Un/Semi/Self-Supervised Learning

Kaihong Wang

Boston University Computer Science Department

A bit about me…

- Fourth-year PhD candidate in CS at BU working with Prof. Betke
- Generally interested in topics that can minimize human annotation efforts, i.e., label-efficient learning
- Enjoy watching cute kitten videos on the internet

Outline

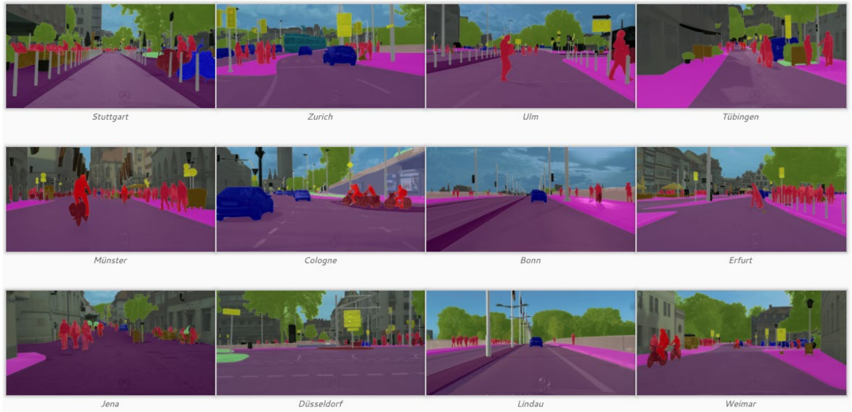
- Motivation & Definition
- Un/Semi-supervised Learning: examples
- Self-supervised Learning: examples
- Summary

Motivation

Motivation

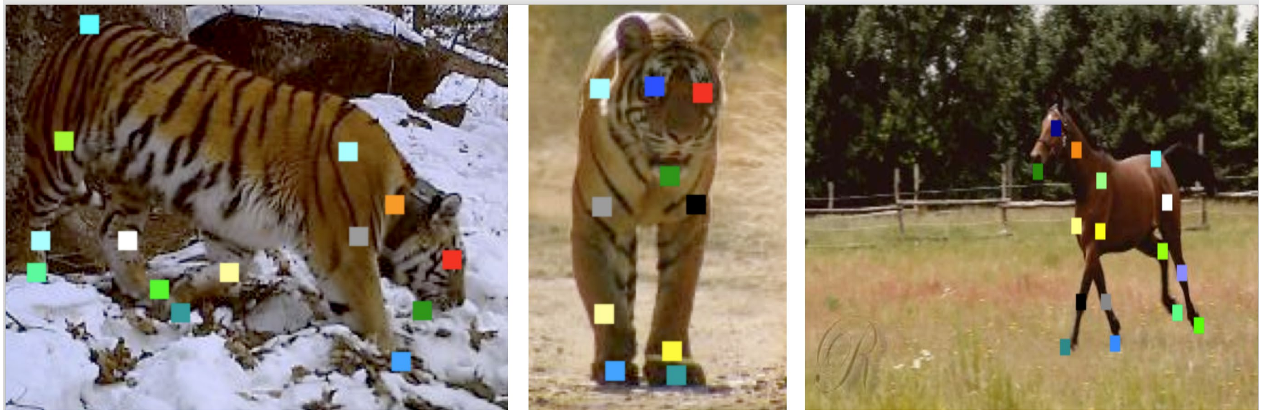
- Manual annotation can be expensive, laborious and subjective

Semantic segmentation[1]



Annotation and quality control in a semantic segmentation dataset required more than 1.5 hour on average for a single image[2]

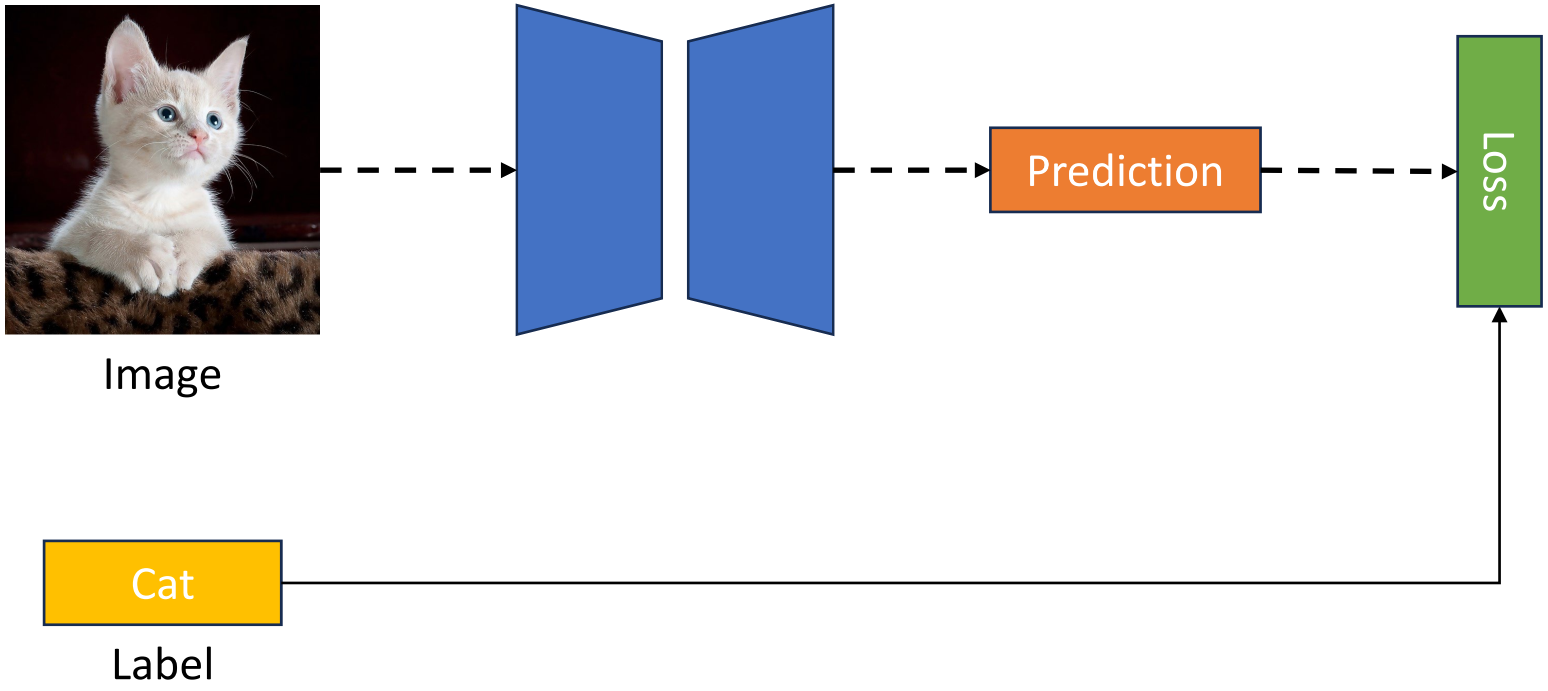
Pose estimation[3]



Manual annotation of key points in pose estimation tasks can be subjective and noisy while the accuracy is barely unverifiable

