Deep Learning (DSE-316) Course Project

Aditya Mishra
aditya21@iiserb.ac.in
Roll Number: 21013

Mohammad Saifullah Khan
mohammad21@iiserb.ac.in
Roll Number: 21169

Sattwik Kumar Sahu
sattwik21@iiserb.ac.in
Roll Number: 21241

April 15, 2024

1 Introduction

Recent advancements in semi-supervised semantic segmentation (SSS) [2, 28, 3] have gained significant attention due to their potential to improve segmentation accuracy while reducing the reliance on large annotated datasets. These methods have shown promising results, but often at the expense of increasing complexity in model design and training procedures. To resolve it, Zhao et al. [29] proposed a novel approach called AugSeg, which prioritized simplicity and efficiency in enhancing SSS performance.

The application of SSS is crucial in various fields such as medical imaging, autonomous driving, and satellite imagery analysis, where accurate segmentation of objects or regions of interest is essential for decision-making processes [29]. Traditional supervised semantic segmentation methods require extensive manual annotation of training data, which is time-consuming and costly [27]. In contrast, SSS methods leverage both labeled and unlabeled data, offering a more cost-effective solution for training segmentation models [29]. AugSeg deviates from the current trend of complex SSS approaches by focusing on data perturbations to boost performance. By simplifying data augmentation techniques and adopting an adaptive approach to inject labeled information into unlabeled samples, AugSeg aims to achieve state-of-the-art performance on SSS benchmarks while maintaining simplicity in design and implementation [29].

Through this report, we aim to reproduce the results of the **Augmentation Matters: A Simple-yet-Effective Approach to Semi-supervised Semantic Segmentation** [29] and explore avenues for introducing novelty into the existing model.

2 Literature Survey

The effectiveness of semi-supervised learning heavily relies on the strategic utilization of unlabeled data [10, 12, 17, 21]. Recently, consistency regularization (CR) [16, 23] has emerged as a fundamental technique for training models with both labeled and unlabeled data concurrently. These CR-based methods, applied in tasks such as classification [7, 22, 24, 30] and segmentation [9, 14], rely on diverse perturbation techniques to induce disagreement on identical inputs. This enables models to be trained by enforcing prediction consistency on unlabeled data without prior knowledge of labeled information. Consequently, numerous semi-supervised (SSS) methods have surfaced, each contributing to the field's advancement.

Based on our review, three primary directions are identified for enhancing SSS performance: **augmentations**, **more supervision**, and **pseudo-rectifying**. Most existing studies employ robust data augmentations to perturb unlabeled data, with some also perturbing inputs at the feature level [18, 25]. Under the **more supervision** approach, multiple training branches, stages, or losses (MBSL) are widely embraced to introduce model perturbations [13]. Given the critical role of pseudo-label quality in semi-supervised training, methods like ECS [19] and ELN [15] introduce additional trainable correcting networks (ACN) to refine pseudo-labels further. While recent state-of-the-art (SOTA) methods [11, 15] yield promising performance, they often combine increasingly complex mechanisms, such as contrastive learning [18] and multiple ensemble models. In contrast, AugSeg's [29] approach aims for simplicity and clarity, primarily leveraging data augmentations to enhance SSS performance.

Data augmentations stand out as the most straightforward and effective means of producing label-preserving perturbations, playing a central role in CR-based semi-supervised studies [9, 20]. Although various auto-augmentation strategies [5, 6] from supervised learning have been adopted in semi-supervised research, direct application poses challenges. Auto augmentations aim to identify optimal augmentation strategies, searching for the best augmentation operations and distortion strengths within a predefined finite discrete space. Conversely, the goal of data augmentation in SSS is to generate diverse inputs without specific objectives and searching spaces. Additionally, direct application of these augmentations may excessively distort unlabeled data and disrupt the data distribution, leading to performance degradation [26]. Instead of employing additional rectifying strategies like distribution-specific batch normalization [1], AugSeg [29] simplify the standard randomAug [6] with a highly random design. By randomly selecting the number of augmentations and uniformly sampling augmentation strength from a continuous interval, their approach promotes better data diversity while minimizing the risk of over-distorting samples.

3 Methodology

The AugSeg methodology for semi-supervised semantic segmentation, as presented in the paper, relies on a simple and clean framework that primarily utilizes data augmentation techniques to enhance the performance of the model. The methodology consists of three main components: weak geometrical augmentation (A_g), random intensity-based augmentation (A_r), and adaptive label-aided augmentation (A_a).

