일반논문 (Regular Paper)
방송공학회논문지 제29권 제1호, 2024년 1월 (JBE Vol.29, No.1, January 2024)
https://doi.org/10.5909/JBE.2024.29.1.10
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

약지도 의미론적 영상 분할을 위한 개별 이미지 별 유사성 강화 프로토타입 기법

임정선[a)], 안수빈[a)], 이수찬[a)‡]

Affinity Enhanced Image-Specific Prototypes for Weakly Supervised Semantic Segmentation

Jungsun Im[a)], Subin An[a)], and Soochahn Lee[a)‡]

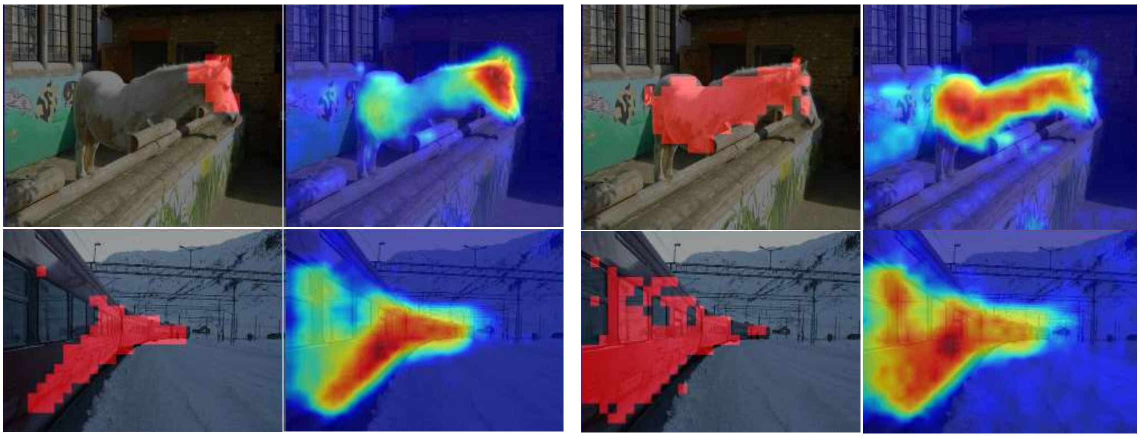
요 약

많은 약지도 학습 의미론적 영상 분할 방법은 이미지 분류를 위해 훈련된 네트워크에서 추출한 픽셀 수준 피처를 활용한다. 클래스 활성화 맵 생성, 픽셀 유사성을 기반으로 유사 피쳐 정의, 특징 군집을 기반으로 클래스 별 특징 프로토타입을 생성하는 데 픽셀 수준 피쳐가 사용 된다. 본 논문은 이전 연구들을 향상시키기 위해 친화도 기반 세분화를 이미지-클래스 별 프로토타입 생성에 통합하는 방법을 제안하며, 이로 인해 클래스 별 특징 프로토타입으로 인한 클래스 활성화 맵 생성 성능이 크게 향상 된다. 이러한 프로토타입은 개선된 의사 레이블을 만들어내며, 궁극적으로 의미론적 영상 분할을 개선한다. 실험 결과는 기준 방법과 비교하여 의미 있는 개선이 있으며, 최신 기법과 유사한 수준의 결과를 가진다. 논문 코드는 https://github.com/IJS1016/AE_SIPE에서 제공한다.

Abstract

Many weakly supervised semantic segmentation methods rely on the pixel-level features extracted from networks trained for image classification. These features can be used to create class activation maps for semantic scores, define pixel affinity as feature similarities, and construct per-class feature prototypes based on feature clustering. This paper proposes a method that enhances previous works by incorporating affinity-based refinement into the generation of image-specific per-class prototypes, resulting in significantly improved representative strength. These prototypes then lead to improved pseudo-labels, ultimately improving segmentations. Experimental results show significant improvements compared to baseline methods, and are on par with recent state-of-the-art methods. The code is available at https://github.com/IJS1016/AE_SIPE.