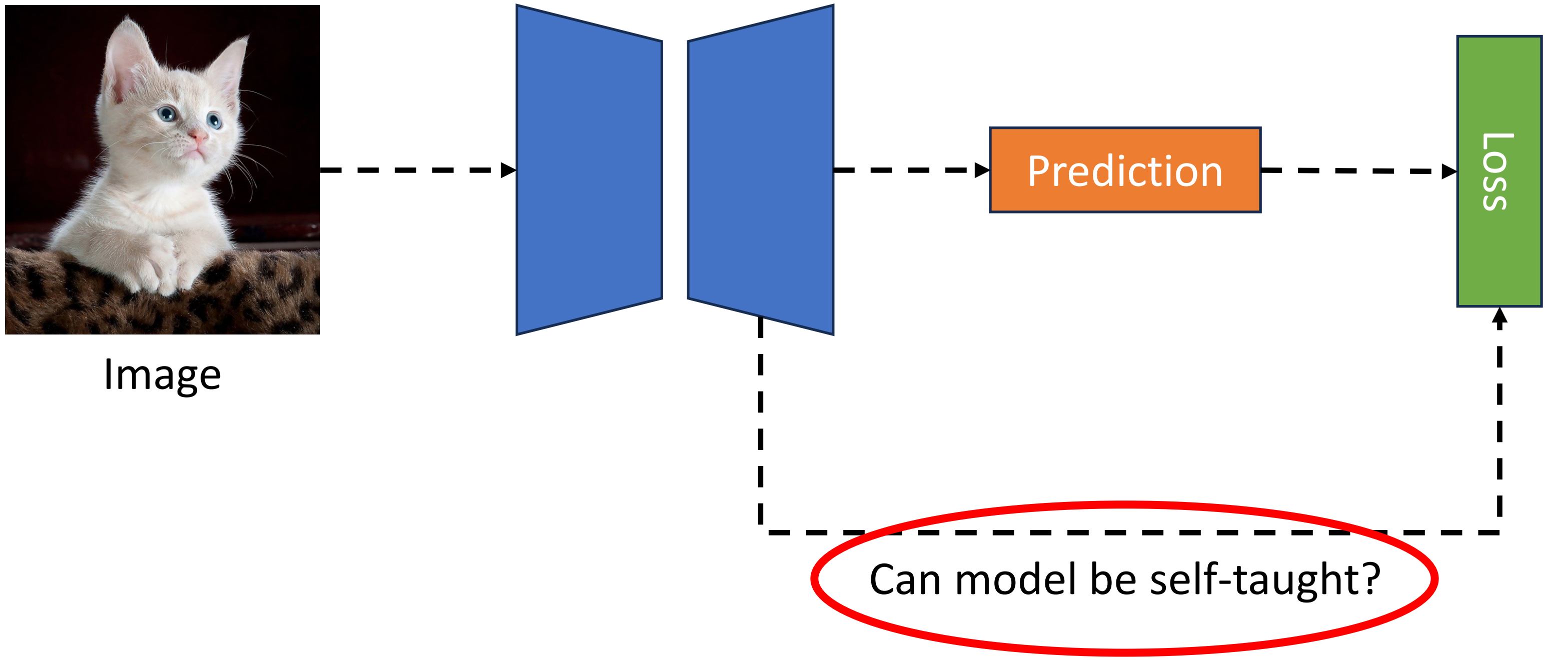
[1] https://www.cityscapes-dataset.com/examples/#fine-annotations
[2] Cordts M, Omran M, Ramos S, et al. The cityscapes dataset for semantic urban scene understanding[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 3213-3223
[3] Del Pero L, Ricco S, Sukthankar R, et al. Articulated motion discovery using pairs of trajectories[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 2151-2160.

Motivation

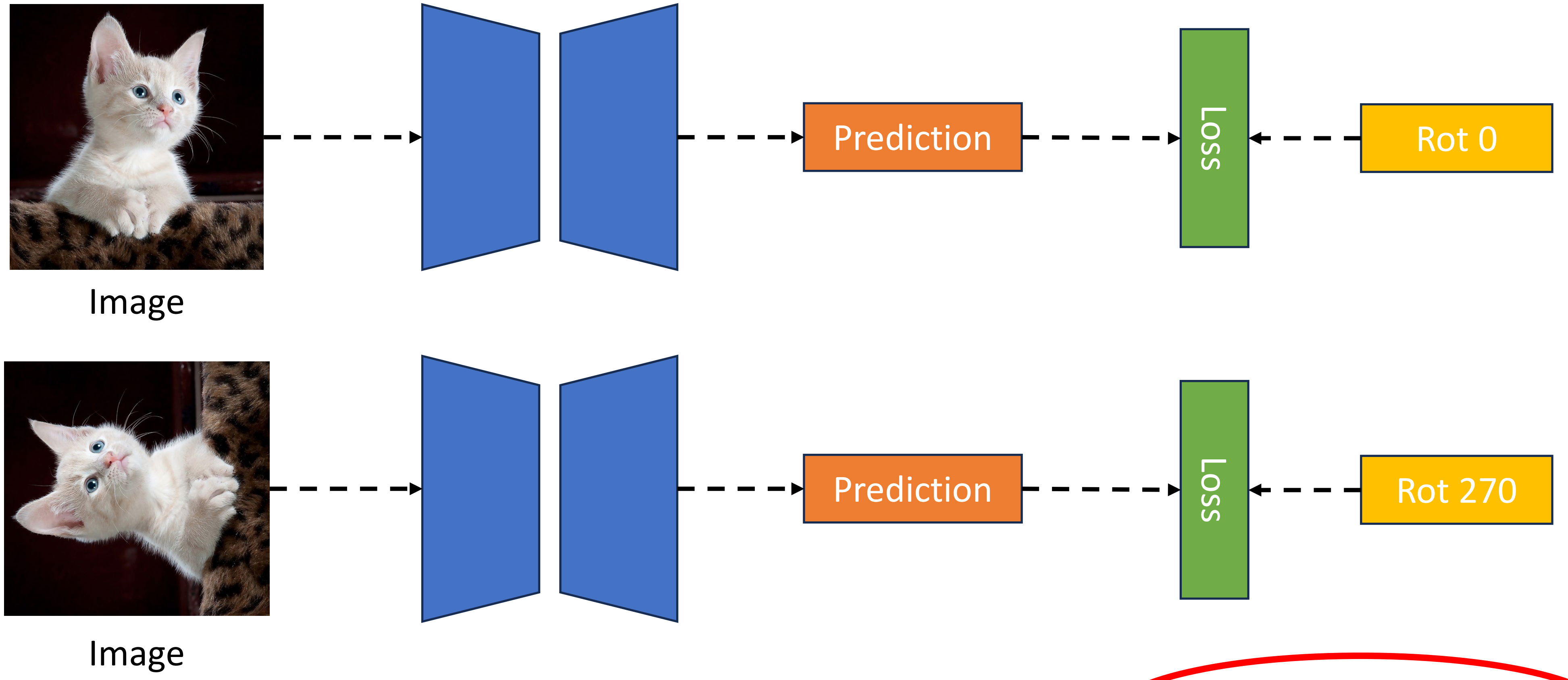


Typical Supervised Learning Pipeline

Motivation: Can we learn with no label?

