Dual-Branch Upsampling Domain Adaptation Network for Unsupervised Segmentation (Postprint)

Authors: Han Zonghuan, Liu Mingguo, Li Shen, Chen Lijia, Tian Min, Lan Tianxiang, Liang Qian

Date: 2022-05-10T11:22:57+00:00

Abstract

In industrial applications, fully supervised semantic segmentation for surface imprinted character images incurs high dataset annotation costs for enterprises. To address this issue, we propose a dual-branch feature fusion domain adaptation segmentation method (Dual-branch Feature Fusion Domain Adaptation, DbFFDA). First, inspired by the skip connection design of U-Net, we propose a residual domain adaptation segmentation network with a dual-branch upsampling structure (Residual Adaptation Network, Res-Adp). Simultaneously, we propose fused feature input to enhance network segmentation performance, overcoming the problem of missing characters. Additionally, we propose a segmentation continuity loss function #1;, which suppresses the generation of noise in segmented images. In unsupervised segmentation experiments on graphite electrode surface imprinted characters, the proposed method achieves an MIoU value of 69.60%. The actual segmentation effect essentially meets character recognition requirements and is expected to be deployed in practical applications in specific industrial scenarios, saving enterprises substantial dataset annotation costs.

Full Text

Preamble

Vol. 39 No. 9
Application Research of Computers
ChinaXiv Partner Journal

Double-Branch Upsampling Domain Adaptive Network for Unsupervised Segmentation

Han Zonghuan[1], Liu Mingguo[1]†, Li Shen[2], Chen Lijia[1], Tian Min[1], Lan Tianxiang[1], Liang Qian[1]
(1. School of Physics & Electronics, Henan University, Kaifeng Henan 475004, China;
2. Kaifeng Pingmei New Carbon Materials Technology Co., Ltd, Kaifeng Henan 475004, China)

Abstract: In industrial applications, fully supervised semantic segmentation of surface-imprinted character images incurs prohibitive dataset annotation costs. To address this challenge, we propose a dual-branch feature fusion domain adaptation segmentation method (DbFFDA). First, drawing inspiration from U-Net's skip connection design, we introduce a residual adaptation network (Res-Adp) with a dual-branch upsampling structure. Simultaneously, we propose fused feature inputs to enhance network segmentation performance and overcome character missing issues. Additionally, we introduce a segmentation continuity loss function that suppresses noise generation in segmented images. In unsupervised segmentation experiments on graphite electrode surface-imprinted characters, our method achieves an MIoU of 69.60%. The actual segmentation quality substantially meets character recognition requirements, demonstrating potential for practical deployment in specific industrial scenarios and offering enterprises substantial savings in dataset labeling costs.

Keywords: surface imprint characters; domain adaptation; semantic segmentation; unsupervised training

0 Introduction

Image semantic segmentation represents a crucial research direction in computer vision with broad application prospects in autonomous driving, character recognition, and other domains. Its objective is to categorize each pixel in an image into classes with specific semantic information. In industrial applications, segmenting surface-imprinted character images proves particularly challenging due to variations in lighting and printer quality, making precise segmentation difficult for traditional algorithms.

Since the advent of deep learning, numerous researchers have explored semantic segmentation, progressively improving both effectiveness and accuracy. In 2014, Long et al. [1] proposed Fully Convolutional Networks (FCN), which consist entirely of convolutional layers without fully connected structures, providing new insights for semantic segmentation network design. In 2015, Ronneberger et al. introduced U-Net [2], featuring a symmetric U-shaped encoder-decoder architecture with skip connections that fuse information across network levels, achieving remarkable success in medical image segmentation. In 2018, Chen et al. [3-6] presented the DeepLab series, whose most significant contribution is the Atrous Spatial Pyramid Pooling (ASPP) module. Convolutional Neural Networks (CNNs) have dramatically enhanced semantic segmentation performance in complex scenes, offering unparalleled advantages over traditional methods.

