RIT
Rochester Institute of Technology

BMVC 2022

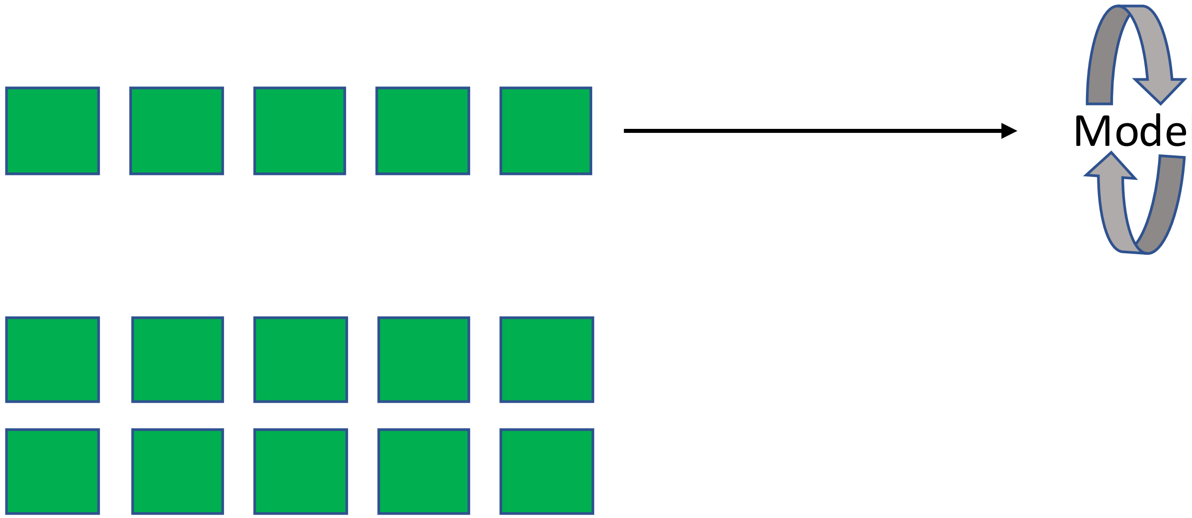
# Semantic Segmentation with Active Semi-Supervised Representation Learning

Aneesh Rangnekar[1], Christopher Kanan[2], Matthew Hoffman[1]