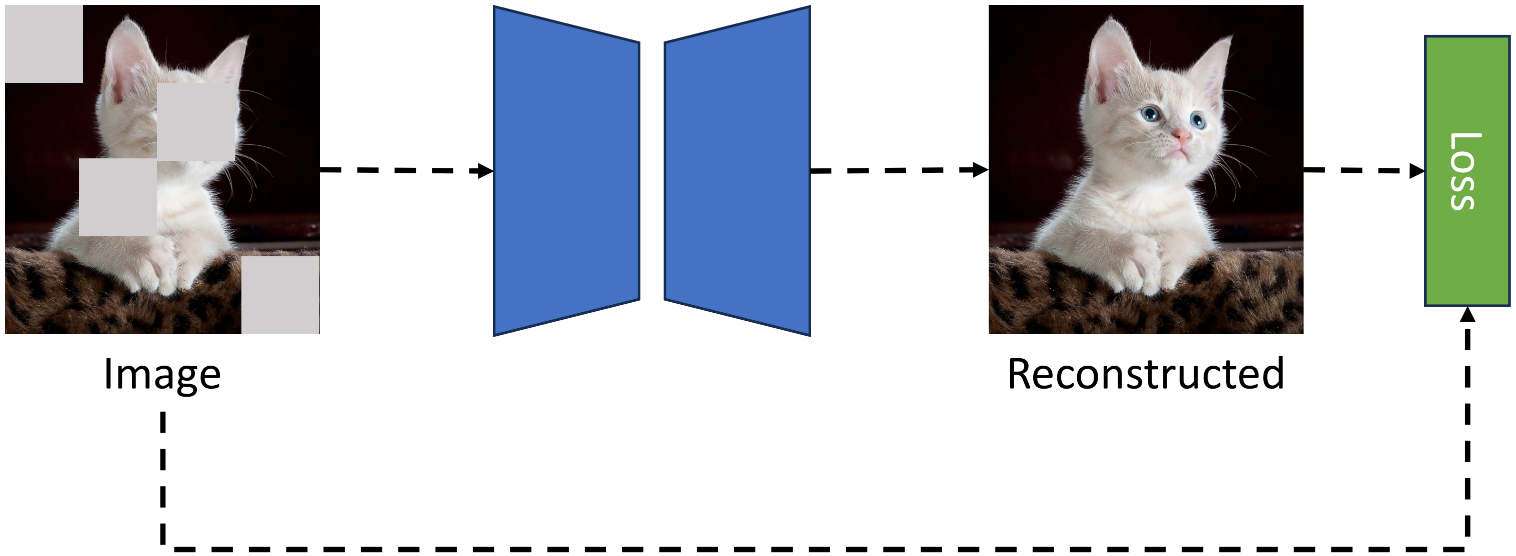


Motivation: Can we learn with no label?



How about creating new tasks?

Motivation: Can we learn with no label?



How about creating new tasks?

Definition & Comparsion

Task	Data	Goal	Example
Supervised Learning	Fully Labeled Data	Make correct prediction	Image classification
Unsupervised Learning	Unlabeled Data	Model underlying distribution of data	Dimensionality reduction
Semi-supervised Learning	Few Labeled + Many Unlabeled Data	Make correct prediction	Image classification with unlabeled training images
Self-supervised Learning	Unlabeled Data	Learn representation	BERT, GPT, CLIP

Un/Semi-supervised Learning

Semi-supervised Learning

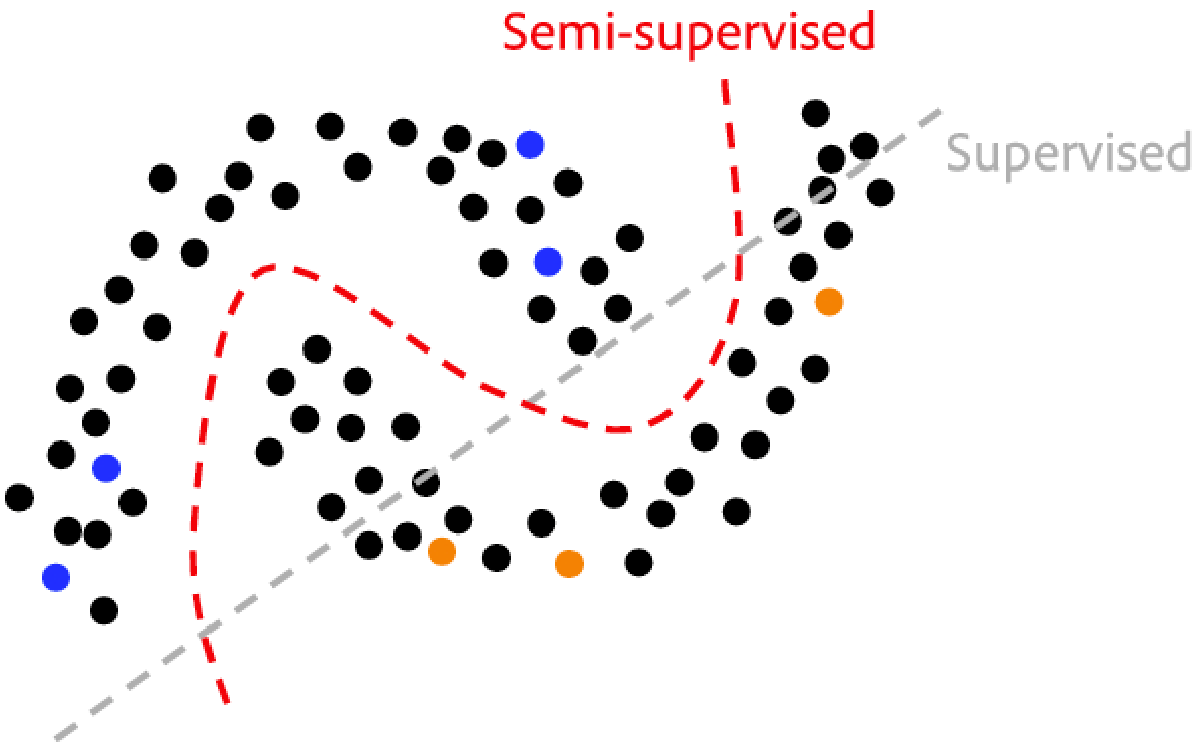


Figure Credit: Burkhart, Michael C., and Kyle Shan. "Deep low-density separation for semi-supervised classification." *International Conference on Computational Science*. Springer, Cham, 2020.
Link to Figure: https://link.springer.com/chapter/10.1007/978-3-030-50420-5_22

Un/Semi-supervised Learning: Examples

- Pseudo-labeling (Self-training)
- Entropy minimization
- Consistency regularization

Semi-supervised Learning: Pseudo-labeling

- Take the class which has the ***maximum predicted probability*** as the ground truth in the future
- Confidence threshold(s) is typically applied to select a ***subset of pseudo-labels*** with higher accuracy
- Pseudo-label can still be noisy and lead to severe ***bias towards major categories*** in the trained model

Lee, Dong-Hyun. "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks." *Workshop on challenges in representation learning, ICML*. Vol. 3. No. 2. 2013.

Semi-supervised Learning: Entropy Minimization

- Instead of learning from predictions in one-hot vectors, minimize the Shannan Entropy:

$$H[p(y|x)] = -\sum_{k=1}^{K} p(y=k|x)\log p(y=k|x)$$

- A “soft-assignment” of pseudo-label that penalizes the model with decision boundary in high-density regions and encourages the model to make predictions with higher confidence

Grandvalet, Yves, and Yoshua Bengio. "Semi-supervised learning by entropy minimization." *CAP* 367 (2005): 281-296.

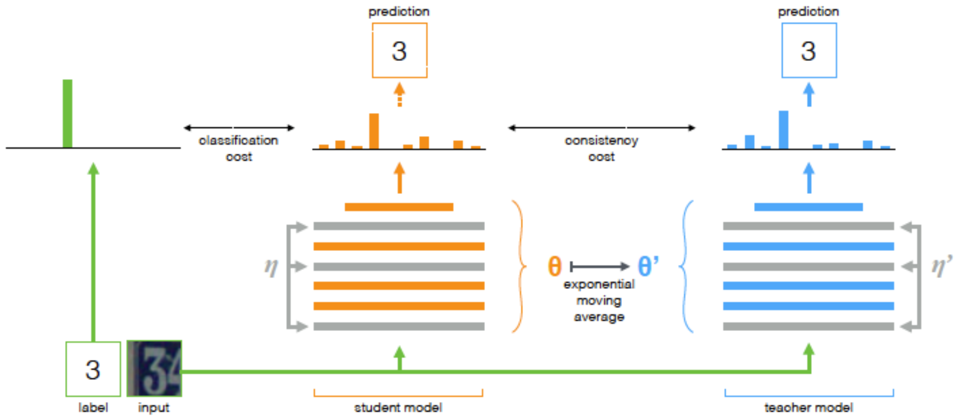
Semi-supervised Learning: Consistency Regularization

- Enforce the network(s) to give consistent predictions for an unlabeled image with different perturbations

- Decision boundary will be pushed towards the low-density regions, as it is sensitive to the difference in the distance to the decision boundary

