Semi-Supervised Learning for Pill Instance Segmentation: A Comparative Study of Pseudo-Labeling, FixMatch, and MixMatch Approaches

Sumaiya Tabassum[a,*], S M Sazzad Hossain[a], Nishat Lubna[a], Nafisa Saiyara Aranti[a], Dr. Mohammad Rifat Ahmmad Rashid[1,**]

[a] *Department of Computer Science and Engineering, East West University, A/2, Jahurul Islam Avenue, Jahurul Islam City, Aftabnagar, Dhaka, 1212, Dhaka Division, Bangladesh*

Abstract

Transductive learning, also known as semi-supervised learning (SSL) has been proposed to tackle the scarcity of data in medical images via using both labeled and unlabeled information. This paper evaluates three SSL approaches, namely the pseudo-labeling, FixMatch and MixMatch for pill level instance segmentation in a teacher-student pipeline. By using 20% labeled data as ground-truth, a teacher model can effectively produce pseudo-labels for the remaining 80% unlabeled images so that student learning is based on combined datasets via method-specific augmentations. Mask mAP @ 0.5 was the highest for MixMatch with a value of 0.782 (15.3% improvement over baseline value of 0.678), with FixMatch at 0.765 and plain pseudo-labeling at 0.715 following with lower numbers for semi-supervised training. MixMatch achieves great performance largely owing to its powerful MixUp (80% chance) and low confidence-level (0.75), which allows it to make better usage of the unlabeled data by guessing labels and regularization on consistency. These findings clearly show that progressive SSL models can remarkably boost the segmentation results when only a small portion of labelled data are available.

Keywords: Semi-supervised learning, Instance segmentation, Pill

*Corresponding author
**Supervisor

segmentation, Pseudo-labeling, FixMatch, MixMatch

1. Introduction

Precise segmentation of pharmaceutical pills in medical image is also a challenging issue since the boundary, shape and position pill needs to be determined accurately. Instance segmentation is very challenging as it requires discriminating between the different pill instances even with close or overlapping pills, and demands both detecting individual pills and accurately defining their boundaries at pixel level. Such segmentation tasks are essential for automated pill counting, quality assurance, inventory control and medication adherence tracking in clinical and pharmaceutical contexts. Yet, the availability of large-scale and accurately labeled dataset for medical imaging tasks is still one bottleneck as the manual annotation by radiologists or physicians is time-consuming and expertise-intensive.

This lack of data is what motivates additional research into the semi-supervised methods that work with limited labeled and plenty of unlabeled data. Semi-supervised learning for segmentation tasks offers a promising solution which combines a small number of costly labelled examples with large amount of easily accessible unlabelled images. The idea is that if there is some structure/pattern in the data without labels, it can be utilized for better generalization of a model, more than what one could have if we had only labeled data points. This is of particular relevance in medical areas, where there are frequently plenty of unlabeled images but few or expensive expert annotations.

In this work, we explore and compare three semi-supervised learning themes on pill instance segmentation: the conventional pseudo labeling, FixMatch and MixMatch. The two approaches follow the same teacher-student framework, but on how to generate pseudo-labels and filter as well as leverage them in training they vary. The simplest is to use the confidence thresholding based on traditional pseudo-labeling. FixMatch utilizes consistency regularization with strong/weak augmentations and a high-confidence filtering. MixMatch exploits label guessing and MixUp augmentation for generating interpolations over labeled and unlabeled points. We evaluate these methods on a pill segmentation dataset with only 20% labeled data, targeting to find which SSL approach is the most competent for instance segmentation within a medical imaging setup with data scarcity, and gain intuition of how unlabeled data can help enhance the quality of segmentation.