[1]Rochester Institute of Technology

[2]University of Rochester

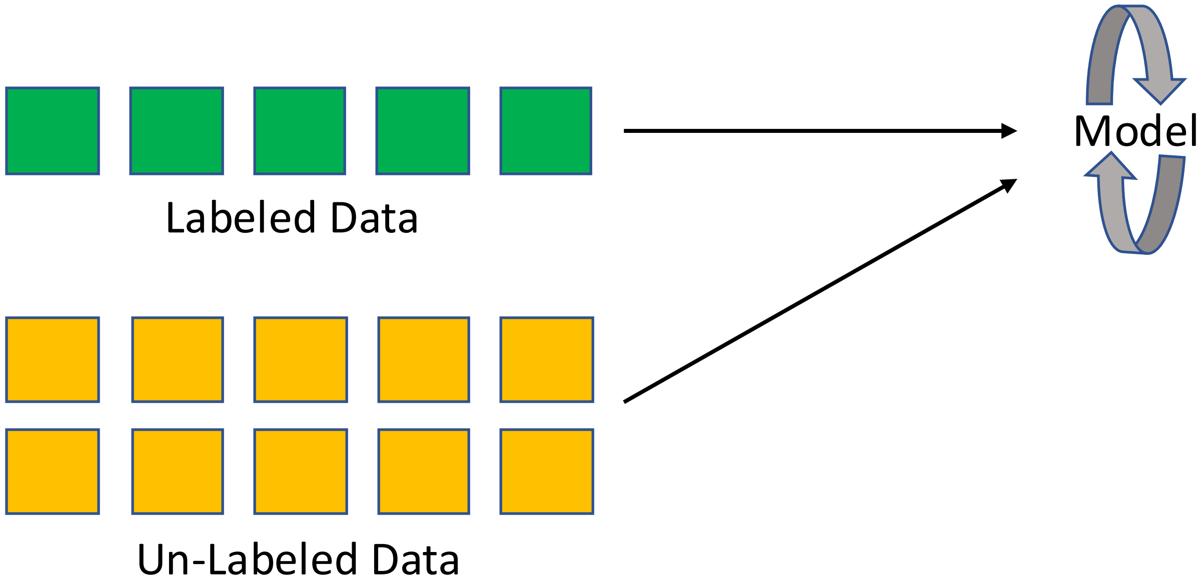
# Terminology Overview




Supervised Learning

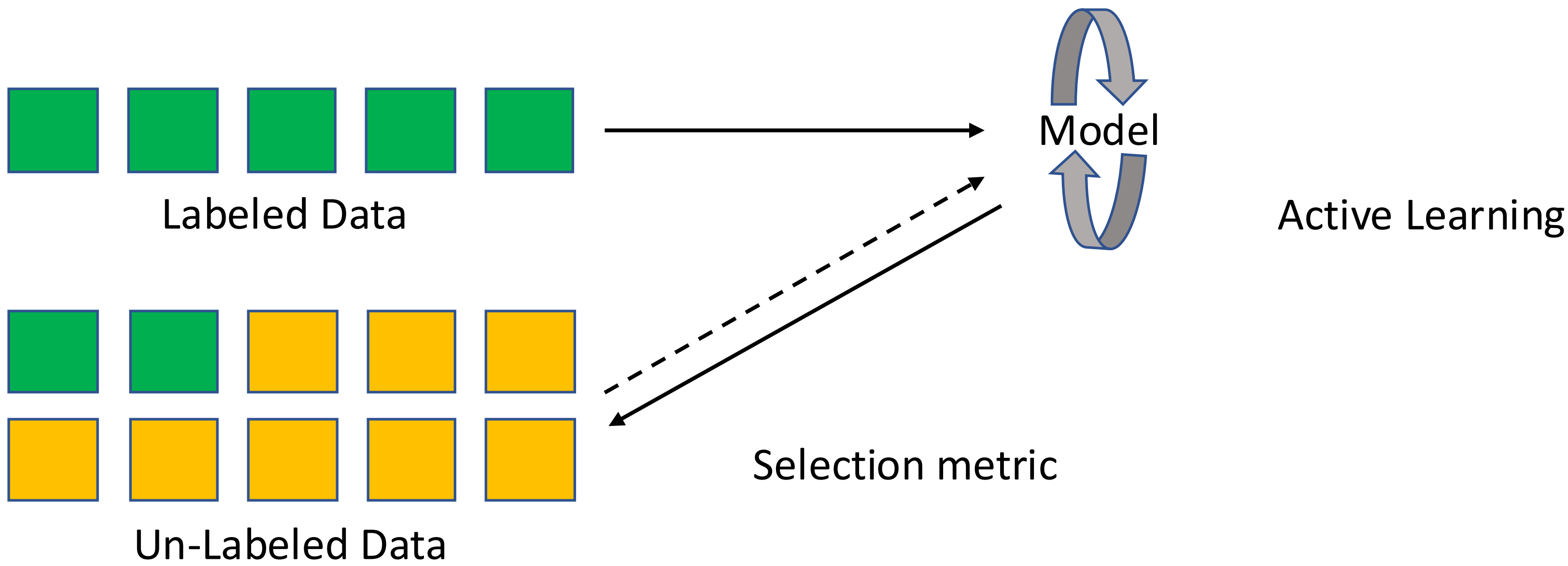
Labeled Data

# Terminology Overview




Semi-Supervised Learning

# Terminology Overview



# Motivation

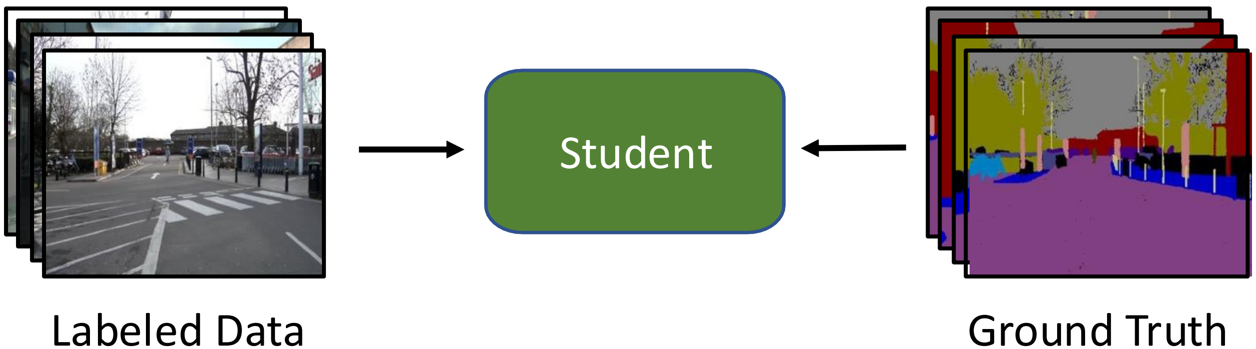
- There were 10K hours for determining the categories present in each image, 20K for using point annotations for each object present, and over 55K for creating segmentation masks

  - Microsoft coco: Common objects in context

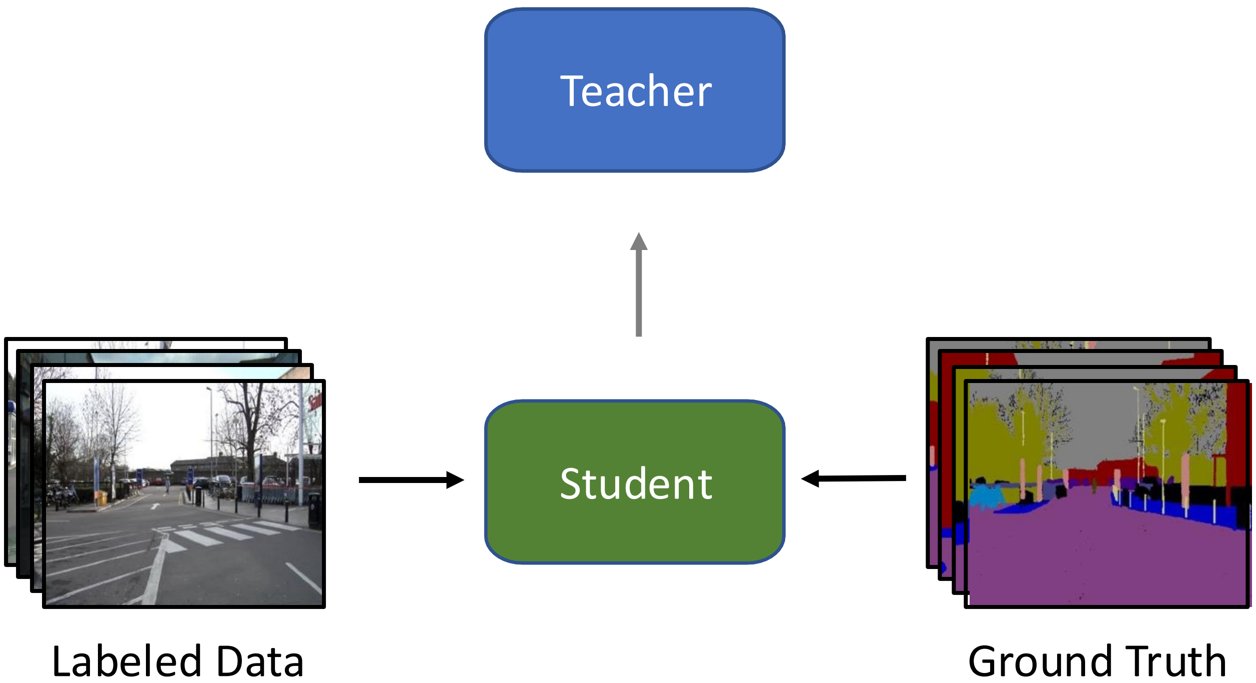
- Annotation and quality control required more than 90 minutes on average for a single image in Cityscapes dataset

  - CityScapes

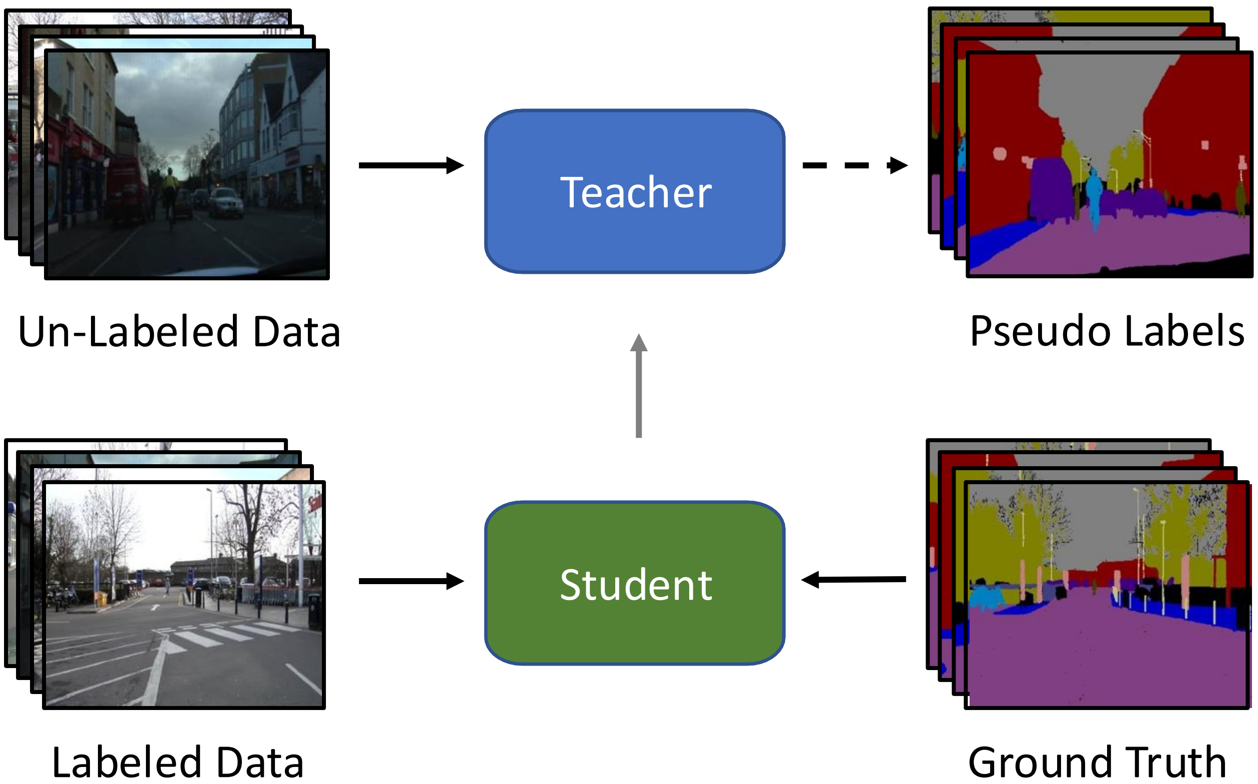
# Semi-Supervised Learning for Semantic Segmentation