1. **Weak Geometrical Augmentation (A_g):** This component includes standard resizing, cropping, and flipping operations. It is applied to the labeled data to generate different views of the same image, which helps the model learn more robust features.

2. **Random Intensity-based Augmentation (A_r):** This component is designed to perturb unlabeled data by sampling the distorting degree uniformly in a continuous space and randomly selecting a number of augmentations from an augmentation pool. The pool is simplified from the pool in RandomAug, and strong intensity-based transformations like Invert operations are removed to prevent over-distortion of the data. Figure 1a depicts how random intensity-based augmentation works.

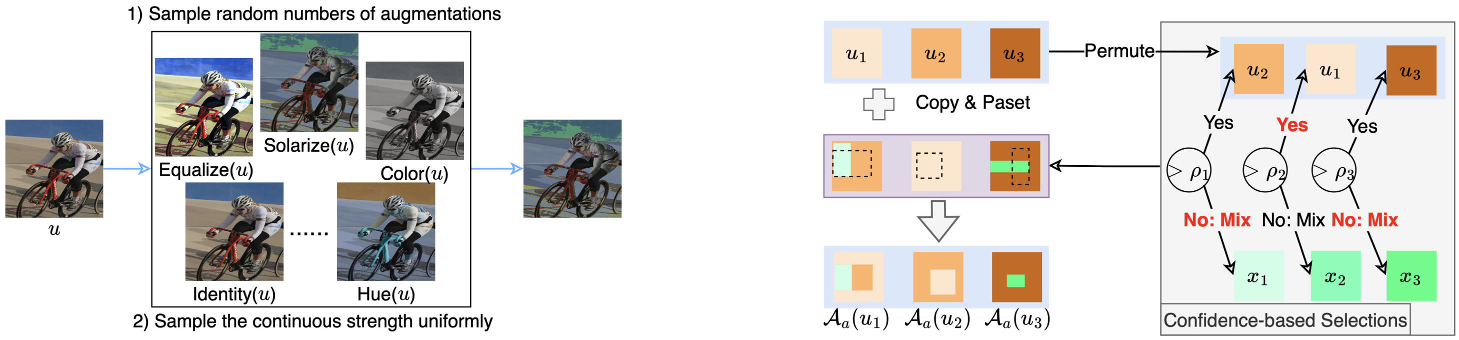
3. **Adaptive Label-aided Augmentation (A_a):** This component is designed to make full use of labeled data to aid the training on unlabeled samples in an instance-specific and confidence-adaptive manner. A confidence score is estimated for each unlabeled instance, and the mixing between labeled and unlabeled instances is randomly applied based on the confidence score. This helps the model to focus more on the less confident unlabeled samples, which are more likely to be aided by the confident labeled samples. Its working is illustrutated by Figure 1b.

The overall framework of AugSeg involves training a student model and a teacher model simultaneously on labeled and unlabeled data. The teacher model is updated gradually via the exponential moving averaging of the student weights. The student model is trained by minimizing a supervised loss (L_x) and an unsupervised consistency loss (L_u) at the same time. The unsupervised loss is formulated based on the prediction disagreement between the student model and the teacher model on augmented unlabeled data. The working of AugSeg can be understood by Figure 2. The teacher model is capable of producing pseudo-labels for training on unlabeled data, and will be updated gradually via the exponential moving averaging of the student weights, i.e.,

$$\theta_t \leftarrow \alpha\theta_t + (1-\alpha)\theta_s \tag{1}$$

The choice of exponential moving average (EMA) for updating the teacher model's weights in the AugSeg framework is based on the idea of gradually evolving the teacher model towards the student model's weights, which are updated using gradient descent. The EMA method is chosen because it provides a smooth and stable update mechanism for the teacher model, allowing it to gradually adapt to the student model's learning. The total training loss for the student model is given by:

$$L = L_x + \lambda_u L_u \tag{2}$$



(a) Random Intensity-based Augmentation

(b) Adaptive Label-aided Augmentation (CutMix)

Figure 1: Augmentations: The sub-figure (a) depicts visualization of random intensity based augmentation. The sub-figure (b) gives visualization of adaptive label-aided CutMix augmentation in a mini-batch. x_i and u_i denote the labeled and unlabeled crops, respectively. ρ_i denote the confidence score for i^{th} unlabeled sample. The core idea of A_a is that, these less confident unlabeled samples, with lower values of ρ_i, are more likely to be aided (mixed) by these confident labeled samples. For the complete understanding of how CutMix works, we request the intrested readers to the work of Zhao et al. [29].

