Self-Supervised Learning for Autonomous Driving

Maciej K. Wozniak, Patric Jensfelt <maciejw@kth.se>

Motivation

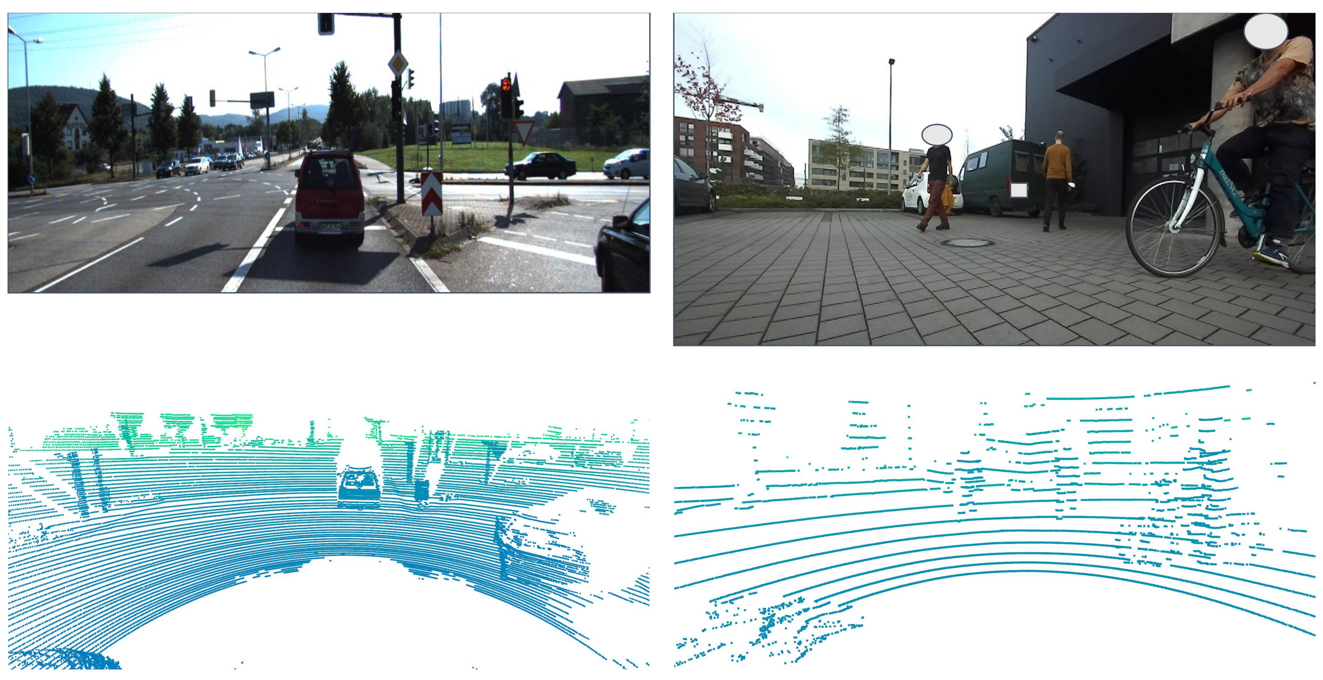
Robot perception fails and it's our task to develop the tools that account for that!

- 3D object detection encounters difficulties with **misaligned sensors, missing data and domain shifts**
- **Perception on sparse LiDAR**, common in mobile robotics, has received insufficient attention in research.
- Every car or robot is different, thus **collecting and labeling** training data for diverse platforms is **costly and labor-intensive**

Unsupervised Domain Adaptation for 3DOD

Motivation

- There is notable research gap when it comes to UDA addressing **larger domain shifts** than those associated with self-driving cars such as:
- self-driving cars to **mobile robots**
- focus on **sparse** (< 32 layers) LiDAR
- street to **sidewalk** environment
- **sim-to-real** on sparse LiDAR
- many current methods rely on **Teacher-Student** approach that **fail domain gap is too large** and teacher successfully cannot distill its knowledge to the student



SOURCE **TARGET**

Fig. 1: Differences between mobile robot and self-driving car LiDAR.

UADA3D

- primary task of $f\theta_f$ and $h\theta_y$ is 3D object detection
- the discriminator $g\theta_D$ aims to classify the domain of each detected instance.
- Discriminator's loss, reversed by GRL, encourages the detector to learn features that are not only effective for object detection but also invariant across domains
- note, we only have access to source domain labels during the training