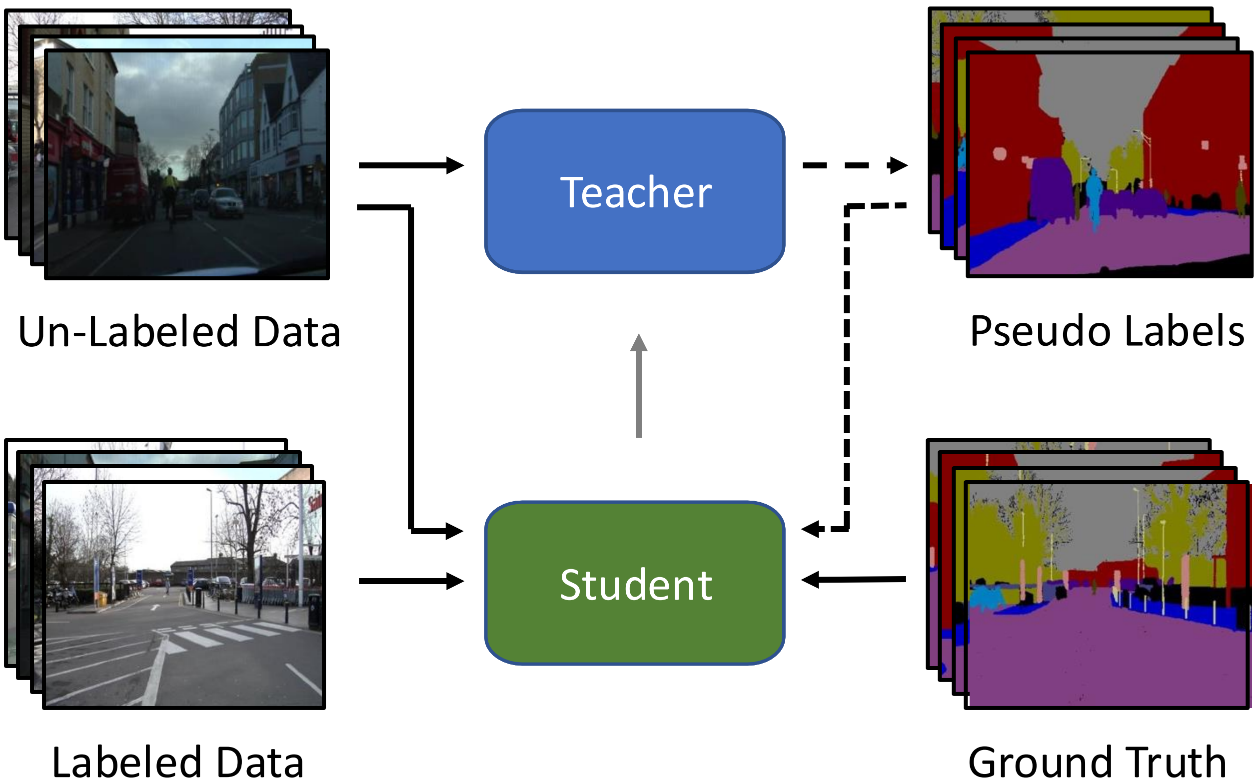
# Semi-Supervised Learning for Semantic Segmentation



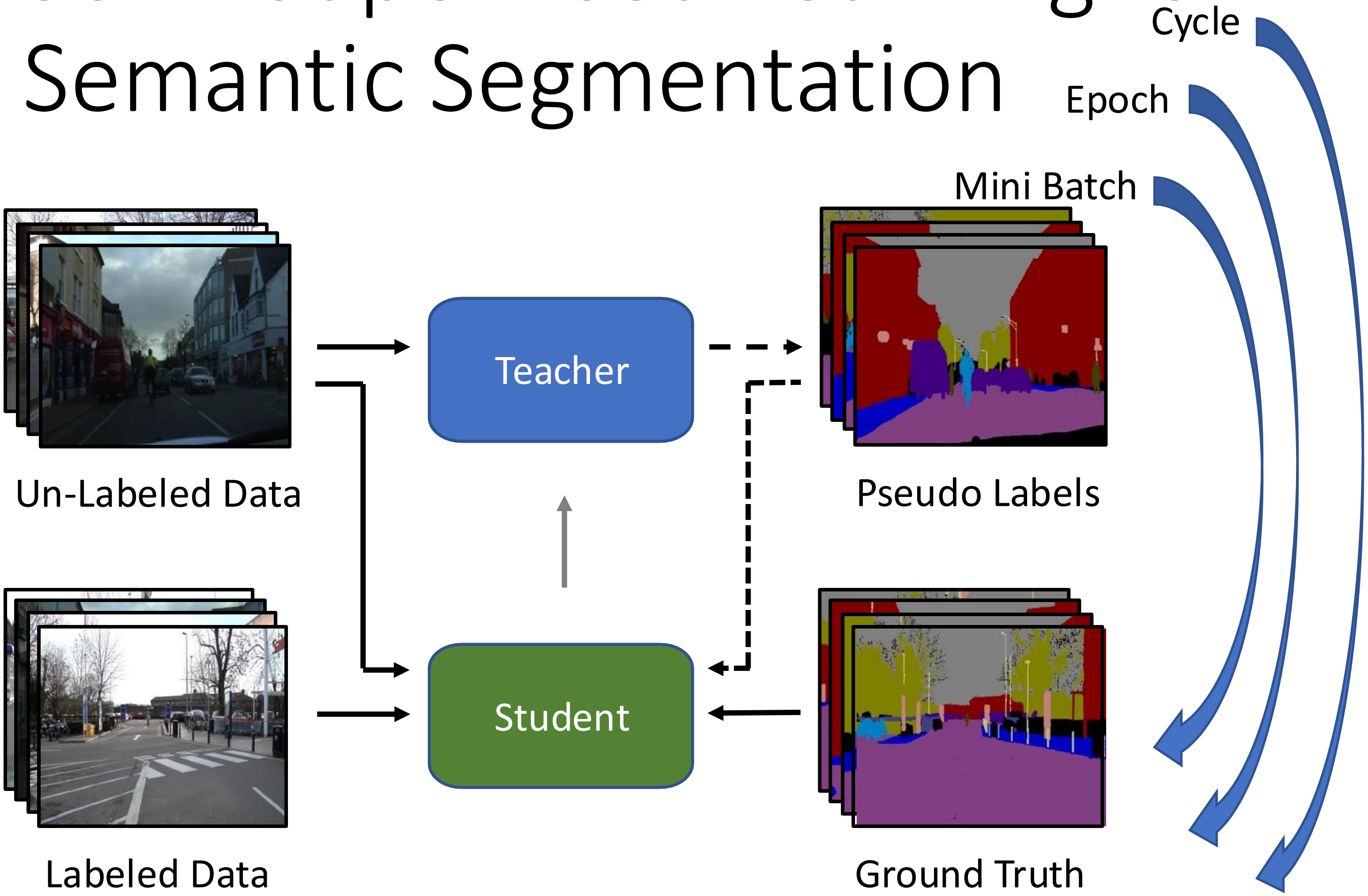
# Semi-Supervised Learning for Semantic Segmentation



# Semi-Supervised Learning for Semantic Segmentation



# Semi-Supervised Learning for Semantic Segmentation



# S4AL+

**Observation 1:** Training with teacher-student pseudo labeling at every mini-batch iteration is very time and GPU resource consuming, especially for semantic segmentation

**Observation 2:** Exponentially moving average based teacher-student learning remains sensitive to underlying class distributions observed by the student

