Learning to Segment Burns with Limited Labels: A Dual-Path Cascaded Network with Uncertainty-Guided Teacher Diffusion Exposure

Joohi Chauhan[1 [0000−0001−7331−851X]] ⋆ and Prabal Pratap Singh[1 [0000−0002−0738−7629]] ⋆

Motilal Nehru National Institute of Technology Allahabad, Prayagraj, Uttar Pradesh, India
joohi@mnnit.ac.in, prabal@mnnit.ac.in

Abstract. Automated burn segmentation from color images remains challenging due to tissue heterogeneity, ambiguous burn boundaries, and limited pixel-level annotations. We propose UADA-MSPCNet, a semi-supervised segmentation framework that improves annotation efficiency and boundary delineation. The method integrates a dual-path cascaded architecture that combines high-resolution spatial encoding with hierarchical atrous contextual aggregation for joint structural and semantic modeling. To leverage unlabeled data, we employ an uncertainty-aware teacher-student module in which predictive uncertainty modulates consistency supervision and cross-scale pseudo-label learning. In addition, diffusion-generated burn images are introduced exclusively through the teacher pathway, enriching feature exposure while avoiding synthetic bias in student optimization. This teacher-side generative exposure improves pseudo-label calibration under limited supervision. Experiments on the EBIS dataset demonstrate consistent improvements over state-of-the-art supervised and semi-supervised methods across multiple labeling settings, achieving superior Dice scores and boundary accuracy. These results highlight the effectiveness of integrating cascaded contextual modeling, uncertainty-aware consistency, and controlled generative augmentation for reliable burn segmentation.

Keywords: Burns Segmentation · SSL · Uncertainty-aware· Diffusion Augmentation · Cascade Network

1 Introduction and Background

Accurate delineation of burns from color images is essential for estimating total burn surface area, guiding decision making, and monitoring therapeutic progression [3,2]. Despite recent advances in medical image segmentation, automated burn analysis remains particularly challenging due to pronounced intra-class variability, irregular morphology, diffuse injury margins, and illumination inconsistencies [4]. These factors complicate boundary localization and increase

⋆ The authors contributed equally to this work and share corresponding authorship

ambiguity in pixel-wise classification. Moreover, expert annotation of burn regions is labor-intensive and clinically demanding, resulting in limited availability of high-quality labeled datasets [3].

Deep convolutional segmentation models have substantially improved dense prediction through multi-scale representation learning. Encoder-decoder frameworks such as U-Net and its derivatives preserve structural detail [16,11], while atrous convolution and pyramid pooling approaches expand receptive fields to capture global semantic context [22,6,7,5]. However, burns exhibit both fine-grained boundary discontinuities and heterogeneous tissue patterns, requiring simultaneous modeling of high-resolution structure and large-scale morphology. Architectures that emphasize contextual aggregation often sacrifice boundary precision, whereas spatially focused models lack sufficient semantic coverage. Effectively balancing these complementary requirements remains an open challenge in burn segmentation. To mitigate annotation scarcity, semi-supervised learning (SSL) has emerged as a practical paradigm for leveraging unlabeled clinical data [20,24,21,14]. Teacher-student consistency frameworks, including Mean Teacher and uncertainty-aware self-ensembling, improve label efficiency by propagating pseudo-supervision under perturbation invariance [24]. Subsequent extensions incorporate cross-branch consistency and confidence-based filtering to suppress noisy pseudo-labels. While these approaches enhance supervision utilization, pseudo-label reliability remains sensitive to ambiguous burn boundaries and distributional variability, limiting performance under low-label regimes.

Generative modeling has emerged as a complementary approach for enriching training data, with models such as GANs, VAEs, and diffusion models [10,15,9] demonstrating the ability to synthesize anatomically plausible medical images. Incorporating these samples into semi-supervised learning can enhance morphological diversity; however, directly using synthetic data in supervised optimization may introduce artifacts and distributional bias. Leveraging generative variability while preserving pseudo-label reliability, therefore, remains an open challenge.

