Active Learning for Video Classification

Outline

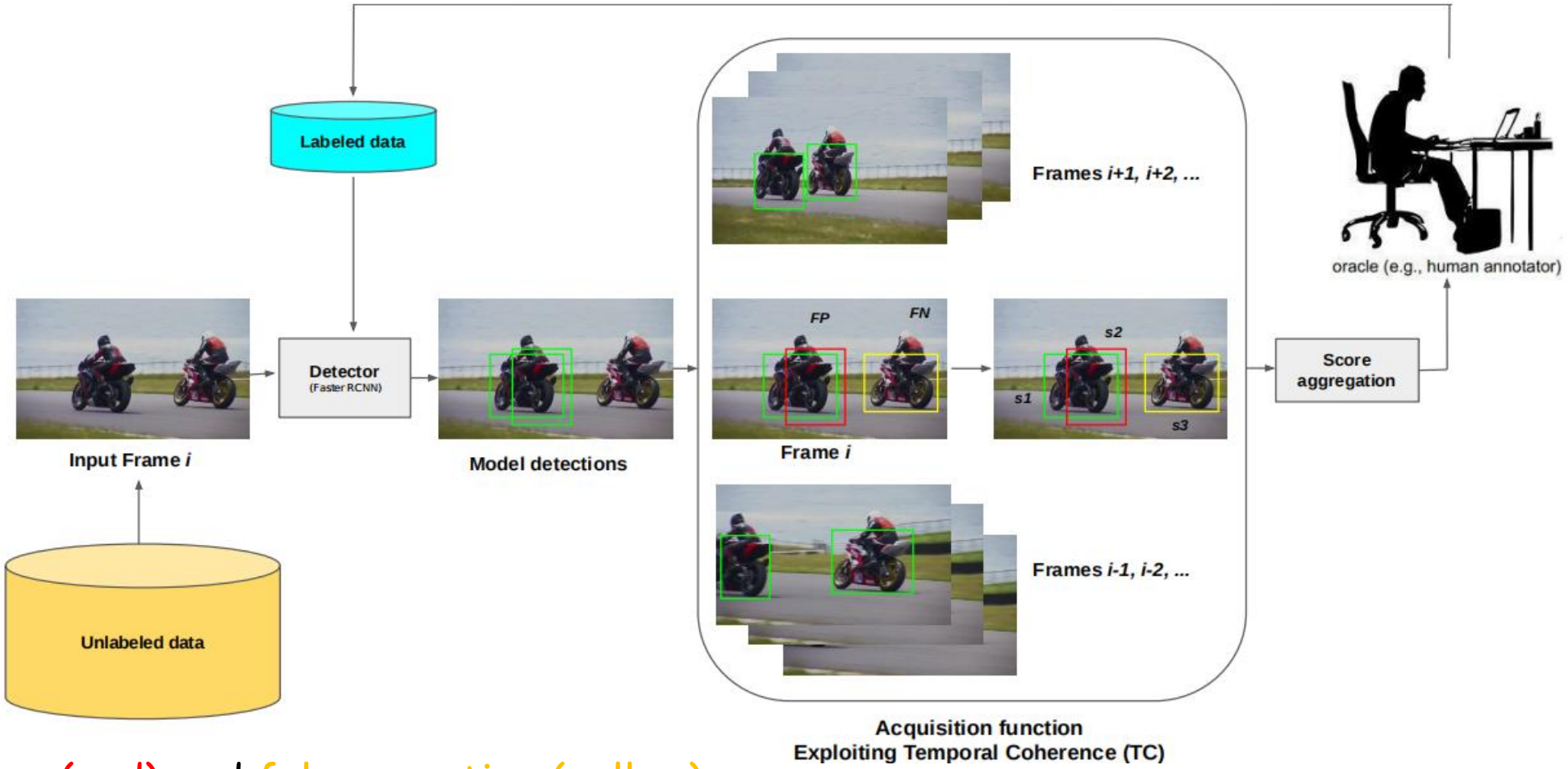
- Active Learning + Video
- Open-set Annotation

Active Learning for Video Classification

- VideoSSL: Semi-Supervised Learning for Video Classification – WACV'21
- Efficient Video Classification Using Fewer Frames – CVPR'19
- Temporal Coherence for Active Learning in Videos – ICCV'19
- Video Annotation and Tracking with Active Learning – NIPS'11

Active Learning for Video Classification

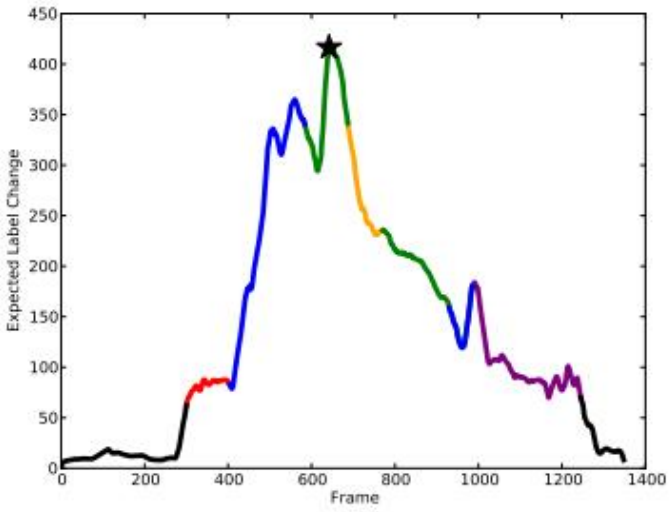
- Temporal Coherence for Active Learning in Videos – ICCV'19
  - 时序相关性（temporal coherence）。相邻的帧之间里所包含的目标是一致的概念，例如相邻帧里没有该目标，当前识别器却错误的识别出当前帧有目标。



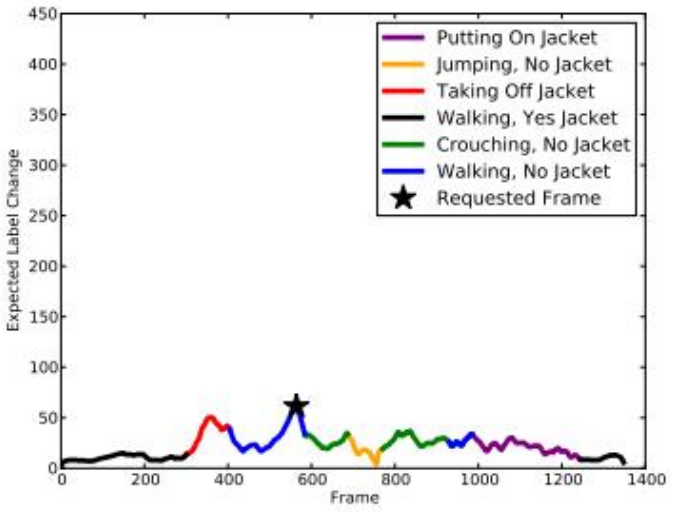
False positive (red) and false negative (yellow) errors.