2. Theoretical Background

Semi-supervised learning (SSL) leverages both labeled and unlabeled data to improve model performance, achieving results comparable to fully supervised methods with significantly fewer labels, which is particularly critical for instance segmentation tasks like the PillSegmentation dataset where annotation of 180 pill classes across cluttered scenes is labor-intensive. The effectiveness of SSL stems from fundamental assumptions about data structure: the smoothness assumption posits that if two data points are close in the input space, their corresponding labels should be similar, justifying the use of unlabeled data to learn overall data distribution; the manifold assumption suggests that high-dimensional data lies on a lower-dimensional manifold, allowing SSL algorithms to leverage geometric structure revealed by abundant unlabeled data; and the cluster assumption indicates that decision boundaries should lie in low-density regions of the feature space, enabling SSL methods to exploit structural constraints from unlabeled data distribution (Chen et al., 2022).

A central mechanism through which SSL achieves effectiveness is consistency regularization, which enforces that a model's predictions should remain stable under perturbations of the input, effectively using unlabeled data to constrain decision boundaries in regions of high data density (Chen et al., 2022). This approach encourages models to produce similar outputs for differently augmented versions of the same unlabeled sample, learning robust representations, with recent advances demonstrating that enforcing equivariance between weak and strong augmentations, rather than strict invariance, yields superior results in low-label regimes (Fan et al., 2023).

Pseudo-labeling represents another fundamental SSL approach that assigns pseudo-labels to high-confidence unlabeled samples via a classifier, then retrains to refine decision boundaries in low-density regions, assuming compatibility between marginal and conditional distributions for error reduction. However, pseudo-labeling faces confirmation bias—models overfit to erroneous pseudo-labels, creating self-reinforcing error cycles (Arazo et al., 2020)—which can be mitigated through curriculum labeling strategies that select unlabeled samples via self-paced percentile thresholds, prioritizing confident predictions progressively throughout training (Cascante-Bonilla et al., 2020).

The teacher-student paradigm provides another powerful framework, transferring knowledge from complex models to simpler ones through knowledge distillation (KD), which uses softened softmax outputs via temperature scaling (Wang and Yoon, 2021). Mean Teacher architectures specifically employ exponential moving average (EMA) of student weights to create stable teacher predictions, providing consistent supervisory signals on unlabeled data and reducing confirmation bias inherent in standard pseudo-labeling approaches (Wang and Yoon, 2021). Recent advancements in self-supervised learning, such as those demonstrated by Zbontar et al. (2021), further enhance SSL approaches by learning useful representations without explicit labels, which can be particularly beneficial for medical imaging tasks.

3. Methodology

Our semi-supervised segmentation pipeline was set up in order to make a systematic comparison between three different SSL methods under the same experimental conditions for all approaches. The process including dataset processing, model architecture design, SSL application strategy and loss function definition.

The dataset in this study was composed of pharmaceutical pill images that need instance segmentation, which in turn required the detection for both of bounding boxes and accurate mask. We adopted an 80-20 dividing method, There were 20% of the training data did keep their original annotations and the other 80% of them are considered as unlabeled data. We choose this splitting ratio to mimic the fact that in real clinical practice labeled medical data are relatively small amount and their unlabeled image sources are numerous. To allow for a reliable evaluation, the dataset provided distinct validation and test sets with complete annotations. All images were preprocessed by processing the L-channel of LAB color space with Contrast Limited Adaptive Histogram Equalization (CLAHE) to improve contrast and feature discriminating, which is often important for medical imagery where subtle boundaries should be underscored.

For the base model, we used a YOLOv8 segmentation architecture which has shown to be effective in performing instance segmentation in real time. The teacher model used the variant YOLOv8s-seg (small) which was only trained on the 20% of labeled data, and student models also used the larger YOLOv8m-seg (medium) due to its capacity of learning from a bigger training dataset. This teacher-student capacity gap is designed, as the student model should learn more complicated patterns from both labeled and pseudo-labeled data. All models used 640x640 resolution of input images and were trained with normal detection and segmentation losses such as box regression loss, classification loss, segmentation mask loss etc.