To address these challenges, we develop a unified semi-supervised segmentation framework called UADA-MSPCNet. This Uncertainty-Aware Diffusion Augmented Multi-scale Multi-Path Cascade Network jointly advances representation learning and supervision reliability when annotations are limited. Our design is motivated by three observations: (1) accurate burn delineation requires simultaneous modeling of fine boundary structure and heterogeneous tissue context, (2) pseudo-label supervision must be calibrated to mitigate uncertainty in ambiguous regions, and (3) generative variability can enhance representation diversity if introduced without inducing optimization bias. Synthetic samples are therefore confined to the teacher pathway to refine pseudo-label formation while preserving stable student learning. This coordinated design enables improved utilization of unlabeled data while preserving boundary fidelity and supervision robustness. Comprehensive evaluation on the Extended Burn Image Segmentation (EBIS) dataset [5] under multiple annotation regimes demonstrates that the proposed framework consistently improves segmentation accuracy and boundary

agreement, particularly in low-label settings. These findings underscore the importance of coupling architectural representation capacity with calibrated semi-supervised learning for clinically reliable burn segmentation.

2 Methodology

Let the labeled dataset be defined as $\mathcal{D}_L = \{(x_i, y_i)\}_{i=1}^{N_L}$, where $x_i \in \mathbb{R}^{H\times W\times 3}$ denotes RGB burn images and y_i denotes pixel-wise burn annotations. We further consider unlabeled samples $\mathcal{D}_U = \{x_j\}_{j=1}^{N_U}$ and diffusion-generated synthetic pairs $\mathcal{D}_G = \{(x_k^{syn}, y_k^{syn})\}_{k=1}^{N_G}$. The objective is to learn a segmentation function $f_\theta : x \to \hat{y}$ that predicts dense burn probability maps. Figure 1 highlights the overview of the proposed framework.

2.1 Dual-Path Multi-Scale Feature Encoding

To jointly capture fine boundary structures and large-scale contextual dependencies, we employ a dual-path encoder consisting of a high-resolution spatial stream and a dense cascade atrous contextual stream (MSPCNet).

(i)Spatial pathway. The spatial stream preserves boundary fidelity through three 3×3 convolutional blocks (channels 64→96→128) with stride-2 downsampling in the first two blocks, producing features $F_s \in \mathbb{R}^{H/4\times W/4\times 128}$. This pathway explicitly maintains high-frequency structural cues critical for irregular burn margins.

(ii)Contextual pathway. Semantic features are extracted using a ResNet-101 backbone, yielding $F_b \in \mathbb{R}^{H/16\times W/16\times 2048}$. A 1×1 projection produces $F_0 \in \mathbb{R}^{H/16\times W/16\times 512}$. Atrous convolution with dilation r_k is denoted as $\mathcal{A}_{r_k}(F) = Conv_{3\times 3}^{d=r_k}(F)$. Unlike parallel ASPP or linear DenseASPP expansions, we employ a regulated cascaded formulation: $F^{(k)} = \mathcal{A}_{r_k}\big(\text{Concat}(F_0, F_g, F^{(<k)})\big)$, where F_g denotes image-level pooled features and $F^{(<k)}$ aggregates previous cascade outputs, also with the dilation set $\{r_k\}_{k=1}^{4} = \{3, 6, 12, 18\}$, where k indexes the cascade stage. Rather than relying on linear receptive-field accumulation, the cascade progressively densifies contextual aggregation by allowing later atrous layers to operate on semantically enriched intermediate representations. This hierarchical staging mitigates sampling sparsity associated with large dilation factors while preserving multi-scale contextual continuity, which is an important property for heterogeneous burn morphology. All contextual outputs are fused: $F_c = Conv_{1\times 1}\big(Concat(F_g, F^{(1)}, F^{(2)}, F^{(3)}, F^{(4)})\big)$, where $F_c \in \mathbb{R}^{H/16\times W/16\times 512}$.

(iii)Cross-path fusion. Spatial and contextual features are concatenated as $F_{fusion} = \text{Concat}(F_s, F_c^{\uparrow}),$, followed by convolutional refinement and upsampling to produce the final burn probability map $\hat{y} \in \mathbb{R}^{H\times W}$. While conceptually related to bilateral networks [23], the proposed design prioritizes dense medical boundary delineation over real-time efficiency, coupling high-resolution structural encoding with deeply cascaded semantic reasoning.