However, all these networks employ fully supervised training, requiring manually annotated labels as training data and incurring enormous annotation costs. The emerging research on unsupervised domain adaptive semantic segmentation networks offers a promising solution to this problem. Domain adaptation focuses on optimizing the alignment of feature distributions between source and target domains, extracting shared features to predict target domain data. Researchers have proposed various domain adaptation methods, most of which utilize Generative Adversarial Networks (GAN) [7] for adversarial training. The segmentation network serves as the generator—source and target domain images are fed into the segmentation network (generator) to produce predictions that are alternately fed to the discriminator. Through adversarial training between the segmentation network and discriminator, the distributions of features from both domains are aligned to achieve domain adaptation segmentation.

Ganin et al. proposed Domain-Adversarial Neural Network (DANN) [8], comprising feature extraction, classification, and domain discrimination modules. This network constructs two loss functions—image classification loss and domain classification loss—simultaneously improving network performance while aligning feature distributions across domains, thereby enhancing classification capability for target domain images. Since single-discriminator adversarial domain adaptation cannot leverage complex multi-modal structures, Pei et al. [9] introduced Multi-Adversarial Domain Adaptation (MADA), constructing multiple discriminators for domain adaptation training. Multiple domain discriminators align distributions across various dimensions, improving adaptation effectiveness. Luo et al. [10] proposed Category-Level Adversarial Network (CLAN), constructing a category-level adversarial network to enforce local semantic consistency within overall alignment. Wang et al. [11] introduced Patch-based Output Space Adversarial Learning (pOSAL), designing a lightweight and efficient segmentation network along with a novel shape-aware segmentation loss to guide the network in generating accurate and smooth predictions. Wang et al. [12] proposed Boundary and Entropy-driven Adversarial Learning (BEAL), a domain adaptation framework that improves segmentation performance in ambiguous boundary regions by encouraging target domain boundary predictions to resemble those of the source domain, thereby generating more accurate boundaries.

Zhu et al. [13] proposed Cycle-Consistent Generative Adversarial Networks (CycleGAN), cascading two GAN networks and introducing cycle consistency loss to construct a bidirectional style transfer network between source and target domain images, enabling generation of simulated samples close to the target domain distribution to assist training. Chen et al. [14] improved the CycleGAN structure by adding a segmentation branch to the target-to-source generator, aligning source and target domain distributions from both image and feature perspectives, achieving remarkable success in domain adaptation segmentation between medical MRI and CT images. Zhang et al. [15] introduced atrous spatial pyramid pooling (ASPP) into the segmentation network to extract multi-scale image features for improved segmentation performance, applying segmentation prediction entropy to adversarial loss to reduce domain shift.

Liu et al. [16] proposed Source-Free Domain Adaptation (SFDA), providing a domain adaptation training approach for scenarios where source domain datasets cannot be publicly released. By recovering and filtering source domain data from the source domain model, domain adaptation can be achieved using only the trained source domain model and target domain dataset. Araslanov et al. [17] proposed a lightweight domain adaptation segmentation method that addresses the complexity and high resource consumption of existing networks by abandoning common approaches like adversarial training and style transfer, instead employing data augmentation techniques such as noise addition, flipping, and scaling to ensure consistency in cross-domain semantic segmentation. Wang et al. [18] introduced Correlation-Aware Domain Adaptation (CorDA), using universal self-supervised depth estimation guidance from both domains to bridge the domain gap. This method explicitly learns task feature correlations with the help of target depth estimation to improve prediction quality. Saha et al. [19] proposed a method for encoding visual task relationships to enhance unsupervised domain adaptation network performance, introducing a Cross-Task Relation Layer (CTRL) that encodes task dependencies between semantic and depth predictions. Liu et al. [20] proposed a two-stage segmentation network —CDR-GANs—where each stage comprises a semantic segmentation network, generator, and discriminator. During training, the discriminator guides the semantic segmentation network and generator to learn the joint probability distribution of original images and segmentation predictions. Li et al. [21] addressed the limitation that existing domain adaptation algorithms use shared source domain networks to learn cross-domain feature representations, which restricts generalization to unlabeled target domain objects, by proposing Transferable Semantic Augmentation (TSA) to implicitly generate source domain features for target domain objects and enhance network adaptability.