**Observation 3:** Training separately on images from the labeled and unlabeled data, while the unlabeled data undergoes heavy augmentations, skews batch normalization

# S4AL+

**Observation 1:** Training with teacher-student pseudo labeling at every mini-batch iteration is very time and GPU resource consuming, especially for semantic segmentation

**Solution:** Approach self-training as a potential solution for semi-supervised learning

**Observation 2:** Exponentially moving average based teacher-student learning remains sensitive to underlying class distributions observed by the student

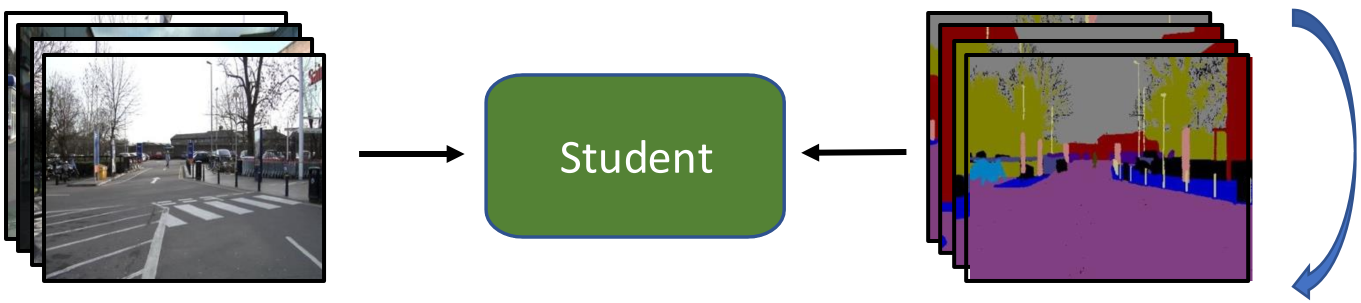
**Solution:** Combat class imbalance by accounting for long-tail classes

**Observation 3:** Training separately on images from the labeled and unlabeled data, while the unlabeled data undergoes heavy augmentations, skews batch normalization

**Solution:** Ensure all images during a mini-batch iteration are seen jointly by the network

# S4AL+

**Self-Training**



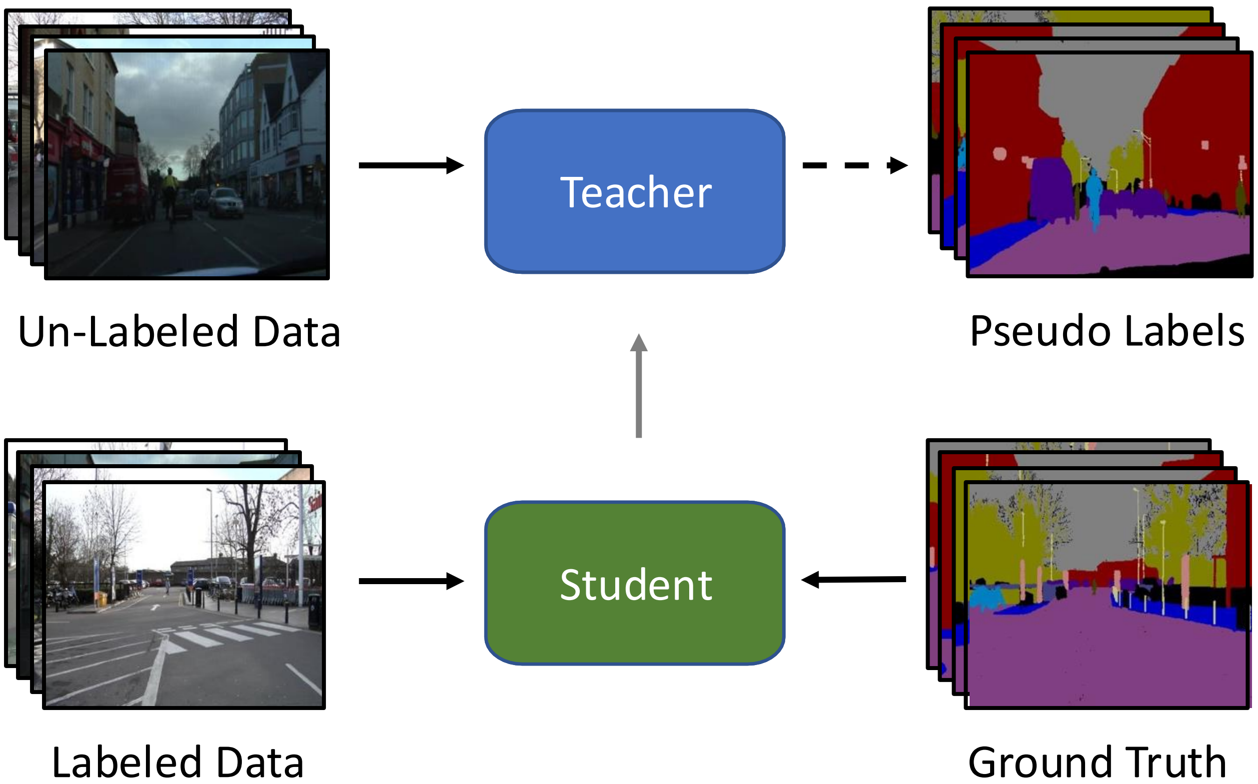

Labeled Data

Ground Truth

Xie, Q., Luong, M.T., Hovy, E. and Le, Q.V., 2020. Self-training with noisy student improves imagenet classification. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10687-10698).