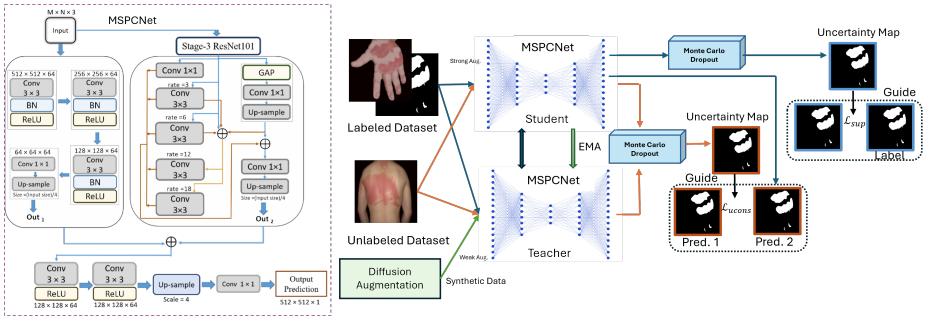

Fig. 1. Overview of the proposed UADA-MSPCNet framework. The dual-path MSPCNet backbone is embedded within an uncertainty-aware teacher–student paradigm, where diffusion-generated samples are introduced exclusively through the teacher pathway for pseudo-label calibration.

2.2 Semi-supervised uncertainty-aware consistency learning.

The dual-path MSPCNet is first established as the core segmentation architecture and subsequently instantiated as both student (f_θ) and teacher ($f_{\theta'}$) networks within a teacher–student semi-supervised framework. The teacher parameters are exclusively updated via an exponential moving average: $\theta' \leftarrow \alpha\theta' + (1-\alpha)\theta, \quad \alpha = 0.99$. However, no gradient updates are applied directly to teacher parameters.

Each training batch contains labeled samples $\mathcal{D}_L$ and real unlabeled samples $\mathcal{D}_U$ in a 1:3 ratio. Weak augmentations (flip, blur, color jitter) are applied to teacher inputs, while strong augmentations (RandAugment, CutOut, elastic deformation) are applied to student inputs to mitigate confirmation bias. Consistency loss is enforced only on real unlabeled samples:

$$L_{cons} = \|f_\theta(T_s(x_u)) - f_{\theta'}(T_w(x_u))\|_2^2. \tag{1}$$

where T_s and T_w denote strong and weak augmentations, respectively.

Predictive uncertainty is estimated via Monte Carlo dropout (rate 0.3, $T = 8$ passes), yielding variance u. Empirically, $T = 8$ provides stable variance estimation with negligible calibration gains beyond $T \geq 12$. The uncertainty-weighted consistency loss becomes: $L_{ucons} = (1-u) \cdot L_{cons}$.

(i) Cross-scale pseudo supervision. To enforce structural agreement between spatial and contextual pathways, cross-scale pseudo supervision (CSPS) is applied to unlabeled samples. Let P_s and P_c denote spatial and contextual predictions. Pseudo-labels are obtained using temperature-scaled logits: $\tilde{P} = \text{Softmax}\left(\frac{z}{\tau_T}\right), \quad \tau_T = 0.5$, which sharpens high-confidence predictions and improves class separability. Supervision is restricted to confident regions ($\delta = 0.7$) and defined as:

$$\mathcal{L}_{csps} = \mathbb{1}(\max(P_c) > \delta)\,\text{CE}_w(P_s, \tilde{y}_c) + \mathbb{1}(\max(P_s) > \delta)\,\text{CE}_w(P_c, \tilde{y}_s), \tag{2}$$

where $\tilde{y}$ denotes the temperature-refined pseudo-labels and CE_w denotes Dice-weighted cross-entropy. Confidence gating and uncertainty weighting suppress unreliable regions, mitigating confirmation bias.

2.3 Diffusion-driven generative augmentation.

To enrich contextual diversity while preventing synthetic artifact propagation to the student network, diffusion-based generative augmentation is incorporated exclusively within the teacher pathway. Synthetic burn images are generated using a mask-conditioned diffusion model [9] trained on the training split only, with conditioning signals derived from burn mask geometry and severity priors. Strict train/test partition is maintained to prevent data leakage, and the diffusion model is trained exclusively on the EBIS training partition without pretraining on external datasets. Given conditioning variable c and noise vector z, synthetic samples are generated as: $x^{syn} \sim G_\phi(z, c)$.

