2022 상반기 세미나

이제임스

Vision & Display Systems Lab.

Dept. of Electronic Engineering, Sogang University

Outline

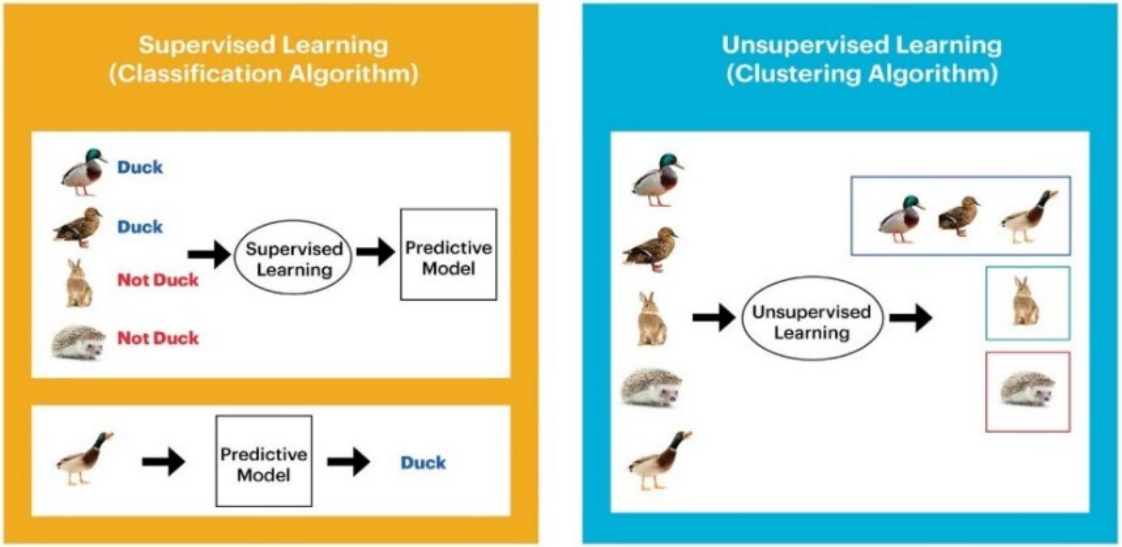
- Introduction
  - Supervised learning, unsupervised learning, semi-supervised learning

- Paper
  - Improving Unsupervised Image Clustering With Robust Learning (CVPR 2021)

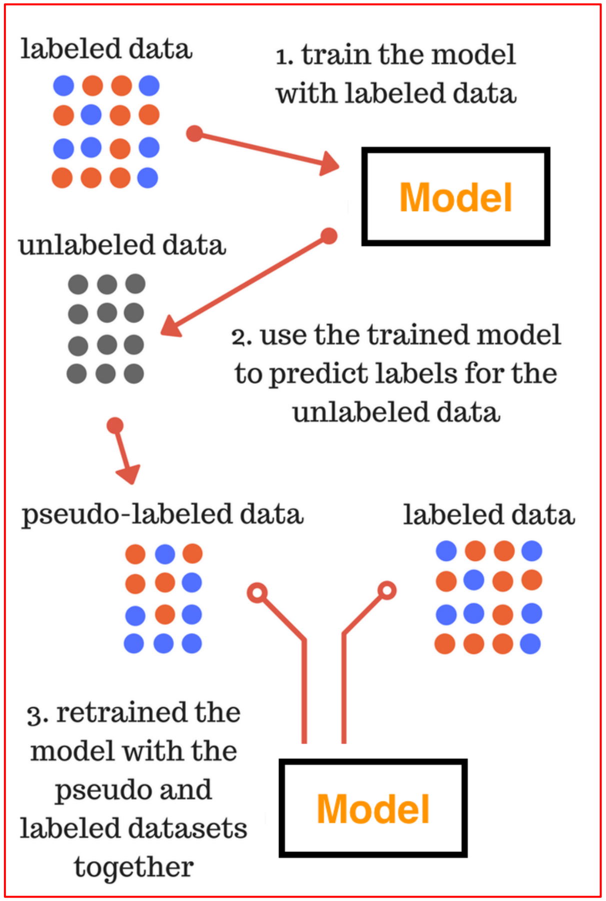
- Application paper
  - Instance-Aware, Context-Focused, and Memory-Efficient Weakly Supervised Object Detection (CVPR 2020)
  - End-to-End Semi-Supervised Object Detection with Soft Teacher (ICCV 2021)

Introduction

- Machine learning
  - Supervised learning
  - Unsupervised learning
  - Semi-Supervised learning