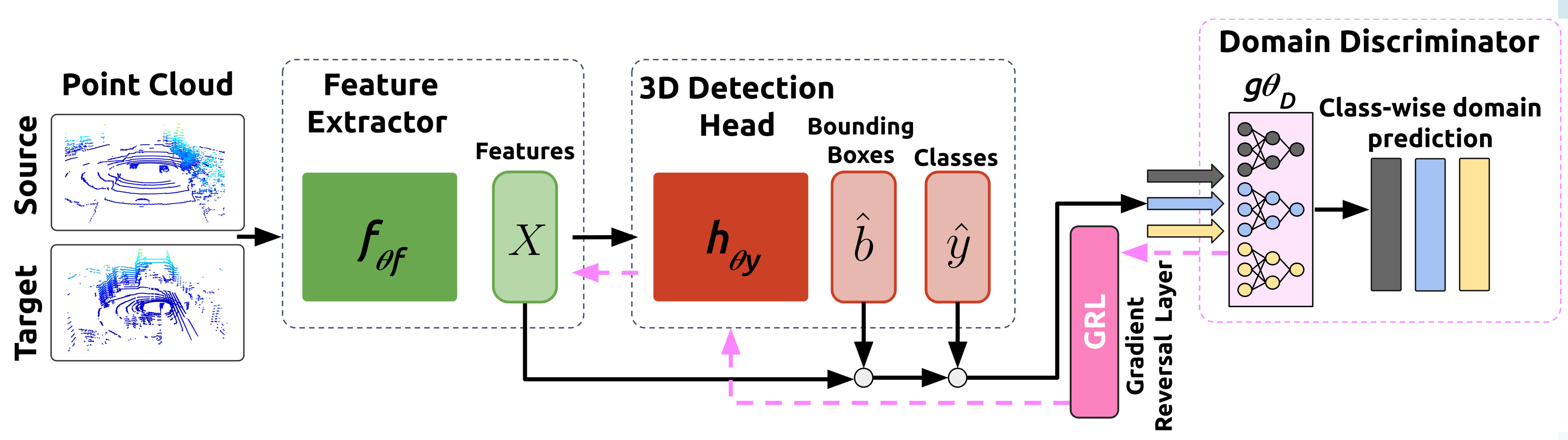


Fig. 2: Overview of UADA3D.

Self-Supervised Pre-Training for AD

Main Contributions

- Overview of feature masking is shown in Figure 3
- The input to the class-wise domain discriminators $g\theta_{D,k}$ is $(x, \hat{b})$, where x are masked features, $\hat{b}$ are predicted bounding boxes
- We obtain masked features x, by masking the feature map $X = f\theta_f(Q)$ with each predicted bounding box $\hat{bn}$ creating corresponding masked features xn.
- Finally, we concatenate xn with the bounding box $\hat{bn}$ and feed to the corresponding class-wise discriminator $g\theta_{D,k}$