# S4AL+

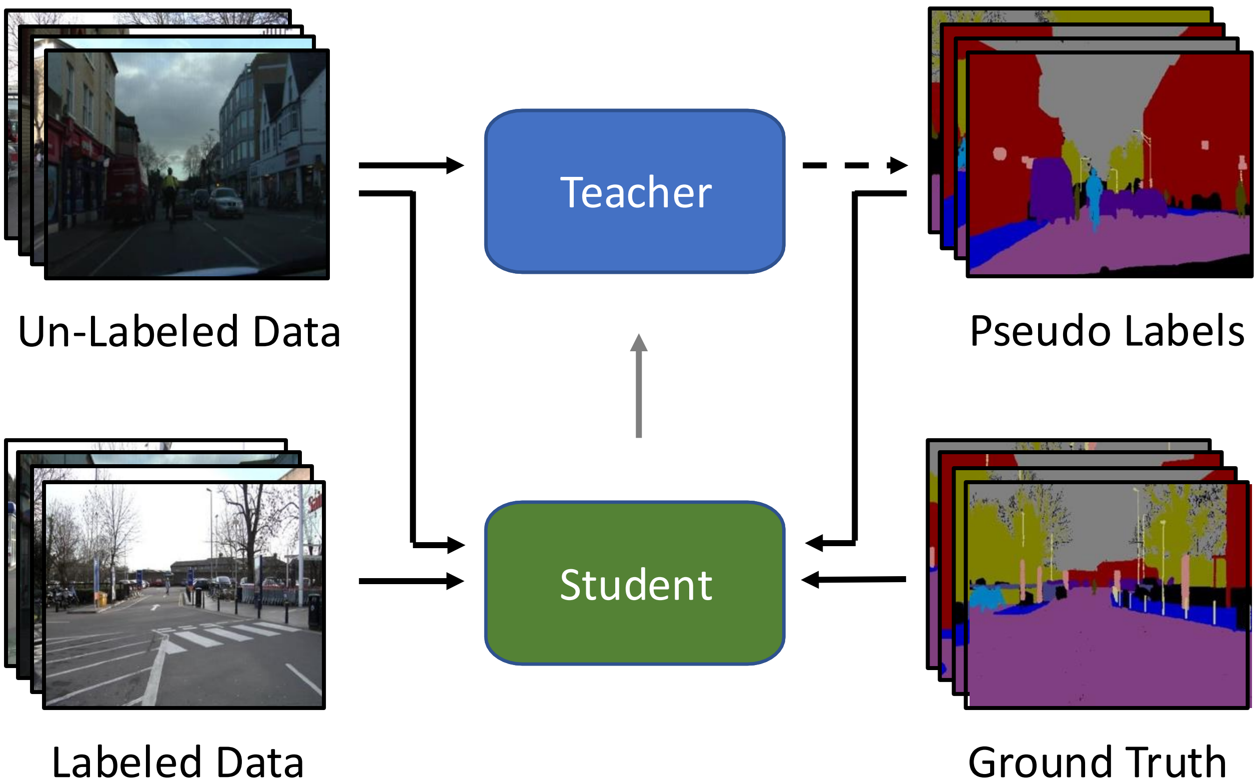
**Self-Training**




Xie, Q., Luong, M.T., Hovy, E. and Le, Q.V., 2020. Self-training with noisy student improves imagenet classification. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10687-10698).

# S4AL+

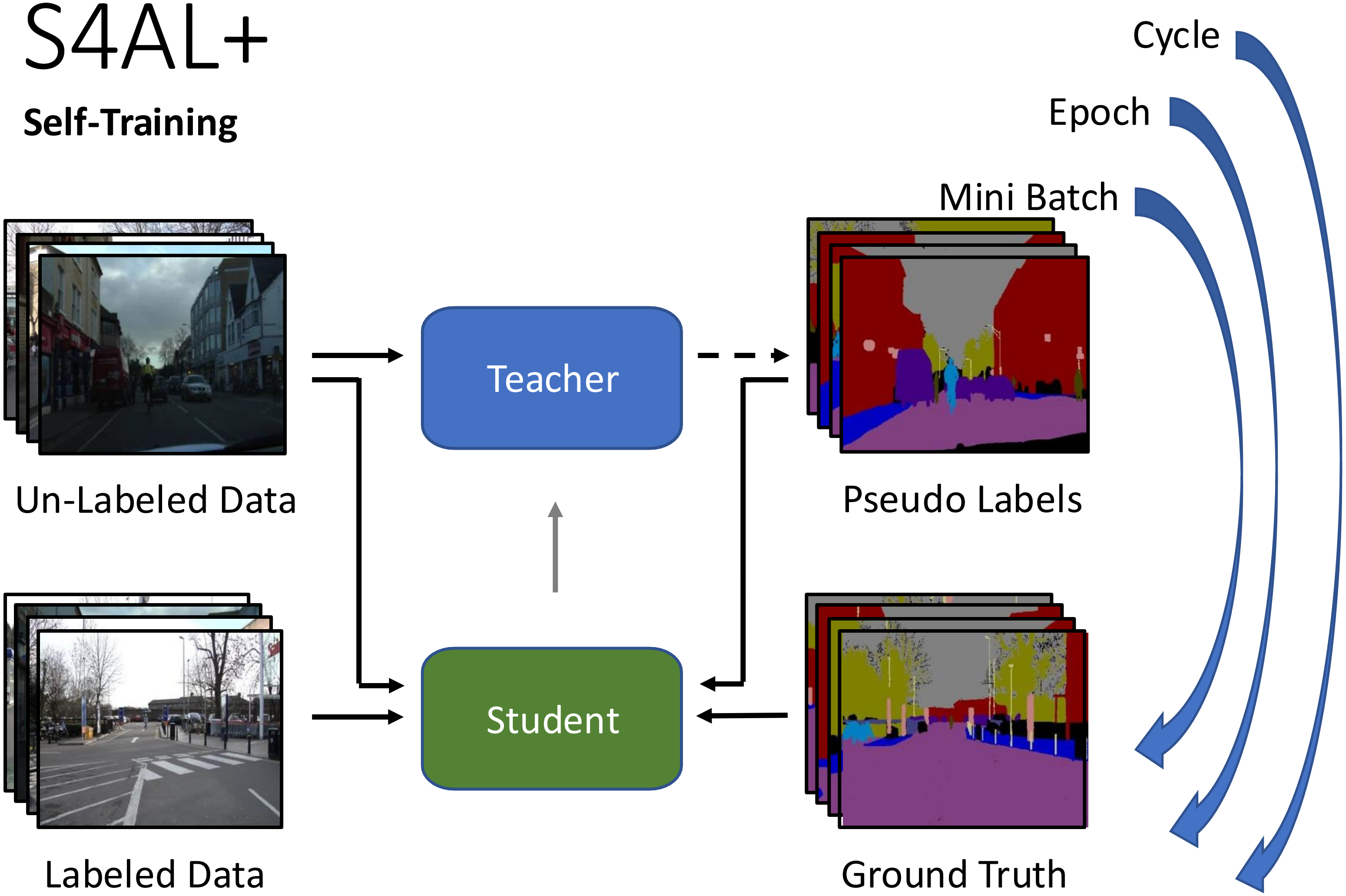
**Self-Training**




Xie, Q., Luong, M.T., Hovy, E. and Le, Q.V., 2020. Self-training with noisy student improves imagenet classification. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10687-10698).

# S4AL+

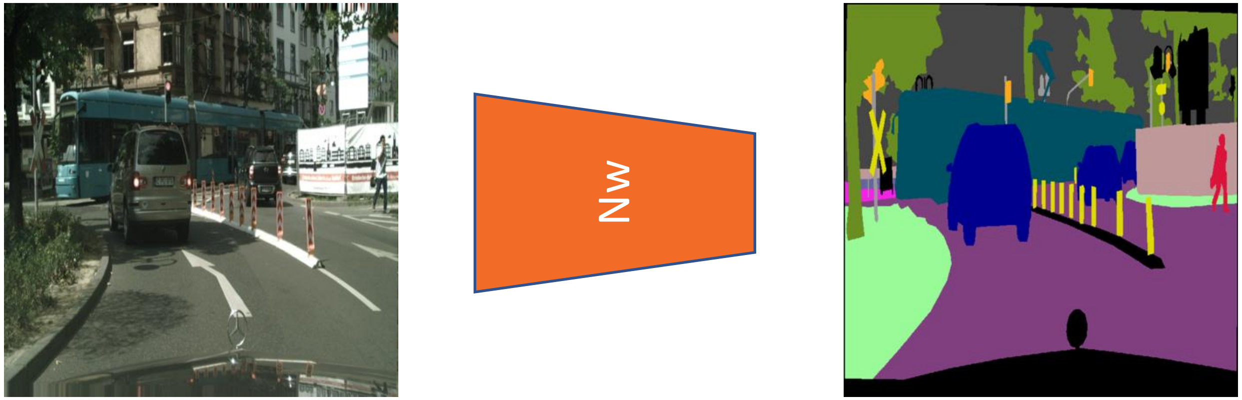
**Self-Training**




Xie, Q., Luong, M.T., Hovy, E. and Le, Q.V., 2020. Self-training with noisy student improves imagenet classification. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10687-10698).