Active Learning for Video Classification

- Video Annotation and Tracking with Active Learning – NIPS'11
  - 对于路径追踪时，其实只需要标注特定的帧就行了。例如追踪器可能以黑色夹克为目标进行追踪，但夹克脱掉后，目标追踪可能会出现问题。因此我们只需要在最关键的帧（脱完夹克后标注人，让追踪器去关注人而不是夹克）标注一下即可。



(a) One click: Initial frame only



(b) Two clicks: Initial and requested frame

(c) Training

(d) Walking, Yes Jacket

(e) Taking Off Jacket

(f) Walking, No Jacket

- **One click**：只标记初始帧，对应为绿色框
- **Two clicks**：标记初始帧以及需求帧，红色框

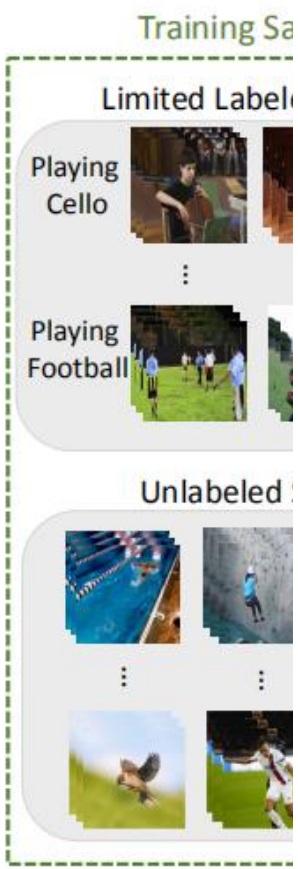
Formalize

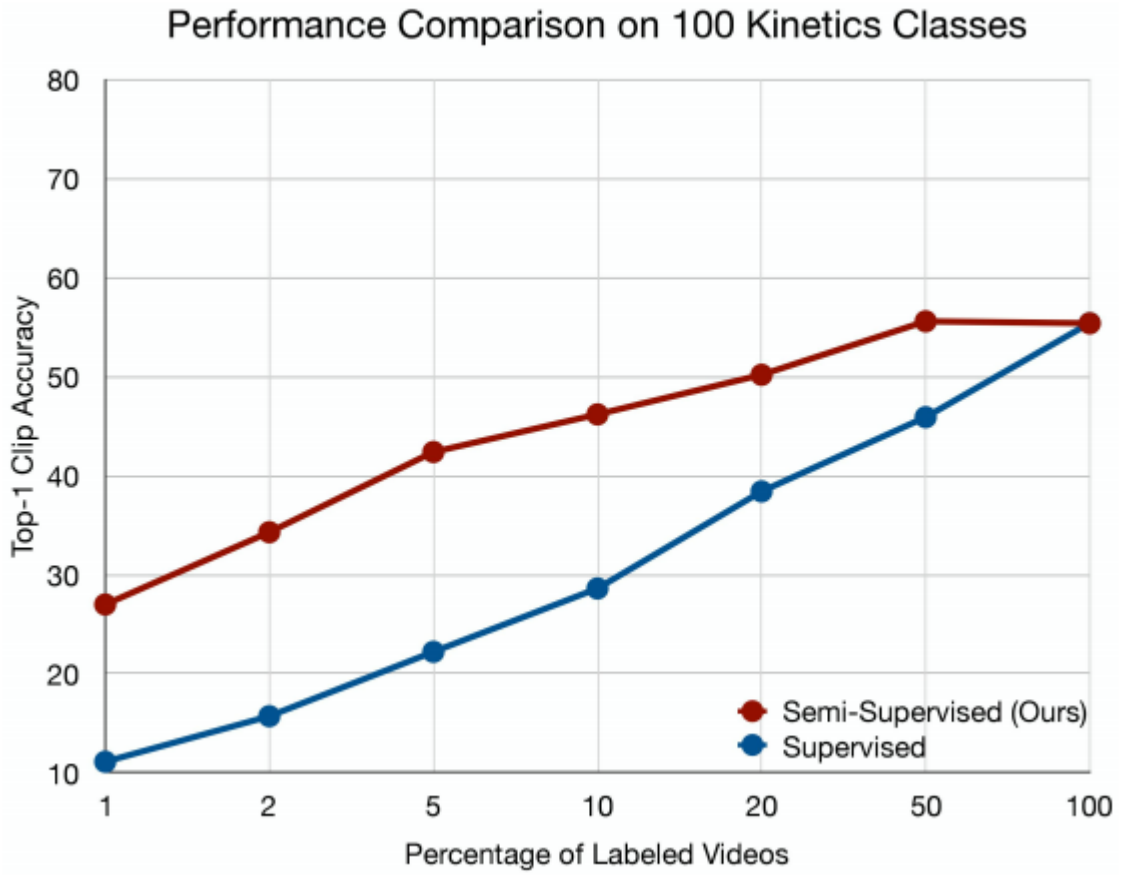
- $curr_{0:T}$: the current best estimation
- ζ: a set of user annotations
- $b_{0:T}^{gt}$: the ground-truth path
- b_t^{gt}: the annotation b_t^{gt} of frame t
- $next_{0:T}(b_t^{gt})$: the next best estimation
- $\zeta' = \zeta \sqcup b_t^{gt}$: the next annotation sets

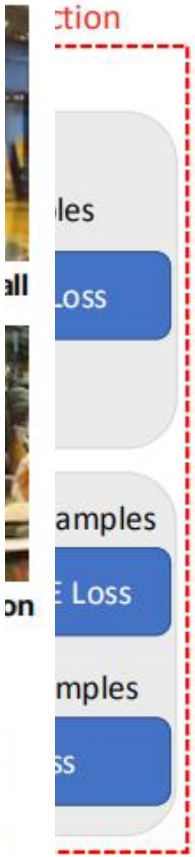
$$t^{opt} = \underset{0 \le t \le T}{\operatorname{argmin}} \sum_{j=0}^{T} \operatorname{err}(b_j^{gt}, \operatorname{next}_j(b_t^{gt}))$$

Active Learning for Video Classification

- VideoSSL: Semi-Supervised Learning for Video Classification – WACV'21







$$L_d = - \sum_{\substack{v \in \{X \cup Z\} \\ a \in v}} \sum_{l=1}^{M} h^l(a) \log q^l(v)$$

$$\hat{y}_i^c = \begin{cases} T, & \text{if } p^c(z_i) \geq \delta \\ p^c(z_i), & \text{otherwise} \end{cases}$$

$$L_u = - \sum_{z_i \in Z} \sum_{c} \hat{y}_i^c \log p^c(z_i).$$

$$L_s = - \sum_{x_i \in X} \sum_{c} y_i^c \log p^c(x_i)$$

Active Learning for Video Classification

- Efficient Video Classification Using Fewer Frames – CVPR'19
  - 知识蒸馏虽然能够降低模型的复杂度，但没法降低计算的FLOPs（the number of float point operations）。动机是为了减少FLOPs，设计进一步的蒸馏方式，想法简单粗暴有效。