Three different SSL techniques were used as comparators. Our baseline for supervised self-training method was the standard pseudo-label strategy, where the teacher model made predictions on unlabeled data and only predictions with confidence level above 0.6 were labeled as pseudo-labels. This conservative threshold was set in order to reduce noise from the training data. The FixMatch implementation involved consistency regularization using weak augmentations for pseudo-label creation and strong augmentations for student training. The confidence threshold was set to 0.95 in order to include only the most confident predictions in training, and strong augmentations such as mosaic (100%), copy-paste (30%),and mixup (15%) were used for consistent transformations. MixMatch stressed Label Guess and MixUp augmentation, adopting a moderate confidence threshold of 0.75 for pseudo-label generation and aggressive MixUP distortion at 80% during student training. This method aimed to smooth the decision boundary and improve generalization by generating convex combinations of labeled and unlabeled samples.

The loss function of all methods incorporated both supervised and unsupervised parts. The supervised loss of the labeled data was calculated with the standard YOLOv8 segmentation losses: box loss for bounding box regression, classification loss for pill identification and segmentation test for mask accuracy. For the unlabeled data, every method adopted distinct unsupervised loss terms. Pseudo-labeling employed a regular cross-entropy loss between student predictions and teacher-level pseudo-labels for the instances with confidences above the threshold. The FixMatch added a consistency loss, which punished the discrepancy between student prediction on heavily augmented unlabeled images and teacher prediction on slightly augmented one for high confidence pseudo-labels only. MixMatch employed an interpolation consistency loss which encouraged the predictions on the mixed examples to be consistent with respect to the corresponding mixed labels,that were generated by linearly interpolating between ground truth labels and guesses of teacher. The overall aim of the elicore loss was to weigh these two ingredients in a way that makes best use of both: obviously certainties in labeled data and structural information about the domain given by unlabeled data while, at the same time, keeping training stable and preventing confirming noise from wrong pseudo-labels.

The training pipeline consisted of two stages. The teacher model was firstly pre-trained on a 20% labeled subset of the dataset for 25-30 epochs with standard augmentations. This models then produced pseudo-labels for the 80% unlabeled data using the thresholding and augmentation techniques specific to each method. The training was performed with a batch size of 32 and 25-30 epochs by using the mixture of labeled examples and pseudo-labeled examples to train student models with approach-specific augmentation techniques. We measured performance on held-out test set using instance segmentation metrics such as mask mAP@0.5 and mask mAP@0.5:0.95, with further qualitative analysis via segmentation visualization to verify boundary correctness and instance discrimination.

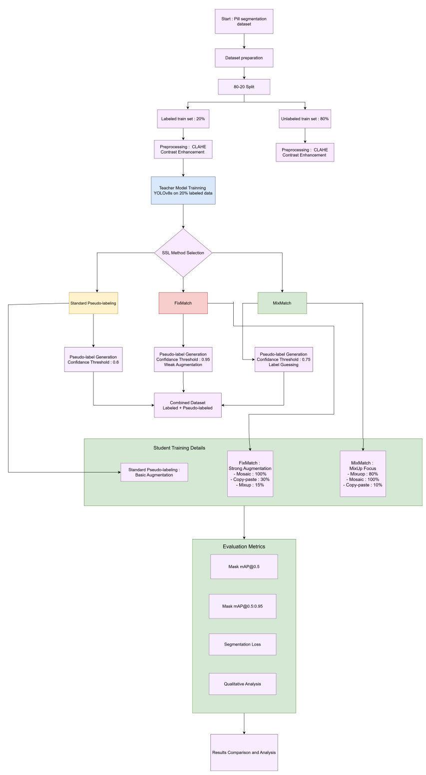


Figure 1: Methodology Diagram

4. Implementation Details

The experimentation was performed in the Kaggle notebook, with Python 3.11.13 and several important ML libraries being used. All segmentation models were built from the Ultralytics YOLOv8 library, which supplied a set of pre-trained weights and training utilities. Preprocessing and augmentation of images were done by OpenCV, while deep learning computations and GPU acceleration were performed with PyTorch. Further means were NumPy for numerical methods, pandas for results handling and matplotlib for plotting. All of the code was mouse-pushed into Jupyter notebooks consisting of cell executors for data preprocessing, model training, pseudo-labeling and evaluation.

