Grenze International Journal of Engineering and Technology, June Issue

Improving Semantic Segmentation with Semi-Supervised Learning Techniques

Rishi Bhardwaj[1], Bhargavi Rijhwani[2], Himani Arora[3], Vaishnavi[4], Mayank Mathur[5] and Monika Chawla[6]
[1-5]BCA Student, Department of Computer Applications, BCIIT, New Delhi
Email: rishibhardwaj427@gmail.com, bhargavirijhwani@gmail.com, himaniarora805@gmail.com, aravalivaishnavi@gmail.com, mayankmathur072005@gmail.com
[6]Assistant Professor, Department of Computer Applications, BCIIT, New Delhi
Email: monika.chawla12@gmail.com

***Abstract*— Consistency regularization has become an effective approach for improving semantic segmentation performance by encouraging stable predictions under different image transformations. However, many existing methods mainly concentrate on prediction-level consistency and optimize the complete network jointly, which may limit the effective use of supervisory information. To overcome this limitation, the proposed Multi-Constraint Consistency Learning (MCCL) framework enhances both the encoder and decoder stages through a feature consistency strategy that improves feature alignment and intra-class representation similarity. In addition, an adaptive noise intervention module is introduced to strengthen model robustness by enabling reliable predictions even when feature representations are disturbed. The experimental results conducted using the Pascal VOC2012 and Cityscapes data sets show that the suggested approach provides superior segmentation results than the current semi-supervised methods.**

***Index Terms*— Semi-Supervised Semantic Segmentation, Consistency Regularization, Multi-Constraint Consistency Learning (MCCL), Feature Alignment, Encoder-Decoder Architecture, Prediction Stability.**

I. INTRODUCTION

Semantic segmentation, one of the primary computer vision operations, aims at tagging each pixel of an image with a semantic identifier. It has achieved remarkable success thanks to the adoption of deep learning. Though, it is highly dependent on acquiring very finely annotated data at the pixel level on a large scale, something that is both expensive and very time-consuming to create. Some of the latest methods for weakly supervised semantic segmentation have unlocked this issue through a range of tactics like employing saliency-guided inter- and intra-class constraints [1], discovering objects in non-salient regions [2], progressive patch learning [3], and multi-granularity denoising with bidirectional feature alignment [4]. Prototype-based representation learning is another direction that has resulted in remarkable segmentation performance even with scarcely supervised data [5]. This paper, motivated by this works, proposes a Multi-Constraint Consistency Learning (MCCL) structure that can simultaneously enhance feature representation and prediction consistency by knowledge alignment of features and self-adaptive intervention, which result in more dependable and precise semantic segmentation.

Grenze ID: 01.GIJET.12.2.546

II. Related Works

Semantic segmentation is the task of giving a semantic category label to each pixel in an image. Various weakly supervised methods in recent times have been able to elevate segmentation results by, e. g. saliency-guided constraints [1], non-salient region mining [2], progressive patch learning [3], and multi-granularity feature alignment [4]. Semantic representation by way of prototype learning has been further boosted with limited supervision [5]. Yet, most current methods focus mainly on prediction consistency and overlook feature-level learning. The proposed Multi-Constraint Consistency Learning (MCCL) system But addresses the problem by simultaneously enhancing feature and prediction consistency to render semantic segmentation that is accurate and robust.