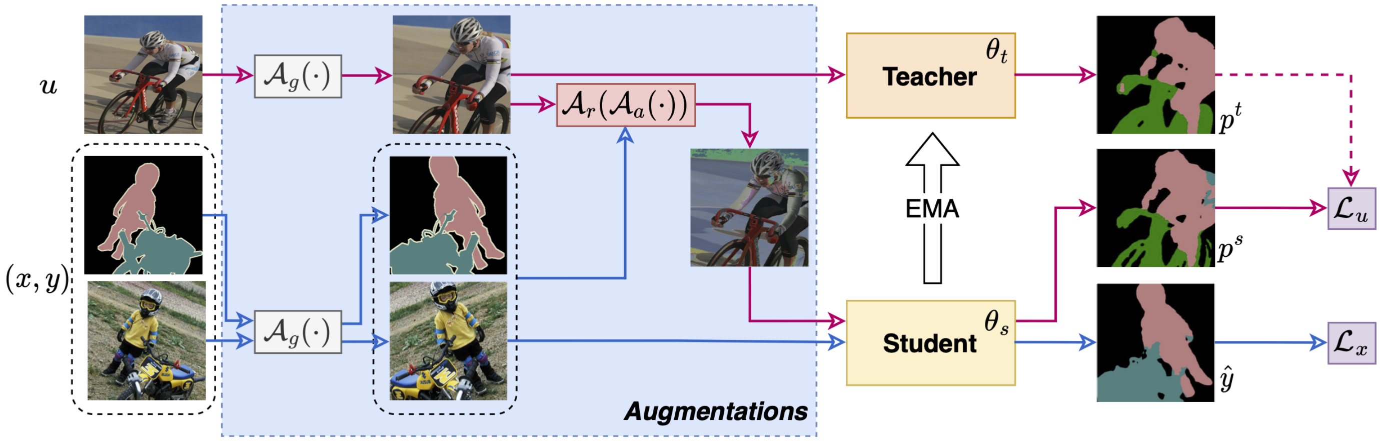


Figure 2: AugSeg operates within a standard teacher-student framework, where the student model (θ_s) is trained on both labeled (x, y) and unlabeled (u) data simultaneously. This training minimizes two key losses: supervised loss (L_x) and unsupervised consistency loss (L_u). Meanwhile, the teacher model (θ_t) is updated through exponential moving averaging (EMA) of θ_s and generates pseudo-labels (p_t) for unlabeled data. AugSeg's core strategy involves applying diverse augmentation techniques to unlabeled samples: weak geometrical augmentation (A_g), random intensity-based augmentation (A_r), and adaptive label-aided augmentation (A_a). The diagram depicts the flow of labeled and unlabeled data (represented by red and blue lines, respectively), with dashed lines indicating the stop gradient operation.

where λ_u is a scalar hyper-parameter to adjust the unsupervised loss weight. Similar to most SSS methods [14, 19], Zhao et al. [29] adopted a standard pixel-wise cross-entropy loss L_{ce} to train on labeled data directly. Mathematically,

$$L_{ce} = \frac{1}{|B_x|} \sum_{i=1}^{|B_x|} \frac{1}{H \times W} \sum_{j=1}^{H \times W} L_{ce}(\hat{y_i}(j), y_i(j)) \tag{3}$$

where $\hat{y} = f(A_g(x_i); \theta_s)$, represents the segmentation result of the student model on the i^{th} weakly-augmented labeled instance, and j represents the j^{th} pixel on the image or the corresponding segmentation mask with a resolution of $H \times W$. As for leveraging the unlabeled data, which is the key to semi-supervised learning, AugSeg mainly on the data perturbation $\tau(\cdot)$ to generate the prediction disagreement. First, segmentation predictions are obtained, p_i^s and p_i^t, of the student model on augmented $\tau(u_i)$ and of the teacher model on augmented $A_g(u_i)$, respectively. Subsequently, the unlabeled loss is formulated as:

$$p_i^t = f(A_g(u_i); \theta_t) \tag{4}$$

$$p_i^s = f(\tau(A_g(u_i)); \theta_s) \tag{5}$$

$$L_u = \frac{1}{|B_u|} \sum_{i=1}^{|B_u|} \frac{1}{H \times W} \sum_{j=1}^{H \times W} L_{ce}(p_i^s(j), p_i^t(j)) \tag{6}$$

In summary, AugSeg is a simple and effective methodology for semi-supervised semantic segmentation that relies on data augmentation techniques to enhance the performance of the model. The methodology is designed to generate different views of the same image and perturb unlabeled data in a controlled manner, which helps the model to learn more robust features and generalize better to unseen data.