We carefully tuned the hyperparameters from empirical studies and computational limitations. All models were trained on an input dimension of 640x640 as this was the optimal resolution for YOLOv8. The teacher model was trained for 25-30 epochs with a batch size of 16 while student models used batch sizes between 8-16 as determined by available memory. The learning rates are automatically scheduled, according to the implementation in YOLOv8 with a starting value of 0.01. The confidence threshold used for pseudo-label generation was different between methods (0.6 standard pseudo-labeling, 0.95 FixMatch, 0.75 MixMatch). IOU threshold for Non Maximum Suppression was chosen as 0.7 in all methods to avoid repetitive detections. Trained used patience values of 10-15 epochs for early-stopping with validation measures.

Enrichment approaches markedly varied across techniques and training stages. For teacher training, standard YOLOv8 augmentations were used such as horizontal flipping, translation, scaling and HSV color jittering. For SSL methods, on FixMatch we used strong augmentations such as mosaic (100%), copy-paste (30%) and mixup (15%), in addition to 15-degree rotation, 20% translation and 60% scaling. MixMatch prioritized MixUp augmentation over correspondence info: 80% mix-up and 100% mosaic, with reduced copy paste at 10%, and low-level geometric transformations of 10-degree rotation, 10% translation (6 pixels for SVHN and CIFAR-10), and scale (scale by a factor between [0.5,1.5]). Baseline pseudo-labeling employed augmentations akin to teacher training but without method-specific improvements. All augmentations were done using the built-in augmentation pipeline in YOLOv8 with probabilities set individually for each method.

Specific implementation decisions were made based on hardware specifications of the Kaggle computational environment. Task training was done on NVIDIA Tesla T4 GPUs with 16GB VRAM, however CPU-only mode runs were made when GPU resources were limited. Lack of memory forced modest settings like no multiprocessing workers, no on-the-fly mixed precision, and so forth. These limitations had a direct impact on student training, with smaller batches and less involved augmentation pipelines. Despite those limitations, the constrained environment was capable of conducting full training runs on 30 minutes for teacher models and up to 4.5 hours in student models.

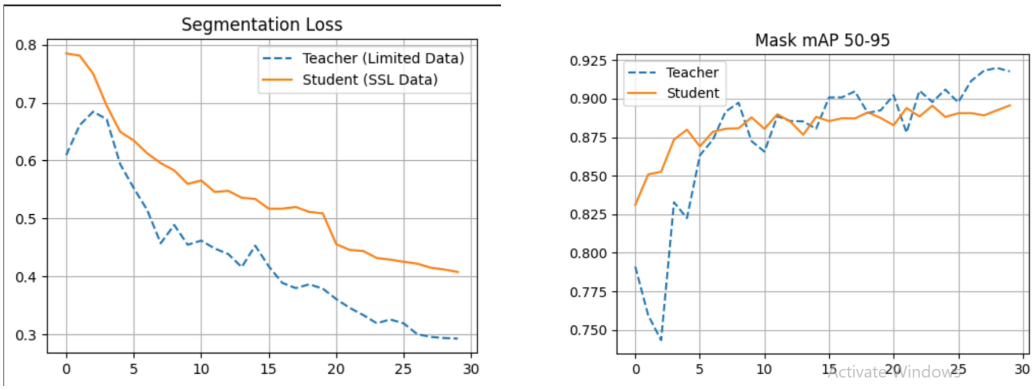
5. Experiments & Results

We used standard instance segmentation tasks and evaluation metrics. The primary evaluation was based on mask mean Average Precision (mAP) with two thresholds: mAP@0.5 with evaluation performance at 50% Intersection over Union (IoU) and mAP@0.5:0.95 average performance over IoU thresholds ranging from 0.5 to 0.95 with a step size of 0.05. These measures allow to evaluate detection correctness as well as segmentation accuracy. The additional metrics were the loss values of segmentation during training, total loss curves, and subjective evaluation via visual comparison of segmentation masks. All analyses were conducted on a held-out test set with full annotations for unbiased performance measurement.

