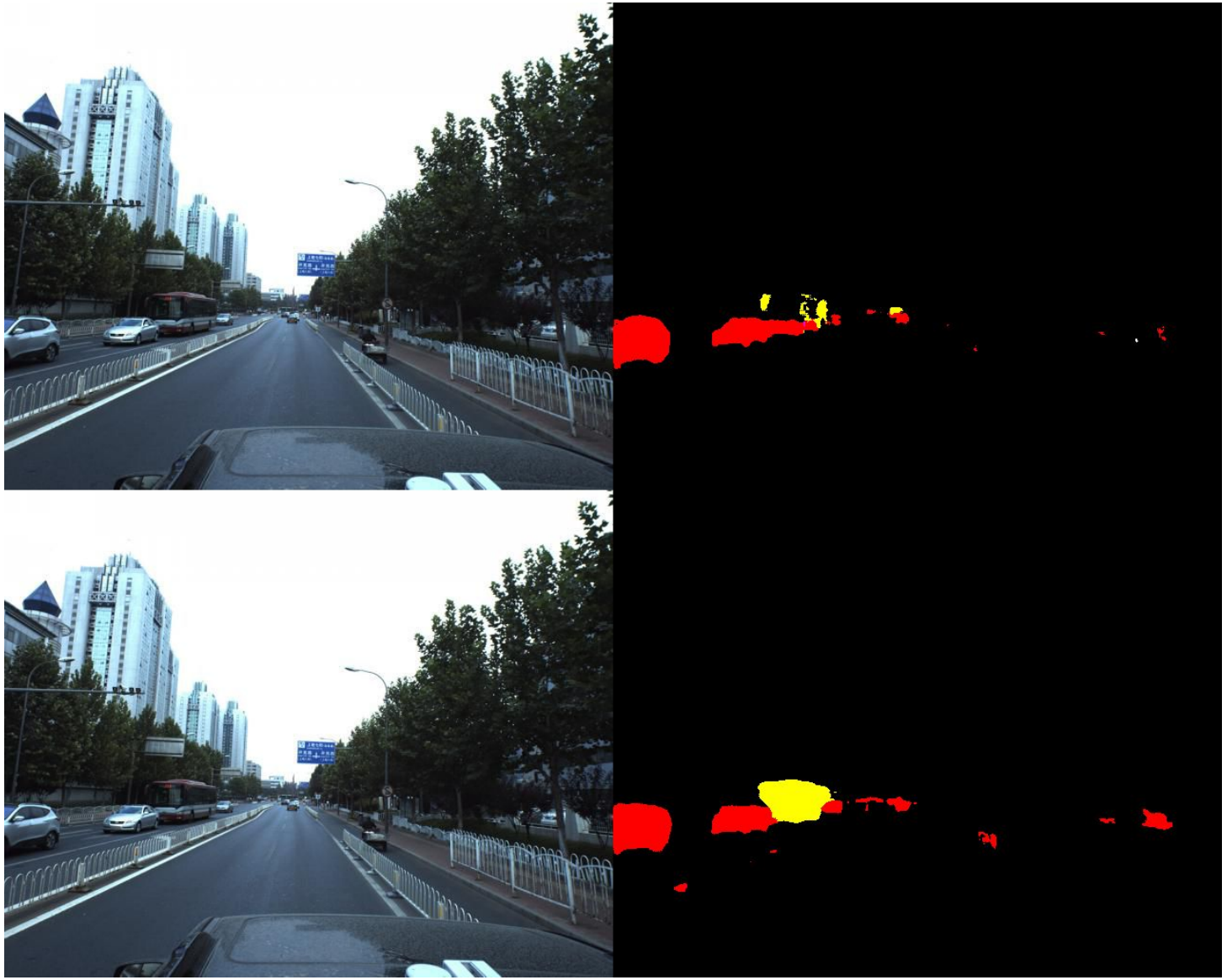
= Car

= Bus

= Person

Top Picture: No coherence
Bottom Picture: With coherence

With coherence > no coherence: large object detection



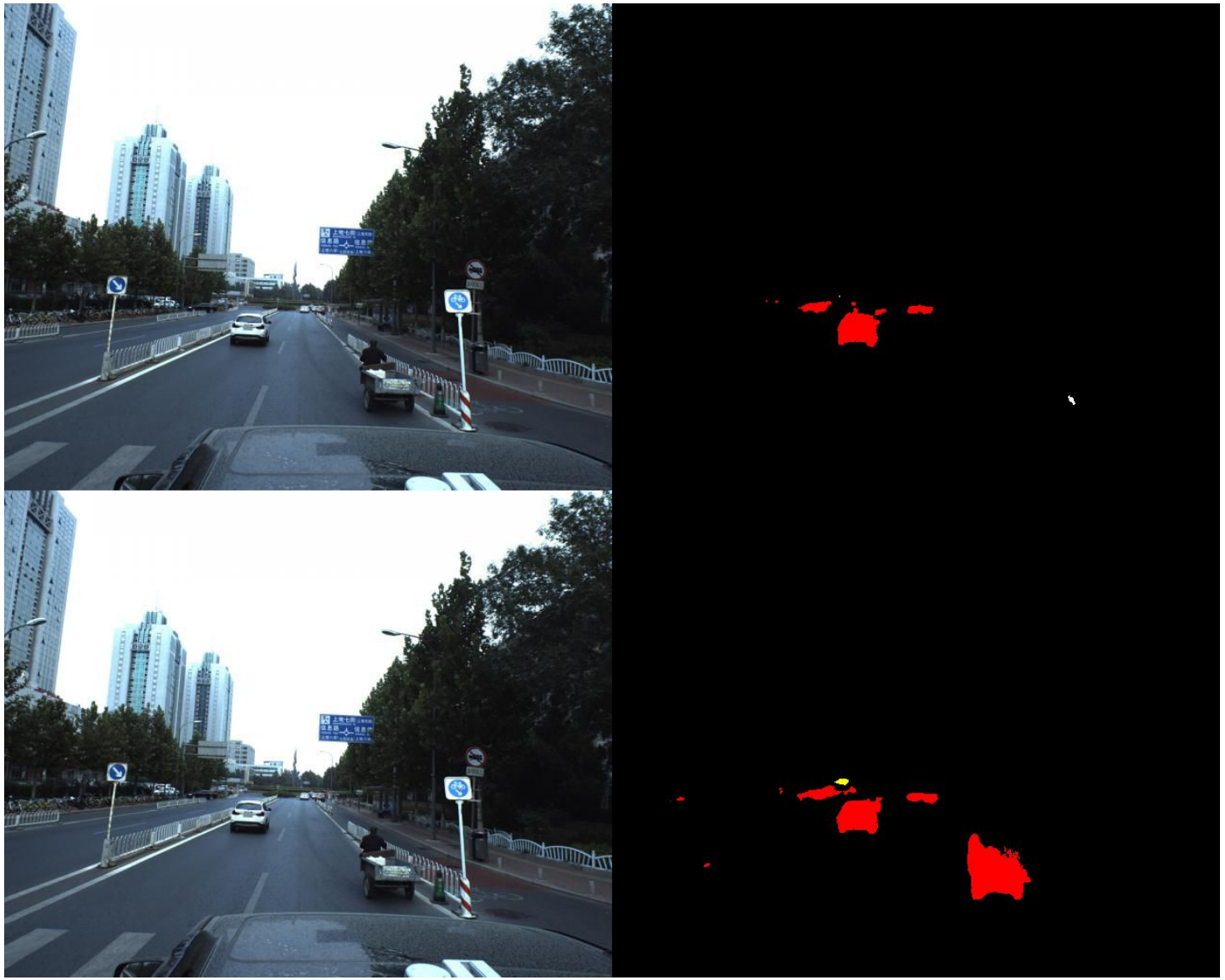
= Car

= Bus

= Person

Top Picture: No coherence
Bottom Picture: With coherence

With coherence > no coherence: large object detection



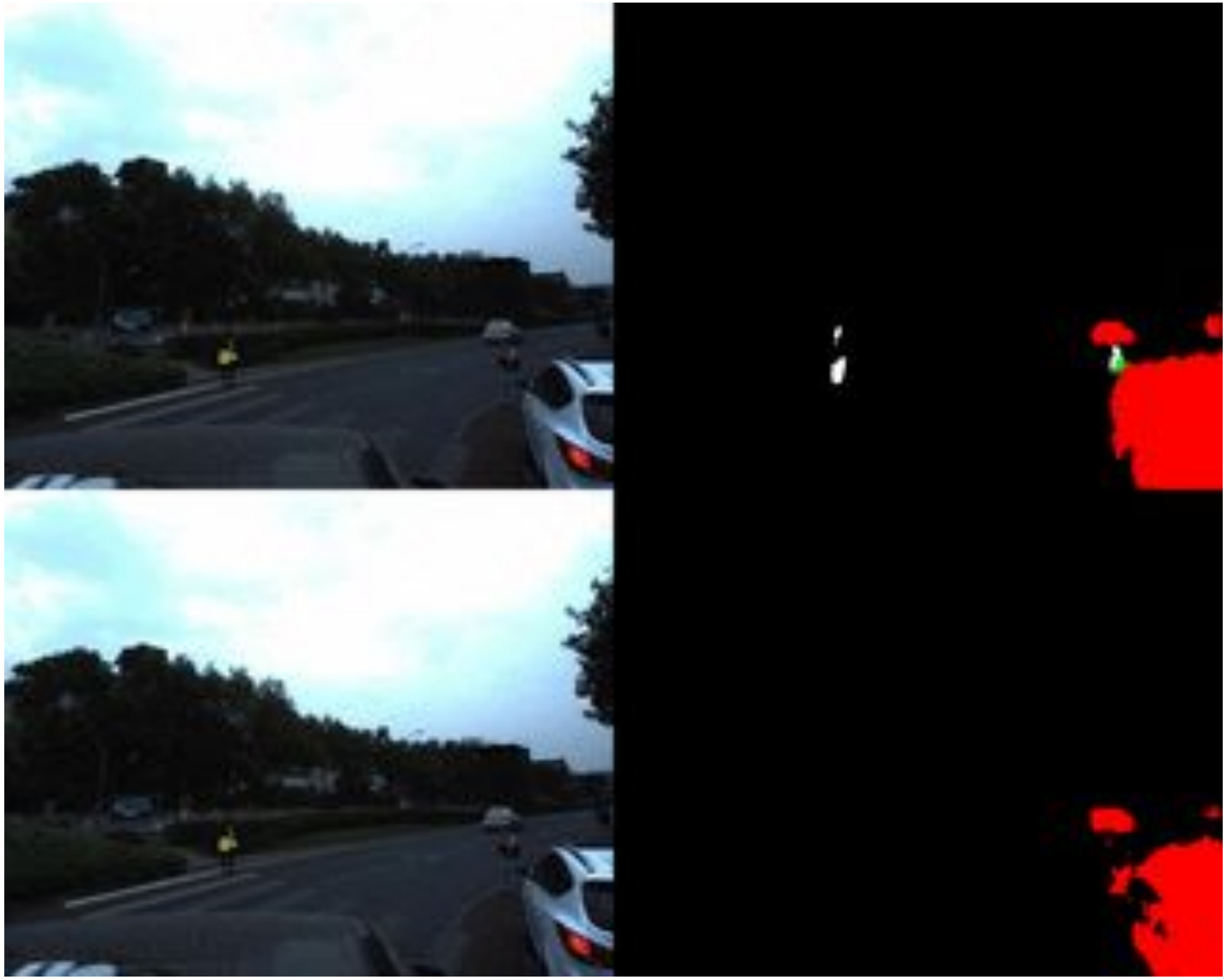
= Car

= Bus

= Person

Top Picture: No coherence
Bottom Picture: With coherence

Video



Top Picture: No coherence
Bottom Picture: With coherence

= Car
= Motorcycle
= Bicycle
= Person
= Nothing
= Truck
= Bus
= Tricycle

Learnt from the project

- Coherence improved classification for large objects but not small objects due to poor frame rate
  - Increase frame rate (currently 15 to 20 fps)
- Improve classification on smaller / more complex objects
  - Don’t down sample images (requires more GPU’s)
- Improve video labeling coherence
  - Implement a complete Siamese network

References and GitHub

- GitHub:
  - https://github.com/connure/wad_ai
- U-Net Paper:
  - https://arxiv.org/pdf/1505.04597.pdf
- Temporal Coherence Paper:
  - https://ronan.collobert.com/pub/matos/2009_video_icml.pdf