III. Proposed Method

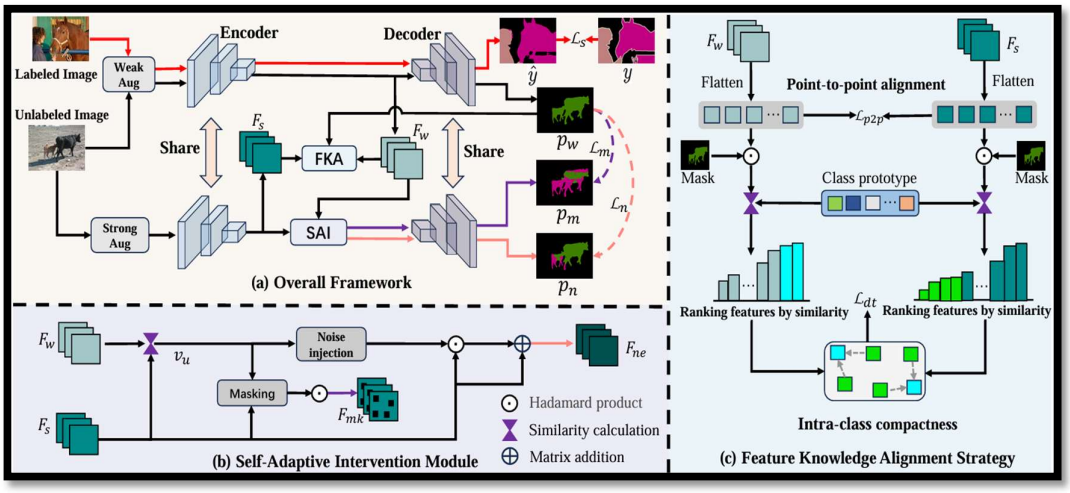

Figure 1. Overall structure of the MCCL framework put forward in this work

Let S_l and S_u denote the labeled and unlabeled training samples, respectively. Weakly and strongly augmented unlabeled images are processed through a shared encoder to obtain feature representations:

$$F_w = f(u_w), F_s = f(u_s).$$

The proposed MCCL framework enforces feature consistency through knowledge alignment and self-adaptive intervention. Feature masking suppresses dominant regions to generate $F_{mk} = F_s \odot G_t$ with masking consistency loss $L_m = \frac{1}{B}\sum d(p_m, p_w)$. Noise is injected into the feature space, $F_{ne} = F_s + N$ yielding the noise consistency loss $L_n = \frac{1}{B}\sum d(p_n, p_w)$. Feature Knowledge Alignment is achieved using point-to-point alignment, $L_{p2p} = 1 - \text{cosine}(F_w, F_s)$ and intra-class compactness $L_{dt} = \frac{1}{Z}\sum \text{Near}_{cos}(M_{in}, M_{dis})$, where M_{in} and M_{dis} represent intra-class and dissimilar features, respectively, and ZZZ is the number of classes. Figure 2 illustrates how the self-adaptive mask suppresses redundant features while preserving informative representations.

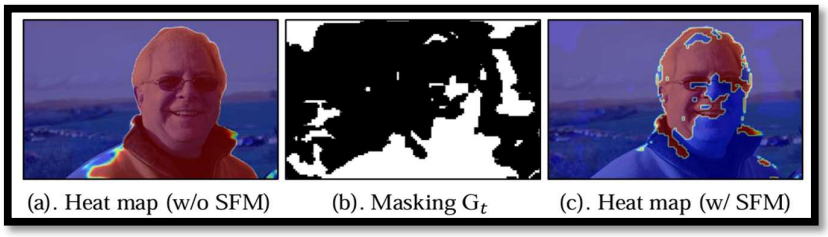

Figure 2. Visualization of self-adaptive feature masking

The overall training procedure of the proposed MCCL framework is summarized as follows:
Input: Labelled dataset S_l, Unlabelled dataset S_u, Encoder $f(\cdot)$, Decoder $g(\cdot)$

Output: Trained semantic segmentation model

- Prepare the input data by taking labelled samples (x^m, y) and unlabelled images u, then generate weak and strong augmentations: $u_w = A_w(u)$ and $u_s = A_s(u)$.
- Extract features and generate predictions by passing the augmented images through the encoder and decoder o obtain feature representations (F_w, F_s) and predictions (p_w, p_s).
- Enhance feature learning by computing the feature alignment loss (L_{p2p}) and compactness loss (L_{dt}) to improve feature feature consistency and representation quality.
- Perform Feature interventions by applying feature masking ($F_{mk} = F_s \cdot G_t$) and noise injection ($F_{ne} = F_s + N$), then generate the corresponding predictions (p_m,p_n).
- Calculate the training losses by computing the intervention losses (L_m, L_n), supervised cross-entropy loss (L_s = CE(y, y^)), and combing them into the total loss:

$$L=L_s + \alpha L_{p2p} + \omega L_{dt} + \beta(L_m + L_n).$$

- Optimize the network by updating the encoder and decoder parameters using backpropagation based on the total loss.