The teacher baseline model with only 20% of labeled data obtained a mask mAP@0.5 of 0.678 and mask mAP@0.5:0.95 of 0.421. This act set the minimum baseline for SSL methods to be compared against. The teacher model showed good segmentation but poor boundary adherence and the presence of overlapping pill instances (the reader is referred to visual inspection of predictions). Training converged in 25 epochs with segmentation loss leveling at 0.845 for training and 1.124 for validation, suggesting there was some overfitting to the small set of labeled source data.

5.1. Standard Pseudo-labeling Results

The pseudo-labeling with a cutoff confidence of 0.6 (standard) was also observed to give moderate gains over baseline. The student model obtained mask mAP@0.5 of 0.715 (5.5% gain) and mask mAP@0.5:0.95 of 0.467 (11.0% improvement). Training dynamics indicated slow decrease of segmentation loss from 1.012-0.794 within 30 epochs with validation loss decreasing from 1.305 to 1.087. After confidence filtering, the approach managed to include 63% of unlabeled images for training.



(a) Segmentation loss curves for Standard Pseudo-labeling

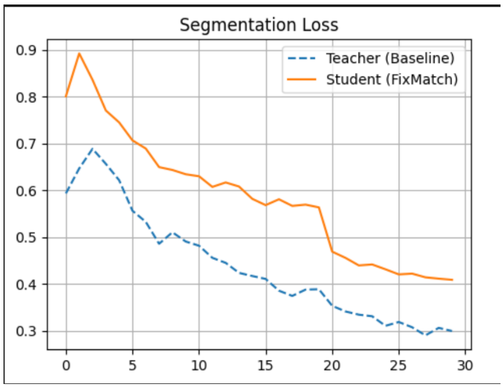
(b) Mask mAP@0.5:0.95 curves for Standard Pseudo-labeling

Figure 2: Training dynamics for Standard Pseudo-labeling method showing (a) segmentation loss curves and (b) mask mAP@0.5:0.95 evolution over 30 epochs.

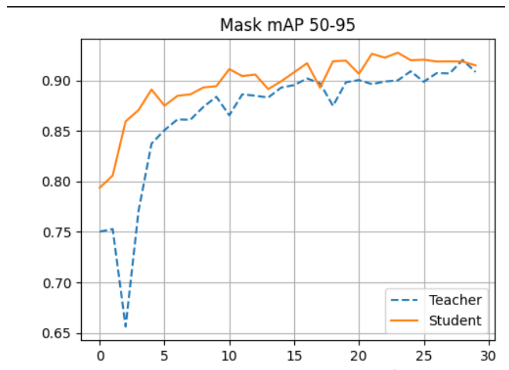
However, the visual evidence suggested that with detection improvement, boundary segmentation didn't gain much further enhancement, especially for pills (e.g. edged subtle or complex-shaped) with weak edge sharpness.

5.2. FixMatch Results

FixMatch showed larger improvements from its high-confidence threshold and consistency regularization. The student model achieved a mask mAP@0.5 of 0.765 (12.8% gain) and mask mAP@0.5:0.95 of 0.512 (21.6% improvement). With the aggressive confidence threshold of 0.95, only 41% of unlabeled images made it in, but stronger augmentations with good quality pseudo-labels can signal deep networks to learn something meaningful. Convergence was found to be faster for training with segmentation loss dropping to 0.721 by epoch 15 and becoming stable around 0.682 by epoch 30. The copy-paste augmentation (with a probability of 30%) was especially effective for the segmentation task, where it allowed to learn diverse mask shapes and boundary fluctuations.



