ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-4485-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICASSP48485.2024.10447006

PROGRESSIVE LEARNING BASED KNOWLEDGE DISTILLATION FOR LOW RESOLUTION CEREBRAL MICROBLEED SEGMENTATION

Tianxiang Xia[1], Rong Zhang[1], Zhenzuo Chen[1], Guomin Xie[2*], Xiping Wu[2], Zhongyue Lv[2], Lijun Guo[1]*

[1] Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo, Zhejiang, China
[2] Department of Neurology, Ningbo Medical Center Li-Huili Hospital, Ningbo, Zhejiang, China

ABSTRACT

This study aims to address key technical issues in the segmentation of Cerebral MicroBleeds (CMBs) based on Low-Resolution (LR) Magnetic Resonance Imaging (MRI) data. There are two challenges in this task. First, the CMB lesions are typically small in size and easily confused with various mimics. Second, anisotropy becomes more prominent and adverse in LR MRI sequences than HR sequences. To address these issues, we propose a Progressive Learning based Knowledge Distillation method. This method progressively transfers knowledge from HR models to their LR counterparts, thereby minimizing the occurrence of false positives attributable to noise from Super-Resolution. To further eliminate the influence of anisotropy, an encoding-enhanced network, called E^{2}U-Net, is proposed in this paper. It can effectively capture anisotropic information and mitigates potential feature loss. The experimental results on multiple publicly accessible CMBs datasets demonstrated the superiority of our proposed approach over existing deep-learning methods.

Index Terms—cerebral microbleeds, low-resolution, magnetic resonance imaging, knowledge distillation

1. INTRODUCTION

Brain Magnetic Resonance Imaging (MRI) plays a crucial role in the clinical diagnosis of various diseases. However, High-Resolution (HR) MRI scanning is relatively scarce due to its expensive cost, especially in many developing countries [1]. At present, Low-Resolution (LR) MRI data remains the primary data source for clinical and academic research [2]. On the other hand, obtaining HR MRI necessitates those patients endure prolonged periods in the gantry. This can increase significant discomfort for patients, especially elderly patients with cognitive dysfunction or those with psychological disorders [3]. In contrast, LR MRI offers a more affordable option with a shorter imaging duration. Therefore, image analysis for LR MRI has important practical significance. This study focuses on the segmentation of cerebral microbleeds based on LR MRI and proposes a new feasible method for the small object segmentation task in LR MRI.

Cerebral MicroBleeds (CMBs) represent hemosiderin deposits resulting from microvascular lesions within the brain, and are frequently encountered in elderly adults with cognitive dysfunction [4]. Currently, the detection of CMBs predominantly relies on MRI, where they typically manifest as hypointense lesions, measuring 2–10 mm in size [5]. Detecting CMBs poses a considerable challenge because of the diminutive size of these lesions. Furthermore, the presence of various CMB mimics, such as calcification deposits and MRI artifacts, contributes to a high incidence of false positives. In recent years, substantial advancements have been made in automated CMB detection methods, primarily leveraging HR MRI [6–9]. The majority of existing methods utilize 3D convolution to extract structural information from consecutively acquired brain image sequences [10,11]. However, when these methods are directly applied to LR data, there is a marked deterioration in performance. One reason for that is the serious anisotropy in LR MRI. As shown in Fig. 1, the in-plane resolution in LR MRI is typically several times higher than its depth [12]. Some methods [11,13] employed Super-Resolution (SR) techniques to generate HR information in the depth direction. Nonetheless, SR also introduces noise that compromises the critical edge information and causes inconsistencies in the small objects [14,15]. The network may erroneously attribute the noise induced by SR as an intrinsic CMBs feature, consequently producing undesirable false positives. Recently, Knowledge Distillation (KD) [16] has shown achievements in the LR face recognition task [17,18] by effectively transferring the prior knowledge from HR images to LR images. This inspired us to adopt KD to address the LR problem in MRI. However, as CMBs are extremely small targets, the feature maps have significant disparities between LR and HR images. Consequently, the model may predict all pixels as backgrounds to minimize distillation losses, leading to model collapse.