During training, synthetic samples are forwarded through the teacher in a forward-pass-only manner and the teacher parameters θ' are updated solely via a non-differentiable exponential moving average of student parameters θ : $\theta' \leftarrow \alpha\theta' + (1-\alpha)\theta$. The teacher network is maintained in training mode with gradients disabled, allowing BatchNorm running statistics to update during synthetic exposure while all weight parameters remain governed solely by EMA. This design enables feature normalization statistics to adapt to generative morphological variability without introducing gradient-based optimization on synthetic data. Synthetic exposure is maintained at a fixed 1:1 ratio with real unlabeled samples per teacher batch. This exposure broadens contextual feature representations learned by the teacher, leading to improved pseudo-label calibration on real unlabeled data, which subsequently supervises student learning via consistency optimization.

Training objective. The overall optimization integrates supervised learning and semi-supervised consistency with teacher-side generative regularization: $L_{total} = L_{sup} + \lambda_1 L_{ucons} + \lambda_2 L_{csps}$. Weights follow ramp-up scheduling to stabilize early training. During inference, segmentation is obtained through a single forward pass of the student network: $\hat{y} = f_\theta(x)$, with optional uncertainty estimation derived from teacher–student variance.

3 Experimental Results and Discussion

We conduct all the experiments on the EBIS dataset [5], comprising burn images of 512×512 resolution with substantial variability in tissue appearance, morphology, and illumination. The dataset includes 392 training and 208 test images. To simulate realistic annotation constraints, semi-supervised evaluation was performed under 5%, 10%, and 20% labeled settings, with the remaining samples treated as unlabeled. Results are averaged across five random splits. The framework is implemented in PyTorch using AdamW optimization for 100 epochs with polynomial learning-rate decay and a batch size of 8 on an NVIDIA

RTX 2080 Ti GPU. Inference relies solely on the student network, ensuring deployment efficiency. As diffusion exposure is restricted to the teacher pathway during training, it does not introduce additional inference overhead during deployment. Also, we reimplemented and evaluated all comparative methods under identical training settings for fairness.

Table 1. Quantitative comparison with supervised and foundation-model baselines on the EBIS test set [5].

Method	Dice ↑	IoU ↑	MCC ↑	HD95 ↓
DenseASPP [22]	67.3 ± 0.41	52.9 ± 0.44	62.5 ± 0.47	18.6 ± 0.36
AdapNet [18]	68.9 ± 0.35	54.2 ± 0.45	63.2 ± 0.33	16.9 ± 0.37
SegNet [1]	69.1 ± 0.48	54.5 ± 0.56	63.2 ± 0.45	16.8 ± 0.52
BiSeNet [23]	69.9 ± 0.33	54.7 ± 0.41	64.4 ± 0.32	17.3 ± 0.25
PSPNet [25]	70.6 ± 0.36	55.8 ± 0.42	65.3 ± 0.38	16.7 ± 0.31
DeepLabV3 [6]	74.7 ± 0.33	63.8 ± 0.37	69.7 ± 0.42	15.2 ± 0.34
UNet [16]	74.9 ± 0.42	64.0 ± 0.31	70.4 ± 0.38	15.4 ± 0.31
DeepLabV3+ [7]	75.8 ± 0.28	65.5 ± 0.35	72.5 ± 0.23	14.3 ± 0.33
RefineNet [12]	76.2 ± 0.32	65.6 ± 0.29	72.9 ± 0.35	13.7 ± 0.26
MSAC-SNet [4]	78.5 ± 0.29	68.3 ± 0.25	76.6 ± 0.35	13.1 ± 0.33
nnUNet [11]	81.3 ± 0.29	75.6 ± 0.23	78.2 ± 0.31	12.2 ± 0.35
BurnsNet [5]	84.7 ± 0.31	73.9 ± 0.36	80.4 ± 0.32	11.6 ± 0.26
HRNetV2 [19]	86.3 ± 0.32	78.8 ± 0.29	82.1 ± 0.35	12.2 ± 0.27
MSPCNet	$\mathbf{88.7 \pm 0.27}$	$\mathbf{82.5 \pm 0.30}$	$\mathbf{85.9 \pm 0.32}$	$\mathbf{9.8 \pm 0.28}$

Under full supervision, Table 1, MSPCNet achieves the best performance across all metrics, attaining Dice 88.7%, IoU 82.5%, MCC 85.9%, and HD95 9.8. Improvements are consistent relative to both task-specific architectures and general-purpose segmentation frameworks, including BurnsNet [5], nnUNet [11], and HRNetV2 [19]. Notably, the reduction in HD95 indicates improved boundary localization rather than solely increased regional overlap, highlighting more precise delineation of irregular burn margins. The simultaneous gains in overlap- and distance-based metrics suggest effective integration of high-resolution spatial encoding with multi-scale contextual aggregation.