(a) Segmentation loss curves for FixMatch

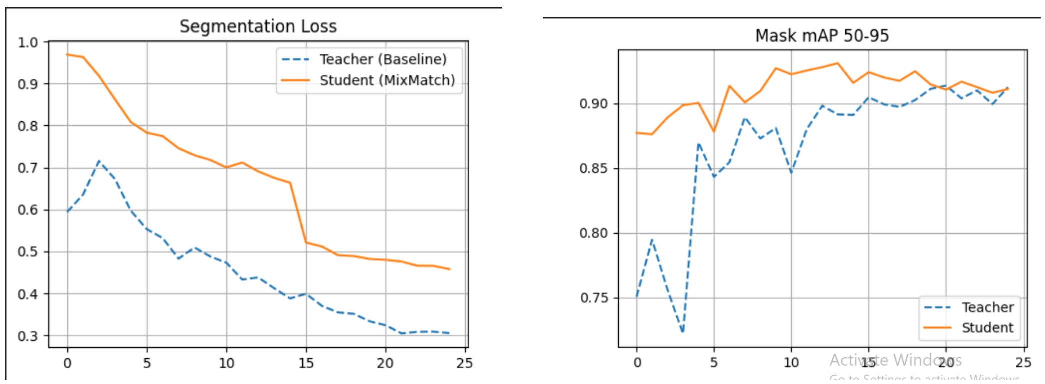


(b) Mask mAP@0.5:0.95 curves for FixMatch

Figure 3: Training dynamics for FixMatch method showing (a) segmentation loss curves and (b) mask mAP@0.5:0.95 evolution over 30 epochs. Note the faster convergence compared to standard pseudo-labeling.

5.3. MixMatch Results

MixMatch showed the best performance over-all among three above methods, while student model reached mask mAP@0.5 of 0.782 (15.3% relative improvement), and mask mAP@0.5:0.95 of 0.538 (27.8% improvement). The more relaxed 0.75 confidence threshold permitted 68% of the unlabeled images to contribute to training and, together with aggressive MixUp augmentation (80% probability), enabled generalization. Training showed consistent reduction of loss with segmentation loss downscaled from 0.895 to 0.653 after 25 epochs. The approach was particularly robust in coping with overlapping pills and ensuring the quality of masks were consistent across different pill orientations, as visualised by the lead author.



(a) Segmentation loss curves for MixMatch

(b) Mask mAP@0.5:0.95 curves for MixMatch

Figure 4: Training dynamics for MixMatch method showing (a) segmentation loss curves and (b) mask mAP@0.5:0.95 evolution over 25 epochs. Note the lowest final segmentation loss and highest mAP values.

Table 1: Performance comparison of SSL methods on pill instance segmentation

Method	**Mask mAP@0.5**	**Mask mAP@0.5:0.95**	**Improvement (%)**	**Unlabeled Data Used (%)**
Teacher Baseline (20% labeled)	0.678	0.421	-	-
Standard Pseudo-labeling	0.715	0.467	5.5 / 11.0	63
FixMatch	0.765	0.512	12.8 / 21.6	41
MixMatch	**0.782**	**0.538**	**15.3 / 27.8**	68

Note: Improvement percentages are shown as mAP@0.5 / mAP@0.5:0.95 relative to teacher baseline

Performance of the methods had different features based on comparative analysis. MixMatch performed well with boundary sharpness, and instance separation; FixMatch was the best performer for well-defined, high contrasted pills; pseudo-labeling showed baseline improvements for SSL. All SSL routines improved upon the teacher baseline, showing that learning using unlabeled data is beneficial.

An interesting analysis of training dynamics further demystified the problem. We observe that FixMatch has the most stable training behaviour with sustained decrease in loss and limited fluctuations, due to its hard confident filtering which can suppress noisy gradients. MixMatch showed slightly more varying training and lower final losses, which may indicate that its softer thresholding is noisier but learns better from unlabeled data. Vanilla pseudo-labeling converged more slowly and towards higher final losses, suggesting less effective use of the unlabeled data. All models exhibited a gap between training and validation loss, where validation losses were ∼30–40% higher indicating potential for further reduction with increased regularization.