This paper proposes the dual-branch feature fusion domain adaptation segmentation method DbFFDA, which improves upon U-Net in network architecture, image preprocessing, and loss function design, achieving satisfactory segmentation results on graphite electrode imprinted character datasets that substantially meet industrial application requirements. DbFFDA' s innovations include three aspects: (a) A dual-branch upsampling unsupervised semantic segmentation network Res-Adp that introduces residual modules into U-Net' s skip connections to build a residual branch for cross-domain feature alignment. During network upsampling, features at each level undergo upsampling through both residual and convolutional branches. The residual branch handles feature alignment using domain-invariant features for image segmentation, while the convolutional branch preserves domain-specific features to supplement segmentation details using unique image characteristics. (b) Fused feature input. Considering that surface-imprinted character images contain numerous noise points and character edges are critically important, we use grayscale images, median-filtered images, and edge detection images as three input channels fused and fed into the network for training. (c) Construction of a segmentation continuity loss function to constrain segmentation network training. Based on the prior knowledge that objects should be internally continuous in segmented images, we propose a segmentation continuity loss function that indirectly improves target domain segmentation by constraining source domain segmentation generation, suppressing both character voids and background noise.

1.1 Domain Adaptation Segmentation Framework

The domain adaptation segmentation framework is illustrated in Figure 1. The segmentation network is our proposed dual-branch upsampling network Res-Adp, shared by both domains. IS represents source domain (simulated) data, which easily obtains labels with rich supervision information and complete label set LS. IT represents target domain (real) data without supervision information.

Figure 1. Domain adaptive segmentation framework

Source domain data consists of computer-generated simulated images whose labels require no manual annotation, while target domain data comprises real images captured by cameras without annotation information. The framework involves two training phases:

(a) **Fully supervised training on source domain data.** Source domain (simulated) images IS are input to the segmentation network to obtain source domain predictions PS. PS can construct a cross-entropy loss function with labels LS to train the segmentation network Res-Adp in a fully supervised manner. The cross-entropy loss function formula is:

$$L_{CE} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(p_i) + (1-y_i)\log(1-p_i)]$$

where p_i represents pixels in the source domain prediction and y_i represents pixels in the source domain label. Target domain (real) images IT are input to the segmentation network to obtain target domain predictions PT. Since target domain images IT lack labels, cross-entropy loss cannot be constructed for segmentation training.

(b) **Adversarial training between segmentation network (generator) and discriminator.** To enable the segmentation network trained on source domain images to accurately segment target domain images, we must align the distributions of both domains through domain adaptation training. Source predictions PS and target predictions PT are fed to the discriminator for adversarial training with the segmentation network (generator). The adversarial loss function L_{Gan} is defined as:

$$L_{Gan} = \mathbb{E}_{x \sim S_{data}}[\log D(S(x))] + \mathbb{E}_{x \sim T_{data}}[\log(1 - D(T(x)))]$$

where S_{data} and T_{data} represent source and target domain images respectively, S represents the segmentation network, and D represents the discriminator.

The overall network loss function L comprises three components:

$$L = L_{Gan} + L_{CE} + L_{Con}$$

where L_{Gan} is the adversarial loss, L_{CE} is the cross-entropy loss, and L_{Con} is our proposed segmentation continuity loss function, detailed in Section 1.5.

The discriminator' s optimization objective is to distinguish whether input segmentation predictions originate from source predictions PS or target predictions PT. During training, the discriminator aims to output an all-ones matrix for source predictions and an all-zeros matrix for target predictions to determine prediction categories. The discriminator constructs L2 norms between its judgment on target predictions and an all-zeros matrix, and between its judgment on source predictions and an all-ones matrix, minimizing the sum of these two losses.

The segmentation network (generator) aims to produce predictions that confuse the discriminator, specifically making the discriminator output an all-ones matrix for target predictions, bringing them close to source predictions. The L2 norm is constructed between the discriminator' s judgment on target predictions and an all-ones matrix, while non-adversarial losses are constructed between source input images and their labels. The segmentation network minimizes the sum of these two losses.

Non-adversarial losses L_{CE} and L_{Con} train segmentation performance, while adversarial loss L_{Gan} trains domain adaptation capability. Under their joint constraints, the segmentation network improves source domain (simulated) image segmentation performance while aligning feature distributions across domains, thereby enhancing target domain (real) image segmentation performance and achieving unsupervised segmentation of target domain images.