Table 2 summarises the learning behavior under annotation scarcity. When trained with labeled data alone, MSPCNet achieves only 41.1% Dice at 5% supervision, reflecting the difficulty of burn segmentation with limited annotations. Conventional semi-supervised methods improve performance to approximately 76.6% Dice under the same setting. Incorporating uncertainty-aware consistency learning in MSPCNet improves the performance to 78.0% Dice while reducing HD95, indicating improved pseudo-label reliability in ambiguous regions. Introducing teacher-side diffusion exposure yields additional gains, reaching 80.5% Dice and reducing HD95 to 9.11 at 5% labeling. Similar improvements persist at higher supervision levels, achieving 85.1% Dice at 10% and 88.9% Dice at 20%, approaching full-supervision performance. These results suggest that con-

Table 2. Semi-supervised segmentation performance and comparison with state-of-the-art SSL methods under varying labeled data ratios. (UA for Uncertainty-aware and UADA for Uncertainty-aware Diffusion augmentation)

Method	% Labeled (L/U)	Dice ↑	IoU ↑	MCC ↑	HD95 ↓
MSPCNet	5% (20/0)	41.1 ± 2.61	36.0 ± 2.54	37.8 ± 2.17	34.7 ± 1.75
MSPCNet	10% (39/0)	53.7 ± 2.99	47.5 ± 2.36	49.0 ± 2.67	27.44 ± 2.33
MSPCNet	20% (78/0)	59.0 ± 2.03	52.7 ± 1.78	55.2 ± 1.84	26.38 ± 1.83
MSPCNet	100% (392/0)	$\mathbf{88.7 \pm 0.79}$	$\mathbf{82.5 \pm 0.88}$	$\mathbf{85.9 \pm 0.85}$	$\mathbf{9.80 \pm 0.68}$
UA-MT [24]	5% (20/372)	73.2 ± 1.42	66.3 ± 1.38	68.4 ± 1.30	19.18 ± 1.11
DTC [13]	5% (20/372)	73.9 ± 1.21	65.4 ± 1.08	68.3 ± 1.16	18.39 ± 1.05
CPS [8]	5% (20/372)	73.6 ± 0.98	66.5 ± 1.04	67.8 ± 0.94	18.01 ± 0.93
UniMatch [21]	5% (20/372)	74.6 ± 0.87	67.8 ± 1.29	70.1 ± 1.05	17.32 ± 0.84
IPixMatch [20]	5% (20/372)	74.9 ± 0.83	67.9 ± 0.89	70.2 ± 0.79	15.54 ± 0.77
MT [17]	5% (20/372)	76.6 ± 0.76	69.2 ± 0.89	71.5 ± 0.69	13.85 ± 0.81
Ours-UA only	5% (20/372)	78.0 ± 0.77	70.3 ± 0.83	73.4 ± 0.79	10.82 ± 0.72
Ours-UADA	5% (20/372)	$\mathbf{80.5 \pm 0.80}$	$\mathbf{72.6 \pm 0.73}$	$\mathbf{77.9 \pm 0.761}$	$\mathbf{9.11 \pm 0.60}$
UA-MT [24]	10% (39/353)	75.1 ± 0.68	69.4 ± 0.65	70.3 ± 0.70	17.11 ± 0.66
DTC [13]	10% (39/353)	75.3 ± 0.69	70.1 ± 0.75	70.6 ± 0.68	16.02 ± 0.71
CPS [8]	10% (39/353)	76.0 ± 0.67	70.4 ± 0.69	70.8 ± 0.62	15.56 ± 0.61
UniMatch [21]	10% (39/353)	76.9 ± 0.58	70.5 ± 0.64	72.0 ± 0.69	15.31 ± 0.58
IPixMatch [20]	10% (39/353)	77.4 ± 0.65	72.5 ± 0.68	74.7 ± 0.58	13.48 ± 0.69
MT [17]	10% (39/353)	78.9 ± 0.66	73.9 ± 0.71	74.7 ± 0.64	12.33 ± 0.57
Ours-UA only	10% (39/353)	82.6 ± 0.68	76.3 ± 0.50	76.8 ± 0.67	9.88 ± 0.58
Ours-UADA	10% (39/353)	$\mathbf{85.1 \pm 0.59}$	$\mathbf{78.7 \pm 0.63}$	$\mathbf{81.9 \pm 0.58}$	$\mathbf{8.30 \pm 0.66}$
UA-MT [24]	20% (72/314)	77.6 ± 0.68	73.1 ± 0.73	75.3 ± 0.75	15.36 ± 0.64
DTC [13]	20% (72/314)	77.8 ± 0.67	73.4 ± 0.72	75.7 ± 0.69	15.02 ± 0.70
CPS [8]	20% (72/314)	78.3 ± 0.67	74.0 ± 0.62	76.0 ± 0.58	14.31 ± 0.68
UniMatch [21]	20% (72/314)	78.6 ± 0.59	75.4 ± 0.60	76.8 ± 0.57	14.33 ± 0.63
IPixMatch [20]	20% (72/314)	81.8 ± 0.62	76.7 ± 0.58	78.4 ± 0.53	12.82 ± 0.66
MT [17]	20% (72/314)	81.3 ± 0.59	77.4 ± 0.63	78.3 ± 0.55	11.10 ± 0.58
Ours-UA only	20% (72/314)	85.8 ± 0.52	79.4 ± 0.57	81.9 ± 0.48	8.69 ± 0.50
Ours-UADA	20% (72/314)	$\mathbf{88.9 \pm 0.43}$	$\mathbf{83.7 \pm 0.52}$	$\mathbf{86.1 \pm 0.48}$	$\mathbf{7.29 \pm 0.46}$