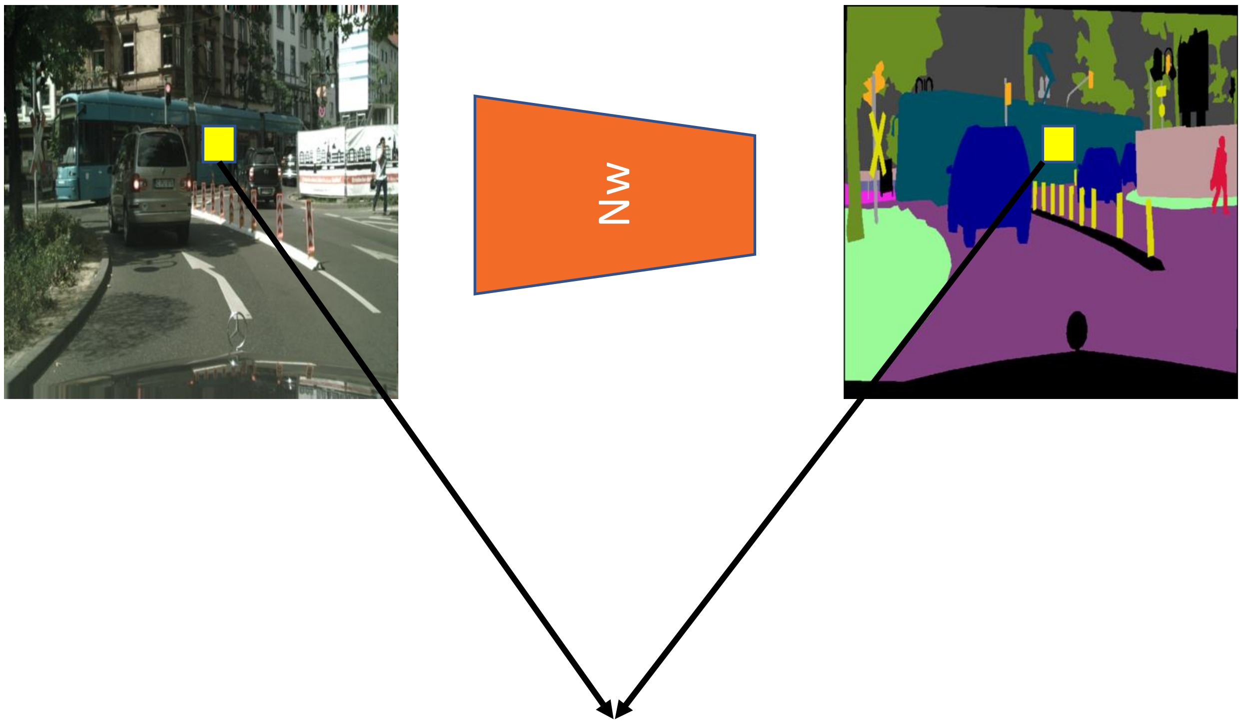
# S4AL+

**Representation Learning**



# S4AL+

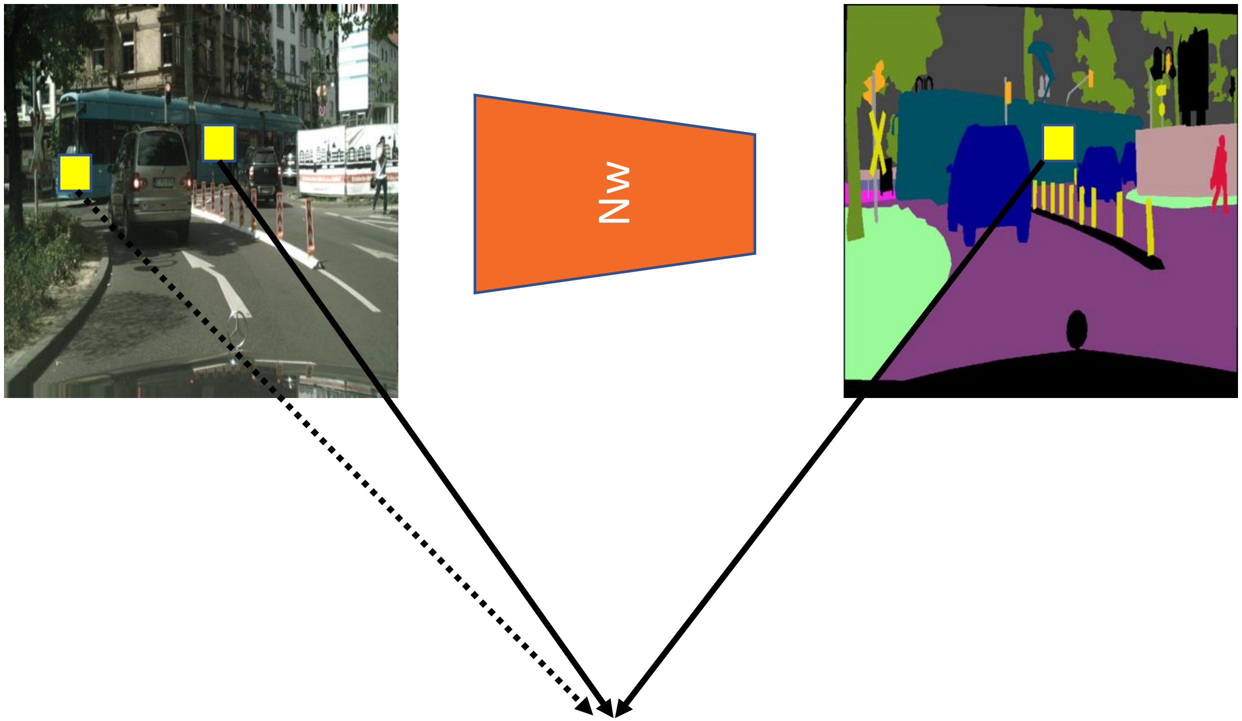
**Representation Learning**




Cross-Entropy Loss

# S4AL+

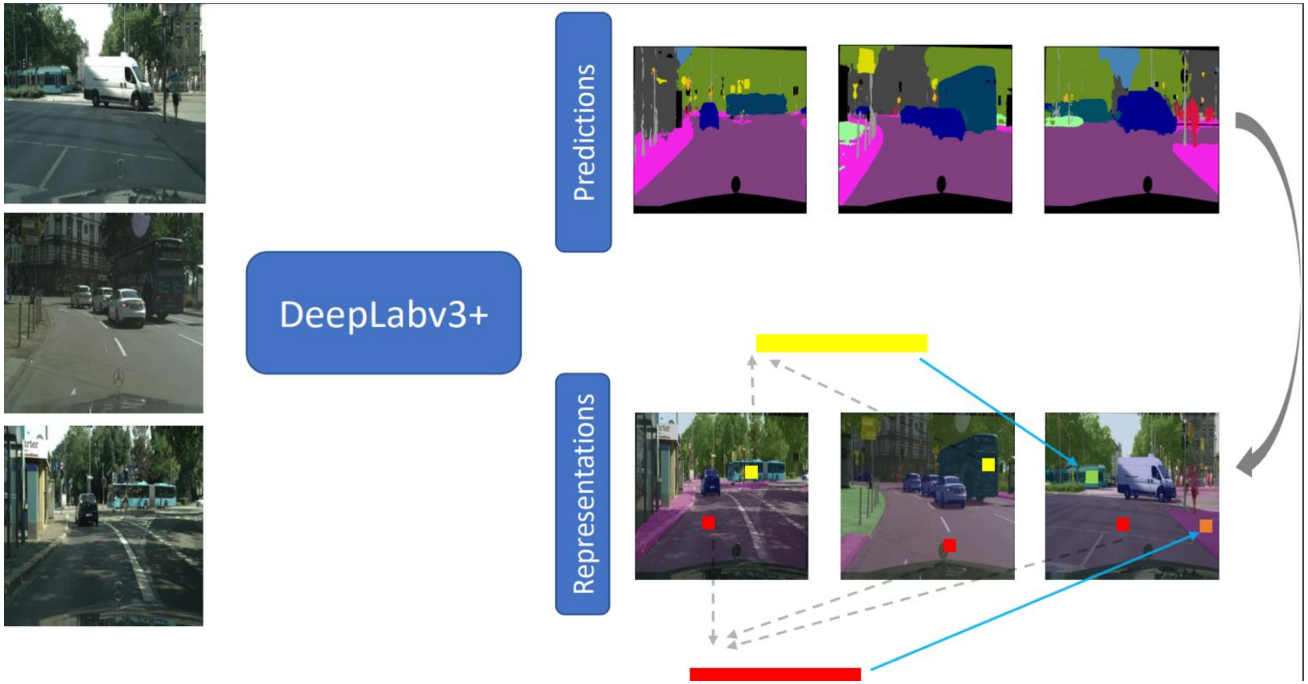
**Representation Learning**




Cross-Entropy Loss ?