1.3 Markov Discriminator

DbFFDA employs a Markov discriminator [24]. As shown in Figure 4, segmentation predictions undergo four downsampling operations followed by one convolutional operation to output a feature map with one channel. Also known as PatchGAN, the Markov discriminator outputs a feature map where each point represents the authenticity of a patch region in the segmentation prediction. The discriminator' s optimization direction is to identify whether input predictions belong to source or target domains, while the segmentation network aims to generate predictions indistinguishable to the discriminator. Through this adversarial game, both components continuously improve their performance to achieve cross-domain distribution alignment.

Figure 4. Markov discriminator structure

1.4 Fusion Feature Input

Under complex and variable natural lighting conditions, constrained by printer quality and image acquisition device performance, surface-imprinted character images exhibit uneven illumination and substantial noise. Directly feeding raw images into the network for training introduces numerous binary noise points into segmented images, severely degrading segmentation purity.

We convert camera-captured color images of surface-imprinted characters to grayscale images as the first input channel. Median filtering is applied to remove noise while preserving foreground information maximally, with the filtered image serving as the second input channel. Character edges in surface-imprinted images constitute critical foreground information essential for segmentation. We apply Sobel operators to median-filtered images for edge extraction, using the resulting images as the third input channel. Figure 5 illustrates the three-channel input data, which are concatenated and fed into the network for training.

Figure 5. Fusion feature input

1.5 Segmentation Continuity Loss Function

Due to substantial noise and partial character distortion in surface-imprinted character images, using only cross-entropy as the non-adversarial loss function leads to character voids and binary noise points in target domain segmentation results.

To address this issue, inspired by traditional computer vision algorithms, we propose a segmentation continuity loss function L_{Con}. In binary label maps, individual characters are internally continuous, as is the background, without interspersed binary noise points—foreground and background remain relatively independent. Therefore, in N-channel labels, points with value 1 and points with value 0 on each channel should be continuous, meaning each class is internally continuous and independent from others.

As shown in Figure 6, except for boundary points, a point' s value should match its neighbors' . Since network segmentation results are obtained by mapping

N-channel feature maps to probability distributions and taking the maximum index, the N-channel feature map output should exhibit the same property—continuous activation values at each point. A point $A(i, j, k)$ on channel k should have activation values similar to its neighbors, enabling construction of the segmentation continuity loss function:

$$L_{Con} = \frac{1}{N} \sum_{k=1}^{c} \sum_{i=1}^{m} \sum_{j=1}^{n} \left| A(i,j,k) - \frac{A(i-1,j,k) + A(i+1,j,k) + A(i,j-1,k) + A(i,j+1,k)}{4} \right|$$

where $A(i, j, k)$ represents the point at position (i, j) on channel k, and $A(i-1, j, k)$, $A(i+1, j, k)$, $A(i, j-1, k)$, $A(i, j+1, k)$ represent its left, right, top, and bottom neighbors respectively. This function effectively suppresses void generation within characters.

Figure 6. Point of a channel in an N-channel label and its four neighborhoods

---

2.1 Experimental Data

Domain adaptation segmentation effectively addresses the high cost of manual semantic segmentation dataset annotation, offering broad application prospects. In industrial applications, numerous scenarios require semantic segmentation algorithms to process captured images for subsequent operations.

In graphite electrode manufacturing, surface-imprinted characters must be recognized for production statistics and management. Semantic segmentation constitutes a critical pre-processing step, yet pixel-level annotation would impose tremendous production costs. This work addresses the problem by using computer-generated source domain data and camera-captured target domain data, effectively saving annotation costs through domain adaptation segmentation.

The experimental dataset comprises source and target domain components. Source (simulated) data consists of screenshots from computer font libraries without manual annotation, featuring complete computer-generated labels with random character fonts, sizes, and spatial positions. Target domain data comprises graphite electrode surface-imprinted character images captured by mobile phone cameras from a carbon materials manufacturing enterprise, as shown in Figure 7.

Figure 7. Real image of graphite electrode

Dataset example images appear in Figure 8. The source domain (computer font library characters) contains 600 images, while the target domain (graphite electrode surface-imprinted characters) contains 550 images, with 440 training images and 110 test images. All dataset images are sized 512$\times$128.

Figure 8. Examples of dataset

2.2 Evaluation Metrics

We evaluate network performance using five metrics: Pixel Accuracy (PA), Mean Pixel Accuracy (MPA), Precision, Recall, and Mean Intersection over Union (MIoU).