trolled generative exposure enhances pseudo-label calibration rather than acting as direct synthetic supervision.

Ablation results as in Table 3 provide further insight into architectural and learning contributions. The spatial pathway alone achieves 78.0% Dice, while the contextual pathway improves performance to 82.4% Dice (+4.4%), indicating the importance of global morphological modeling. Their fusion yields the largest gain, reaching 88.7% Dice (+6.3% over contextual alone), demonstrating complementary roles in boundary preservation and contextual reasoning. Among dilation configurations, the regulated cascade $(3, 6, 12, 18)$ performs best. Removing smaller dilation factors reduces Dice to 86.2% (–2.5%), suggesting that early-stage receptive fields contribute to structural continuity, whereas extending dilation ranges introduces redundancy without measurable benefit.

Table 3. Ablation study examining the impact of spatial–contextual fusion, dilation configurations, semi-supervised learning, and diffusion-based teacher calibration.

Configuration	% images	Dice ↑	IoU ↑	MCC ↑	ECE ↓
Spatial Pathway	100%	78.0 ± 0.38	71.6 ± 0.43	74.2 ± 0.39	-
Contextual Pathway	100%	82.4 ± 0.35	75.0 ± 0.40	78.1 ± 0.32	-
Fusion (3, 6, 12)	100%	86.8 ± 0.37	80.0 ± 0.35	83.9 ± 0.36	-
Fusion (6, 12, 18)	100%	86.2 ± 0.36	80.4 ± 0.41	84.2 ± 0.37	-
Fusion (6, 12, 18, 24)	100%	86.2 ± 0.41	80.8 ± 0.38	84.4 ± 0.40	-
Fusion (3, 6, 12, 18, 24)	100%	86.1 ± 0.37	80.8 ± 0.36	84.4 ± 0.34	-
Fusion (3, 6, 12, 18)	100%	$\mathbf{88.7 \pm 0.27}$	$\mathbf{82.5 \pm 0.30}$	$\mathbf{85.9 \pm 0.32}$	-
+ SSL(Real images only)	20%	85.8 ± 0.52	79.4 ± 0.57	81.9 ± 0.48	0.047 ± 0.004
+ SSL + Diffusion Both	20%	84.4 ± 0.51	77.3 ± 0.48	79.5 ± 0.56	0.064 ± 0.006
+ SSL + Teacher Diffusion	20%	$\mathbf{88.9 \pm 0.43}$	$\mathbf{83.7 \pm 0.52}$	$\mathbf{86.1 \pm 0.48}$	$\mathbf{0.042 \pm 0.003}$