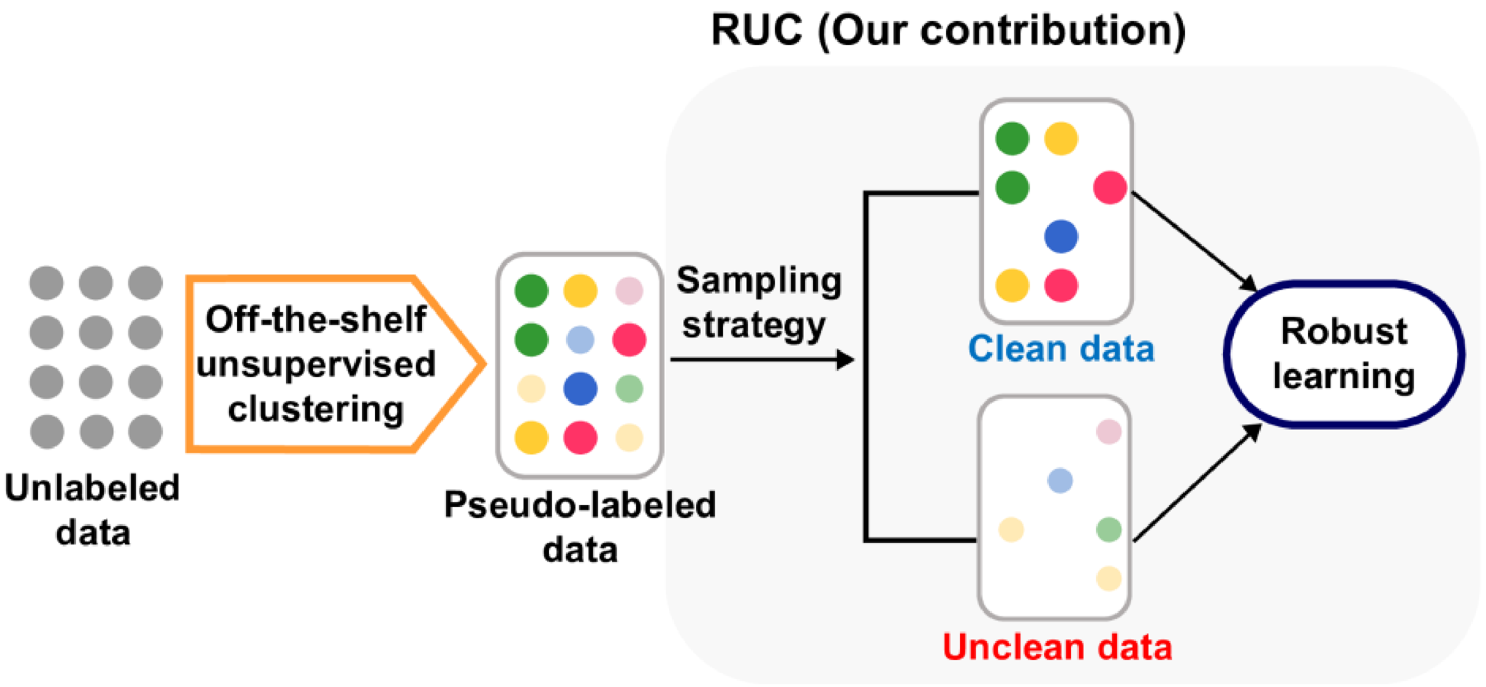


<Pseudo labeling>



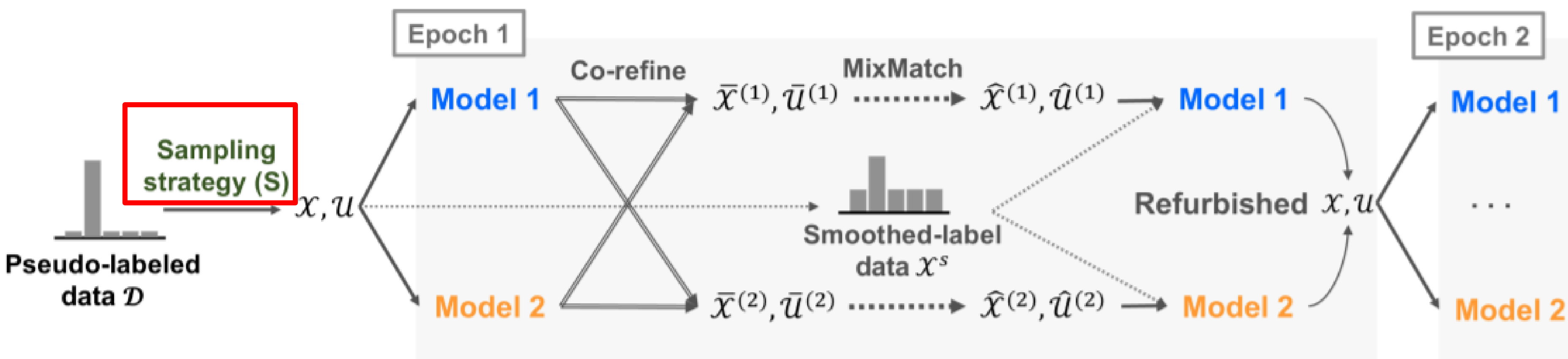
Paper

- Improving Unsupervised Image Clustering With Robust Learning (CVPR 2021)
  - Extracting Clean Samples
  - Retraining via Robust Learning



Method

- Extracting Clean Samples
  - Confidence-based strategy
    - Set a sufficiently high threshold for confidence score
  - Metric-based strategy
    - Leverages an additional embedding network learned in an unsupervised manner (SimCLR)
    - Non-parameteric classifier based on k-Nearest Neighbor
  - Hybrid strategy

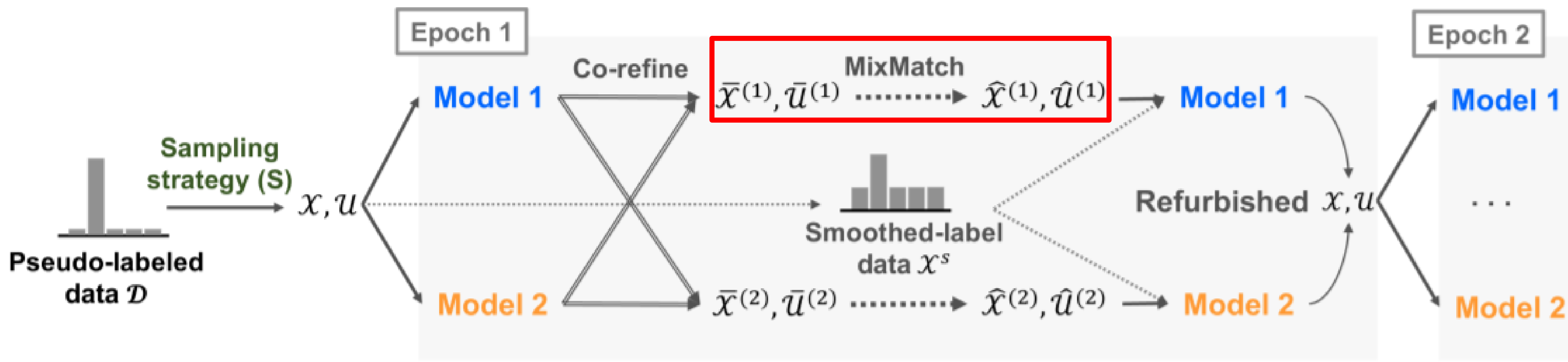


Method

- Extracting Clean Samples
  - SCAN: Learning to Classify Images without Labels
    - Unsupervised Image classification
      - Pretext task
        - ✓Image representation learning without label (SimCLR)
        - ✓Learning semantic features that do not change according to image transformation
      - Scan clustering
        - ✓Learn to predict the same cluster by maximizing the similarity of the closest neighbors by image
      - Self-labeling
        - ✓Proceed with supervised learning based on the confidence of the well-clustered image