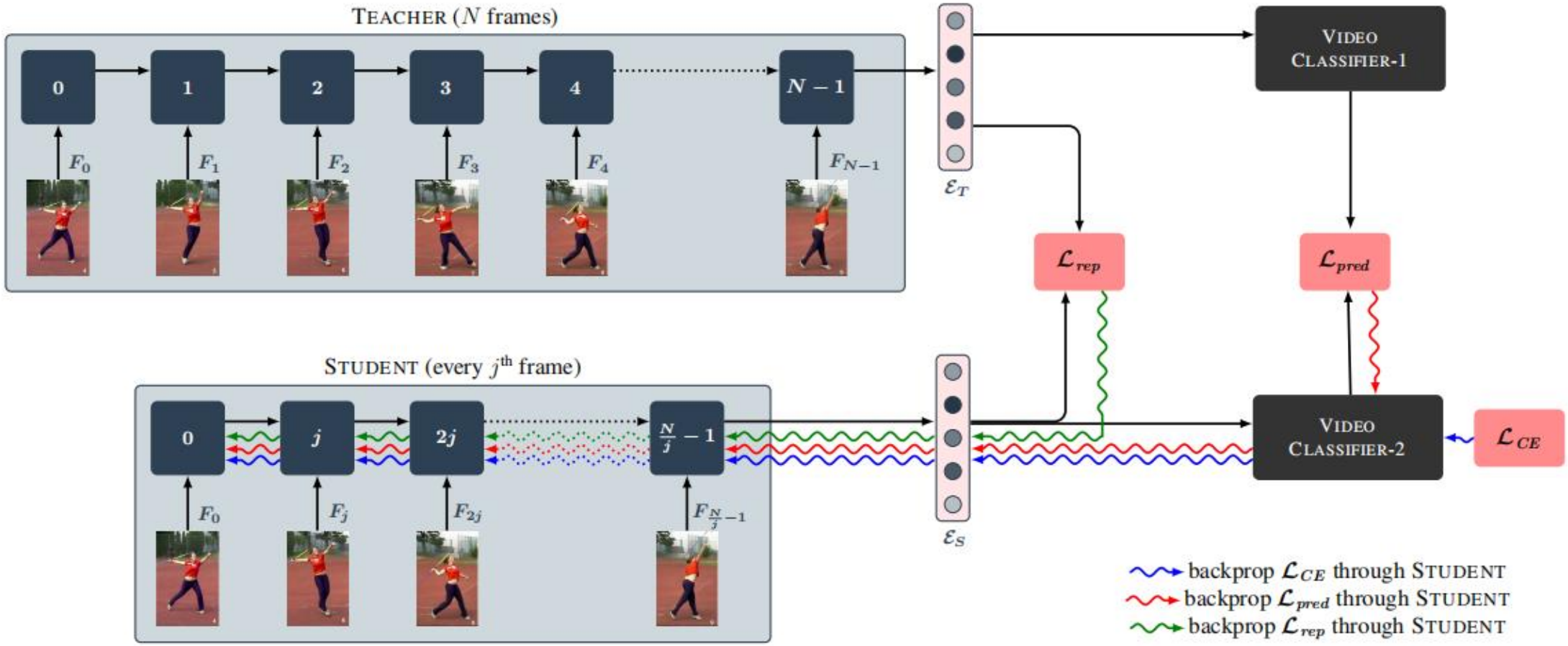


Figure 1: Architecture of TEACHER-STUDENT network for video classification

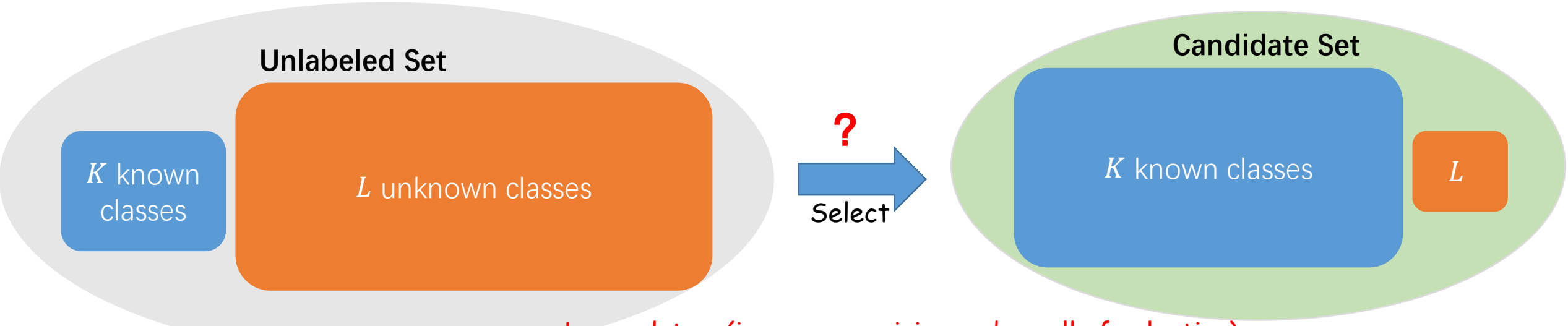
Close-set vs. Open-set

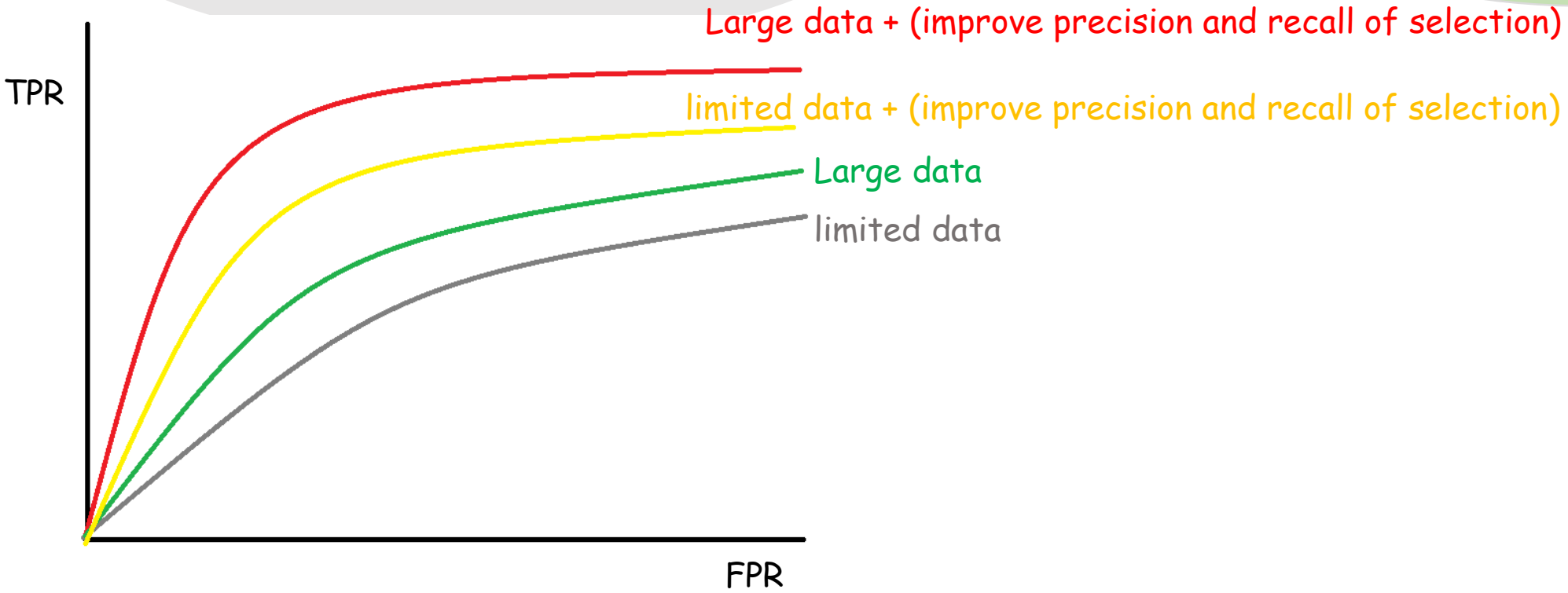
Close-set Classification

Open-set Classification

Open-set Annotation

Open-set Annotation

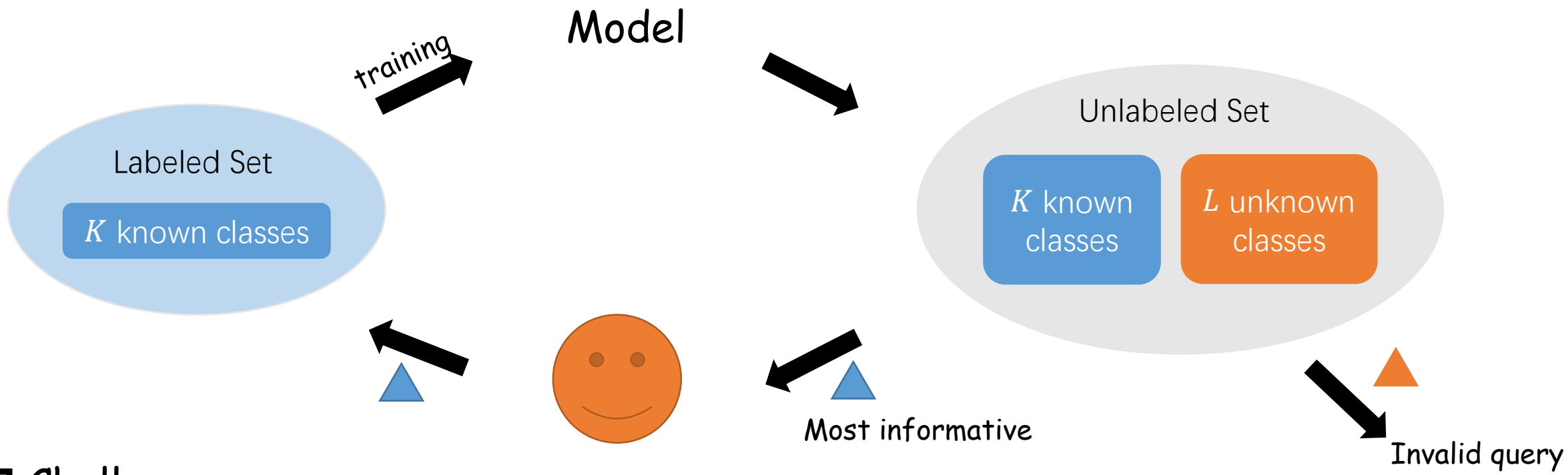




Open-set Annotation + AL

Motivation