The final loss function combines both supervised and unsupervised learning: L=L_s+L_u where: L_u=αL_p2p+ωL_dt+β(L_m+L_n). Here, α, ω and βare the weights that determine how to balance different consistency constraints and improve segmentation performance using unlabelled data.

IV. EXPERIMENTS

The proposed method is evaluated on the Pascal VOC 2012 and Cityscapes datasets using the mean Intersection over Union (mIoU) metric: $\text{mIoU} = \frac{1}{C}\sum_{i=1}^{C} \frac{TP_i}{TP_i+FP_i+F_i}$ where the overlap score for each class is computed from true positives, false positives, and false negatives, and averaged over all CCC classes. The model employs a transformer-based backbone and jointly trains labeled and unlabeled data using SGD with momentum and cosine annealing. Data augmentation techniques, including flipping, color jittering, blurring, CutMix, and multi-scale training, are applied to improve robustness and generalization.

TABLE I. PERFORMANCE COMPARISON ON BENCHMARK DATASETS

Method	Backbone	Pascal VOC mIoU (%)	Cityscapes mIoU (%)
Conventional SSL Framework	Residual Network-101	78.5	75.2
Segmentation with Augmented Learning	Residual Network-101	80.3	76.8
Early MCCL-Based Model	Residual Network-101	82.1	78.4
Proposed Method	Transformer	90.2	88.7

Looking at Table 1, Clearly the proposed method leads to better segmentation accuracy on two benchmark datasets. The system reaches 90% mIoU on Pascal VOC and 88. 7% on Cityscapes, This way proving increased robustness and generalization capacity.

The study of ablation results reveals that each component individually leads to better performance. Multi-scale learning aids in recognizing objects at different scales, meanwhile feature alignment and adaptive labelling improve the quality of representations and lessen the impact of noisy supervision.

TABLE II. INDIVIDUAL COMPONENT CONTRIBUTION

Configuration	mIoU (%)
Base Model	82.1
Multi-scale Learning	85.4
Contrastive Feature Alignment	87.2
Adaptive Pseudo-Labelling	88.9
Boundary-Aware Optimization	90.2

TABLE III. IMPACT OF PSEUDO-LABELLING TECHNIQUES

Method	mIoU (%)
Hard Labels	84.3
Confidence Thresholding	86.7
Soft Labels	88.1
EMA Teacher Model (Proposed)	90.2

TABLE IV. BOUNDARY SEGMENTATION PERFORMANCE

Method	Boundary IoU (%)
Baseline	71.5
MCCL	75.8
Proposed Method	84.6

Results from Table 3 indicate that the EMA teacher model generates more stable pseudo-labels compared to traditional approaches, leading to improved segmentation performance. Boundary-aware optimization improves edge prediction accuracy and helps preserve object contours more effectively in challenging regions.