4 Contributions

4.1 Different Means to Update Teacher Weights

Different types of means can be used to update the teacher weights from student weights. This can be done to check what method of averaging the series of weights of the students yields the best results and provides the best stability and robustness to the teacher weights.

4.1.1 Weighted Moving Average (WMA)

A Weighted Moving Average (WMA) assigns different weights to each data point within the window. The window below is the set of weights of the student in the last k time steps $\theta_s[t-(k-1)]$ to $\theta_s[t]$. The weights w_i can be chosen to match the required task

$$\theta_t \leftarrow \frac{\sum_{i=0}^{k-1} \theta_s[t-i] \cdot w_i}{\sum_{i=0}^{k} w_i} \tag{7}$$

4.1.2 Simple Moving Average (SMA)

A simple average over the weights of the student in the last k time steps, obtained by setting all $w_i = 1$ in Equation (7)

4.1.3 EMA with Variable Decay Rate

Varying the decay factor α in (1) according to a window of the student weights θ_s in the last k time steps helps the decay to be adjusted according to the variation in the window, providing more adaptability and robustness to the averaged weights.

For example,

$$\begin{aligned} \Theta_s &= \{\theta_s[t], \theta_s[t-1], \ldots, \theta_s[t-(k-1)]\} \\ \sigma &= \text{std}(\Theta_s) \\ \alpha_{max} &= \max \Theta_s \\ \alpha_{min} &= \min \Theta_s \end{aligned}$$

We define α as

$$\alpha := \alpha_{min} + (\alpha_{max} - \alpha_{min}) \cdot e^{-\lambda \times \sigma} \tag{8}$$

Where λ is a scaling factor which controls how much the α responds to changes in volatility $(\alpha_{max} - \alpha_{min})$

4.2 Multi Teacher Knowledge Distillation

In the literature, multiple teachers can distill knowledge into one student, called Multi-Teacher Knowledge Distillation [31]. The proposed method involves a hierarchical structure where multiple larger models are trained by the student-teacher architecture proposed by AugSeg [29], with each of them following different teacher weight update strategies as described in Section. These ”students” then act as teachers, distilling knowledge into a smaller model through Multi-Teacher Knowledge Distillation, as seen in Section 4.1. The hierarchy is shown in Figure 3.