- Reduce annotation cost. (especially in video tasks)
- More practical. (Unlabeled data collected from the Internet usually include non-object classes.)



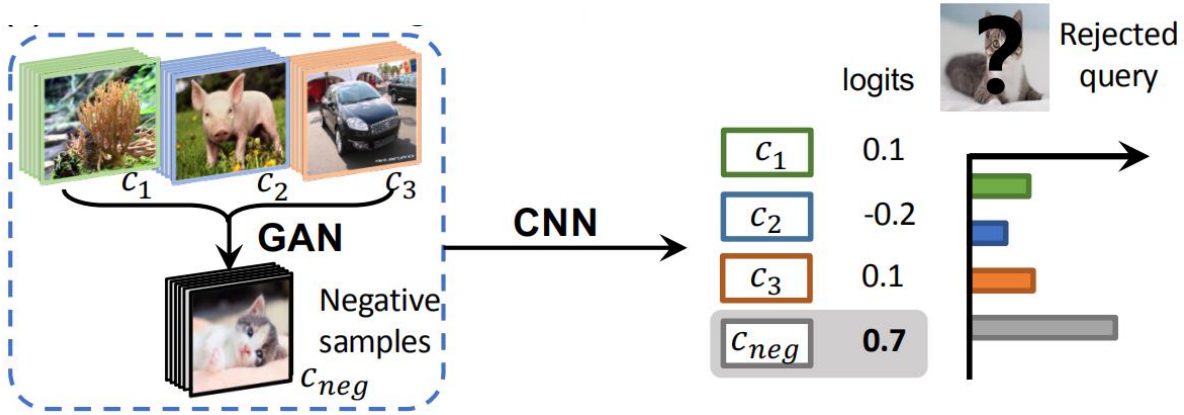
Challenges

- Common AL sampling methods tend to select unknown examples. (Uncertainty-based)
- How to select the most informative example from unlabeled set with K known classes.
- How to improve the performance of recall and precision.