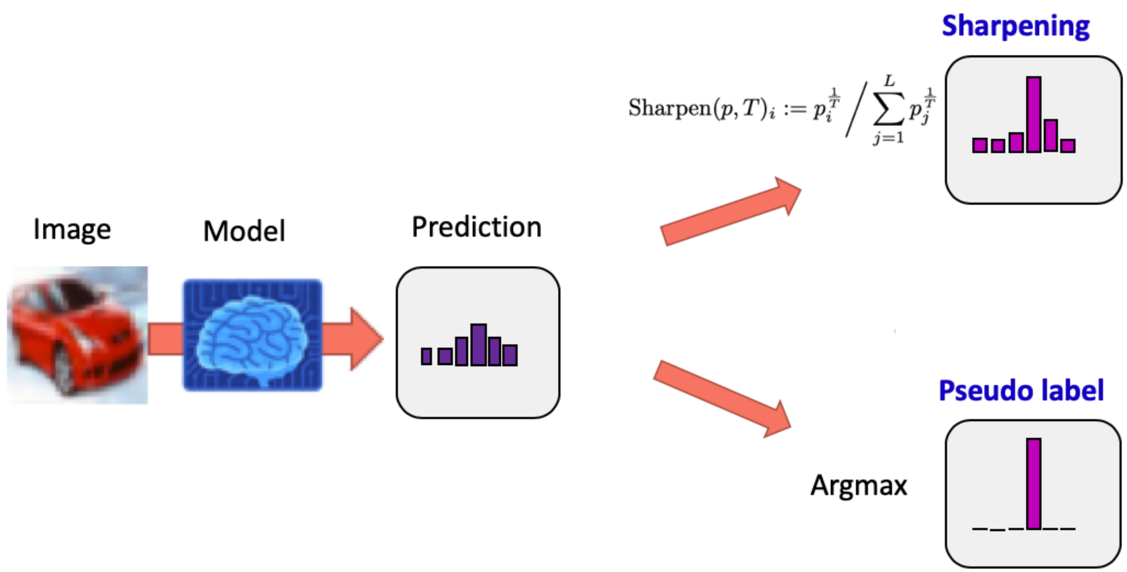
Method

- Retraining via Robust Learning
  - Vanilla semi-supervised learning
    - MixMatch
      - Estimates low-entropy mixed labels from unlabeled examples using MixUp augmentation
      - Bring additional resistance against noisy labels
        - ✓Consistency Regularization
        - ✓Entropy Minimization
        - ✓Mix up



Method

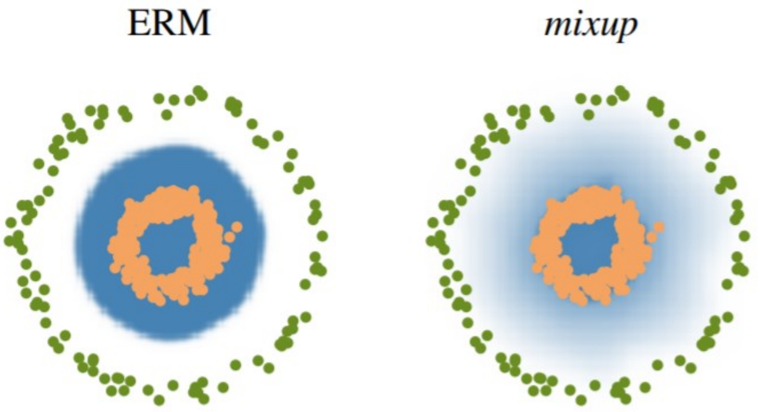
- Consistency Regularization
  - Data augmentation
  - Minimizing the pairwise difference of output
- Entropy Minimization
  - Distinguish the ambiguous ones near the decision boundary
  - MixMatch ➔ sharpening, FixMatch ➔ pseudo labeling



Method

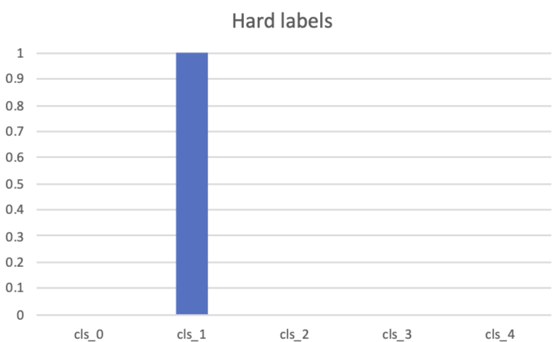
- Mixup
  - The solution for problem of overfitting training data
    - Use the distribution of training datasets in the vicinity ➔ data augmentation
  - Labeled data
    - Augment one by data + label ➔ convex combination ➔ supervised loss
  - Unlabeled data
    - Augment K by data + guessed label ➔ convex combination➔ consistency loss

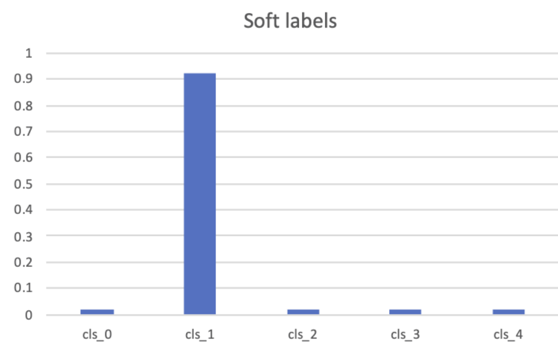
$$\lambda \sim \mathrm{Beta}(\alpha, \alpha)$$
$$x' = \lambda\, x_1 + (1-\lambda\,)x_2$$
$$p' = \lambda\, p_1 + (1-\lambda\,)p_2$$



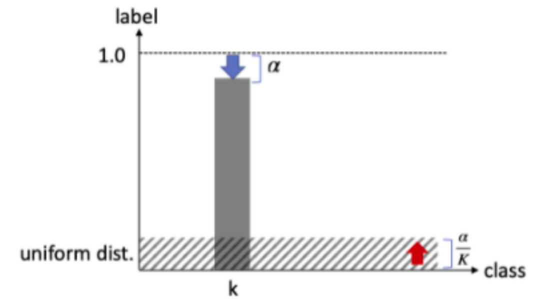
Method

- Retraining via Robust Learning
  - Label smoothing
    - Regularize the model from being overconfident to noisy predictions
    - Hard label ➔ soft label
      - Make a uniform distribution of classes except for the correct answer
        - ✓Inject uniform noise to all classes
      - Compute cross-entropy soft label and predicted label





$$y_k^{LS} = y_k(1-\alpha) + \alpha/K$$



Method

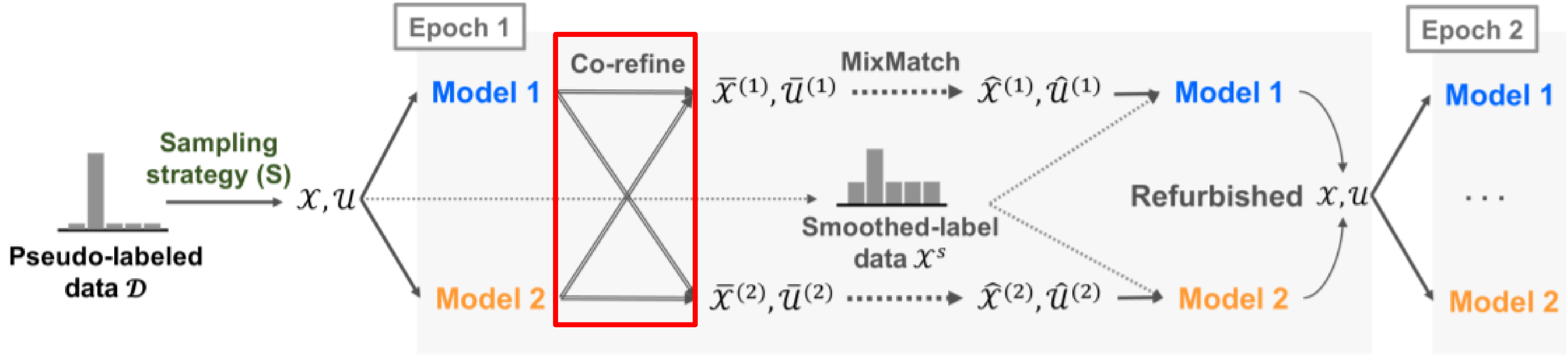
- Retraining via Robust Learning
  - Co-training
    - Single network ➔ vulnerability of overfitting to incorrect pseudo-labels
    - Two networks f1 , f 2 are trained in parallel and exchange their guesses

Labeled

$$\bar{\mathbf{y}} = (1 - w^{(2)}) \cdot \mathbf{y} + w^{(2)} \cdot f_{\theta^{(2)}}(\mathbf{x})$$
$$\bar{\mathbf{y}} = \text{Sharpen}(\bar{\mathbf{y}}, T),$$

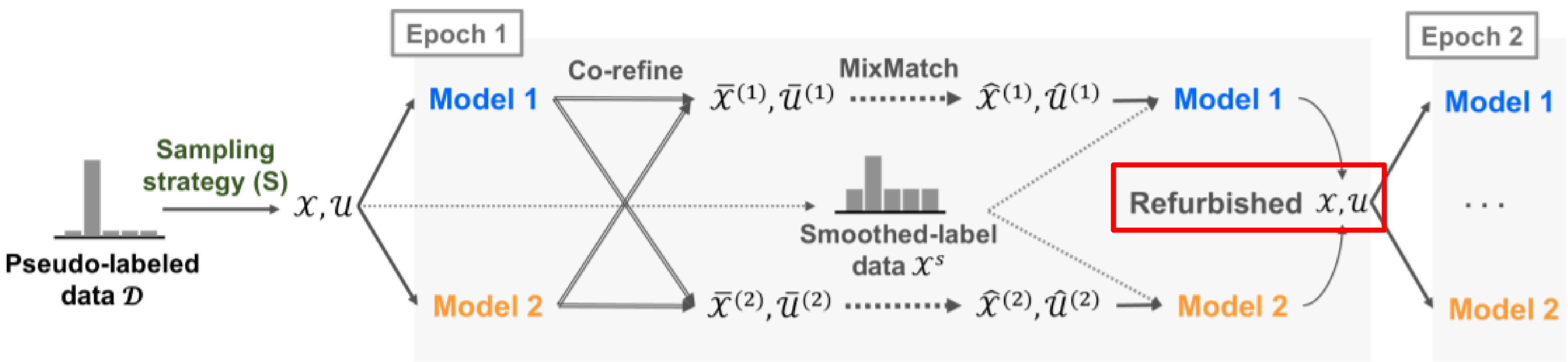
Unlabeled

$$\bar{\mathbf{q}} = \frac{1}{2M}\sum_m (f_{\theta^{(1)}}(\mathbf{u}_m) + f_{\theta^{(2)}}(\mathbf{u}_m))$$
$$\bar{\mathbf{q}} = \text{Sharpen}(\bar{\mathbf{q}}, T),$$



Method

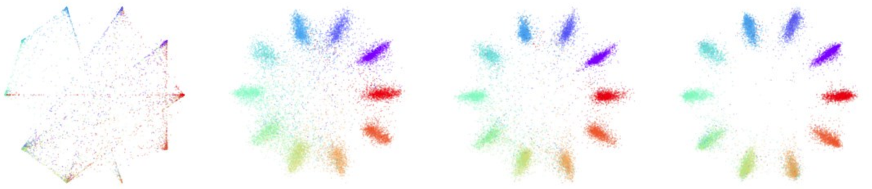
- Retraining via Robust Learning
  - Co-refurbishing
    - Unclean data
      - Refurbish the noise samples at the end of every epoch to deliver the extra clean samples
      - The prediction probability of network exceeds the threshold value
        - ✓Add to the clean data and appended to the labeled set X



Experiments

- Unsupervised Image Clustering

Method	CIFAR-10	CIFAR-20	STL-10
k-means [45]	22.9	13.0	19.2
Spectral clustering [55]	24.7	13.6	15.9
Triplets [37]	20.5	9.9	24.4
Autoencoder (AE) [4]	31.4	16.5	30.3
Variational Bayes AE [24]	29.1	15.2	28.2
GAN [36]	31.5	15.1	29.8
JULE [52]	27.2	13.7	27.7
DEC [49]	30.1	18.5	35.9
DAC [8]	52.2	23.8	47.0
DeepCluster [6]	37.4	18.9	33.4
ADC [15]	32.5	16.0	53.0
IIC [22]	61.7	25.7	49.9
TSUC† [17]	80.2	35.5	62.0
SCAN† [42]	88.7	50.6	81.4
TSUC + RUC (Confidence)	81.8 / 82.5	39.6 / 40.6	65.1 / 65.5
TSUC + RUC (Metric)	82.5 / 82.9	39.5 / 40.4	66.3 / 66.6
TSUC + RUC (Hybrid)	82.1 / 82.8	39.5 / 40.6	66.0 / 66.8
SCAN + RUC (Confidence)	**90.3** / 90.3	53.3 / 53.5	**86.7** / 86.8
SCAN + RUC (Metric)	89.5 / 89.5	53.9 / 53.9	84.7 / 85.1
SCAN + RUC (Hybrid)	90.1/ 90.1	**54.3** / 54.5	86.6 / 86.7



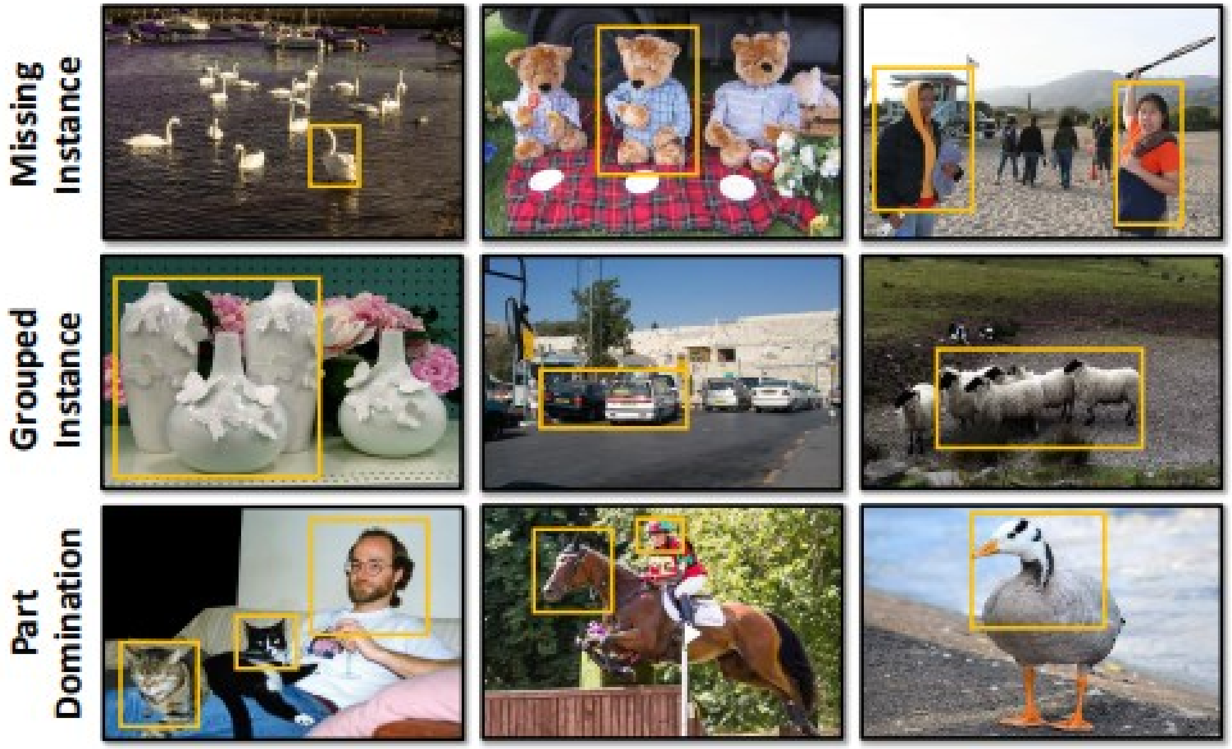
(a) SCAN (b) SCAN+RUC (epoch 50) (c) SCAN+RUC (epoch 100) (d) SCAN+RUC (epoch 200)

Method	SCAN (Best)	SCAN + RUC (Last / Best accuracy)
ImageNet-50	76.8	**78.5** / 78.5

Setup	Last Acc	Best Acc
RUC with all components	**86.7**	**86.8**
without co-training	86.2	86.4
without label smoothing	85.5	85.8
with MixMatch only	85.2	85.4

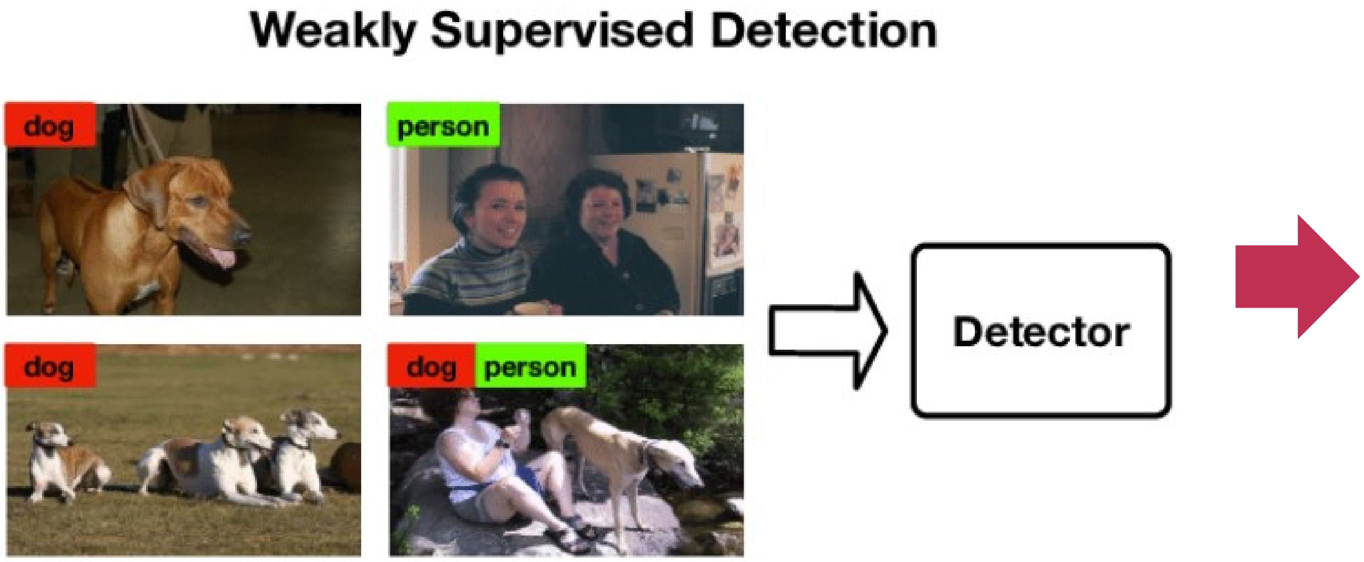
Paper

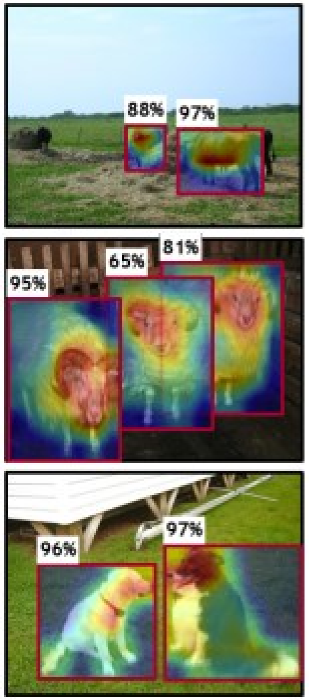
- Instance-Aware, Context-Focused, and Memory-Efficient Weakly Supervised Object Detection (CVPR 2020)



Method

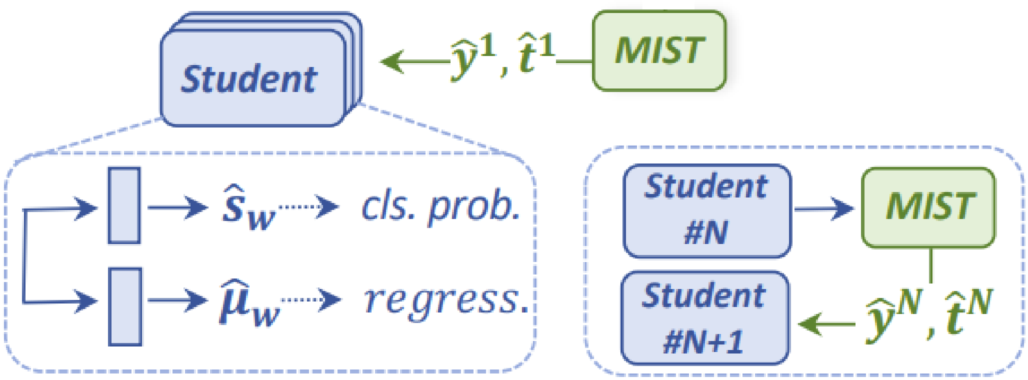
- Weakly Supervised Object Detection
  - Multiple instance learning (MIL)
    - Select major instances among them and accurately predict the label of the bundle
  - Training based on classification loss
    - To select the most confident positive proposals