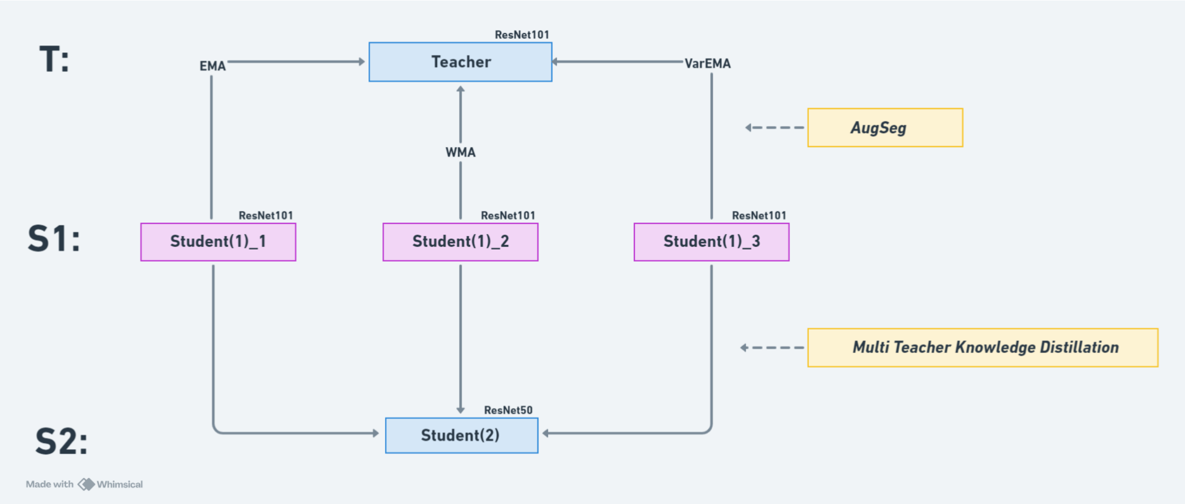


Figure 3: This diagram represents a knowledge distillation framework where a ’Teacher’ neural network, specifically a ResNet101 architecture, is used to train multiple ’Student’ models. The Teacher model’s knowledge is transferred to Student models at the first stage (S1), each a ResNet101 architecture, through processes labeled EMA, WMA, and VarEMA, which stand for Exponential Moving Average, Weighted Moving Average and Exponential Moving Average with variable decay rate respectively. The dotted lines represent the AugSeg strategy. At the second stage (S2), the knowledge from the first stage students is further distilled into a ’Student(2)’ model with a ResNet50 architecture. The box labeled ’Multi Teacher Knowledge Distillation’ implies this process involves multiple teacher models or a complex distillation process from a single teacher model to multiple student models.

4.2.1 T and S1

The root teacher model T is the teacher for the first level students in S_1. Models in S_1 are trained using the method described in [29], with each student has a different weight update strategy for the teacher T. The teachers weights are updated by averaging over the weight updates $\theta_T^{(1,i)}$ from all n_{S_1} models in S_1 level[1].

$$\theta_T \leftarrow \frac{1}{n_{S_1}} \sum_{i=0}^{n_{S_1}} \theta_T^{(1,i)} \tag{9}$$

4.2.2 S_1 and S_2

This is the step involving Multi-Teacher Knowledge Distillation. In this method, pseudo-label logits $\hat{y}_{S1}^{(1)} \ldots \hat{y}_{S1}^{(n_{S_1})}$ are generated by the S_1 models. These labels are in the form of real numbers. A parameter called temperature *tau* is also chosen, which determines how harshly lower values are rejected by the softmax function. The logits output from model S_2 are $\hat{y}_{S2}^{(i)}$. On each

[1] This is a very primitive strategy and can be refined to include probabilistic selection and weighting of the $\theta_T^{(1,i)}$ from each student model, or also learnable weighting of the updates

of these $\hat{y}_{S1}^{(i)}$ we calculate

$$p_{S1}^{(i)} = \text{softmax}(\hat{y}_{S1}^{(i)}) \tag{10}$$

$$p_{S2} = \text{softmax}(\hat{y}_{S2}) \tag{11}$$

Then we calculate the cross-entropy loss of the S_2 model as usual (y is the ground truth)

$$\mathcal{L}_{ce} = -\sum_{i} y_i \log(p_{S2})$$

We then calculate the average KL-Divergence of output of student with the teacher models

$$\mathcal{L}_{kl} = \sum_{i=0}^{n_{S_1}} \text{KL}(p_{S2}, p_{S1}^{(i)}) \tag{12}$$

The total loss for the student model S_2 is given by

$$L_{S_2} = \mathcal{L}_{ce} + \mathcal{L}_{kl}$$

This loss is then backpropagated through the S_2 student model to update its weights. Thus the root teacher is used to train the S_1 models which in turn train the S_2 model.

4.3 Advantages

1. The teacher model in T could be a very large pre-trained model (example in Figure 3). The models in S_1 also have the same size as T. However, S_2 could be a much smaller model, which, through knowledge distillation learns to emulate the larger and more powerful models, providing compression - a smaller model can provide performance similar to a larger one.

2. The models in S_1 have different weight update strategies for T, which ensures robustness of the weight update for T, not relying solely on one student model, as shown in AugSeg [29].

3. The Multi-Teacher Knowledge Distillation provides ensemble learning, combining the knowledge of multiple larger powerful models into a smaller, faster model.

4. The student model is less prone to adversarial attacks, because the student model S_2 learns to mimic the S_1 teachers' decision-making process, which may be more resilient to perturbations.