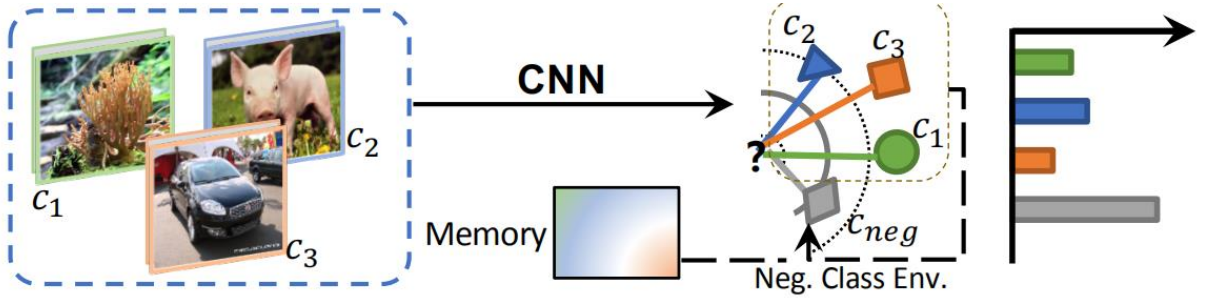
Few shot open-set recognition

Generative methods

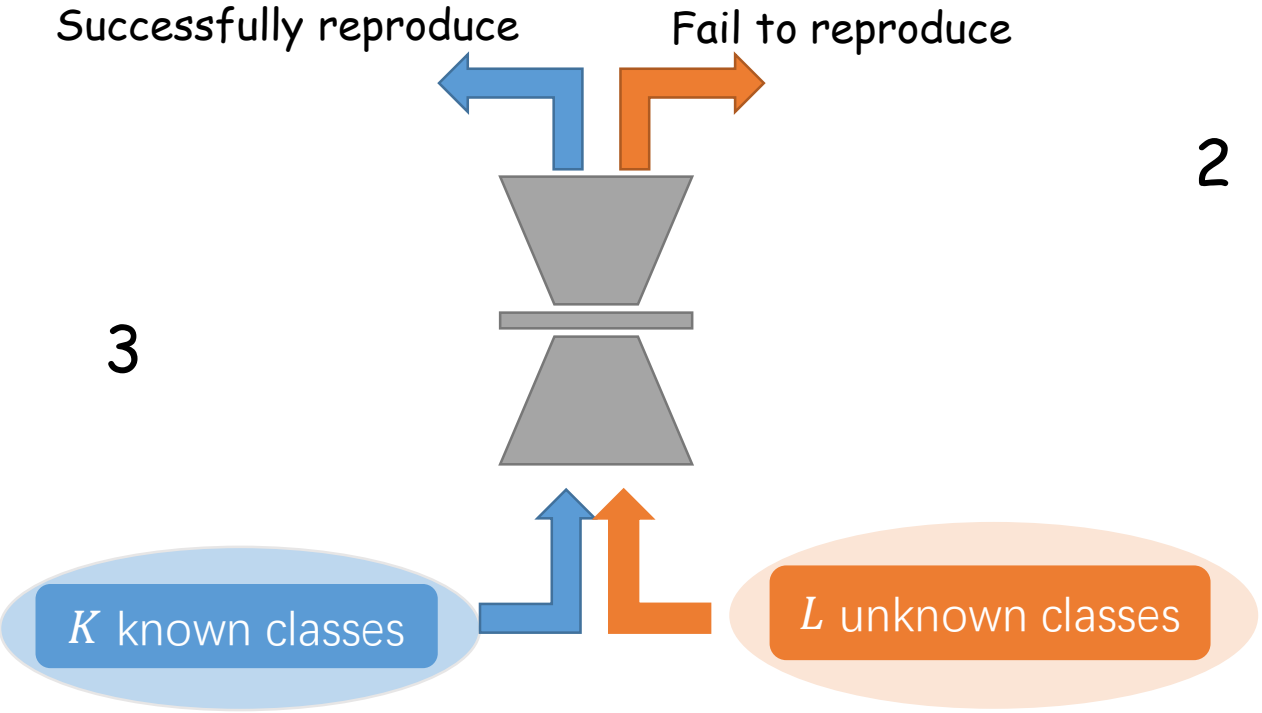
1



2

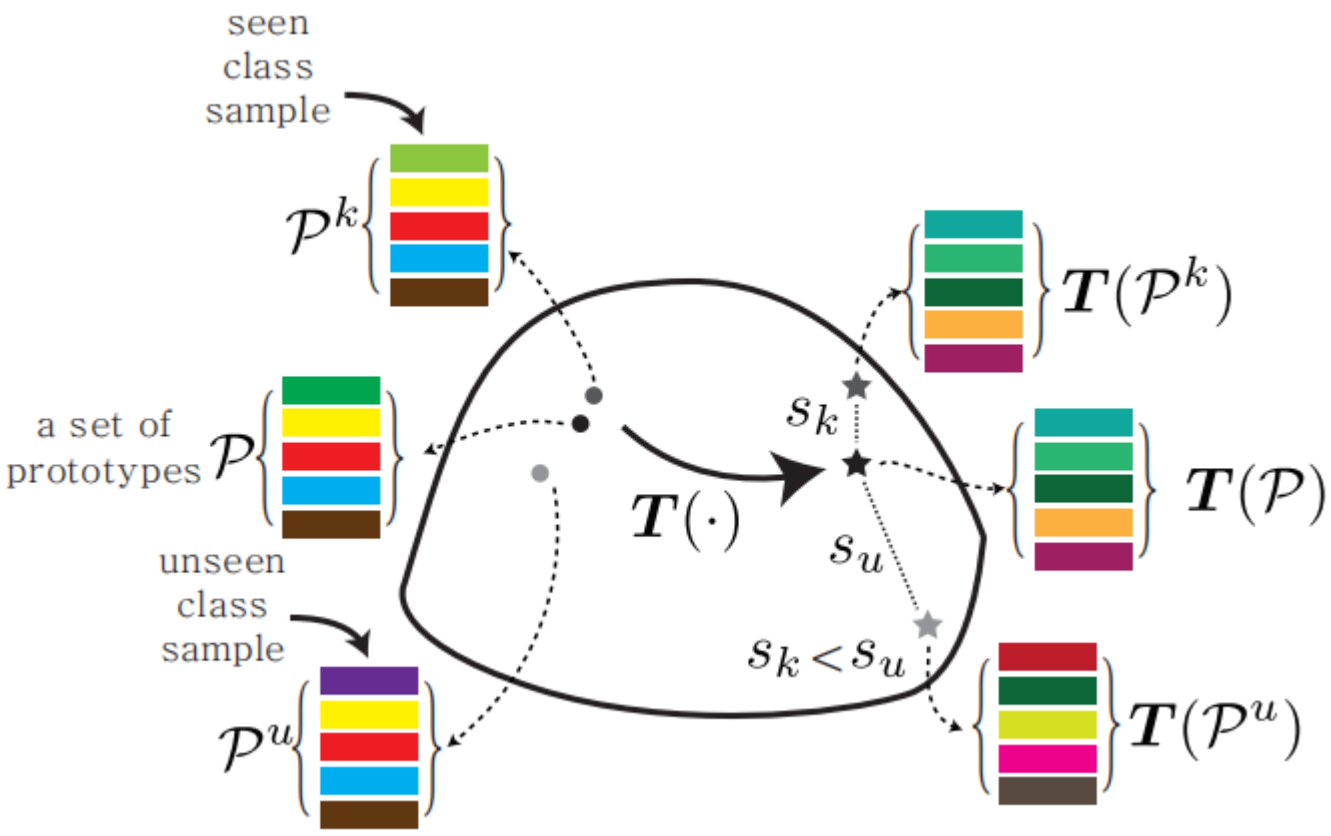


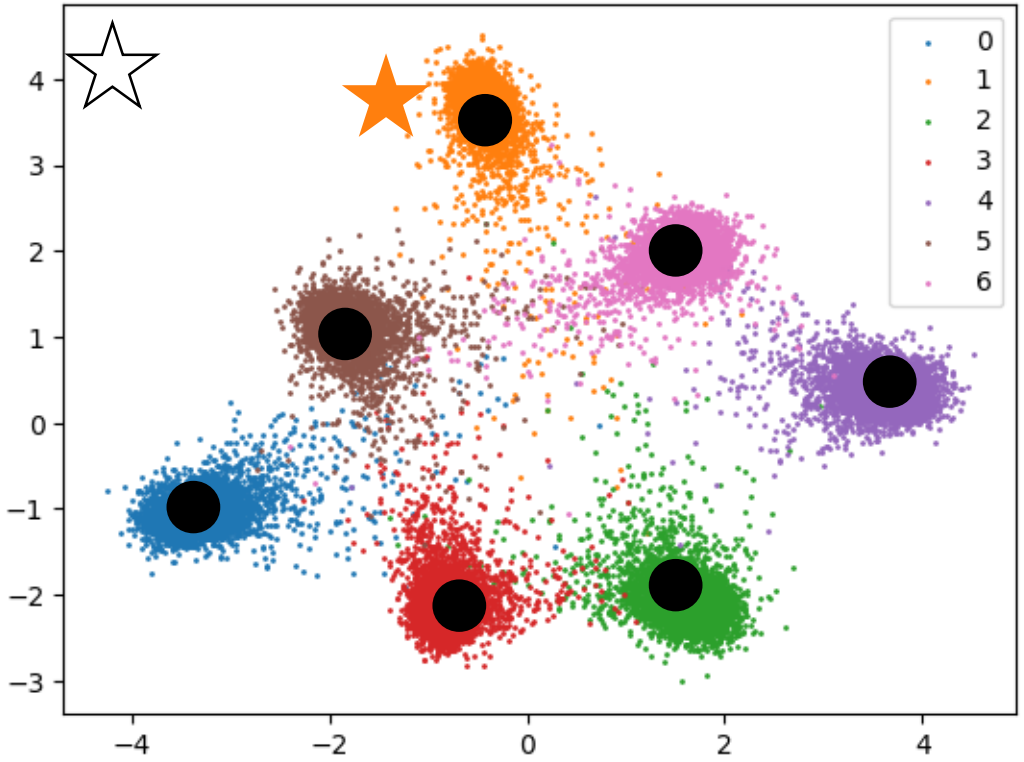
3



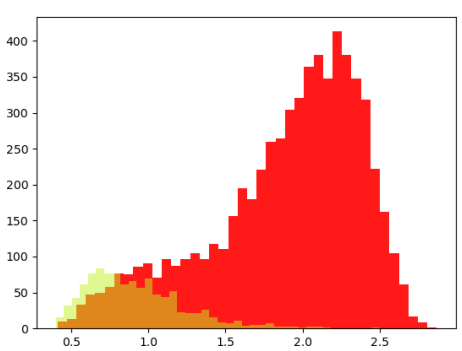
Few shot open-set recognition

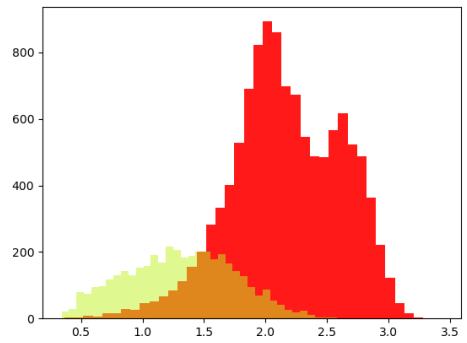
- Transformation Consistency

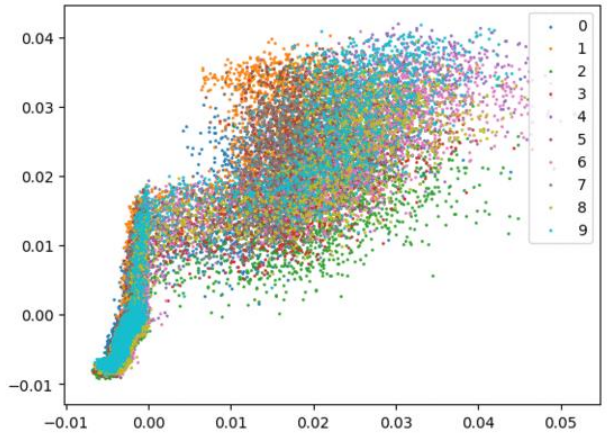