Consistency Regularization: Mean Teacher

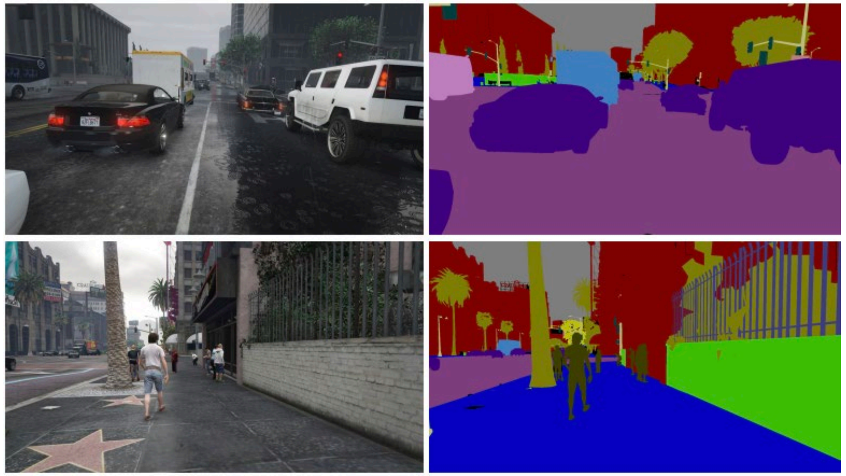
- Includes a student network and its temporal ensemble, a teacher network
- Input image is randomly perturbed
- Consistency loss is calculated between the output of the networks to train the student network
- Regular supervised loss is also applied for the labeled data



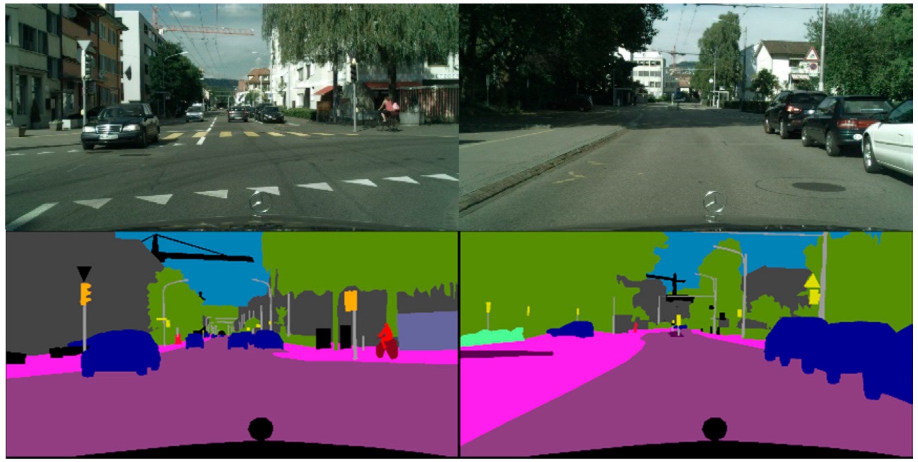
Tarvainen, Antti, and Harri Valpola. "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results." *arXiv preprint arXiv:1703.01780* (2017).

An Alternative Solution: Learning from Synthetic Data

- Although the manual annotation for real images could be expensive and unreliable, its barely costly in virtual domains



Synthetic data
(GTAV)

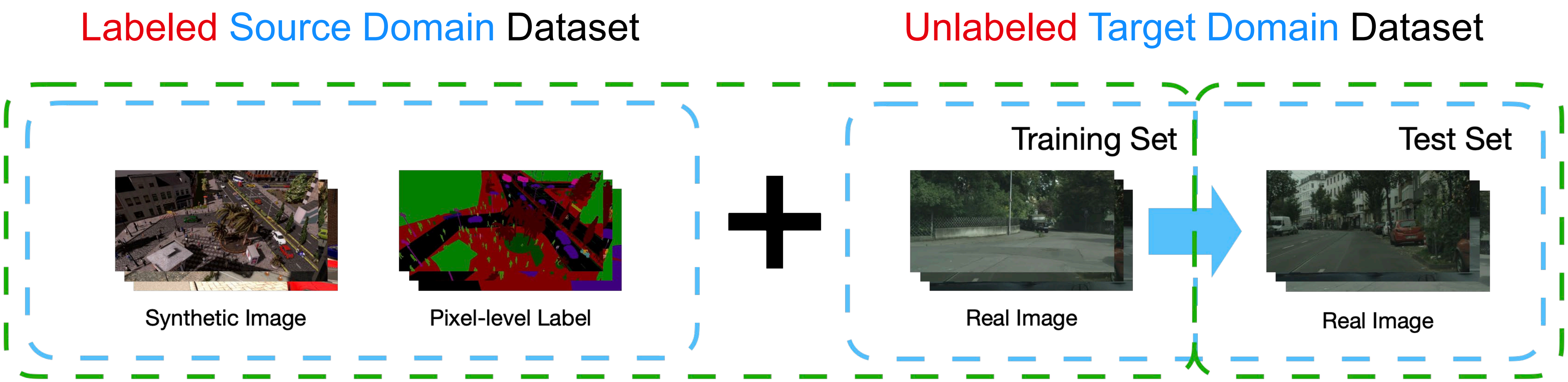


Real data
(Cityscapes)

Richter, Stephan R., et al. "Playing for data: Ground truth from computer games." *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*. Springer International Publishing, 2016.
Cordts, Marius, et al. "The cityscapes dataset." *CVPR Workshop on the Future of Datasets in Vision*. Vol. 2. 2015.

Definition: Unsupervised Domain Adaptation

- Learning from a **labeled** source domain dataset (synthetic) and an **unlabeled** target domain dataset (real-world)



Semi-Supervised Learning + Domain Adaptation -> Unsupervised Domain Adaptation

Definition: Unsupervised Domain Adaptation

- Learning from a **labeled** source domain dataset (synthetic) and an **unlabeled** target domain dataset (real-world)

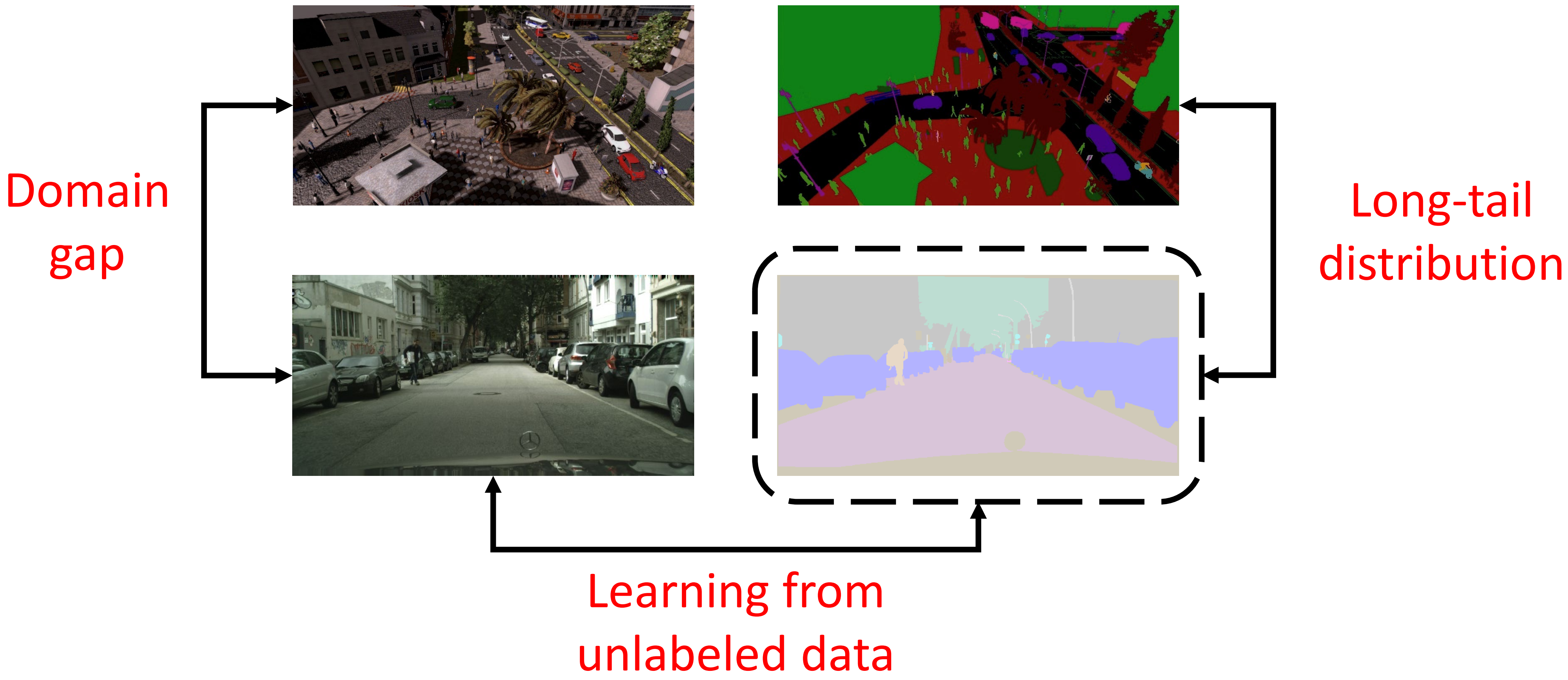
Given:

- Source dataset $D^s = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$ from a source domain S
- Target dataset $D^t = \{(x_i^t)\}_{i=1}^{N_t}$ from a target domain T

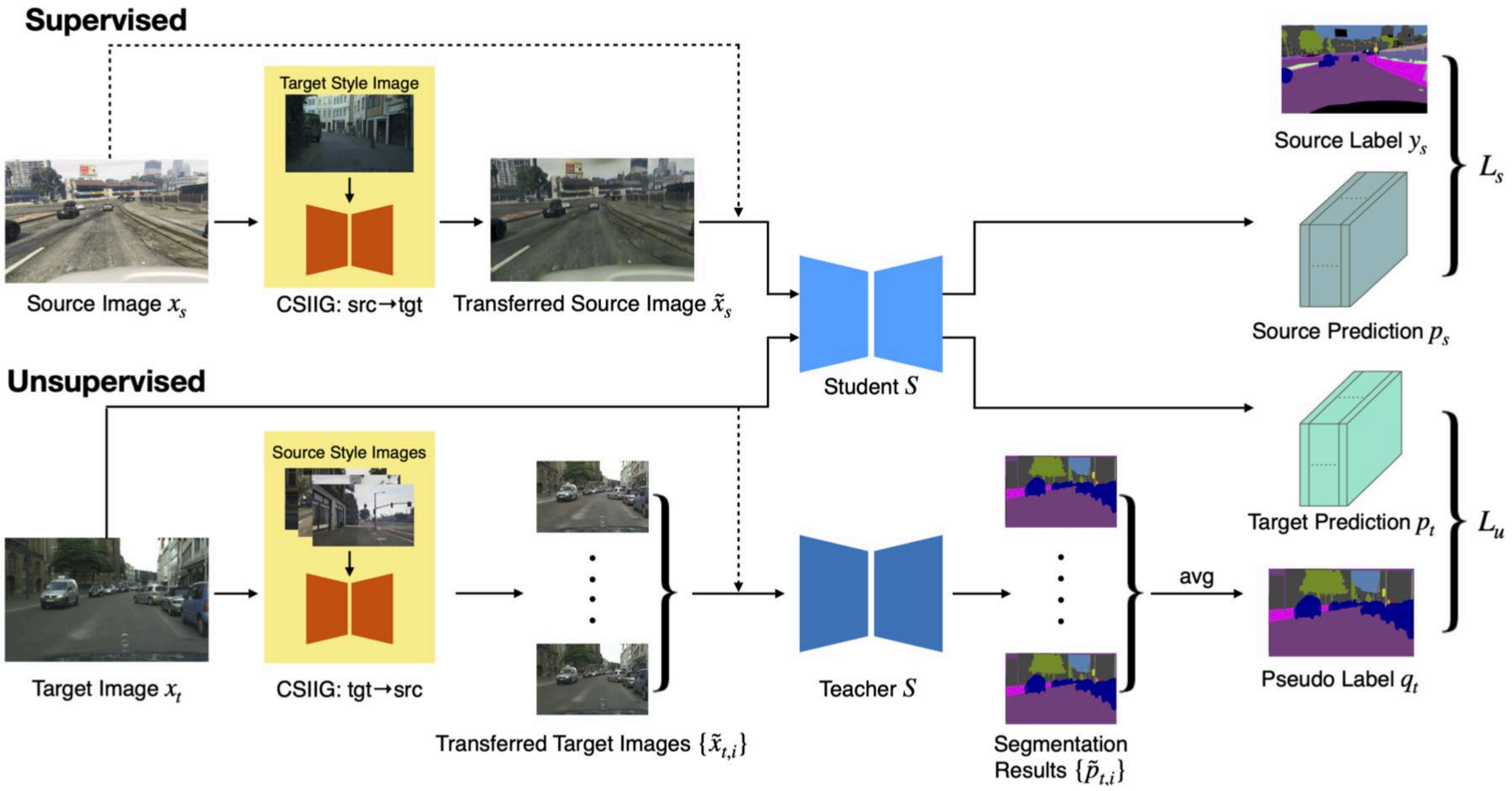
Aim:

- Optimize the performance of a task model F on target domain T

Challenge

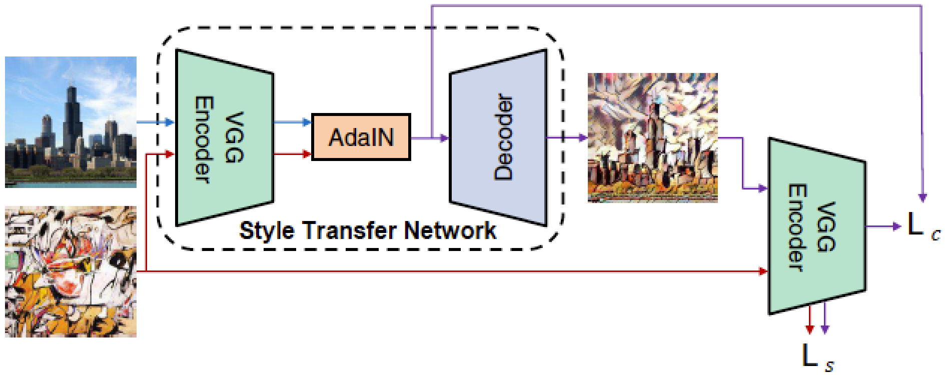

Domain gap
Long-tail distribution
Learning from unlabeled data

BiSIDA



Wang, Kaihong, Chenhongyi Yang, and Margrit Betke. "Consistency Regularization with High-dimensional Non-adversarial Source-guided Perturbation for Unsupervised Domain Adaptation in Segmentation." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. No. 11. 2021.