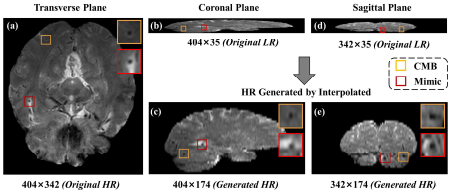

Fig. 1. Illustration the anisotropy of the three viewing planes in LR MRI. The limited resolution in the depth direction poses a challenge in distinguishing between CMBs and mimics. Although SR techniques can enhance certain information, they may also introduce noise, leading to undesirable false positives.

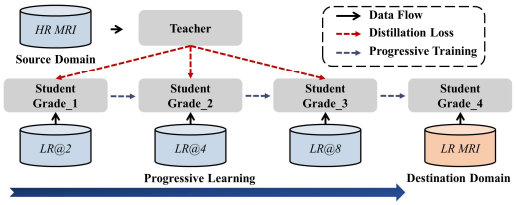

Fig. 2 Concept of the progressive learning based knowledge distillation. PLKD is a progressive approach to educate the student. Under the guidance of the teacher, a student gradually encounters LR data and accepts increasingly difficult learning challenges. Consequently, the ability of the student is gradually enhanced and ultimately is able to cope with LR data in the destination domain. LR@2, LR@4 and LR@8 are derived from HR datasets with resolution from high to low.

To address these issues, we propose Progressive Learning based Knowledge Distillation (PLKD) to progressively distill well-constructed feature maps from an HR network into an LR network. As shown in Fig. 2, a student starts learning from relatively high LR samples and gradually comes into contact with increasingly LR samples. This strategy of progressive learning can avoid the problem of model collapse in knowledge distillation due to significant differences in features between the student and the teacher. By leveraging the HR model, the LR model can mitigate the noise sensitivity and learn discriminative CMBs features. In addition, in order to increase the ability to extract small target features in LR MRI, we introduced a novel encoding-enhanced network architecture called as E2U-Net, which synergistically combines the strengths of 2D and 3D convolutions to effectively capture anisotropic information. Furthermore, we incorporated a Feature Compensation Module (FCM) to mitigate potential information loss during training. Through extensive experimental evaluations on multiple publicly CMBs datasets, our approach consistently demonstrated significant advancements over existing deep-learning approaches.

2. METHODOLOGY

Recently, several studies on detecting CMBs have adopted two-stage frameworks to improve efficiency [6–11]. Our framework follows a similar process, which consists of a screening stage and a segmentation stage. As shown in Fig. 3, in the screening stage, we employ an FCN with six convolutional modules and one pooling layer. This FCN processes the entire brain MRI as an input, producing 3D score volumes. The volume of each value represents the probability of the presence of CMB in the corresponding region. In the segmentation stage, we utilize the proposed E^{2}U-Net, which accepts the candidate volumes as input and generates segmentation-level results, which further refine the outcomes by eliminating false-positive candidate volumes. The work in this study is focusing on the segmentation network of the second stage. The remainder of this section will provide a detailed introduction to our segmentation network and its training method in the case of small targets segmentation for LR MRI.

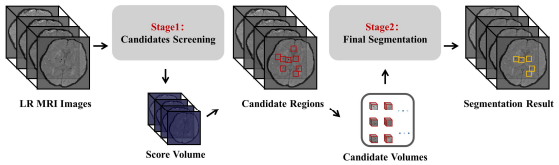

Fig. 3 Pipeline of general two-stage framework for CMBs detection and segmentation using MRI.

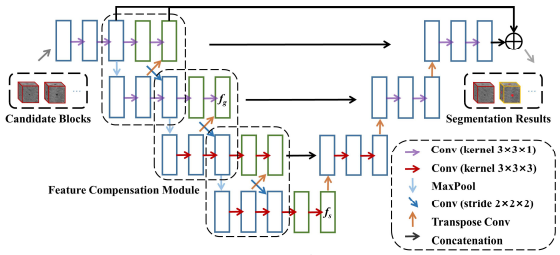

Fig. 4 Illustration of the proposed E^{2}U-Net.

2.1. Encoding-Enhanced Network

As shown in Fig. 4, the proposed E^{2}U-Net is meticulously designed as a discriminative network explicitly tailored for the segmentation of small targets in LR MRI, in which relatively LR in depth leads to significant anisotropy.