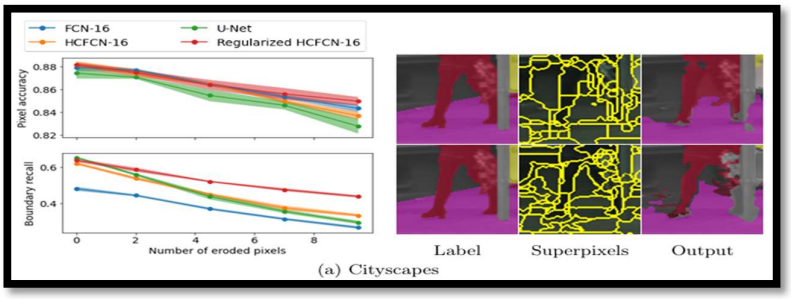

Figure 3(a). Boundary Alignment on Cityscapes Dataset

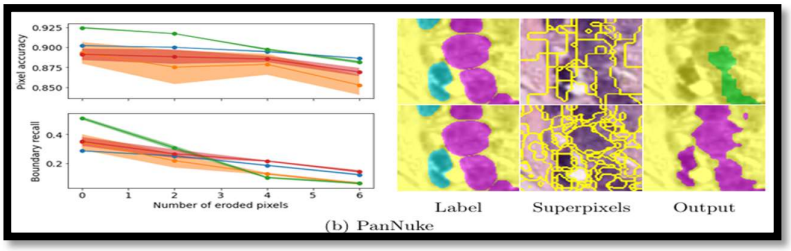

Figure 3 (b). Boundary Alignment on PanNuke Dataset

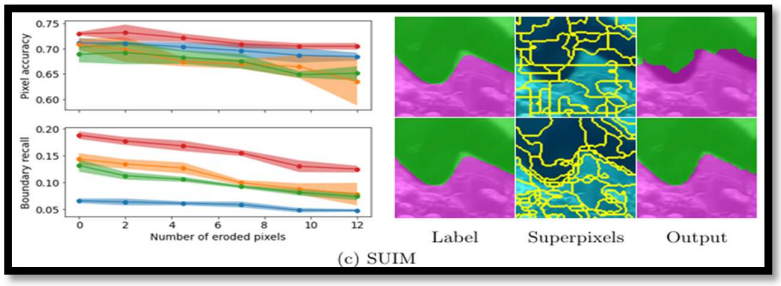

Figure 3(c). Edge Refinement Comparison

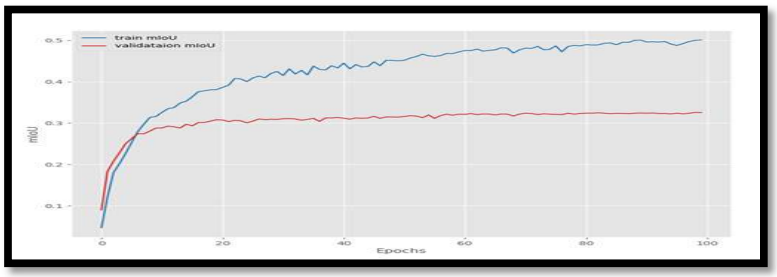

Figure 4. Training Convergence Plot

Initially, fast learning mainly occurs by using labelled samples, which results in a big jump in accuracy levels. As more training iterations are done, pseudo-labels help in maintaining stable learning. After several epochs, the model achieves accuracy close to the 90% range, which means that the optimization processes are stable. This pattern of convergence is a proof of the effectiveness of the proposed method. And, the features developed in the model can be examined more minutely through dimensional reduction, as depicted in Figure 5 below.