Keywords : weakly supervised semantic segmentation, affinity enhancement, prototype exploration, self-supervised learning, image-specific



Prototype Seeds IS-CAM

Affinity Enhanced IS-CAM (Proposed)

그림 1. 자기 지도 이미지별 프로토타입 탐색(SIPE) 내에서 이미지별 CAM(IS-CAM)을 생성할 때 픽셀 수준의 유사성을 향상함
Fig. 1. We enhance pixel-level affinity when generating image-specific CAM (IS-CAM) within the self-supervised image-specific prototype exploration (SIPE)[1]

I. INTRODUCTION

The goal of weakly supervised semantic segmentation (WSSS) is to learn how to generate pixel-level labels from limited supervision, usually in the form of image-level class labels[2]. The introduction of the Class Activation Map (CAM)[3] was a significant advancement towards achieving this goal, as it provides a means of generating pixel-level per-class scores based on image classification. However, it has been observed that meaningful CAM scores are often only assigned to a selective number of the most discriminative pixels, leading to limitations in directly using CAM as a segmentation solution.

Nonetheless, CAM proves to be a highly efficient technique for utilizing image-level annotations to make pixel-level predictions. It has frequently served as a base upon which multiple methods have been proposed to enhance and optimize the acquisition of pixel-level class probabilities.

One approach is to erase[5] or suppress[6] the more discriminative regions, further mine discrimenative pixels. Another approach is to assign the limited discriminative regions as seeds and expand them into full segmentation labels using conventional region growing algorithms[7,8], based on the similarities of local pixel values. Further methods extended this approach by incorporating pixel adaptive refinement[4], random walks on semantic features[9], or multitask inference of displacement and class boundary[9,10].

Many recent methods are based on self-supervised learning. A contrastive learning framework, with positive image pairs defined by pairing an image with its linear transform and negative pairs of different images, were applied in[11,12]. Another approach uses network features to create a per-class feature prototype-based alternative score map, providing supervision to guide the network towards generating consistent features with pixel affinities and image-level class labels[1]. Combining these methods with

a) 국민대학교(Kookmin University)
‡ Corresponding Author : 이수찬(Soochahn Lee)
E-mail: sclee@kookmin.ac.kr
Tel: +82-2-910-4837
ORCID: https://orcid.org/0000-0002-2975-2519

※ This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021R1A2C2095452)
· Manuscript September 4, 2023; Revised January 9, 2024; Accepted January 9, 2024.

others has shown benefits, as seen in recent works[11,12,1]. The improved CAM-like score maps generated by these methods are used to enhance pixel affinities and generate pseudo-labels[13,10], which are used to train a fully supervised semantic segmentation network[14].

In this paper, we propose a method to incorporate pixel-adaptive mask refinement (PAMR)[4] so that pixel affinity is maximized when generating score maps within the self-supervised image-specific prototype exploration (SIPE) method[1]. Experimental results demonstrate that our proposed method provides substantial improvements over the baseline method SIPE. We also propose additional modifications that further improve quantitative results.

II. RELATED WORK

Using image-level labels in WSSS tasks has the advantage of lower label generation burden compared to tasks using other labels. Consequently, research on learning image segmentation using image-level labels is actively progressing. Most existing techniques apply the CAM to generate semantic pseudo-masks. However, the conventional use of CAM in WSSS is limited to representing only distinctive parts. To address this limitation in WSSS, various methods have been proposed, such as Growing Seed Regions with Constraints, Erasing, Self-supervised Manner, and Prototyping.

Growing Seed Regions with Constraints Growing Seed Regions with Constraints involves expanding regions around seeds that represent the object's location and refining them to predict pixel-wise labels that closely match the actual class object. In SEC[7], seeds are used to expand regions, and a Conditional Random Field (CRF) is employed to explore object boundaries based on the probability of object existence and color information per pixel. This process aims to predict masks close to the actual objects.

Erasing Conventional CAM represents only the most distinctive parts of an image. To broaden the CAM area, the Erasing method involves regenerating CAM by covering the initially created CAM area with masked images. This method, known as Adversarial Erasing (AE)[5], expands the expected object area.

Self-Supervised Manner Most WSSS techniques utilize only one pre-processed image for learning. The Self-supervised Equivariant Attention Mechanism (SEAM)[11] improves performance by using CAM results obtained from both the original and transformed images together.

Prototype Several techniques, such as Pixel-to-Prototype Contrast (PPC)[12], SIPE[1], define CAM-based prototypes for each class and utilize them for learning. PPC[12], based on the SEAM[11] technique, defines prototypes for each class using the top CAM-scored pixels, enhancing SEAM[11] performance. Regional Semantic Contrast and Aggregation (RCA)[25] defines class prototypes based on CAM and uses a memory bank containing training information from the entire dataset to improve CAM. SIPE [1] defines prototypes for each class per image, improving IS-CAM generation from conventional CAM. Defining prototypes for each class allows recognizing features not captured by CAM and removing noise.