5 Datasets

5.1 Cityscapes

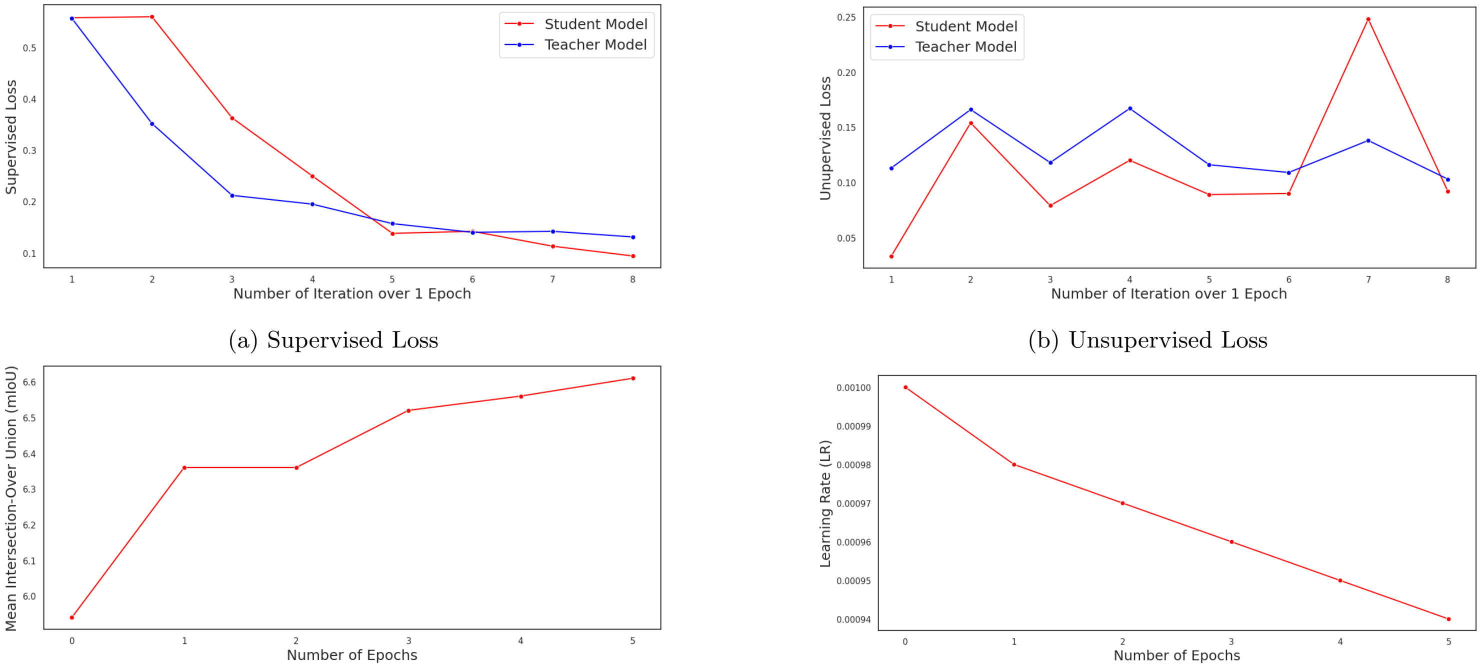
The Cityscapes [4] dataset, used in the AugSeg paper for evaluating semi-supervised semantic segmentation methods, comprises 2,975 training images and 500 validation images, all finely annotated across 19 semantic urban scene classes. This dataset is crucial for understanding urban scenes and is employed to demonstrate the effectiveness of the AugSeg method under various training partitions. AugSeg leverages both labeled and unlabeled data, showcasing its ability to perform robustly even with limited labeled data, a common scenario in real-world applications. Performance on the Cityscapes dataset is assessed using the mean intersection-over-union (mIoU), a standard metric for semantic segmentation, which quantifies the accuracy of the model across different classes.

5.2 Pascal VOC

The Pascal VOC [8] 2012 dataset, as detailed in the paper, is a standard benchmark for evaluating semi-supervised semantic segmentation methods and includes 1,464 fine-labeled training images and 1,449 validating images, annotated across 21 semantic classes including the background. Additionally, the study incorporates 9,118 coarsely-labeled images from the Segmentation Boundary dataset (SBD), bringing the total training dataset to 10,582 images. This dataset is pivotal in testing the AugSeg method under various training partitions, demonstrating its effectiveness in environments with both densely and sparsely labeled data. The performance on the Pascal VOC 2012 dataset is assessed using the mean intersection-over-union (mIoU), a common metric for semantic segmentation accuracy. Results show that AugSeg achieves state-of-the-art performance, significantly improving outcomes especially in scenarios with limited labels, thus highlighting its capability to leverage both labeled and unlabeled data effectively.

6 Training Plots

Figure 4 presents a comprehensive visualization of the training process. It includes plots representing the loss function evolution, mean Intersection-over-Union (mIoU) across epochs, and the learning rate dynamics. These visualizations offer valuable insights into the performance and optimization trajectory of the model during training.