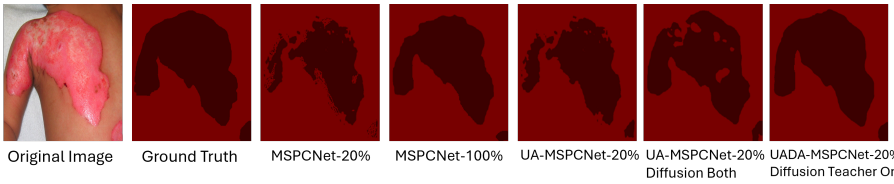

Fig. 2. Qualitative results on a test image under varying supervision levels, highlighting improved boundary fidelity of the proposed method.

When semi-supervised learning (Ours-UA only) is introduced under 20% labeled supervision, performance improves from 59.0% Dice (supervised only) to 85.8% Dice with real-image SSL (+26.8%). Applying diffusion augmentation to both teacher and student degrades performance to 84.4% Dice (–1.4%), indicating distributional bias from direct synthetic supervision. In contrast, restricting diffusion exposure to the teacher (Ours-UADA) increases Dice to 88.9% (+3.1% over SSL-only), supporting the hypothesis that generative data is most effective for representation enrichment and pseudo-label calibration rather than gradient-based optimization. Qualitative results on a single test image are shown in figure 2. Furthermore, teacher-restricted diffusion exposure improves predictive calibration, reducing Expected Calibration Error (ECE) from 0.047 ± 0.004 (SSL only) to 0.042 ± 0.003, whereas applying diffusion to both networks degrades calibration (0.064 ± 0.006), indicating that generative variability is most effective when confined to teacher-side pseudo-label formation.

Overall, results demonstrate that segmentation performance under limited annotations depends jointly on representation capacity and supervision reliability. The proposed framework improves boundary delineation while maintaining annotation efficiency without increasing inference complexity. Although performance approaches full supervision at moderate label ratios, further validation on larger multi-center and demographically diverse datasets is necessary to assess generalization robustness.

4 Conclusion

We presented a semi-supervised framework, UADA-MSPCNet, that improves burn segmentation by jointly optimizing representation learning and pseudo-label calibration. Experimental results demonstrate enhanced boundary delineation and annotation efficiency, with performance approaching that of full supervision with limited labels. Future work will focus on multi-center and demographic validation to further assess clinical generalization. To ensure reproducibility, the code for this work will be made publicly available, and the EBIS dataset is available at https://github.com/VEDAs-Lab/EBIS.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Badrinarayanan, V., Kendall, A., Cipolla, R.: SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **39**(12), 2481–2495 (Dec 2017). https://doi.org/10.1109/TPAMI.2016.2644615
2. Chauhan, J., Goswami, R., Goyal, P.: Using deep learning to classify burnt body parts images for better burns diagnosis. In: Lepore, N., Brieva, J., Romero, E., Racoceanu, D., Joskowicz, L. (eds.) Processing and Analysis of Biomedical Information. pp. 25–32. Springer International Publishing, Cham (2019)
3. Chauhan, J., Goyal, P.: Bpbsam: Body part-specific burn severity assessment model. Burns **46**(6), 1407–1423 (2020). https://doi.org/https://doi.org/10.1016/j.burns.2020.03.007, https://www.sciencedirect.com/science/article/pii/S030541791930782X
4. Chauhan, J., Goyal, P.: Convolution neural network for effective burn region segmentation of color images. Burns **47**(4), 854–862 (Jun 2021). https://doi.org/10.1016/j.burns.2020.08.016
5. Chauhan, J., Rosin, P.L., Goyal, P.: Burnsnet: Burn Region Segmentation Network From Color Images With Two-Way CNN. In: 2024 IEEE International Conference on Image Processing (ICIP). pp. 3172–3178 (Oct 2024). https://doi.org/10.1109/ICIP51287.2024.10647789
6. Chen, L.C., Papandreou, G., Schroff, F., Adam, H.: Rethinking Atrous Convolution for Semantic Image Segmentation (Dec 2017). https://doi.org/10.48550/arXiv.1706.05587
7. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) Computer Vision – ECCV 2018. pp. 833–851. Springer International Publishing, Cham (2018). https://doi.org/10.1007/978-3-030-01234-2_49
8. Chen, X., Yuan, Y., Zeng, G., Wang, J.: Semi-Supervised Semantic Segmentation with Cross Pseudo Supervision. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2613–2622. IEEE, Nashville, TN, USA (Jun 2021). https://doi.org/10.1109/CVPR46437.2021.00264