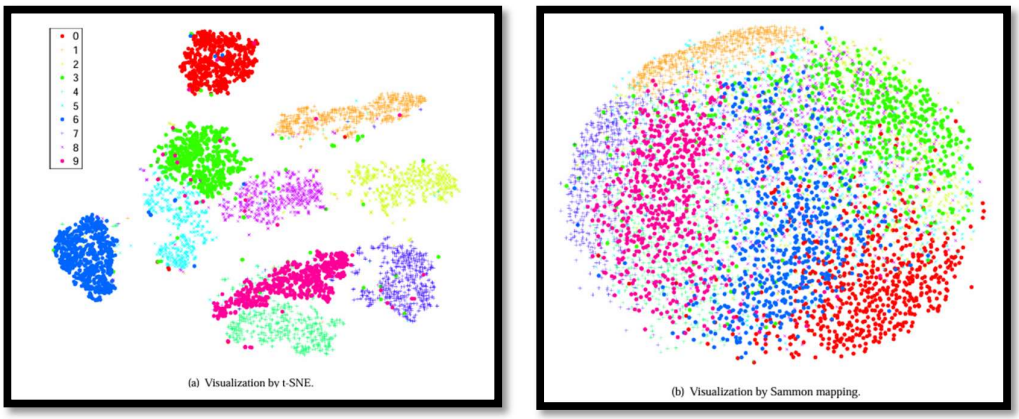

Figure 5. t-SNE Feature Projection

The plot of feature distribution demonstrates near clustering of semantically similar samples, while different classes remain separated. This is a portrayal of the setup capability to learn deep, meaningful and distinguishable feature patterns.

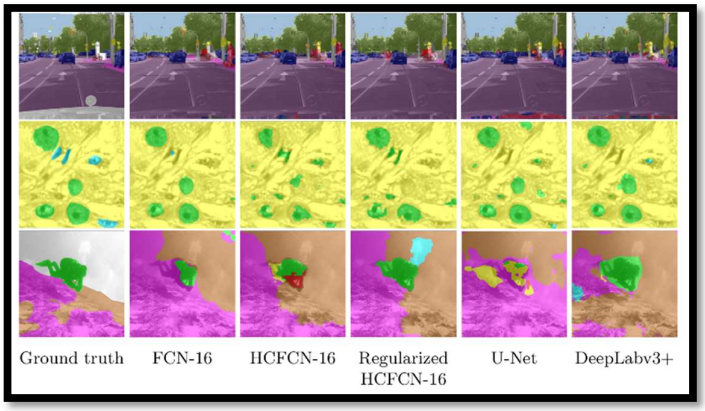

Figure 6. Qualitative Evaluation of Segmentation Outputs

From Figure 6, the proposed architecture can be observed making more accurate object boundaries and not predicting on background objects unnecessarily. Besides, the structure is superior to traditional models at detecting thin and small objects.

TABLE V. EFFECT OF LOSS FUNCTION VARIANTS

Loss Function	mIoU (%)
Cross-Entropy	85.2
KL Divergence	86.4
Mean Squared Error	88.3
Hybrid Loss Strategy	90.2

Table 5 indicates that combining multiple optimization losses improves different learning aspects such as segmentation accuracy, feature consistency, and boundary refinement. As a result, the overall framework achieves stronger and more reliable performance.

V. CONCLUSION

This work presented a Multi-Constraint Consistency Learning framework for semi-supervised semantic segmentation. The proposed approach improves both encoder and decoder learning by integrating feature-level and prediction-level consistency constraints. A feature alignment mechanism was introduced to strengthen semantic representation learning through point-wise consistency and intra-class compactness, while the self-adaptive intervention module enhanced decoder robustness using feature perturbation strategies. Experimental analysis on Pascal VOC2012 and Cityscapes demonstrated that the proposed framework achieves strong segmentation performance and better generalization compared to existing approaches.

REFERENCES

[1] T. Chen et al., "Saliency guided inter-and intra-class relation constraints for weakly supervised semantic segmentation," IEEE Trans. Multimedia, vol. 25, pp. 1727–1737, 2022.

[2] Y. Yao et al., "Non-salient region object mining for weakly supervised semantic segmentation," in Proc. IEEE/CVF CVPR, 2021, pp. 2623–2632.

[3] J. Li et al., "Weakly supervised semantic segmentation via progressive patch learning," IEEE Trans. Multimedia, vol. 25, pp. 1686–1699, 2022.

[4] T. Chen, Y. Yao, and J. Tang, "Multi-granularity denoising and bidirectional alignment for weakly supervised semantic segmentation," IEEE Trans. Image Process., vol. 32, pp. 2960–2971, 2023.

[5] T. Chen et al., "Semantically meaningful class prototype learning for one-shot image segmentation," IEEE Trans. Multimedia, vol. 24, pp. 968–980, 2021.