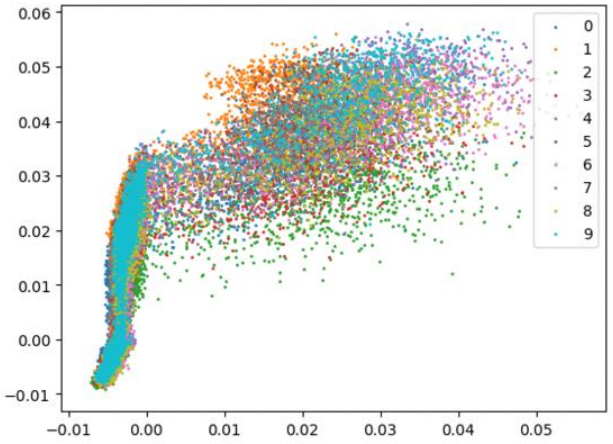


Experiments - MNIST

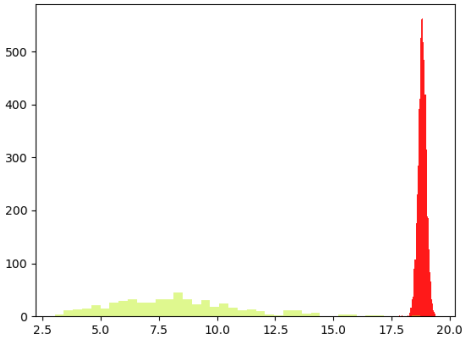


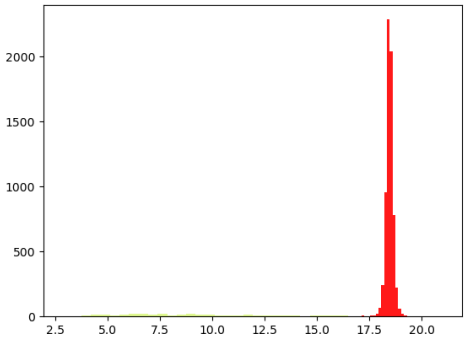


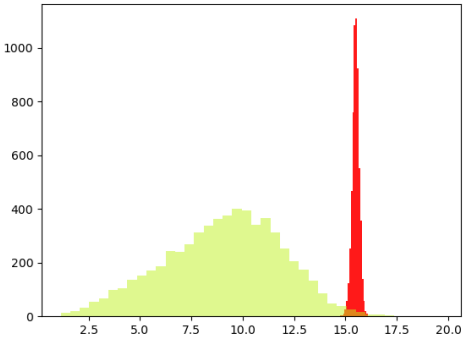


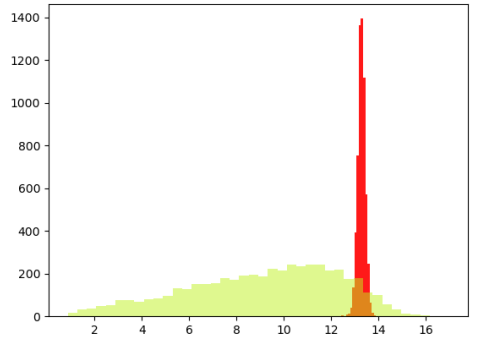


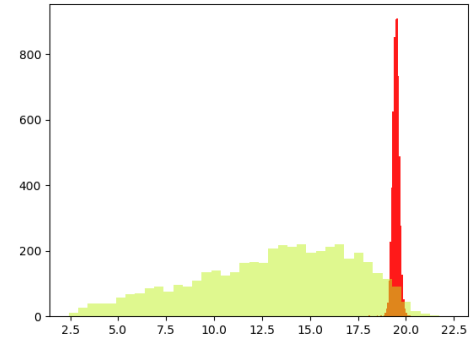
Mini - Center Loss



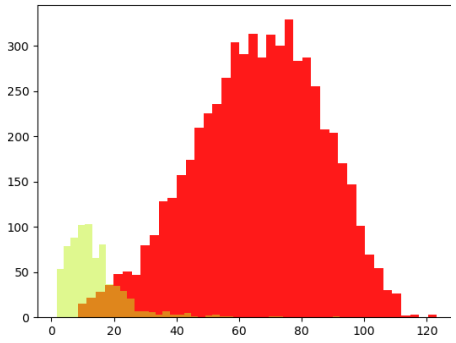


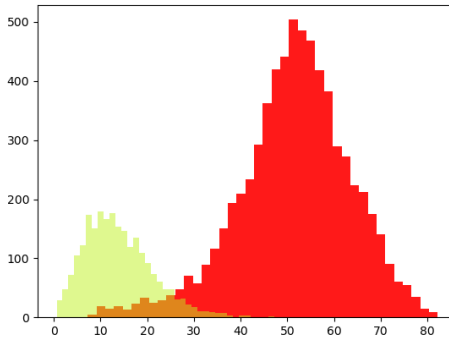


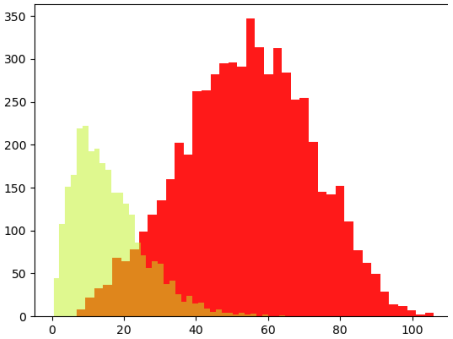