BiSIDA

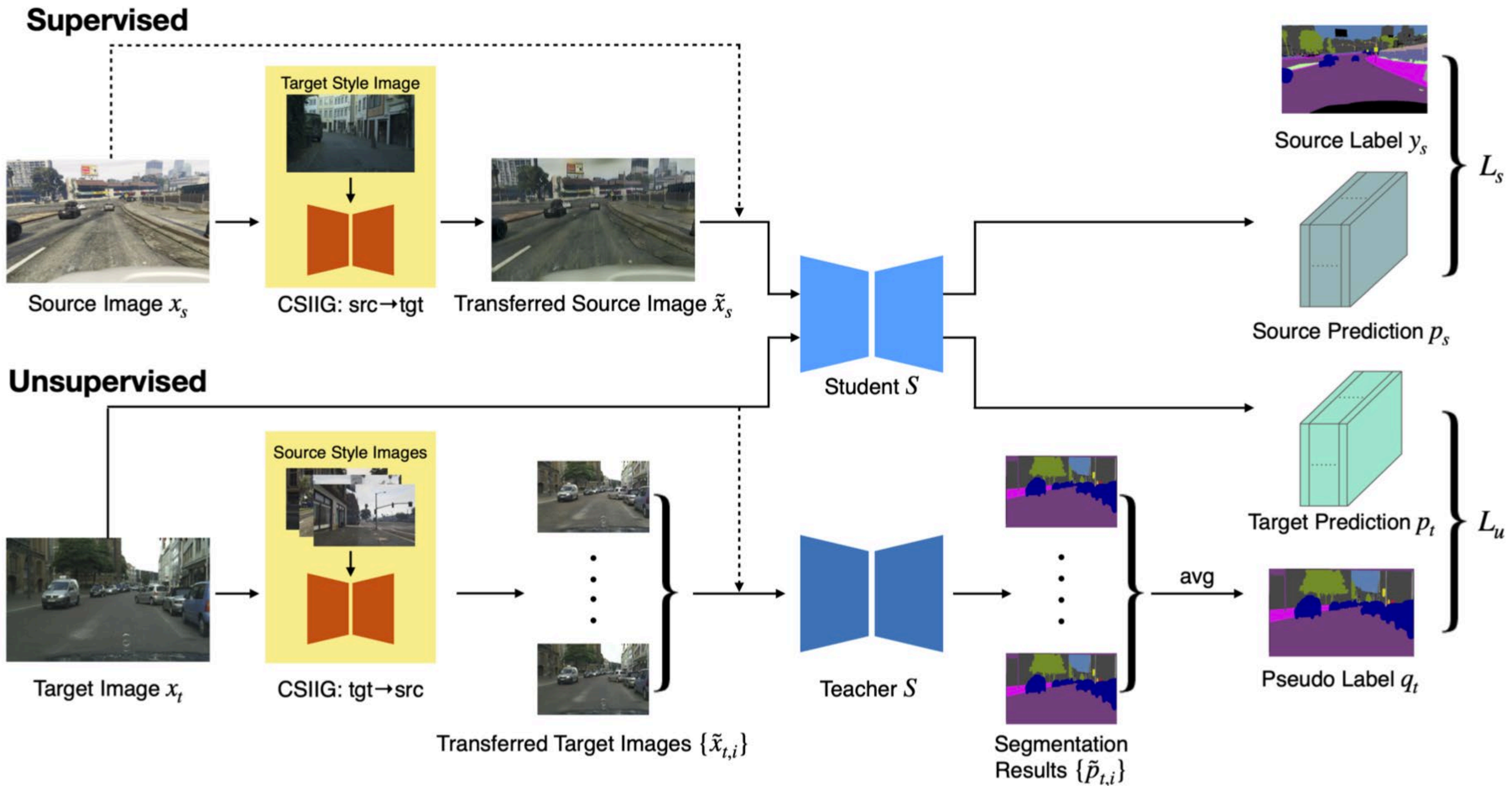


Given a content image $\boldsymbol{c}$, style image $\boldsymbol{s}$, a content-style trade-off parameter $\boldsymbol{\alpha}$, the intermediate feature after Adaptive Instance Norm $\hat{\boldsymbol{t}}$, and the content feature $\boldsymbol{t_c}$, and the decoder $\boldsymbol{g}$, we have the output:

$$G(c, s, \alpha) = g(\alpha\hat{t} + (1-\alpha)t_c)$$

Wang, Kaihong, Chenhongyi Yang, and Margrit Betke. "Consistency Regularization with High-dimensional Non-adversarial Source-guided Perturbation for Unsupervised Domain Adaptation in Segmentation." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. No. 11. 2021.

BiSIDA



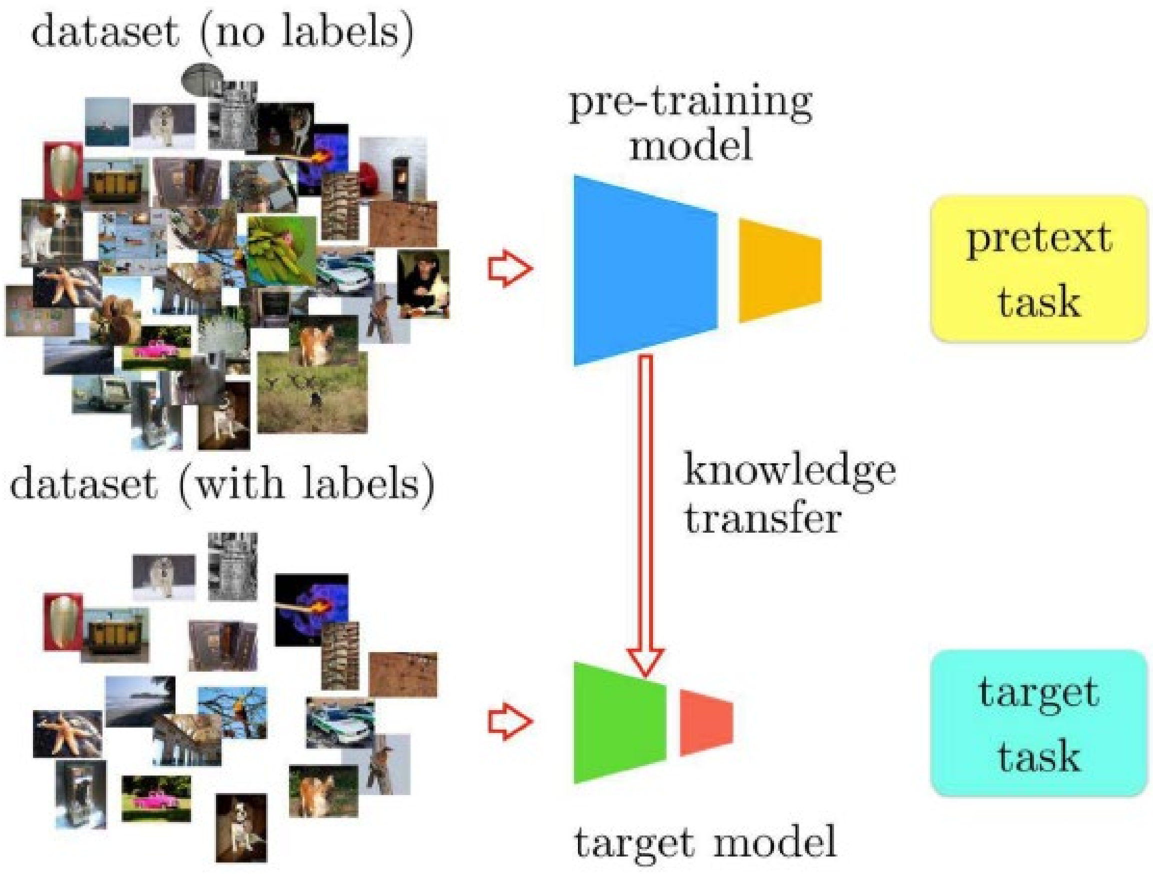
Wang, Kaihong, Chenhongyi Yang, and Margrit Betke. "Consistency Regularization with High-dimensional Non-adversarial Source-guided Perturbation for Unsupervised Domain Adaptation in Segmentation." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. No. 11. 2021.