# S4AL+

## Representation Learning



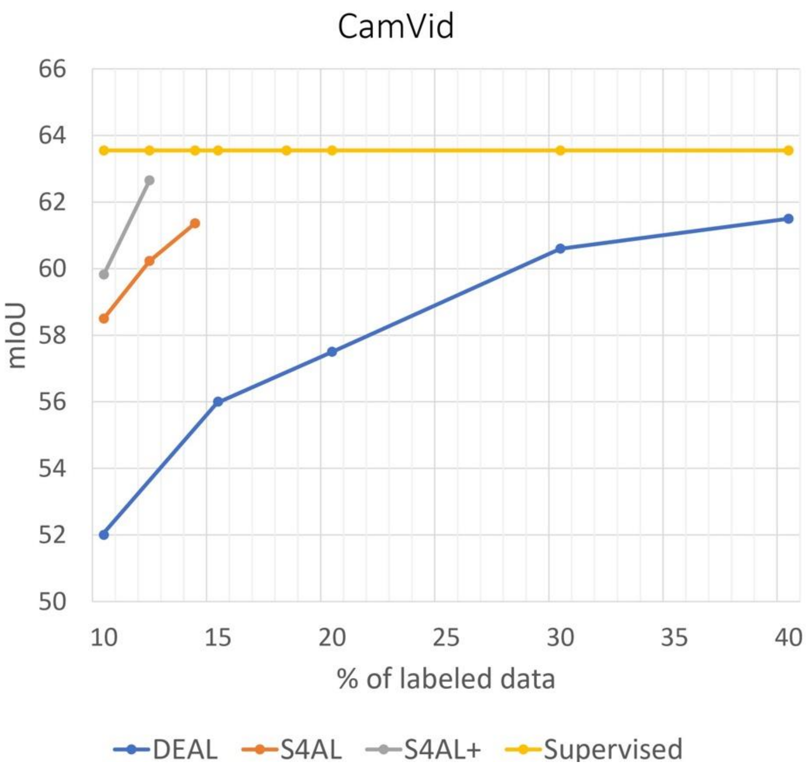

- Cross-Entropy Loss
- Regional Contrast Loss (ReCo)

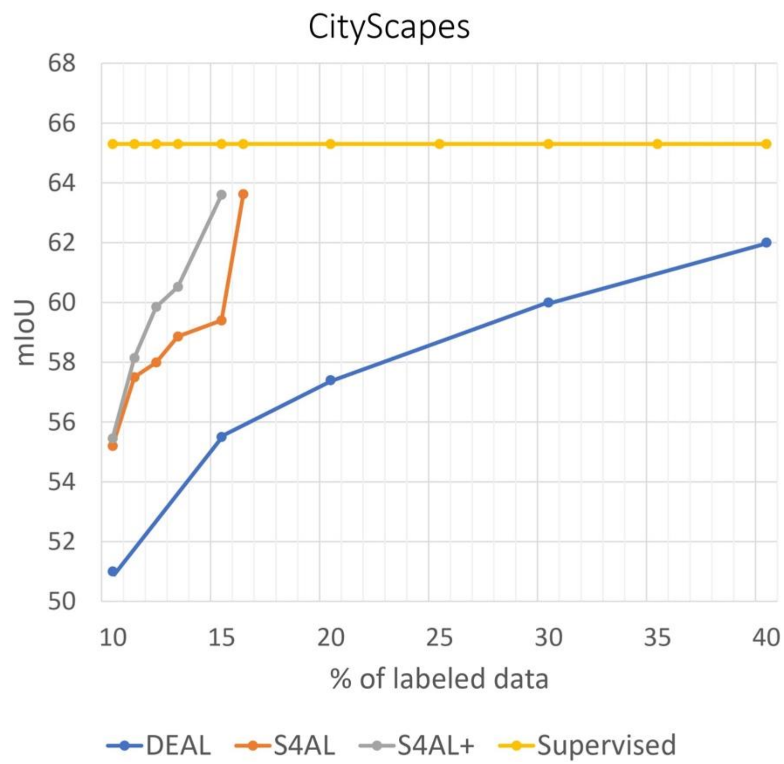
| | | | | |
|---|---|---|---|---|
| void | road | sidewalk | building | wall |
| fence | pole | traffic light | traffic sign | vegetation |
| terrain | sky | person | rider | car |
| truck | bus | train | motorcycle | bicycle |

Liu, S., Zhi, S., Johns, E. and Davison, A.J., 2021. Bootstrapping semantic segmentation with regional contrast. *arXiv preprint arXiv:2104.04465*.

# Results

- Active Learning




Aneesh Rangnekar, Christopher Kanan, and Matthew Hoffman. Semantic segmentation with active semi-supervised learning. IEEE/CVF Winter Conference on Applications of Computer Vision, 2023
Xie, S., Feng, Z., Chen, Y., Sun, S., Ma, C. and Song, M. Deal: Difficulty-aware active learning for semantic segmentation. Asian Conference on Computer Vision, 2020