(a) Supervised Loss

(b) Unsupervised Loss

(c) Mean Intersection-Over Union (mIoU)

(d) Learning Rate

Figure 4: Training Plots: Sub-figure (a) illustrates the supervised loss of the student and teacher models over the number of iterations for an epoch, while sub-figure (b) depicts the unsupervised loss of the student and teacher models over the number of iterations for an epoch. Sub-figure (c) shows the mean Intersection-over-Union (mIoU) over 5 epochs, and sub-figure (d) illustrates the learning rate over 5 epochs.

7 Evaluation Metrics and Comparison

In evaluating the performance of the AugSeg methodology for semi-supervised semantic segmentation, we employ standard evaluation metrics to assess its effectiveness compared to existing methods. The primary evaluation metric utilized is the mean Intersection-over-Union (mIoU), which quantifies the accuracy of semantic segmentation models by measuring the overlap between predicted and ground truth segmentation masks across different classes.

To ensure a fair comparison, Zhao et al.[29] benchmarked the performance of AugSeg against state-of-the-art semi-supervised semantic segmentation methods on established datasets such as Cityscapes and Pascal VOC 2012. AugSeg's ability to leverage both labeled and unlabeled data to improve segmentation accuracy, particularly in scenarios with limited labeled data.

8 Discussion

During the experimentation phase, we encountered several challenges and issues that influenced our choice of dataset and model configuration. One notable issue arose when attempting to run the ResNet50 and ResNet101 models on the Cityscapes [4] dataset. Despite efforts to debug the code and adjust parameters such as kernel size and image input size, we encountered a persistent error related to the mismatch between kernel size and input size. Despite our best efforts, we were unable to resolve this error, prompting us to explore alternative datasets and model configurations.

Subsequently, we opted to utilize the Pascal VOC [8] dataset with the ResNet50 and ResNet101 models. However, upon accessing the dataset provided in the GitHub repository by Zhao et al.[29], we encountered corruption issues and missing files. To address this issue, we promptly raised an issue in the repository and requested the missing files from the authors. Upon receiving the necessary files, we proceeded to run the ResNet101 model on the Pascal dataset.

Despite the challenges encountered during the experimental process, we were able to successfully execute the ResNet101 model on the Pascal dataset and obtain meaningful results. These results are discussed in detail in this report, providing valuable insights into the performance and efficacy of the AugSeg methodology in the context of semi-supervised semantic segmentation.

9 Conclusion

In this report, we presented an in-depth exploration of the AugSeg methodology for semi-supervised semantic segmentation. By leveraging data augmentation techniques and adaptive label-aided augmentation, AugSeg offers a simple yet effective approach to improving segmentation accuracy while reducing the reliance on large annotated datasets. Through extensive experimentation and evaluation on datasets such as Cityscapes and Pascal VOC 2012, Zhao et al.[29] demonstrated AugSeg's capability to achieve state-of-the-art performance in various training scenarios.

Our investigation revealed that AugSeg excels in environments with both densely and sparsely labeled data, showcasing its robustness and versatility across different segmentation tasks. By prioritizing simplicity and efficiency in model design and training procedures, AugSeg presents a promising solution for real-world applications where annotated data may be limited or costly to obtain.

In this report, we introduced innovative approaches for updating teacher weights, including Weighted Moving Average (WMA), Simple Moving Average (SMA), and Exponential Moving Average (EMA) with Variable Decay Rate, providing insights into their impact on model stability and robustness. Additionally, we investigated the application of Multi-Teacher Knowledge Distillation, leveraging hierarchical structures to distill knowledge from multiple teacher models into a single student model. These contributions expand the methodological repertoire for SSS and offer valuable insights into optimizing model performance in real-world scenarios.

References

[1] Woong-Gi Chang, Tackgeun You, Seonguk Seo, Suha Kwak, and Bohyung Han. Domain-specific batch normalization for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, pages 7354–7362, 2019.

[2] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017.

[3] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018.

[4] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[5] Ekin D Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V Le. Autoaugment: Learning augmentation policies from data. *arXiv preprint arXiv:1805.09501*, 2018.

[6] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 702–703, 2020.

[7] Yue Duan, Zhen Zhao, Lei Qi, Lei Wang, Luping Zhou, Yinghuan Shi, and Yang Gao. Mutexmatch: semi-supervised learning with mutex-based consistency regularization. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[8] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111(1):98–136, January 2015.