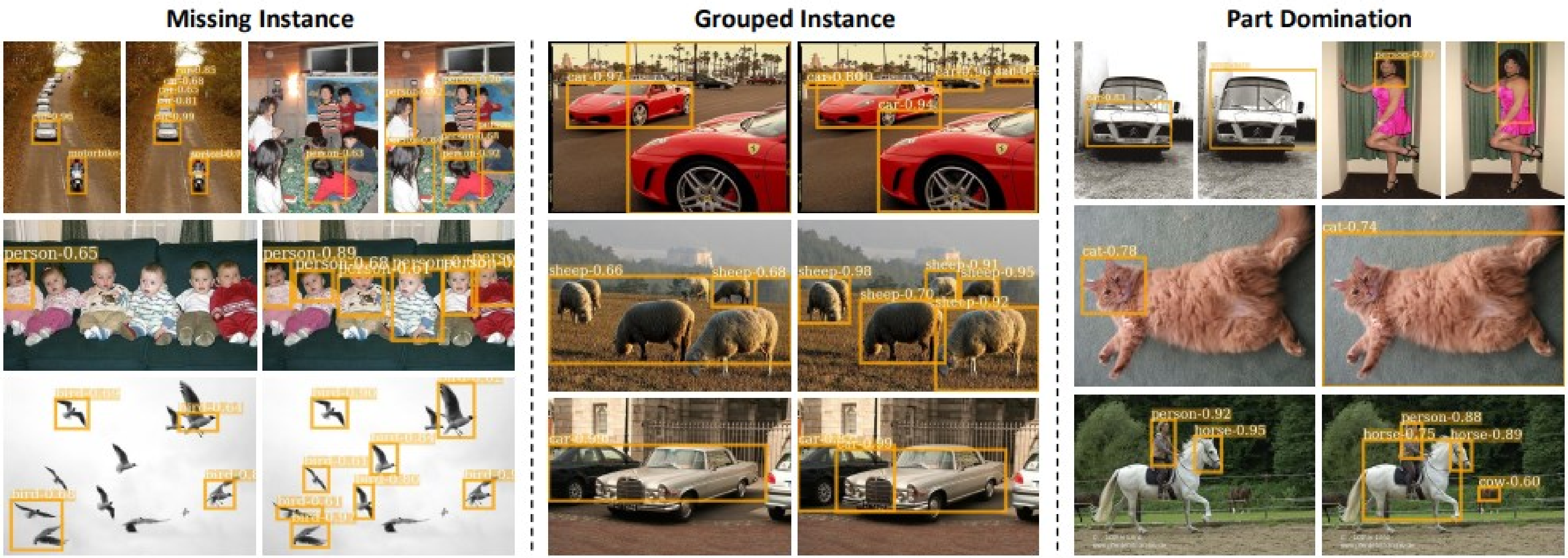
Method

- Multiple instance self-training (MIST)
  - Input image ➔ Pretrained network ➔ proposal R (instance)
  - Sort the score of each class
    - High-scoring non-overlapping regions R’ (pseudo label)
      - Instance-level regression and classification label
      - (Classification logits + detection) logits score
  - Teacher-student repetitive distilling process proceeds (self-training)



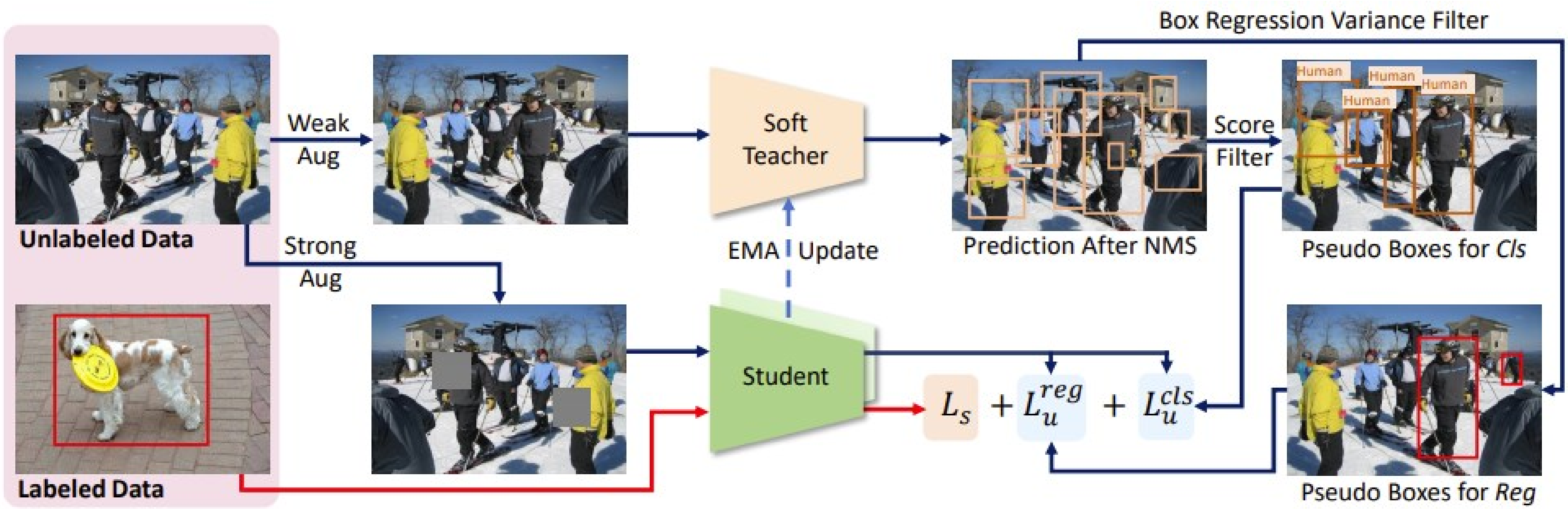
Experiments

- Comparison of the model to the baseline
  - Left: baseline, right: model results