# Results

- Active Learning

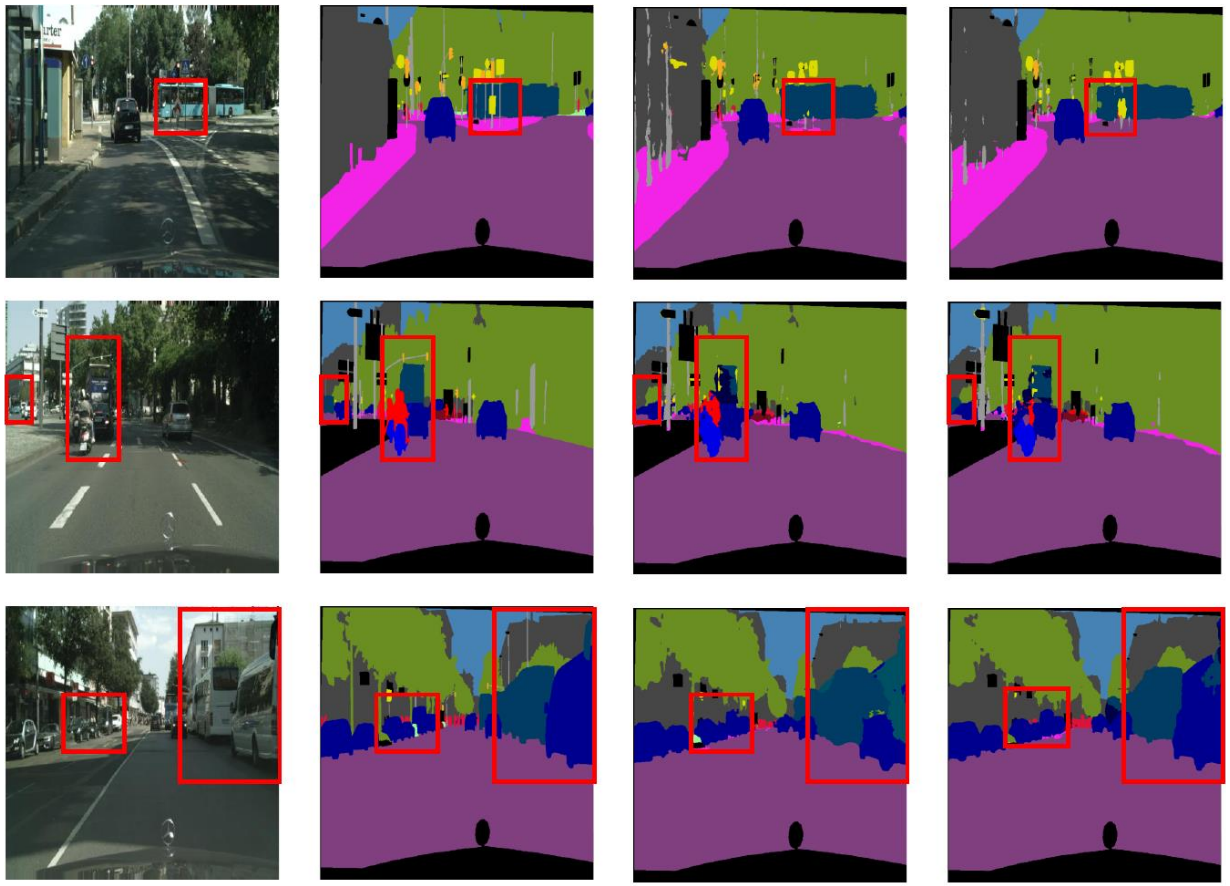
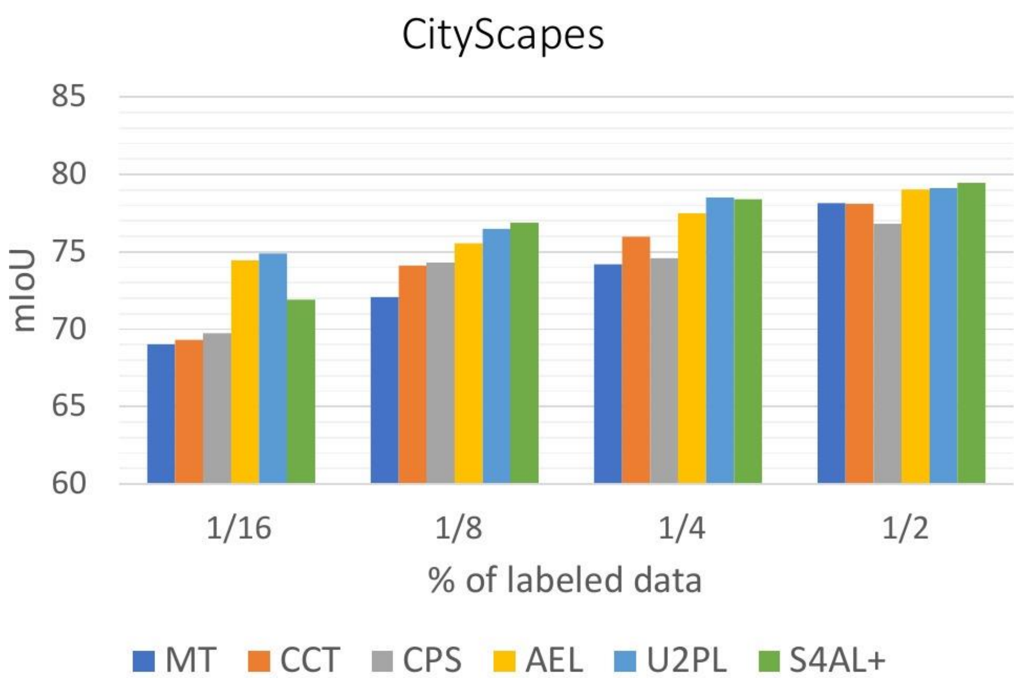


Image Ground Truth S4AL S4AL+

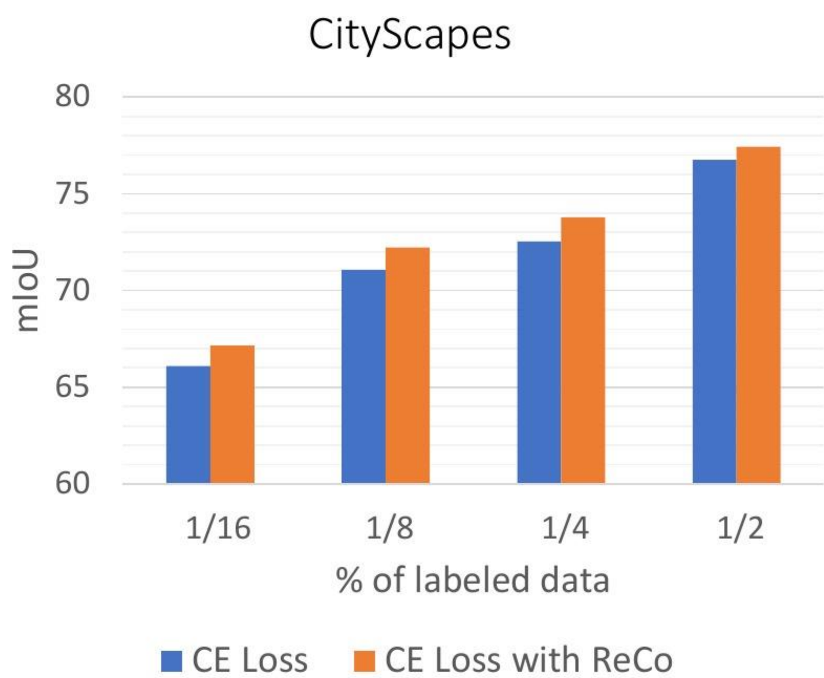
# Results

- Semi-Supervised Learning:
  - CityScapes



# Results

- Does representation learning help?
  - CityScapes



# Results

- Future Work
  - Knowledge Distillation

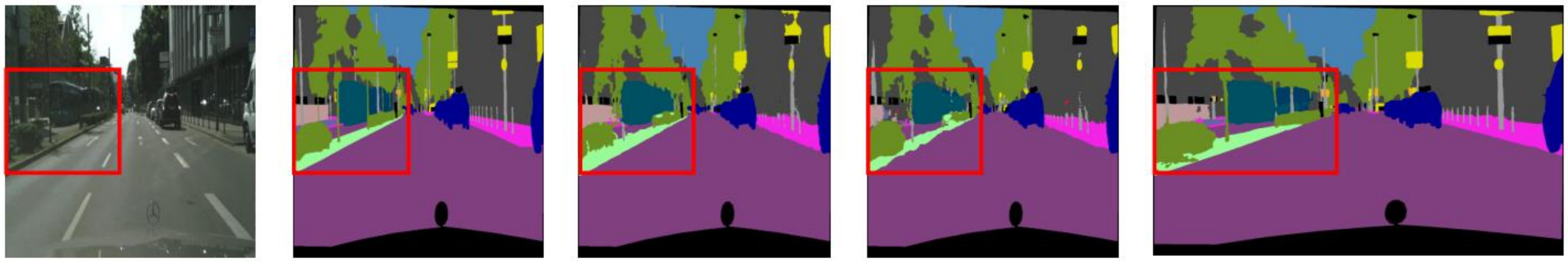


Image | Ground Truth | S4AL | S4AL+ | S4AL+ (SSL_R101D)

# Thank you for watching!