[9] Geoff French, Timo Aila, Samuli Laine, Michal Mackiewicz, and Graham Finlayson. Semi-supervised semantic segmentation needs strong, high-dimensional perturbations. 2019.

[10] Yves Grandvalet and Yoshua Bengio. Semi-supervised learning by entropy minimization. *Advances in neural information processing systems*, 17, 2004.

[11] Dayan Guan, Jiaxing Huang, Aoran Xiao, and Shijian Lu. Unbiased subclass regularization for semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9968–9978, 2022.

[12] Guan Gui, Zhen Zhao, Lei Qi, Luping Zhou, Lei Wang, and Yinghuan Shi. Improving barely supervised learning by discriminating unlabeled samples with super-class. *Advances in Neural Information Processing Systems*, 35:19849–19860, 2022.

[13] Lin Hu, Jiaxin Li, Xingchen Peng, Jianghong Xiao, Bo Zhan, Chen Zu, Xi Wu, Jiliu Zhou, and Yan Wang. Semi-supervised npc segmentation with uncertainty and attention guided consistency. *Knowledge-Based Systems*, 239:108021, 2022.

[14] Wei-Chih Hung, Yi-Hsuan Tsai, Yan-Ting Liou, Yen-Yu Lin, and Ming-Hsuan Yang. Adversarial learning for semi-supervised semantic segmentation. *arXiv preprint arXiv:1802.07934*, 2018.

[15] Donghyeon Kwon and Suha Kwak. Semi-supervised semantic segmentation with error localization network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9957–9967, 2022.

[16] Samuli Laine and Timo Aila. Temporal ensembling for semi-supervised learning. *arXiv preprint arXiv:1610.02242*, 2016.

[17] Dong-Hyun Lee et al. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on challenges in representation learning, ICML*, volume 3, page 896. Atlanta, 2013.

[18] Yuyuan Liu, Yu Tian, Yuanhong Chen, Fengbei Liu, Vasileios Belagiannis, and Gustavo Carneiro. Perturbed and strict mean teachers for semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4258–4267, 2022.

[19] Robert Mendel, Luis Antonio De Souza, David Rauber, Joao Paulo Papa, and Christoph Palm. Semi-supervised segmentation based on error-correcting supervision. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIX 16*, pages 141–157. Springer, 2020.

[20] Takeru Miyato, Shin-ichi Maeda, Masanori Koyama, and Shin Ishii. Virtual adversarial training: a regularization method for supervised and semi-supervised learning. *IEEE transactions on pattern analysis and machine intelligence*, 41(8):1979–1993, 2018.

[21] Avital Oliver, Augustus Odena, Colin A Raffel, Ekin Dogus Cubuk, and Ian Goodfellow. Realistic evaluation of deep semi-supervised learning algorithms. *Advances in neural information processing systems*, 31, 2018.

[22] Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in neural information processing systems*, 33:596–608, 2020.

[23] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems*, 30, 2017.

[24] Qizhe Xie, Zihang Dai, Eduard Hovy, Thang Luong, and Quoc Le. Unsupervised data augmentation for consistency training. *Advances in neural information processing systems*, 33:6256–6268, 2020.

[25] Lihe Yang, Lei Qi, Litong Feng, Wayne Zhang, and Yinghuan Shi. Revisiting weak-to-strong consistency in semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7236–7246, 2023.

[26] Jianlong Yuan, Yifan Liu, Chunhua Shen, Zhibin Wang, and Hao Li. A simple baseline for semi-supervised semantic segmentation with strong data augmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8229–8238, 2021.

[27] Man Zhang, Yong Zhou, Jiaqi Zhao, Yiyun Man, Bing Liu, and Rui Yao. A survey of semi-and weakly supervised semantic segmentation of images. *Artificial Intelligence Review*, 53:4259–4288, 2020.

[28] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017.

[29] Zhen Zhao, Lihe Yang, Sifan Long, Jimin Pi, Luping Zhou, and Jingdong Wang. Augmentation matters: A simple-yet-effective approach to semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11350–11359, 2023.

[30] Zhen Zhao, Luping Zhou, Yue Duan, Lei Wang, Lei Qi, and Yinghuan Shi. Dc-ssl: Addressing mismatched class distribution in semi-supervised learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9757–9765, 2022.

[31] Konrad Zuchniak. Multi-teacher knowledge distillation as an effective method for compressing ensembles of neural networks, 2023.