PA is the ratio of correctly classified pixels to total pixels. MPA is the mean of per-class ratios of correctly classified pixels to total pixels of that class. Precision is the proportion of correctly predicted positive examples among all predicted positives. Recall is the proportion of correctly predicted positives among all actual positives. Intersection over Union (IoU) is the ratio of intersection pixel count to union pixel count for a predicted and labeled object class, while MIoU represents the mean IoU across all classes. The five evaluation metrics are formulated as:

$$\mathrm{PA} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\mathrm{MPA} = \mathrm{mean}\left(\frac{TP}{TP + FN + TN + FP}\right)$$

$$\mathrm{Precision} = \frac{TP}{TP + FP}$$

$$\mathrm{Recall} = \frac{TP}{TP + FN}$$

$$\mathrm{MIoU} = \mathrm{mean}\left(\frac{A \cap B}{A \cup B}\right)$$

where True Positive (TP) denotes pixels correctly predicted as positive, True Negative (TN) denotes pixels correctly predicted as negative, False Positive (FP) denotes pixels incorrectly predicted as positive, and False Negative (FN) denotes pixels incorrectly predicted as negative.

2.4.2 Ablation Experiment Results and Analysis

Ablation experiments compare five configurations: U-Net (fully supervised) [2], U-Net (unsupervised) [2], Res-Adp (innovation 1), Res-Adp + fusion feature input (innovations 1+2), and Res-Adp + fusion feature input + L_{Con} (DbFFDA, innovations 1+2+3). These experiments quantify the performance gap between DbFFDA and fully supervised U-Net while validating each proposed innovation. For each configuration, we identify the training epoch with maximum MIoU and report its metrics. Segmentation comparisons appear in Figure 11, detail comparisons in Figure 12, and quantitative evaluations in Table 2.

Figure 11. Image of segmentation effect comparison of ablation experiment

Figure 12. Segmentation detail comparison of ablation experiments

Table 2. Quantitative evaluation of ablation experiment

Approach	PA	MPA	Precision	Recall	MIoU
U-Net (fully supervised) [2]	-	-	-	-	-
U-Net (unsupervised) [2]	-	-	-	-	-
Res-Adp	-	-	-	-	-
Res-Adp + fusion feature input	-	-	-	-	-
Res-Adp + fusion feature input + L_{Con}	-	-	-	-	69.60%

As shown in Figures 11-12 and Table 2: (a) Unsupervised U-Net suffers from severe character missing and numerous noise points. Res-Adp (innovation 1) significantly improves this issue through its dual-branch upsampling structure, which aligns cross-domain features while preserving domain-specific features, yielding continuous characters without missing parts and improving MIoU by 1.52% over unsupervised U-Net. (b) Res-Adp + fusion feature input (innovations 1+2) adds median-filtered and Sobel edge detection images to network inputs, suppressing noise and enhancing character edge information. This produces smoother character edges, reduces burrs, and decreases noise points compared to Res-Adp alone, improving MIoU by 0.43%. (c) Res-Adp + fusion feature input + L_{Con} (DbFFDA, innovations 1+2+3) generates characters that are continuous, have smooth edges without burrs, and contain minimal background noise. The segmentation continuity loss function L_{Con} further suppresses background noise and character voids, achieving the best segmentation performance with MIoU reaching 69.60%.

Although DbFFDA still trails fully supervised U-Net in objective metrics, its actual segmentation quality substantially meets character recognition requirements and shows promise for practical industrial deployment.

3 Conclusion

Addressing the difficulty of obtaining labels for surface-imprinted character images in industrial applications, this paper proposes the dual-branch feature fusion domain adaptation segmentation method DbFFDA. First, inspired by U-Net' s skip connections, we introduce the dual-branch upsampling segmentation network Res-Adp. Additionally, we fuse grayscale, median-filtered, and edge detection images as network input to suppress noise and enhance character edge information. Furthermore, based on the prior that object interiors should be continuous in labels, we propose the segmentation continuity loss function L_{Con}. This function indirectly improves target domain segmentation by constraining source domain segmentation generation, further suppressing character voids and background noise.