Research utilizing prototypes for WSSS is actively progressing, and this paper enhances performance by generating improved prototypes based on SIPE[1].

III. PROPOSED METHOD

1. Framework

A visual summary of the proposed method is presented in Fig. 2. The baseline method, SIPE[1], comprises 1) the encoder module which generates pixel-level features, 2) the image classification module, which provides image-level supervision and generates the CAM, 3) the structure analysis module, which generates semantic structure seeds, and 4)

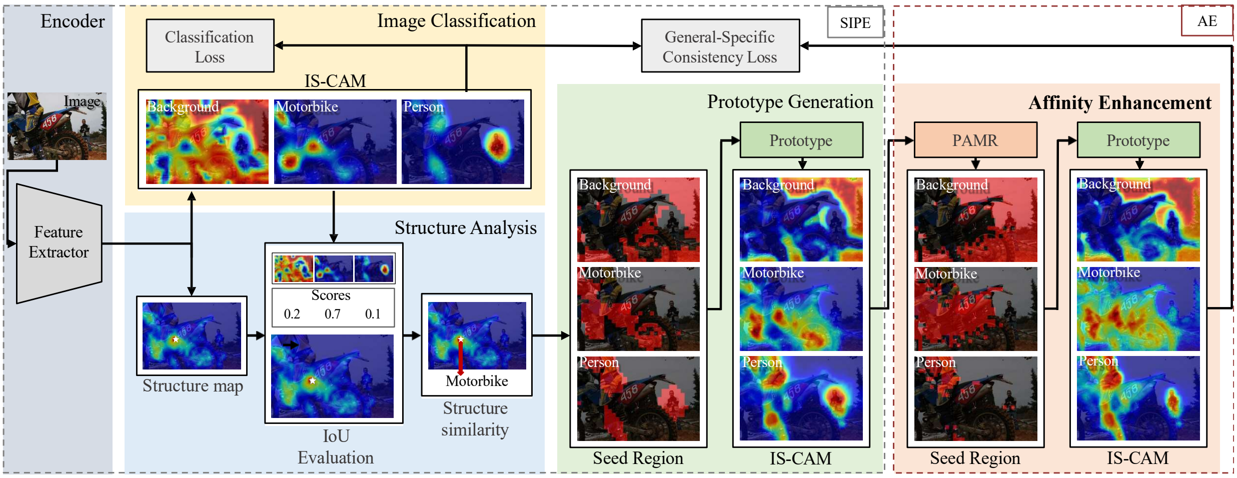


그림 2. 제안된 방법의 요약, 자기 지도 이미지별 프로토타입 탐색 (SIPE)[1] 프레임워크 기반으로 픽셀 적응형 마스크 정제 (PAMR)[4]를 사용해 유사성 기반으로 프로토타입을 강화하여 성능을 향상함

Fig. 2. Visual summary of the proposed method. We build upon the previous framework of self-supervised image-specific prototype exploration (SIPE) method[1] to enhance prototypes based on affinity using pixel-adaptive mask refinement (PAMR)[4], leading to substantial improvements in quantitative evaluations

the prototype module, which generates image-specific per-class prototype features and pixel-level per-class scores, denoted as image-specific CAM (IS-CAM). We note that we are using our own terminology, which we believe provides a more intuitive understanding of the framework.

In the proposed method, we incorporate the affinity enhancement (AE) module to the framework. In the AE module, a refined IS-CAM is generated, which is then used to generate refined region seeds, which are used to refine the prototypes and generate an improved IS-CAM. To aid the description of the AE module in 2.2, we provide a brief summary of the modules of SIPE[1] as follows:

Encoder comprises a backbone CNN, pre-trained on image classification. The feature tensor generated from this encoder $\mathcal{E}$ for the input image $\mathcal{I}$ is denoted as $\mathcal{F}=\mathcal{E}(\mathcal{I})$, and each feature vector at grid coordinate (i,j) is denoted as f_{ij}.

Classification comprises a layer to compute the CAM, and the global average pooling layer to connect the CAM with the image-level supervision through the classification loss.