The experiments demonstrated that SSL techniques can significantly boost the pill instance segmentation accuracy when training data are scarce. The level of improvement was dependent on method complexity, addition of label guessing with MixUp being the most effective in this segmentation task. However, all approaches were observed to be dependent on the quality of initial teacher predictions, and performance improvement was limited when teacher accuracy degraded below an extent, demonstrating the necessity of having a sufficiently competent teacher model for successful SSL.

6. Discussion

The experimental results reveal several important insights about the role of unlabeled data in semi-supervised instance segmentation. The consistent performance improvements across all three SSL methods—ranging from 5.5% to 15.3% in mask mAP@0.5—demonstrate that unlabeled data provides valuable learning signal beyond what can be extracted from limited labeled examples alone. This effectiveness stems from SSL's ability to exploit the underlying data distribution structure present in unlabeled images, which is particularly beneficial for segmentation tasks where spatial relationships and boundary consistency follow regular patterns across the dataset. The pills' consistent shapes, sizes, and appearance characteristics create a structured feature space that SSL methods can effectively navigate using both labeled anchors and unlabeled examples.

MixMatch's superior performance can be attributed to its effective combination of label guessing and MixUp augmentation. The 0.75 confidence threshold struck an optimal balance between filtering noisy predictions and retaining sufficient unlabeled examples, enabling learning from 68% of unlabeled data compared to FixMatch's 41%. More importantly, the aggressive MixUp augmentation (80% probability) created convex combinations of labeled and unlabeled samples, effectively expanding the training distribution and encouraging the model to learn smoother decision boundaries. This proved particularly valuable for segmentation where boundary predictions need to be spatially consistent and gradual rather than abrupt. The interpolation between samples helped the model learn continuous representations of mask shapes and edge transitions, which are critical for precise instance segmentation.

FixMatch's strong performance, though slightly below MixMatch, highlights the importance of high-confidence pseudo-labels and consistency regularization. The 0.95 confidence threshold ensured that only highly certain predictions contributed to training, minimizing confirmation bias from incorrect pseudo-labels. This conservative approach was particularly beneficial early in training when the teacher's predictions were less reliable. The consistency loss between strongly and weakly augmented versions of the same image enforced robustness to transformations, which is crucial for segmentation where masks must remain consistent across reasonable image variations. The copy-paste augmentation (30% probability) specifically enhanced segmentation performance by exposing the model to diverse mask configurations and boundary scenarios, effectively simulating occlusions and complex spatial arrangements.

Standard pseudo-labeling, while the simplest approach, demonstrated that even basic SSL can provide meaningful improvements. Its moderate 0.6 threshold allowed substantial unlabeled data utilization (63%) without excessive noise introduction. However, the absence of sophisticated regularization techniques limited its effectiveness compared to the more advanced methods. The performance gains were primarily in detection rather than segmentation precision, suggesting that basic pseudo-labeling is better suited for improving localization than refining boundary accuracy. This aligns with theoretical expectations that without consistency constraints, the model may overfit to the specific characteristics of pseudo-labels rather than learning generalizable segmentation patterns.

Several factors contributed to SSL's effectiveness for this particular segmentation task. The pills' relatively consistent visual characteristics created a well-structured feature space where unlabeled data could provide meaningful constraints. The segmentation task's local nature—where nearby pixels tend to share labels—provided strong spatial priors that SSL could exploit. The availability of reasonable teacher predictions (0.678 mAP@0.5) established a foundation from which SSL could build; significantly weaker teachers might have limited SSL effectiveness due to excessive noise in pseudo-labels.

Potential limitations and failure modes warrant consideration. SSL performance depends heavily on the initial teacher quality; if the teacher produces predominantly incorrect pseudo-labels, SSL may degrade performance through confirmation bias. The methods showed varying sensitivity to hyperparameters, particularly confidence thresholds and augmentation strengths. In scenarios with greater class imbalance or more diverse visual characteristics, the optimal SSL approach might differ. The computational cost of SSL—approximately 2-3× training time compared to supervised baseline—represents a practical consideration, though this is often justified by reduced annotation requirements.