DbFFDA produces segmented images with complete characters, smooth edges, and minimal noise, achieving 69.60% MIoU that substantially meets industrial character recognition requirements. For images suffering from excessive noise in segmentation predictions due to uneven illumination, future work will explore discriminator structure improvements to better constrain generator optimization, suppress noise generation, and enhance segmentation quality.

References

[1] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE Press, 2015: 3431-3440.

[2] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation [J]. Medical Image Computing and Computer-Assisted Intervention, 2015: 1520-1528.

[3] Chen L C, Papandreou G, Kokkinos I, et al. Semantic image segmentation with deep convolutional nets and fully connected crfs [C]// International Conference in Learning Representations, 2015.

[4] Chen L C, Papandreou G, Kokkinos I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2018, 40 (4): 834-848.

[5] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation [J]. arXiv preprint arXiv: 1706. 05587, 2017.

[6] Chen L C, Zhu Yukun, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C]// Proceedings of the European Conference on Computer Vision (ECCV). 2018: 801-818.

[7] Goodfellow I J, Pouget A J, Mirza M, et al. Generative adversarial networks

[C]// Proceedings of Annual Conference on Neural Information Processing Systems. Montreal, Quebec, Canada: ACM. 2014. 2672–2680.

[8] Ganin Y, Lempitsky V S. Unsupervised domain adaptation by backpropagation [C]// Proceedings of the 32nd International Conference on Machine Learning. Lille, France: JMLR, 2015. 1180–1189.

[9] Pei Zhongyi, Cao Zhangjie, Long Mingsheng, et al. Multi-adversarial domain adaptation [C]// Proceedings of the 32nd Conference on Artificial Intelligence, New Orleans, Louisiana, USA: AAAI, 2018. 3934–3941.

[10] Luo Yawei, Zheng Liang, Guan Tao, et al. Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 2507-2516.

[11] Wang Shujun, Yu Lequan, Yang Xin, et al. Patch-based output space adversarial learning for joint optic disc and cup segmentation [J]. IEEE Trans on Medical Imaging, 2019, 38 (11): 2485-2495.

[12] Wang Shujun, Yu Lequan, Li Kang, et al. Boundary and entropy-driven adversarial learning for fundus image segmentation [C]// International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2019: 102-110.

[13] Zhu Junyan, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks [C]// 2017 International Conference on Computer Vision (ICCV). 2017: 2242-2251.

[14] Chen Cheng, Dou Qi, Chen Hao, et al. Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation [J]. IEEE Trans on Medical Imaging, 2020, 39 (7): 2494-2505.

[15] Zhang Xunhui, Zhou Yong, Zhao Jiaqi, et al. Entropy enhanced unsupervised domain adaptive remote sensing image semantic segmentation [J]. Application Research of Computers, 2021, 38 (09): 2852-2856.

[16] Liu Yuang, Zhang Wei, Wang Jun. Source-free domain adaptation for semantic segmentation [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 1215-1224.

[17] Araslanov N, Roth S. Self-supervised augmentation consistency for adapting semantic segmentation [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 15384-15394.

[18] Wang Qin, Dai Dengxin, Hoyer L, et al. Domain adaptive semantic segmentation with self-supervised depth estimation [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 4235-4245.

[19] Saha S, Obukhov A, Paudel D P, et al. Learning to relate depth and semantics for unsupervised domain adaptation [C]// Proceedings of the IEEE/CVF

Conference on Computer Vision and Pattern Recognition. 2021: 8197-8207.

[20] Liu Shaopeng, Hong Jiaming, Liang Jiepeng, et al. Medical image segmentation using semi-supervised conditional generative adversarial nets [J]. Journal of Software, 2020, 31 (8): 2588−2602.

[21] Li Shuang, Xie Mixue, Gong Kaixiong, et al. Transferable semantic augmentation for domain adaptation [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 11516-11525.

[22] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Deep residual learning for image recognition [C]// Proceedings of Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016: 770-778.

[23] Hu Yangtao, Pei Yang, Lin Chuan, et al. Atrous residual u-net for retinal vessel segmentation [J]. Computer Engineering and Applications, 2021, 57 (07): 185-191.

[24] Isola P, Zhu Junyan, Zhou Tinghui, et al. Image-to-image translation with conditional adversarial networks [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 1125-1134.

Note: Figure translations are in progress. See original paper for figures.