Structure analysis relates the spatial distribution of features to the CAM to create seeds for prototypes. The pixel-wise spatial structure of grid coordinate (i,j) is first defined as $\mathrm{S}_{\mathrm{ij}}=\mathrm{ReLU}(S_C(f_{ij},\mathcal{F}))$, with the cosine similarity function $\mathrm{S}_{\mathrm{C}}(\cdot)$ being broadcast for the elements of $\mathcal{F}$. This is then compared to CAM to determine the semantic structure seed label, $SS_{ij}=\underset{k}{\mathrm{argmaxIoU}}\left(S_{ij},\mathcal{M}^{k}\right)$ denoting intersection-over-union with the $\mathcal{M}^{k}$ for the kth class.

Prototypes p^{k} are generated as $p^{k}=\frac{1}{|\mathcal{S}^{k}|}\Sigma_{(i,j)\in\mathcal{S}^{k}}f_{ij}$ denotes the set of coordinates with $\mathcal{S}^{k}=\{(\mathrm{i,j})\mid \mathrm{SS}_{ij}==\mathrm{k}\}$That is, p^{k} is the mean of the features with seed label k. The IS-CAM $\tilde{\mathcal{M}}^{\mathrm{k}}$ is defined as $\tilde{\mathcal{M}}^{\mathrm{k}}_{\mathrm{ij}}=\mathrm{ReLU}(\mathrm{Sc}(\mathrm{f}_{\mathrm{ij}},\mathrm{p}^{\mathrm{k}}))$.

Training loss comprises the classification loss, defined as the cross-entropy between ground truth and the inferred image-level labels, and the general-specific consistency (GSC) loss, defined as the pixel-level L1 distance between the initial CAM $\mathcal{M}^{k}$ and refined IS-CAM $\tilde{\mathcal{M}}^{k}$, for all classes k.

2. Affinity Enhanced Image-specific CAM

Within the AE module, PAMR[4], which is essentially bilateral filtering[15] on the semantic labels, is applied to the refined IS-CAM˜Mk. PAMR is defined as follows:

$$\mathrm{PAMR}(\tilde{\mathcal{M}}_{ij}^{k})= \textstyle\sum_{(p,q)\in\mathcal{N}}\alpha_{ij,pq}\tilde{\mathcal{M}}_{pq}^{k} \tag{1}$$

where the affinity kernel α_{pq} is a function based on the differences in image pixel values $\alpha_{ij,pq}= \frac{1}{W}exp\left(-\frac{||I_{ij}-I_{pq}||_2}{\sigma^2}\right)$, with a normalization term W ensuring that $\sum_{(p,q)\in\mathcal{N}}\alpha_{pq}= 1$. N denotes the local neighborhood of (i, j), which is defined as a combination of multiple 3×3 windows with varying dilation rates.

The further refined CAM $\hat{\mathcal{M}}_{ij}^{k}$ is obtained by iteratively applying PAMR n times, as $\hat{\mathcal{M}}_{ij}^{k}= \mathrm{PAMR}^{n}(\tilde{\mathcal{M}}_{ij}^{k})$. Using $\hat{\mathcal{M}}_{ij}^{k}$, we redefine seed labels as $\hat{SS}_{ij}= \underset{k}{\mathrm{argmax}}\hat{\mathcal{M}}_{ij}^{k}$ to compute AE prototypes as $\hat{p}^{k}= \frac{1}{|\hat{\mathcal{S}}^{k}|}\sum_{(i,j)\in\hat{\mathcal{S}}^{k}}f_{ij}$. The final affinity-enhanced IS-CAM, which we term AE-IS-CAM, is computed as $\check{\mathcal{M}}_{ij}^{k}= \mathrm{ReLU}(Sc(f_{ij},\hat{p}^{k}))$. Examples that highlight the improvements from $\mathcal{S}^{k}$ to $\hat{\mathcal{S}}^{k}$ and from $\tilde{\mathcal{M}}^{k}$ to $\check{\mathcal{M}}^{k}$ are depicted in Fig. 1.

3. Additional Modifications

We also apply further minor modifications regarding the details of encoded features, normalization of refined (AE) IS-CAM, and rescaling of background scores. We observed these modifications result in small improvements in the quantitative evaluations.

Structure Analysis with Hierarchical Features: In SIPE [1], only features from the last layer (semantic features) are used in structure analysis, while the concatenation of projected features generated from all internal layers (hierarchical features) are used in prototype and IS-CAM generation. However, we use hierarchical features for structure analysis as well as prototype, IS-CAM and AE-IS-CAM generation.

IS-CAM Normalization: As the cosine similarities between features and prototypes may not range from the full range of [0, 1], we apply min-max normalization on the AE-IS-CAM.

Rescaling of Background Scores: We observed background scores to be generally higher than the foreground class, as background regions may be more diverse in appearance. We thus rescale the background class activations by a factor of 0.8.

Ⅳ. EXPERIMENTS

1. Experimental Settings

Implementation: The experiments were conducted on two Titan RTX GPUs, using an implementation based on the source code provided by the authors of[1], built on the PyTorch framework. The encoder module utilized a pre-trained ResNet-101[16] as the backbone network. Training employed standard SGD optimization with a momentum of 0.9 and weight decay of 1e-4. The learning rate was set to 1e-2 for the pre-trained layers and 1e-1 for the layers in feature concatenation and the final classification layer. The PAMR process underwent 10 iterations, and a set of {1, 2, 4, 8, 12, 24} dilation rates defined N.

Multi-stage Pipeline: The complete segmentation pipeline consisted of three stages: 1) the proposed method for initial pseudo label construction, 2) the IRN[10] for refining the initial pseudo labels, 3) the DeepLabV3[17] trained using the refined pseudo labels.

Dataset: The PASCAL VOC 2012 segmentation dataset, widely recognized as the standard benchmark for WSSS, was used. This dataset comprises 21 classes, including the background, with 1,464, 1,449, and 1,456 images in the train, validation, and test sets, respectively. To enhance training, the augmented train set containing 10,582 images[18] was used. Performance was evaluated using mean Intersection over Union (mIoU), and the mIoU score on the VOC test set was computed through the official evaluation server.

2. Comparative Evaluation

Quantitative evaluation results are summarized in Table 1. The incorporation of the AE module alongside the introduced modifications leads to improvements of 2.2% and 1.4% points over the baseline[1] on the validation and test sets, respectively. Qualitative comparisons against the baseline are depicted in Figures 1 and 3. These examples illustrate instances where the proposed method yields enhanced segmentations by more effectively distinguishing both the background and the semantic classes of foreground objects.

표 1. 제안된 기법 (AE-SIPE)과 SOTA 모델 간의 PASCAL VOC 2012 데이터셋 결과 비교, 공정한 비교를 위해 이미지 수준 지도 사용한 모델들과 비교함

Table 1. Comparative evaluation of proposed AE-SIPE with SOTA on PASCAL VOC 2012 dataset. Models that rely only on image-level supervision are included for fair comparison

Model	Pub.	Backbone	Val	Test
SSWS[4]	CVPR'20	WideResnet38	62.7	64.3
SEAM[11]	CVPR'20	ResNet38	64.5	65.7
AdvCAM[19]	CVPR'21	ResNet101	68.1	68.0
CSE[20]	ICCV'21	ResNet38	68.4	68.2
CPN[21]	ICCV'21	ResNet38	67.8	68.5
PPC[12]	CVPR'22	ResNet38	67.7	67.4
AMN[22]	CVPR'22	ResNet101	69.5	69.6
RecurSeed[23]	ArXiv'22	ResNet101	72.8	72.8
SIPE[1]	CVPR'22	ResNet38	68.2	69.5
SIPE[1]	CVPR'22	ResNet101	68.8	69.7
AE-SIPE	Proposed	ResNet101	71.0	71.1

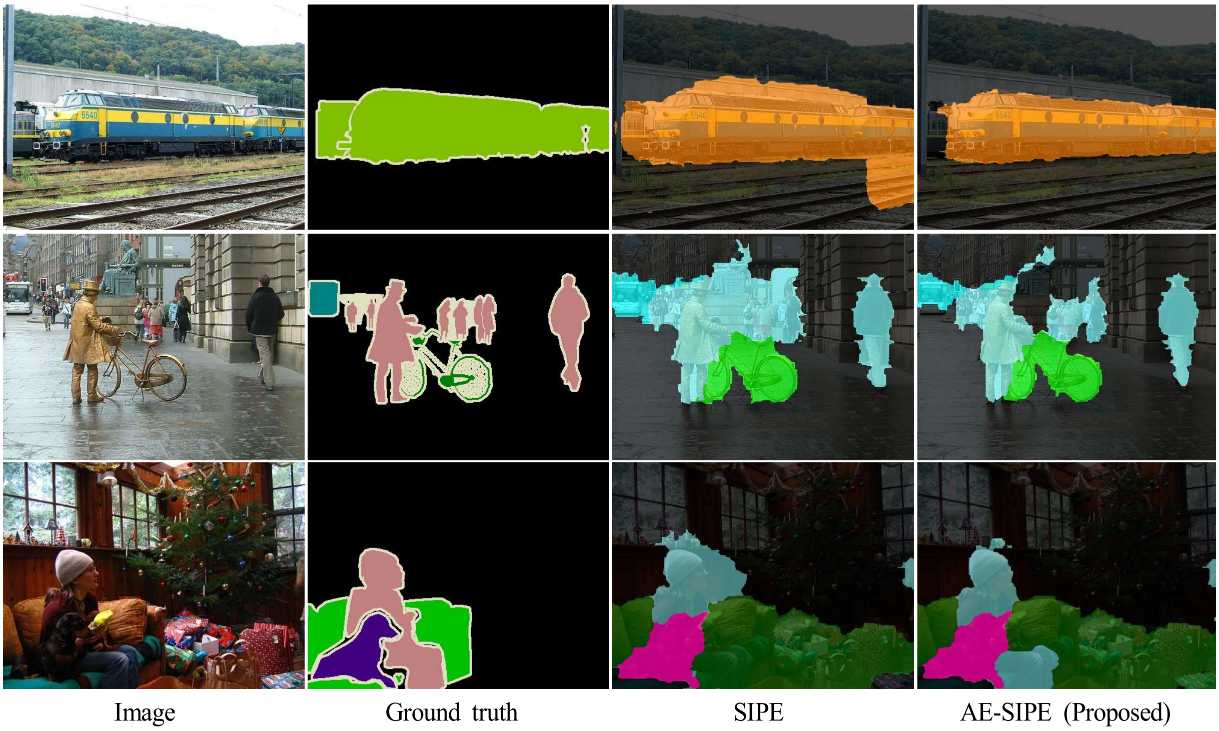


그림 3. 기준 SIPE[1] 및 제안된 방법의 PASCAL VOC 2012 데이터셋 이미지 세그멘테이션 레이블 결과

Fig. 3. Qualitative results of segmentation labels for sample images of the PASCAL VOC 2012 dataset for the baseline SIPE[1] and the proposed method

3. Ablative Study

표 2. 제안된 방법의 기준 IS-CAM 및 AE-IS-CAM의 PASCAL VOC 2012 훈련 세트에서의 제거 성능(mIoU %), [13]에 의해 정제된 결과
Table 2. Ablation performance (mIoU %) of the baseline IS-CAM and AE-IS-CAM of the proposed method on the PASCAL VOC 2012 train set, refined by [13]

Model	Train	Train+CRF[13]
Baseline[1]	58.6	64.7
+Affinity Enhancement	64.2	66.6
+Hierarchical Features	65.4	66.9
+IS-CAM Normalization	65.4	66.9
+Background Rescaling	65.8	67.8

표 3. PASCAL VOC 2012 훈련 세트에서 [13]에 의해 정제된 PAMR, 프로토타입(Pr) 및 IS-CAM(IS)의 다양한 조합에 대한 유사성 향상(AE) 비교.
Table 3. Comparison of various combinations of PAMR, prototype (Pr) and IS-CAM (IS) comprising affinity enhancement (AE) on the PASCAL VOC 2012 train set, refined by [13].

Module	Train	Train+CRF[13]
(Pr, IS) (Baseline[1])	58.6	64.7
Baseline+(Pr, IS)	59.5	65.1
Baseline+(Pr, IS, PAMR, Pr, IS)	56.6	58.7
Baseline+(PAMR, Pr, IS, Pr, IS)	62.3	65.4
Baseline+(PAMR, Pr, IS) (AE)	64.2	66.6

In this section, we delve into the specific effects of each proposed components: AE, structure analysis with HF, IS-CAM normalization, and background rescaling, as part of ablative analysis, presented in Table 2. Our observations highlight that the primary improvements from the AE module, with marginal enhancements arising from supplementary modifications. We also provide results from various combinations of PAMR, prototype generation, and IS-CAM generation, which constitute the submodules of the AE module, in Table 3. Notably, iterations of PAMR or prototype, and IS-CAM generation did not consistently yield improvements. The optimal results were achieved through the proposed AE module.