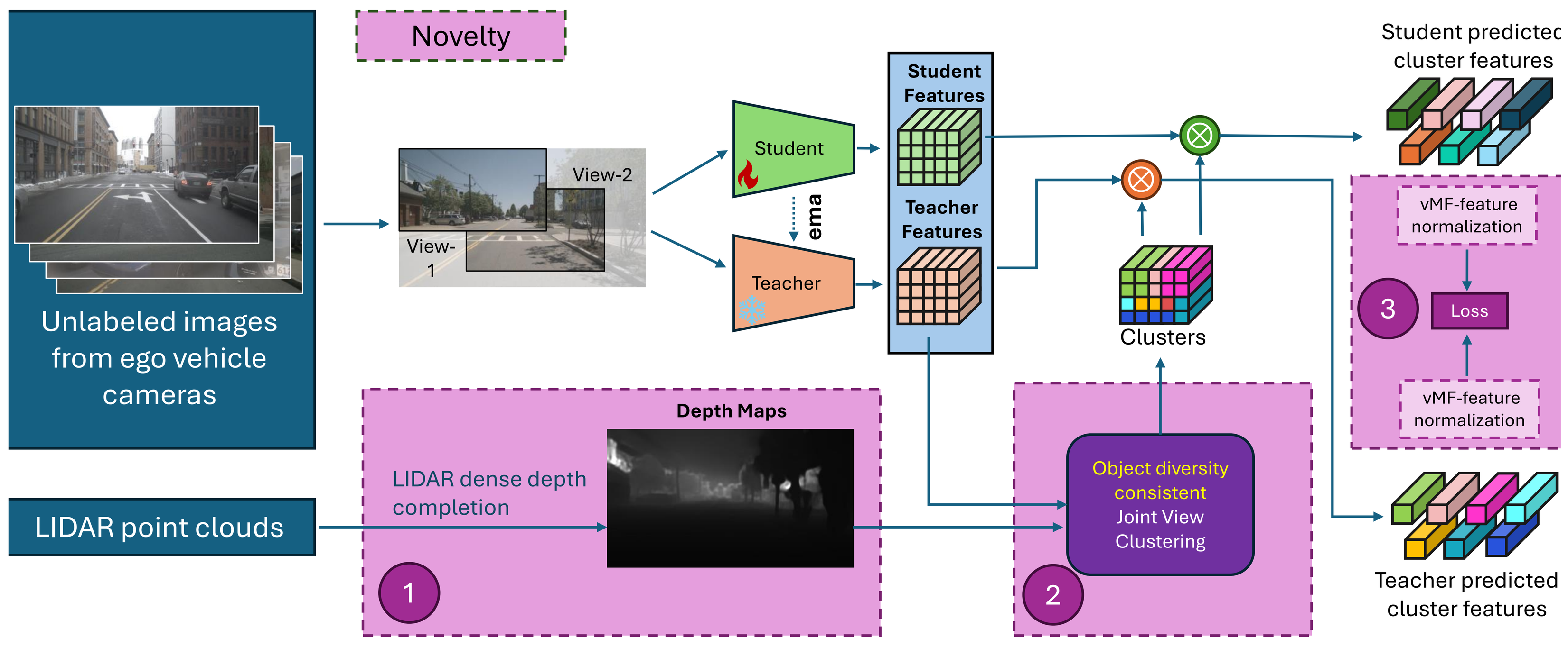
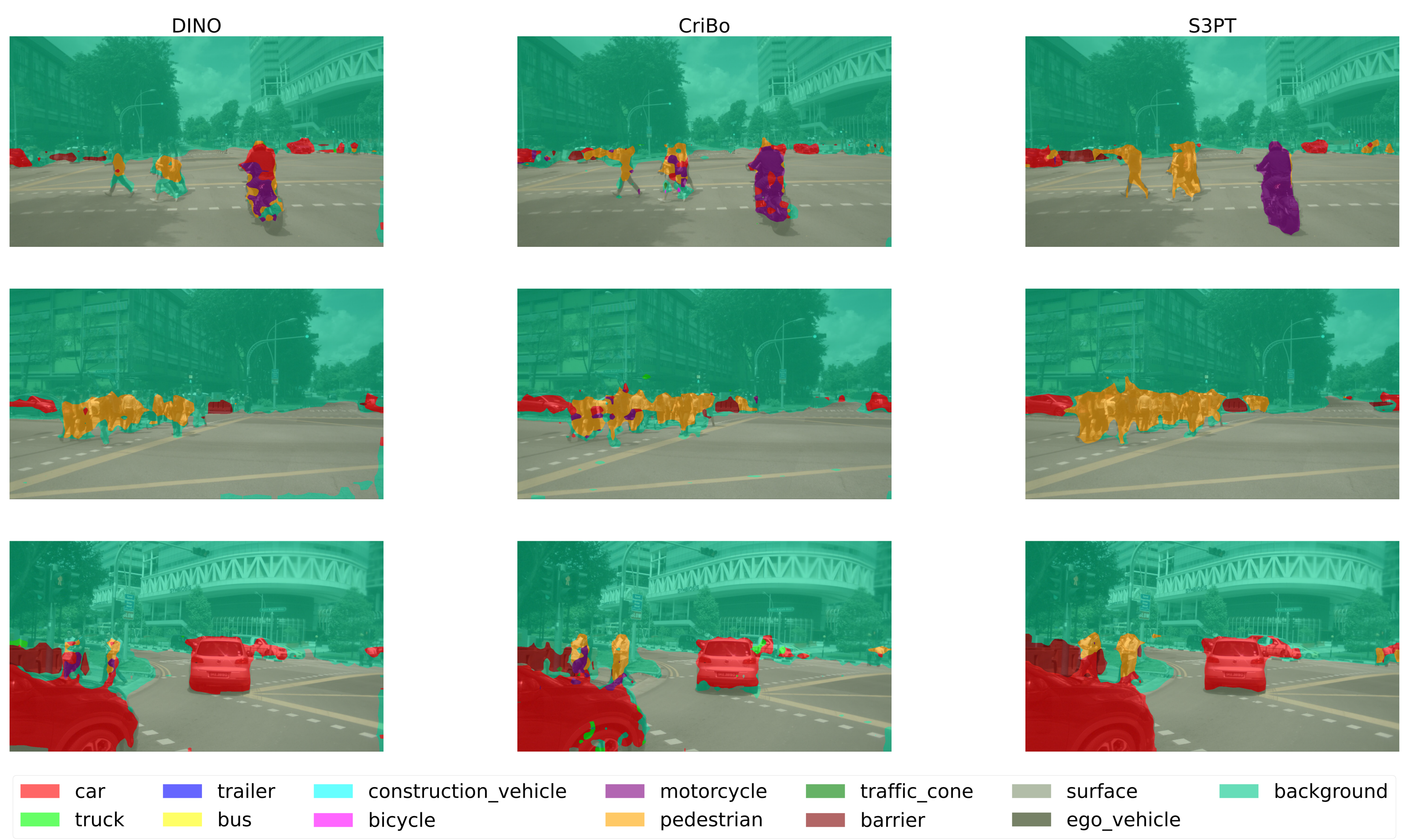


Fig. 3: Feature masking.

Results

- **UADA3D** performs the best across different classes
- We can see our method being especially superior when it comes to **larger domain gaps** (e.g. adaptation to sparse robot data)
- Please refer to **our paper for more details**



WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM