Pietraru Alfred-Andrei
Popescu Justin-Marian

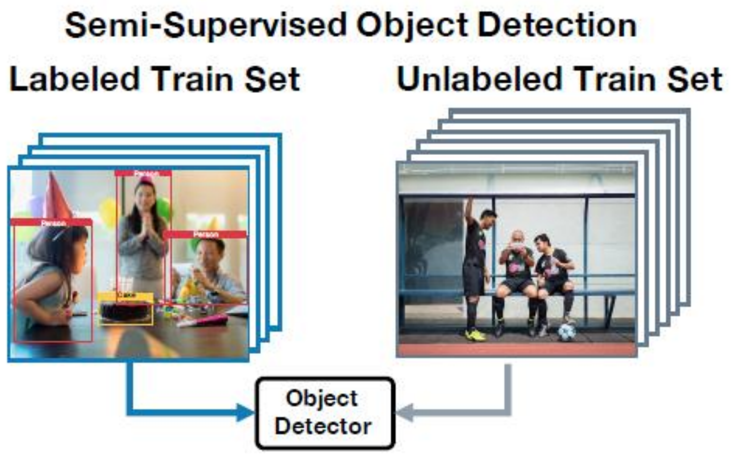
Semi-Supervised Object Detection

Unbiased Teacher-Student

Our framework enhances object detection performance with limited labeled data by leveraging a dynamic Teacher-Student training strategy. This approach is designed for robustness and efficiency in environments where extensive annotation is impractical or costly.

Core Idea (Unbiased Teacher)

- Teacher generates **pseudo-labels** for unlabeled images.
- Student is trained on strong augmentations using these pseudo-labels.
- Student's learned weights feed back into the teacher through



Teacher: YOLOn11 model fine-tuned on labeled set, federated-pretrained.
Student: Grad-CAM++ enhanced localization head (Multi-Scale CAM Head)

- **ResNet50** pretrained on Imagenet dataset.
- Uses **activation heatmaps** to derive bounding boxes.
- Lightweight and stable under strong augmentation.

EMA (Exponential Moving Average): handles class imbalance & false positives more robustly than classical pseudo-labeling.

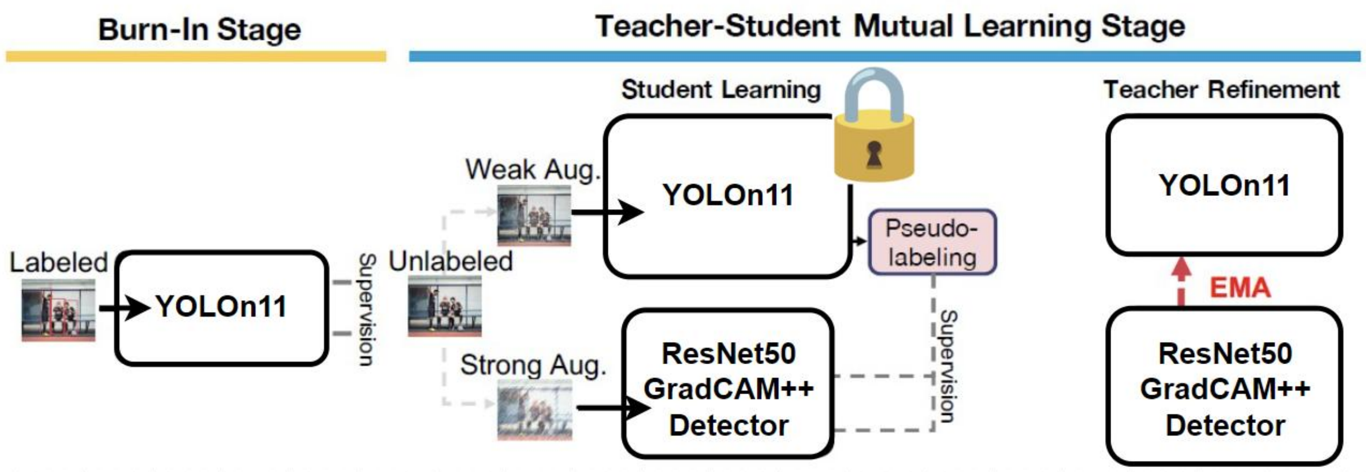
Pietraru Alfred-Andrei
Popescu Justin-Marian

Teacher-Student Ensemble:

- YOLOn11(teacher)
- Resnet50 with Grad-CAM++ detector(student)

Why CAM-Head Instead of RPN/ROI?

- Simpler and better suited for unlabeled data.
- Avoids anchor tuning and region proposals.
- Directly exploits **class activation localization**.



Procedure steps:

1. Student backbone, *pretrained ResNet50* → class activation maps (CAMs).
2. Threshold CAM values to isolate high-saliency regions.
3. Convert connected high-activation clusters into raw boxes.
4. Apply NMS to merge overlapping CAM clusters.
5. Supervision uses *teacher YOLOn11 boxes* to guide CAM localization.

Phase 1 Burn-In (Supervised Stage)

- Train YOLO-N11 teacher on the labeled subset.
- Ensures stable bounding-box priors before entering SSL.

$$\mathcal{L}_{sup} = \sum_i \mathcal{L}_{cls}^{rpn}(\boldsymbol{x}_i^s, \boldsymbol{y}_i^s) + \mathcal{L}_{reg}^{rpn}(\boldsymbol{x}_i^s, \boldsymbol{y}_i^s) + \mathcal{L}_{cls}^{roi}(\boldsymbol{x}_i^s, \boldsymbol{y}_i^s) + \mathcal{L}_{reg}^{roi}(\boldsymbol{x}_i^s, \boldsymbol{y}_i^s)$$

Phase 2 Teacher-Student Mutual Learning

Teacher Inference (Weak Aug):

- YOLO-N11 predicts boxes + class scores on weakly augmented images.
- Apply **NMS** (to remove duplicates) + **confidence threshold 0.5.**

$$\theta_s \leftarrow \theta_s + \gamma \frac{\partial(\mathcal{L}_{sup} + \boldsymbol{\lambda}_u \mathcal{L}_{unsup})}{\partial \theta_s}, \quad \mathcal{L}_{unsup} = \sum_i \mathcal{L}_{cls}^{rpn}(\boldsymbol{x}_i^u, \hat{\boldsymbol{y}}_i^u) + \mathcal{L}_{cls}^{roi}(\boldsymbol{x}_i^u, \hat{\boldsymbol{y}}_i^u)$$

Student Learning (Strong Aug):

- Student CAM-head receives the strong-augmented version.
- Loss for Student:
  - Classification loss (CAM heatmap → bboxes)
  - Object detection loss offered by mIoU
  - Box regression loss using teacher pseudo-labels

Phase 3 Teacher Refinement (EMA)

- **Smooting Teacher weights ← 0.6·Teacher + 0.4·Student**
- ths training, prevents pseudo-label drift, reduces bias
- EMA gives best mAP and balanced pseudo-label distribution.

$$\theta_t \leftarrow \alpha\theta_t + (1-\alpha)\theta_s$$

Quantitative Results (Training Metrics) 1st plot

Validation Loss Trends

- Loss stabilizes ~0.28 with fluctuations which are normal for an SSL.
- Indicates consistent pseudo-label quality and stable teacher updates.

Detection Metrics Over Epochs

(Evaluation Metrics) 2nd plot

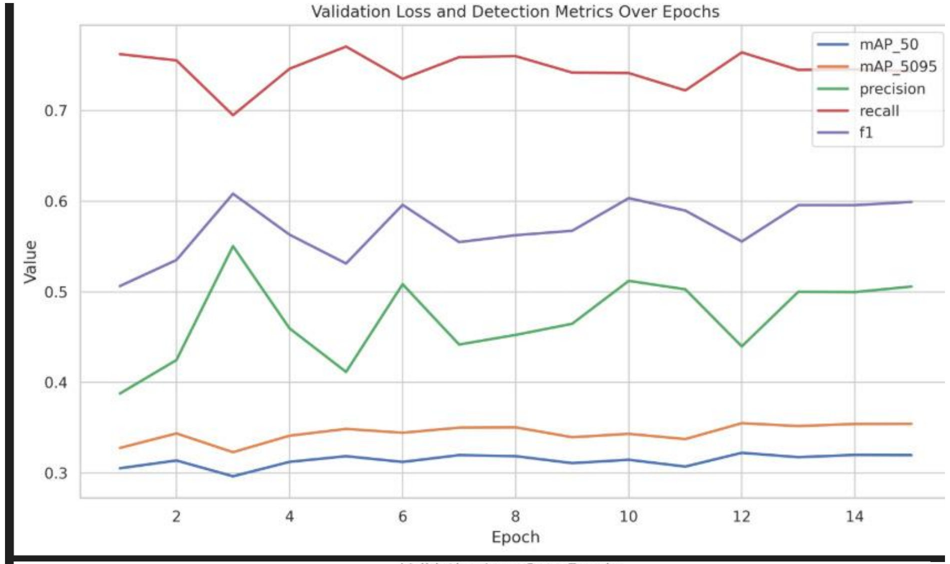
- **mAP@50:** ~0.32
- **mAP@50–95:** ~0.35
- **Precision:** ~0.50
- **Recall:** ~0.74
- **F1 Score:** ~0.60