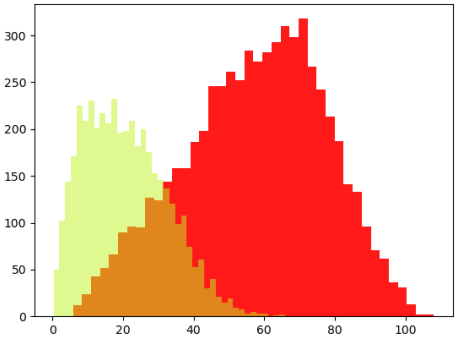


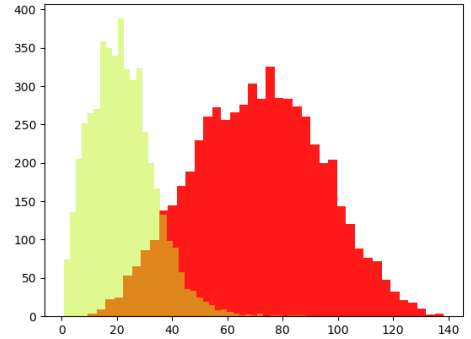
Center Loss





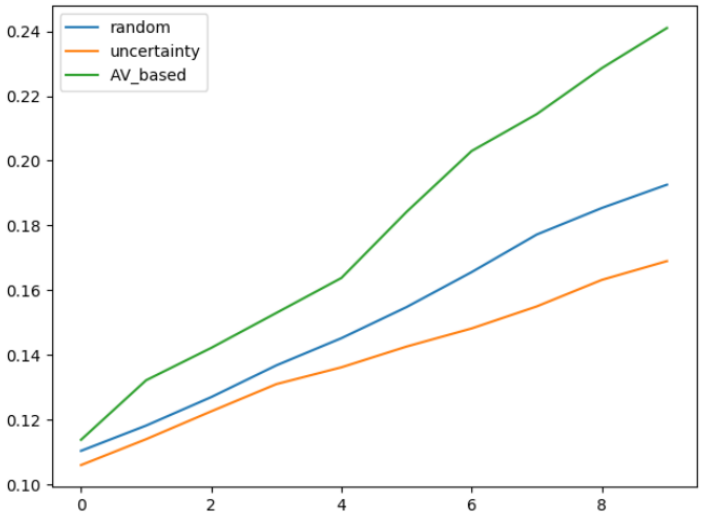


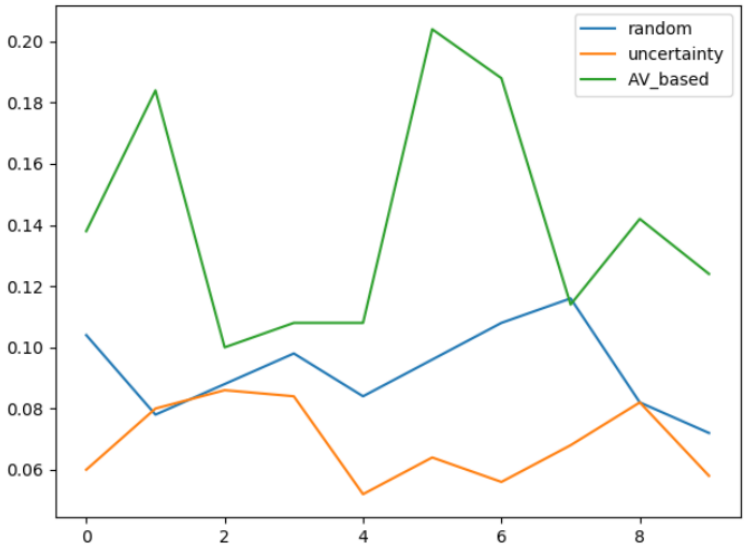


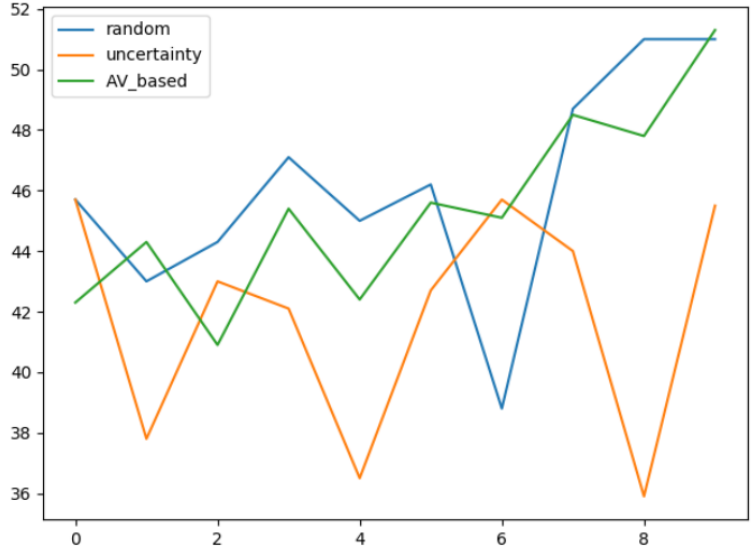


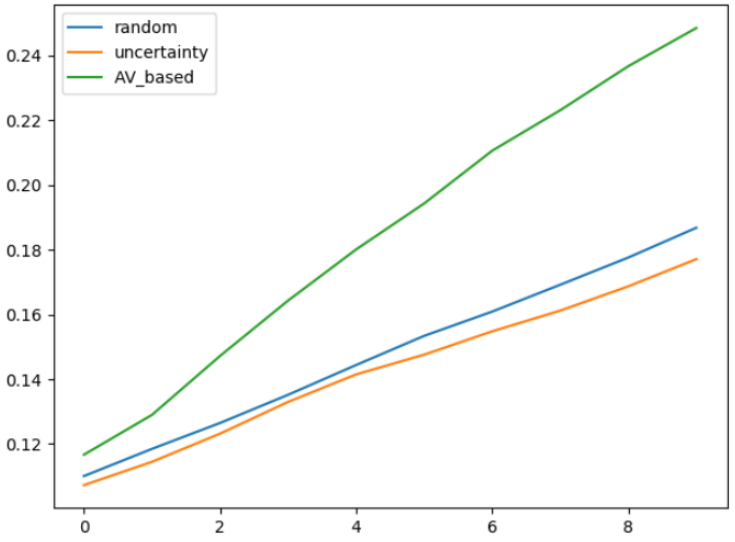
CE Loss

Experiments - CIFAR100

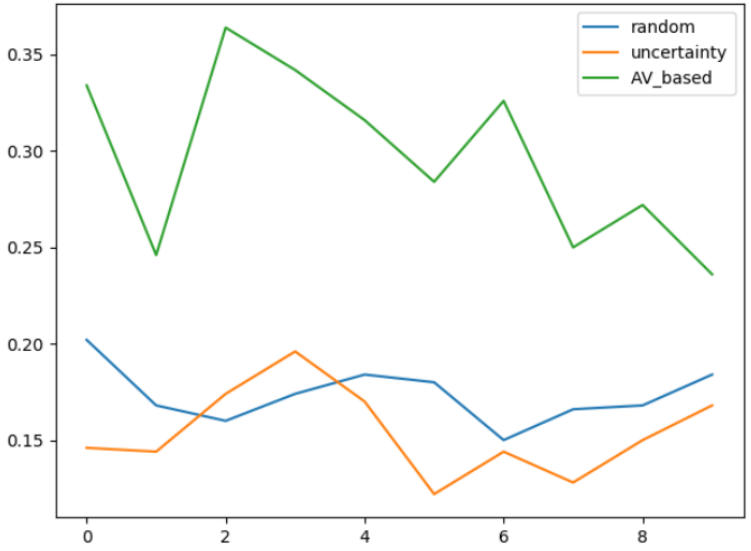




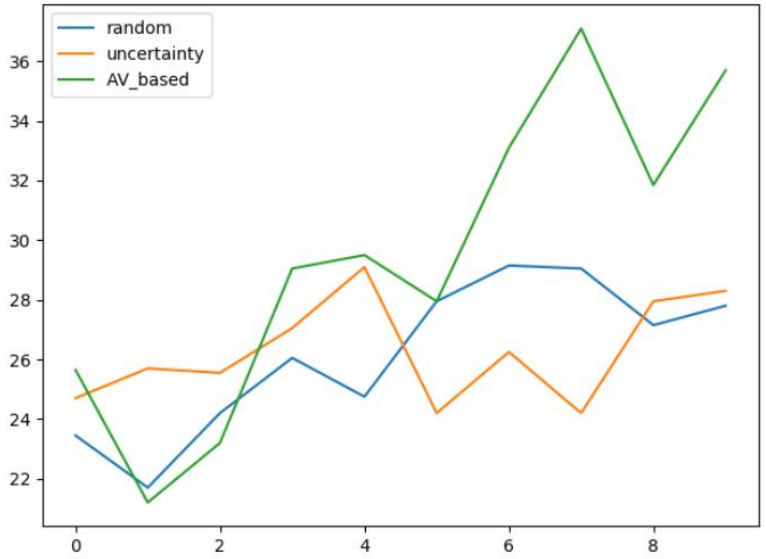




Recall



Precision



Accuracy

Future works

Idea

- The information of activation layer (activation vectors) is effective to recognize unknown data. (From Experiment)
- Pre-trained model (a better feature representation).
- Provide some unknown data or generate some negative examples.
- 训练多个类prototypes，unknown样本相较于known样本肯定要离对应的prototype更远

$$L = L_{ce} + \lambda * \sum_{i \in known} \frac{1}{\mathrm{E}_k[h_k(x_i)]} + \gamma * \sum_{j \in unknown} \mathrm{E}_k[h_k(x_j)]$$

Thanks