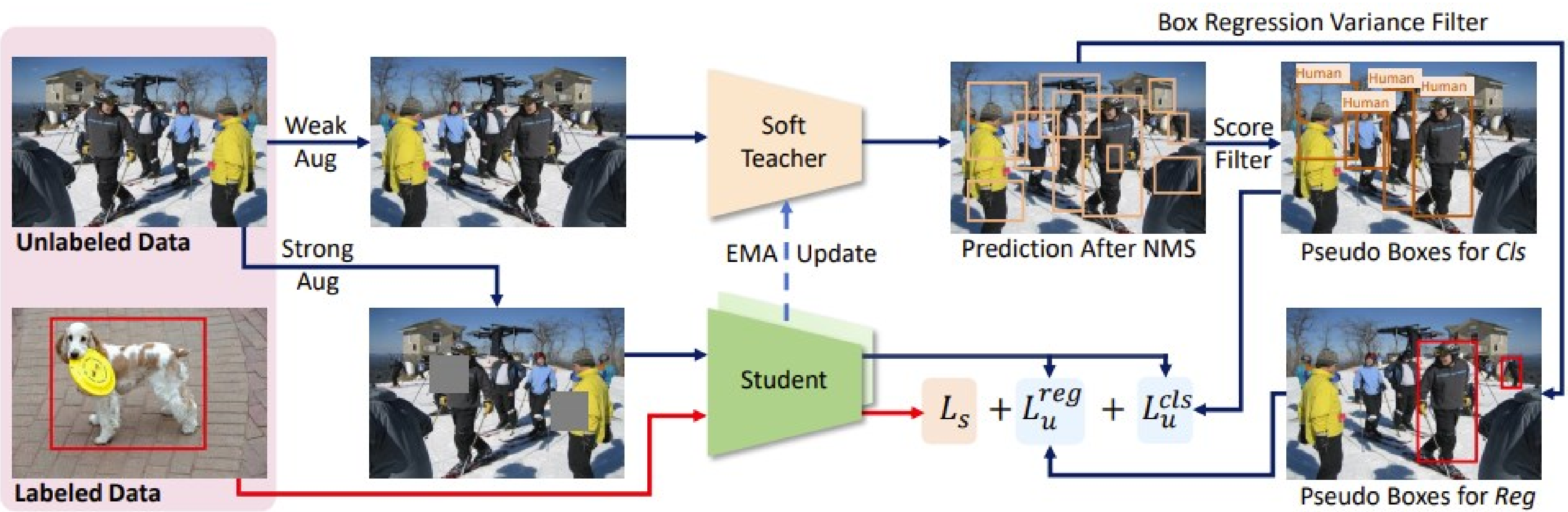
Paper

- End-to-End Semi-Supervised Object Detection with Soft Teacher (ICCV 2021)
  - Method
    - End-to-End Pseudo-Labeling Framework
    - Soft Teacher



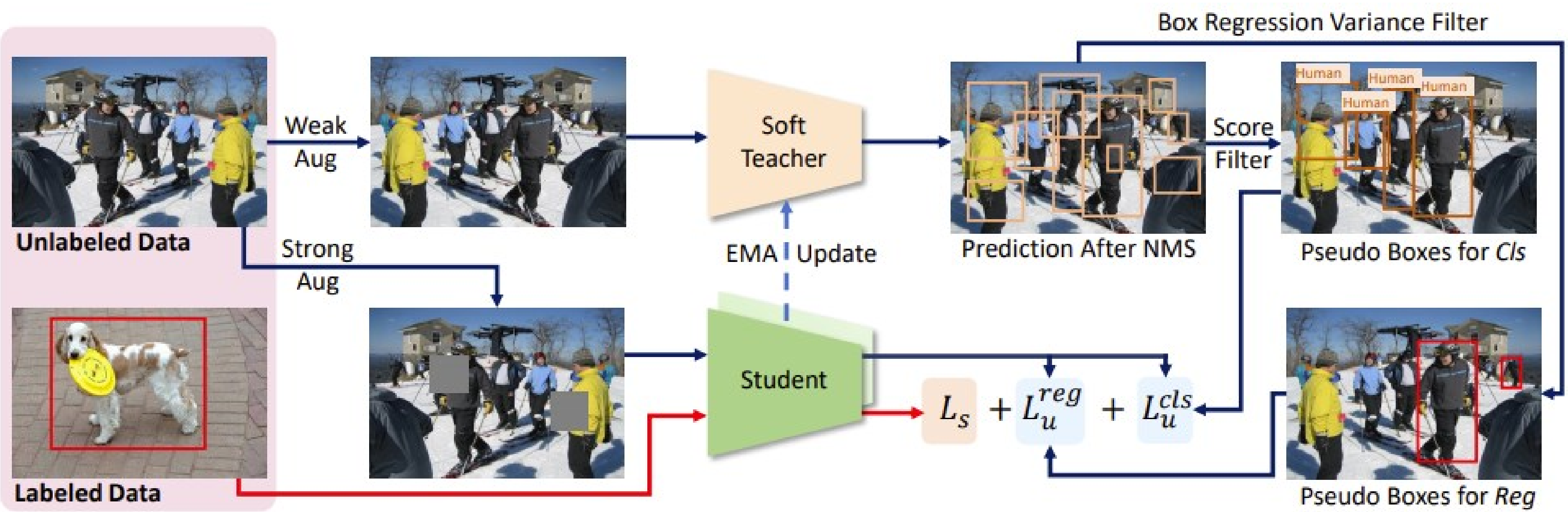
Method

- End-to-End Semi-Supervised Object Detection with Soft Teacher (ICCV 2021)
  - Soft Teacher
    - Teacher model
      - Weak augmentation (unlabeled data) ➔ Soft Teacher ➔ NMS ➔ high-confidence score ➔ Box regression variance filter
      - EMA update (student ➔ teacher)



Method

- End-to-End Semi-Supervised Object Detection with Soft Teacher (ICCV 2021)
  - Soft Teacher
    - Student model
      - Labeled data + strong augmentation (unlabeled data) ➔ prediction
      - Comparison prediction of student and pseudo labeled data produced by Soft teacher



Experiments

- Qualitative results
  - Left: baseline, right: model results

References

- Lee, Pilhyeon, Youngjung Uh, and Hyeran Byun. "Background suppression network for weakly-supervised temporal action localization." Proceedings of the AAAI conference on artificial intelligence. Vol. 34. No. 07. 2020.

- Zhao, Ting, and Xiangqian Wu. "Pyramid feature attention network for saliency detection." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.

- Zhang, Can, et al. "CoLA: Weakly-Supervised Temporal Action Localization with Snippet Contrastive Learning." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

Thank you!