9. Croitoru, F.A., Hondru, V., Ionescu, R.T., Shah, M.: Diffusion Models in Vision: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(9), 10850–10869 (Sep 2023). https://doi.org/10.1109/TPAMI.2023.3261988
10. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative Adversarial Nets. In: Advances in Neural Information Processing Systems. vol. 27. Curran Associates, Inc. (2014)
11. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (Feb 2021). https://doi.org/10.1038/s41592-020-01008-z
12. Lin, G., Liu, F., Milan, A., Shen, C., Reid, I.: RefineNet: Multi-Path Refinement Networks for Dense Prediction. IEEE Transactions on Pattern Analysis and Machine Intelligence **42**(5), 1228–1242 (May 2020). https://doi.org/10.1109/TPAMI.2019.2893630
13. Luo, X., Chen, J., Song, T., Wang, G.: Semi-supervised Medical Image Segmentation through Dual-task Consistency. Proceedings of the AAAI Conference on Artificial Intelligence **35**(10), 8801–8809 (May 2021). https://doi.org/10.1609/aaai.v35i10.17066
14. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**(1), 654 (Jan 2024). https://doi.org/10.1038/s41467-024-44824-z
15. Rezende, D.J., Mohamed, S., Wierstra, D.: Stochastic Backpropagation and Approximate Inference in Deep Generative Models. In: Proceedings of the 31st International Conference on Machine Learning. pp. 1278–1286. PMLR (Jun 2014)
16. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer International Publishing, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
17. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. pp. 1195–1204. NIPS'17, Curran Associates Inc., Red Hook, NY, USA (Dec 2017)
18. Valada, A., Vertens, J., Dhall, A., Burgard, W.: AdapNet: Adaptive semantic segmentation in adverse environmental conditions. In: 2017 IEEE International Conference on Robotics and Automation (ICRA). pp. 4644–4651 (May 2017). https://doi.org/10.1109/ICRA.2017.7989540
19. Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., Liu, W., Xiao, B.: Deep High-Resolution Representation Learning for Visual Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence **43**(10), 3349–3364 (Oct 2021). https://doi.org/10.1109/TPAMI.2020.2983686
20. Wu, K., Li, W., Xiao, X.: IPixMatch: Boost Semi-supervised Semantic Segmentation with Inter-Pixel Relation (Apr 2024). https://doi.org/10.48550/arXiv.2404.18891
21. Yang, L., Zhao, Z., Zhao, H.: UniMatch V2: Pushing the Limit of Semi-Supervised Semantic Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **47**(4), 3031–3048 (Apr 2025). https://doi.org/10.1109/TPAMI.2025.3528453

22. Yang, M., Yu, K., Zhang, C., Li, Z., Yang, K.: DenseASPP for Semantic Segmentation in Street Scenes. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3684–3692 (Jun 2018). https://doi.org/10.1109/CVPR.2018.00388
23. Yu, C., Wang, J., Peng, C., Gao, C., Yu, G., Sang, N.: BiSeNet: Bilateral Segmentation Network for Real-Time Semantic Segmentation. In: Computer Vision – ECCV 2018: 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part XIII. pp. 334–349. Springer-Verlag, Berlin, Heidelberg (Sep 2018). https://doi.org/10.1007/978-3-030-01261-8_20
24. Yu, L., Wang, S., Li, X., Fu, C.W., Heng, P.A.: Uncertainty-Aware Self-ensembling Model for Semi-supervised 3D Left Atrium Segmentation. In: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.T., Khan, A. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. pp. 605–613. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-32245-8_67
25. Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid Scene Parsing Network. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6230–6239 (Jul 2017). https://doi.org/10.1109/CVPR.2017.660