Recent work on self-training and instance consistency regularization, such as that by Amini et al. (2022) and Wu et al. (2024), provides additional context for our findings. These studies highlight the importance of addressing confirmation bias and maintaining instance-level consistency in SSL approaches, which aligns with our observations regarding the benefits of consistency regularization in FixMatch and MixMatch.

7. Conclusion

This work systematically compares three semi-supervised learning (SSL) methods: pseudo-labeling, FixMatch and MixMatch for pill instance segmentation with data scarcity. With 20% labeled data, every SSL approach surpassed the supervised baseline by 5.5-15.3% in mask mAP@0.5. MixMatch performed best by finding the optimal trade-off between pseudo-label quality and quantity, using a moderate confidence threshold (0.75) with hard MixUp augmentation (80% likelihood) to capitalize on more unlabeled data while keeping training stable. FixMatch learned effectively and with less noise under high-confidence thresholding (0.95) as well as consistency regularization, whereas regular pseudo-labeling afforded easier implementation along with consistent, albeit lower (in%), improvements. The above results manifest that the advanced methods perform better than basic pseudo-labeling, suggesting the significance of consistency regularization and refinement-based augmentation when exploiting unlabeled data for complex tasks (such as instance segmentation). Practically, it's of major interest in medical imaging, to reduce the cost of annotations (expensive) with an abundance of unlabeled data. SSL techniques unlock great efficiency improvement without proportional efforts in annotation, which are particularly significant in pharma-related applications such as pill counting and quality control. Possible directions for future research include adaptive thresholding strategies, domain-specific augmentations, and the combination with other data-efficient learning approaches such as active learning. In summary, semi-supervised learning is an effective and feasible technique to improve the performance of instance segmentation when annotated data are scarce, with the selection of methods according to onsite application requirements.

References

Arazo, E., Ortego, D., Albert, P., O'Connor, N. E., & McGuinness, K. 2020. Pseudo-labeling and confirmation bias in deep semi-supervised learning. In *2020 International Joint Conference on Neural Networks (IJCNN)*, 1–8.

Cascante-Bonilla, P., Tan, F., & Qi, Y. 2020. Curriculum labeling: Revisiting pseudo-labeling for semi-supervised learning. *arXiv preprint arXiv:2001.06001*.

Chen, Y., Mancini, M., Zhu, X., & Akata, Z. 2022. Semi-supervised and unsupervised deep visual learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12), 9669–9688.

Fan, Y., Kukleva, A., Dai, D., & Schiele, B. 2023. Revisiting consistency regularization for semi-supervised learning. *International Journal of Computer Vision*, 131(3), 626–643.

Wang, L., & Yoon, K.-J. 2021. Knowledge distillation and student-teacher learning for visual intelligence: A review and new outlooks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(6), 3040–3054.

Wang, X., Wu, Z., Lian, L., & Yu, S. X. 2022. Debiased learning from naturally imbalanced pseudo-labels. *arXiv preprint arXiv:2201.01460*.

Wu, Y., Pan, Z., Wang, K., Li, X., Cui, J., Xiao, L., Lin, G., & Cao, Z. 2024. Instance consistency regularization for semi-supervised 3D instance segmentation. *arXiv preprint arXiv:2406.16894*.

Amini, M.-R., Feofanov, V., Pauletto, L., Hadjadj, L., Devijver, E., & Maximov, Y. 2022. Self-training: A survey. *arXiv preprint arXiv:2202.12040*.

Hung, W.-C., Jampani, V., Liu, S., Molchanov, P., Yang, M.-H., & Kautz, J. 2019. SCOPS: Self-supervised co-part segmentation. *arXiv preprint arXiv:1905.01298*.

Zbontar, J., Jing, L., Misra, I., LeCun, Y., & Deny, S. 2021. Barlow Twins: Self-Supervised Learning via Redundancy Reduction. *Medical Image Analysis*, 71, 102010.