Ⅴ. DISCUSSION

In the case of AE-SIPE, the utilization of the conventional prototype generation technique serves to eliminate less significant areas, while simultaneously leveraging the PAMR[4] that utilizes the RGB characteristics of prominent areas. These complementary operations effectively enhance performance.

In the comparative evaluation in Table 1, highlights that the RecurSeed method[23] attains the highest performance. This method employs PAMR[4] for refining pseudo-labels, coupled with a self-correlation map generation(SCG) module[24]. Coincidentally, this SCG process, initially proposed for weakly supervised object localization, bears resemblance to the structure analysis module in SIPE[1]. Upon Further comparison revealed that while prototypes are used to generate pseudo-semantic segmentation labels in the proposed AE-SIPE, RecurSeed employs a decoder to infer these pseudo-labels. Additionally, while iterations improve results in RecurSeed, they do not consistently do so in the proposed method.

Also, the impact of AE is not always beneficial. When the initial Seed Region misclassifies areas outside the actual class existence region as the class, and these areas are extensive, there is a tendency for AE to further expand the erroneously predicted areas. Examining the sofa class in the third row of Fig. 3, it is evident that the initial Seed Region misclassified the gift box area as the sofa class. Through AE, the misclassified area as a sofa is expanded, resulting in a broader misclassification area compared to the original SIPE[1]. Additional techniques should be considered to address and improve this phenomenon.

We believe that the prototype approach offers simplicity, while the decoder approach may offer greater capacity. There exists a relative scarcity of works that explicitly address the decoder structure within the self-supervised framework for WSSS, warranting further research. Additionally, we aim to identify refinement processes amenable to iteration for enhanced performance improvements.

참 고 문 헌 (References)

[1] Q. Chen, L. Yang, J. Lai, and X. Xie, "Self-supervised image-specific prototype exploration for weakly supervised semantic segmentation," in CVPR, pp. 4278-4288, 2022.
doi: https://doi.org/10.48550/arXiv.2203.02909

[2] S. Hong, S. Kwak, and B. Han, "Weakly supervised learning with deep convolutional neural networks for semantic segmentation: Understanding semantic layout of images with minimum human supervision," IEEE Signal Processing Magazine, vol. 34, no. 6, pp. 39-49, 2017.
doi: https://doi.org/10.1109/MSP.2017.2742558

[3] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in CVPR, pp. 2921-2929, 2016.
doi: https://doi.org/10.48550/arXiv.1512.04150

[4] N. Araslanov and S. Roth, "Single-stage semantic segmentation from image labels," in CVPR, pp. 4252-4261, 2020 .
doi: https://doi.org/10.48550/arXiv.2005.08104

[5] Y. Wei, J. Feng, X. Liang, M.-M. Cheng, Y. Zhao, and S. Yan, "Object region mining with adversarial erasing: A simple classification to semantic segmentation approach," in CVPR, pp. 6488-6496, 2017.
doi: https://doi.org/10.48550/arXiv.1703.08448

[6] B. Kim, S. Han, and J. Kim, "Discriminative region suppression for weakly-supervised semantic segmentation," AAAI, vol. 35, no. 2, pp. 1754-1761, May 2021.
doi: https://doi.org/10.48550/arXiv.2103.07246

[7] A. Kolesnikov and C.H. Lampert, "Seed, expand and constrain: Three principles for weakly-supervised image segmentation," in ECCV, pp. 695&711, 2016.
doi: https://doi.org/10.48550/arXiv.1603.06098

[8] Z. Huang, X. Wang, J. Wang, W. Liu, and J. Wang, "Weakly-supervised semantic segmentation network with deep seeded region growing," in CVPR, pp. 7014-7023, 2018.
doi: https://doi.org/10.1109/CVPR.2018.00733

[9] J. Ahn and S. Kwak, "Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation," in CVPR, pp. 4981-4990, 2018.
doi: https://doi.org/10.48550/arXiv.1803.10464

[10] J. Ahn, S. Cho, and S. Kwak, "Weakly supervised learning of instance segmentation with inter-pixel relations," in CVPR, June 2019.
doi: https://doi.org/10.48550/arXiv.1904.05044