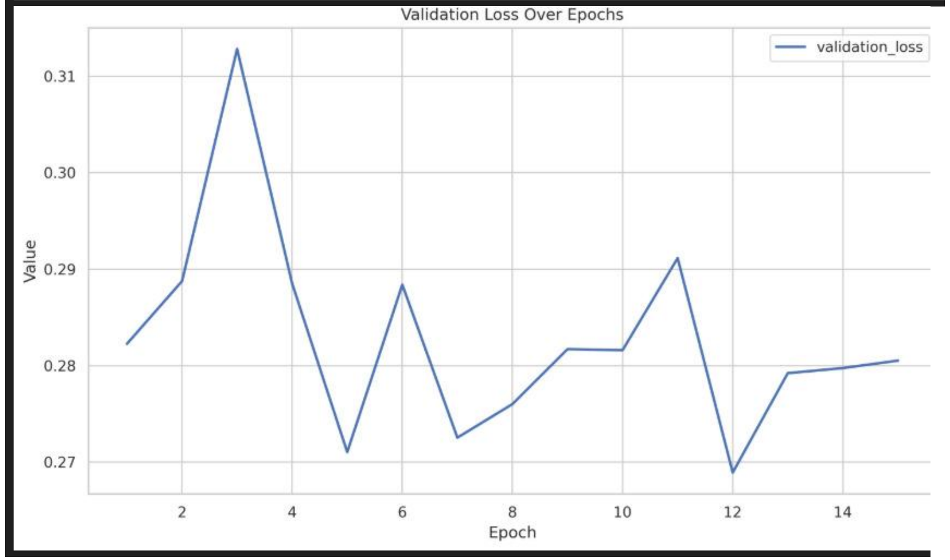
Interpretation

High Recall (0.74):
Student CAM-head captures most objects → less under-detection.

Good Precision (0.50):
YOLO teacher generates reasonably clean pseudo-labels.

Balanced F1 (0.60):
System is neither overly conservative nor overly permissive.





- **Training Loss Components**
- Teacher losses (YOLO):

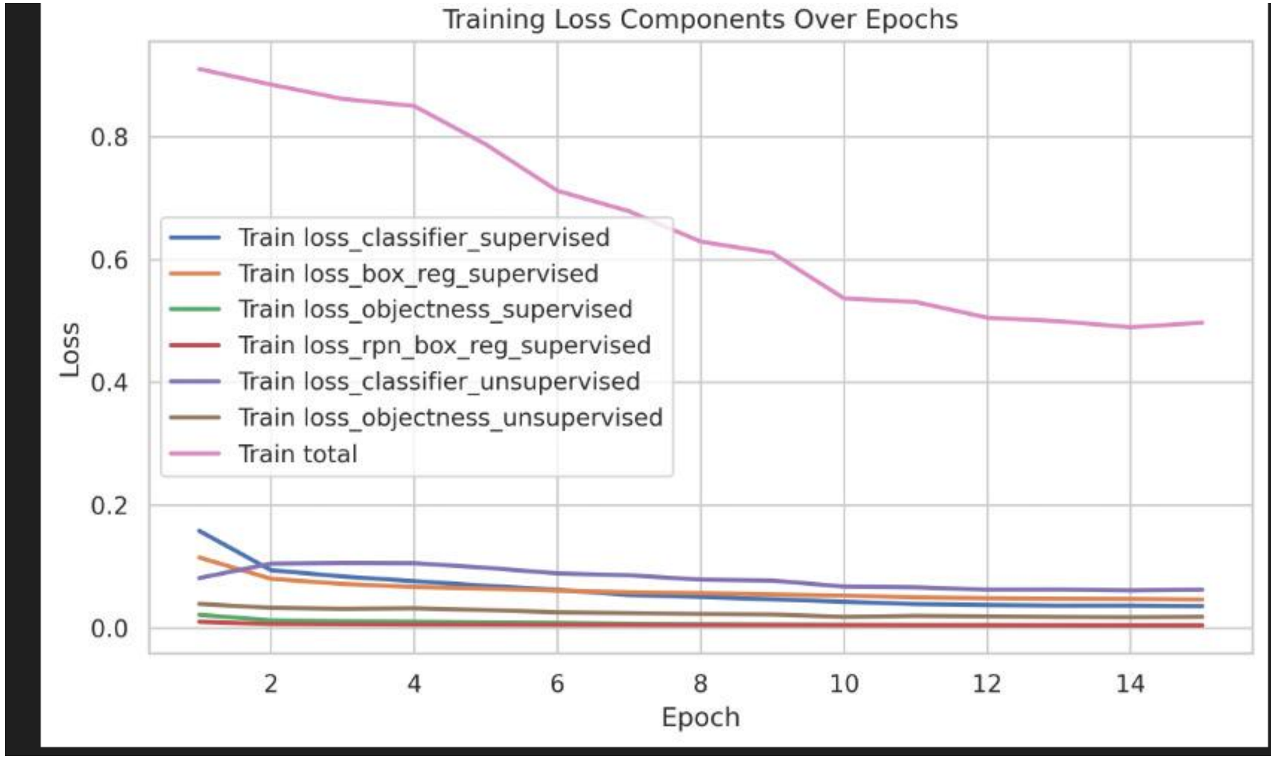
Classifier + objectness + box regression all steadily decreasing, showing consistent pseudo-label.

- **Student losses (ResNet50 + GradCAM++):**

Lower magnitude → smoother gradients → acts as regularization for teacher.

Stable convergence despite noisy unlabeled data due to:

- EMA Teacher refinement
- NMS + confidence threshold
- CAM-based localization robustness
- Balanced supervised + unsupervised loss strategy



[Creating Bounding Boxes from Class Activation Maps]

- CAM intensity ∝ object likelihood.
- Spatial gradients enforce tighter localization over epochs.
- Student becomes a strong regularizer (CAM localization reduces teacher false positives).