Self-supervised Learning

Self-supervised Learning

- Relation to unsupervised learning?
  - Self-supervised learning can be considered a ***subset*** of unsupervised learning as “unsupervised feature learning”
  - Unsupervised learning algorithms are more ***data-centric***: focus on uncovering underlying distribution and relationship in data
  - Self-supervised learning algorithms are more ***model-centric***: focus on pretraining a model for following downstream tasks that need extra supervised finetuning
  - Self-supervised learning is an emerging, trending topic in recent years arise from the rapid development in deep learning

Self-supervised Learning: a novel paradigm



Noroozi, Mehdi, et al. "Boosting self-supervised learning via knowledge transfer." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

Self-supervised Learning: Examples

- Rotation Prediction
- Masked Autoencoders
- SimCLR

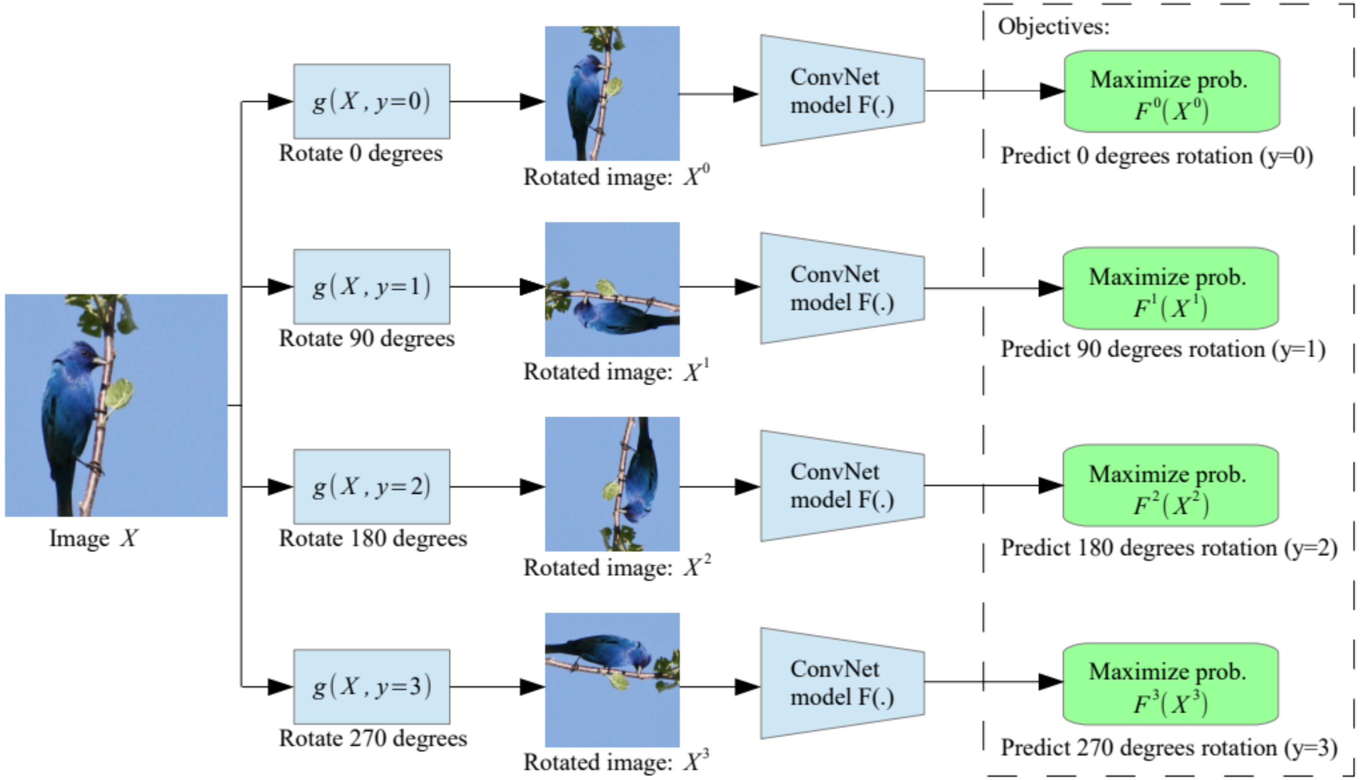
Self-supervised Learning: Rotation Prediction

- Rotate images by a predefined angles

- Teach the model identify the rotation

Gidaris, Spyros, Praveer Singh, and Nikos Komodakis. "Unsupervised representation learning by predicting image rotations." *arXiv preprint arXiv:1803.07728* (2018).

Self-supervised Learning: Rotation Prediction



Gidaris, Spyros, Praveer Singh, and Nikos Komodakis. "Unsupervised representation learning by predicting image rotations." *arXiv preprint arXiv:1803.07728* (2018).

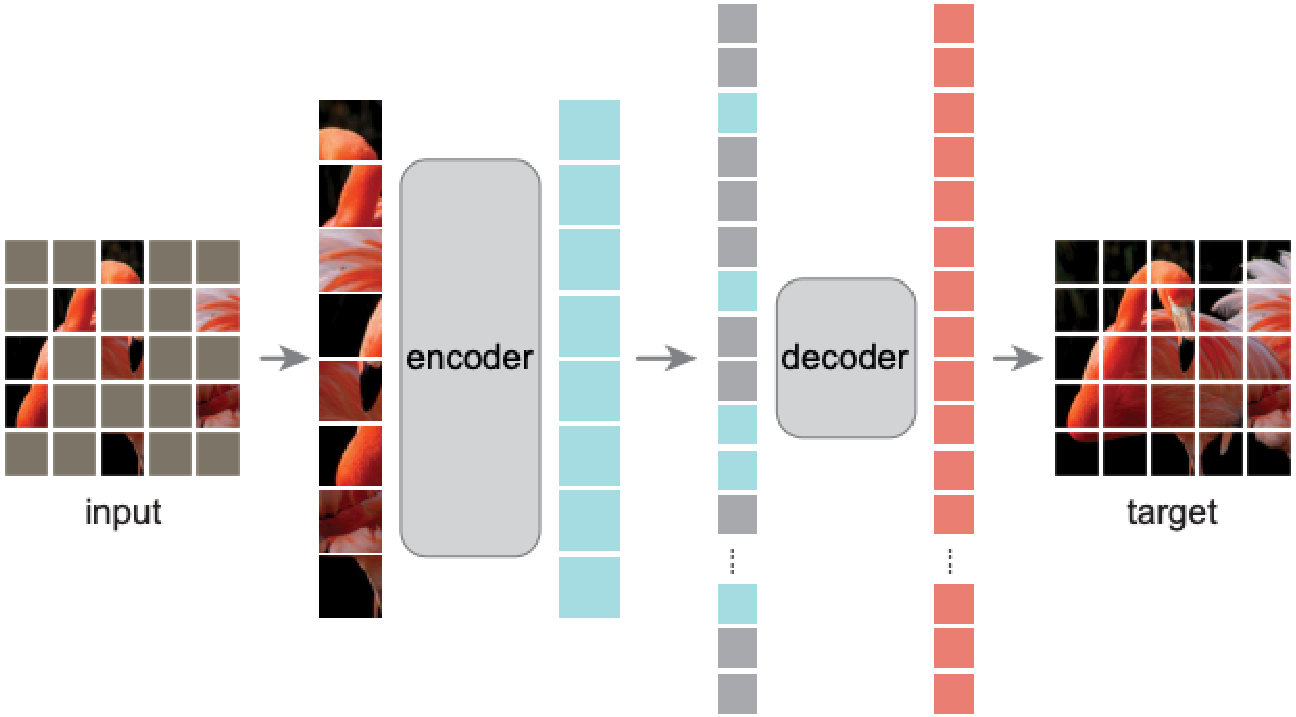
Self-supervised Learning: MAE

- Mask the input image and input only the visible parts

- Teach the model to reconstruct the masked region

He, Kaiming, et al. "Masked autoencoders are scalable vision learners." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022.

Self-supervised Learning: Mask Autoencoders



He, Kaiming, et al. "Masked autoencoders are scalable vision learners." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022.