[11] Y. Wang, J. Zhang, M. Kan, S. Shan, and X. Chen, "Self-supervised equivariant attention mechanism for weakly supervised semantic segmentation," in CVPR, pp. 12272-12281, 2020.
doi: https://doi.org/10.48550/arXiv.2004.04581

[12] Y. Du, Z. Fu, Q. Liu, and Y. Wang, "Weakly supervised semantic segmentation by pixel-to-prototype contrast," in CVPR, pp. 4310-4319, 2022.
doi: https://doi.org/10.48550/arXiv.2110.07110

[13] P. Kr¨ahenb¨uhl and V. Koltun, "Efficient inference in fully connected crfs with gaussian edge potentials," in NeurIPS, vol. 24, 2011.
doi: https://doi.org/10.48550/arXiv.1210.5644

[14] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A.L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," IEEE Trans. PAMI, vol. 40, no. 4, pp. 834-848, 2018.
doi: https://doi.org/10.48550/arXiv.1606.00915

[15] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in ICCV, pp. 839-846, 1998.
doi: https://doi.org/10.1109/ICCV.1998.710815

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in CVPR, pp. 770&778, 2016.
doi: https://doi.org/10.48550/arXiv.1512.03385

[17] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017.
doi: https://doi.org/10.48550/arXiv.1706.05587

[18] B. Hariharan, P. Arbel`aez, L. Bourdev, S. Maji, and J. Malik, "Semantic contours from inverse detectors," in ICCV, pp. 991-998, 2011.
doi: https://doi.org/10.1109/ICCV.2011.6126343

[19] J. Lee, E. Kim, and S. Yoon, "Anti-adversarially manipulated attributions for weakly and semi-supervised semantic segmentation," in CVPR, pp. 4070-4078, 2021.
doi: https://doi.org/10.48550/arXiv.2103.08896

[20] H. Kweon, S.-H. Yoon, H. Kim, D. Park, and K.-J. Yoon, "Unlocking the potential of ordinary classifier: Class-specific adversarial erasing framework for weakly supervised semantic segmentation," in ICCV, pp. 6974-6983, 2021.
doi: https://doi.org/10.1109/ICCV48922.2021.00691

[21] F. Zhang, C. Gu, C. Zhang, and Y. Dai, "Complementary patch for weakly supervised semantic segmentation," in ICCV, pp. 7222-7231, 2021.
doi: https://doi.org/10.48550/arXiv.2108.03852

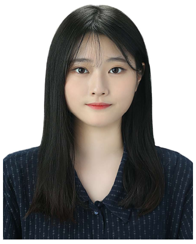
[22] M. Lee, D. Kim, and H. Shim, "Threshold matters in wsss: Manipulating the activation for the robust and accurate segmentation model against thresholds," in CVPR, pp. 4320-4329, 2022.
doi: https://doi.org/10.48550/arXiv.2203.16045

[23] S. Jo, I.-J. Yu, and K. Kim, "Recurseed and edgepredictmix: Single-stage learning is sufficient for weakly supervised semantic segmentation," 2022.
doi: https://doi.org/10.48550/arXiv.2204.06754

[24] X. Pan, Y. Gao, Z. Lin, F. Tang, W. Dong, H. Yuan, F. Huang, and C. Xu, “Unveiling the potential of structure preserving for weakly supervised object localization,” in CVPR, pp. 11637-11646, 2021.
doi: https://doi.org/10.48550/arXiv.2103.04523

[25] Zhou, Tianfei, et al. "Regional semantic contrast and aggregation for weakly supervised semantic segmentation." In CVPR, 2022.
doi: https://doi.org/10.48550/arXiv.2203.09653

저자소개



임정선

- 국민대학교 인공지능 연구실 석사
- 국민대학교 전자공학과 학사
- ORCID : https://orcid.org/0009-0007-4839-9419
- 주관심분야 : Computer Vision, Weakly supervised semantic segmentation

안수빈

- 국민대학교 인공지능 연구실 석사
- 국민대학교 전자공학과 학사
- ORCID : https://orcid.org/0009-0007-2921-2284
- 주관심분야 : Computer Vision, Machine Learning

이수찬

- 국민대학교 전자공학부 조교수
- 서울대학교 전기컴퓨터공학부 박사
- ORCID : https://orcid.org/0000-0002-2975-2519
- 주관심분야 : Computer Vision, Medical Imaging, Machine Learning