Self-supervised Learning: SimCLR

- Each image produces a pair of samples via separate augmentations
- Deploys ***contrastive loss***
- Repel pairs from different origins and attract pairs from the same

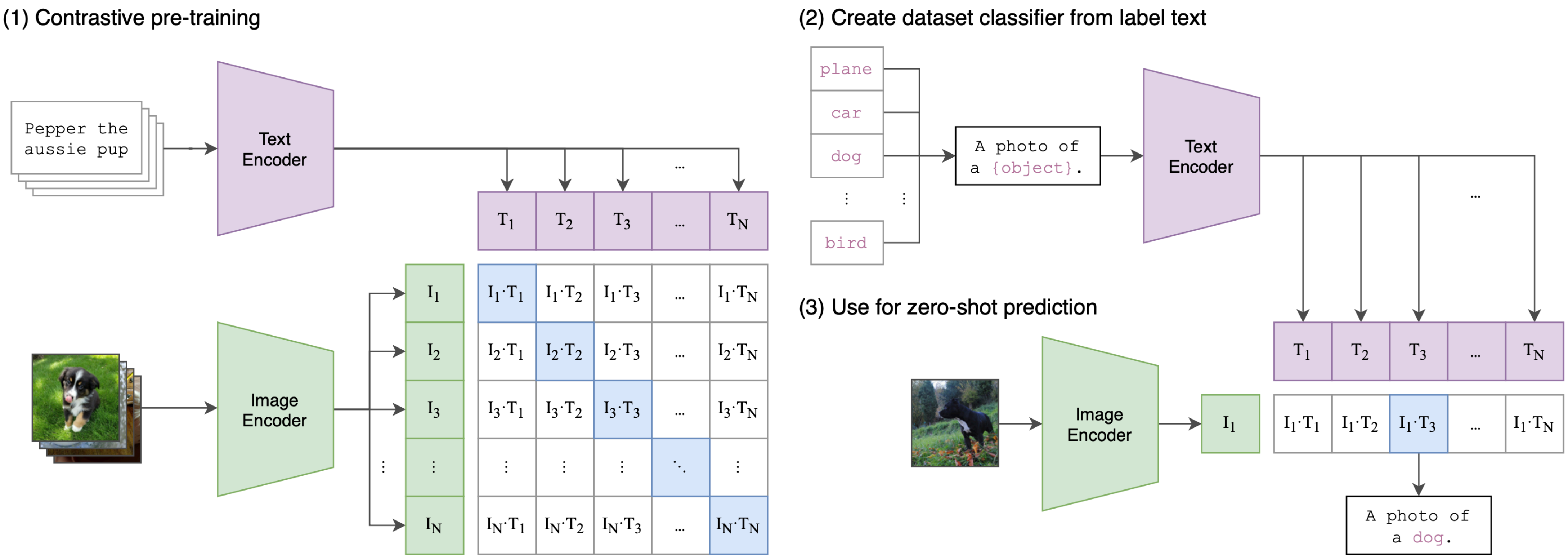
$$\ell_{i,j} = -\log \frac{\exp(\mathrm{sim}(\boldsymbol{z}_i, \boldsymbol{z}_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(\mathrm{sim}(\boldsymbol{z}_i, \boldsymbol{z}_k)/\tau)},$$

Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." *International conference on machine learning*. PMLR, 2020.

Self-supervised Learning: SimCLR

https://medium.com/one-minute-machine-learning/simclr-explained-in-simple-terms-3fa69af45ff9

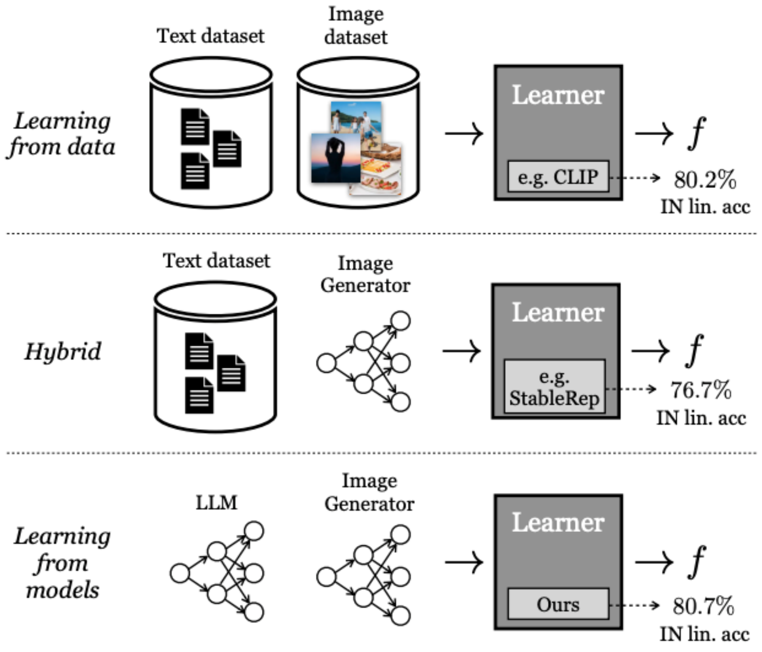
Contrastive learning extension: CLIP

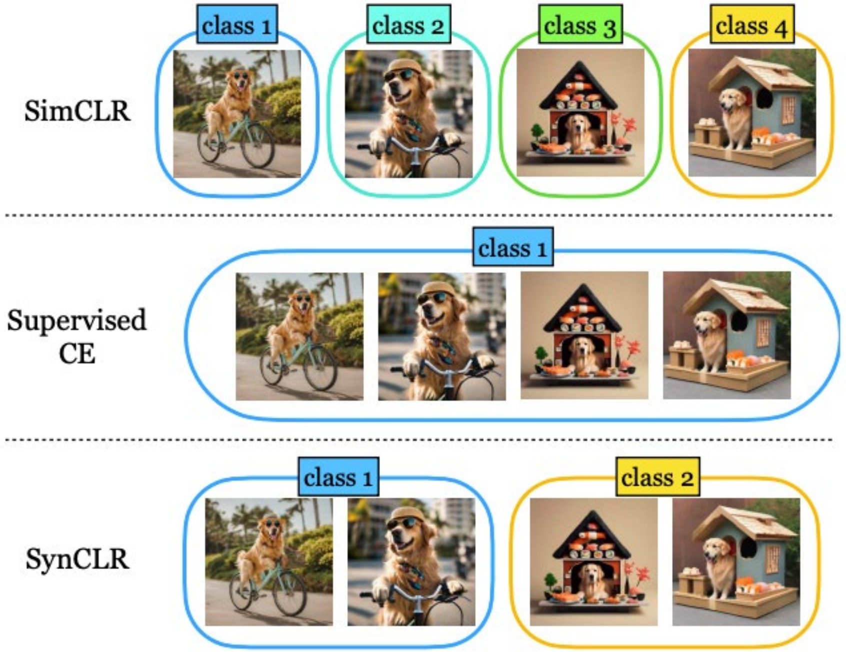


https://lilianweng.github.io/posts/2021-05-31-contrastive/
Radford, Alec, et al. "Learning transferable visual models from natural language supervision." *International conference on machine learning*. PMLR, 2021.

An Alternative Solution: Learning from Synthetic Data (Again!)

- Synthetic data can be used for self-supervised learning now, too





Tian, Yonglong, et al. "Learning vision from models rivals learning vision from data." *arXiv preprint arXiv:2312.17742* (2023).

Summary

- Un/semi-supervised learning
  - Task specific approaches: learn to achieve a task with few/no labels
  - A series of well-established frameworks with a longer history even before deep learning

- Self-supervised learning
  - A more upstream topic that aims to enhance model's capacity to extract feature rather than tackling a specific task
  - An active research region where a lot of impactful works that lead to the "big model era"

Learning Objectives

- Motivations behind label-efficient learning
- Different topics and directions for label-efficient learning problems
- Relation between un/semi/self-supervised learning
- Representative approaches in un/semi/self-supervised learning
- Common design choices under different scenarios

Thank You!