2D+3D Hybrid Convolution. To solve the problem of anisotropy in LR MRI, we flexibly use 2D and 3D convolutions to extract spatial information at different levels. To leverage the rich spatial information in the transverse plane with relatively HR, we employ 2D convolutions in the first two layers to extract geometric features f_g, as shown in Fig. 4. Additionally, we employ 3D convolutions in the subsequent layers to extract semantic features f_s. This approach offers the distinct advantage of allocating a more significant share of computational resources to critical features, thus solving the problem of anisotropy and reducing the impact of SR noise.

Feature Compensation Module. In the segmentation of small targets, gradients and edge information are more susceptible to loss, particularly during the downsampling process. To address this concern, we add a Feature Compensation Module (FCM) into the encoder to enhance the information transmission and interaction between two adjacent layers, as shown in Fig. 4. The supplementary downsampling branches help complement and preserve important features that may have been lost during downsampling. Additionally, the upsampling branches also provide multi-scale information to the model, enhancing its awareness of small targets.

2.2. Progressive Learning based Knowledge Distillation

To address the challenges associated with segmenting small objects in LR imagery, we introduce a PLKD method to train the model effectively. Direct distillation from HR to LR could lead to significant feature disparities and, consequently, ineffective knowledge transfer. Instead, PLKD guides students

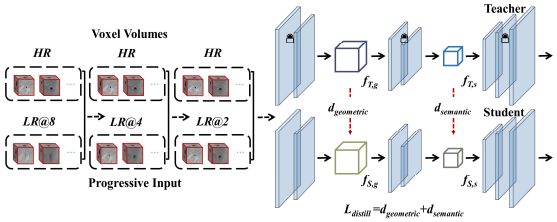

Fig. 5 Flow chart of progressive learning knowledge distillation.

through a step-by-step learning process from simple to complex tasks. In Fig. 5, the student network is initially fed with relatively HR samples, which are then progressively reduced in resolution to match the LR of the target domain. This strategy allows for the full utilization of HR sample information and mitigates the risk of transfer failures due to the stark resolution contrast between the source and target domains.

Low Resolution MRI Generation. We utilize pairs of HR and LR MRI to train the LR network. Specifically, we downsample HR MRI with ratios of ×2, ×4, and ×8. Subsequently, the reduced-size images are upsampled back to their original sizes using bicubic interpolation.

Progressively Learning. As depicted in Fig. 5, we first input pairs of voxel volumes of HR and LR with a 2× ratio to furnish a primary student model with preliminary HR knowledge. Subsequently, we train the student model using HR and LR with ×4 and ×8 ratios to learn middle and senior student models. This enables the student model to gradually comprehend and grasp the subtle features present in HR voxel volumes. Finally, to ensure the adaptability of models, we conduct further fine-tuning on the student model. In summary, our PLKD process facilitates an enhanced alignment of features between HR and LR images, effectively guiding the student model in attaining refined HR knowledge.

Loss. We adopt a distance function based on cosine similarity [21] to minimize the angular separation between the feature vectors associated with both the teacher and student models. This cosine distance is calculated between the feature maps of the teacher and the student models using the following equation:

$$d = 1 - \frac{f_T \cdot f_S}{\|f_T\| \|f_S\|}, \tag{1}$$

where f_T and f_S denote the features from the teacher and the student, respectively. Throughout the encoding stage, low-level features contain geometric details, while high-level features contain semantic information [22]. Therefore, we employ two distilled layers, as shown in Fig. 5. The distillation loss is the combined value of $d_{geometric}$ and $d_{semantic}$, which are calculated by between $f_{T,g}$ and $f_{S,g}$, and between $f_{T,s}$ and $f_{S,s}$, respectively.

$$L_{distill} = \lambda_1 d_{geometric} + \lambda_2 d_{semantic}, \tag{2}$$

where λ_1 and λ_2 denote the weights of each loss, set to 0.5 and 0.5. In addition, the soft dice loss [23] is employed as another loss function for the segmentation. Therefore, the total loss is the sum of the soft dice loss and distillation loss:

$$L_{total} = \alpha L_{SoftDice} + \beta L_{distill}, \tag{3}$$

where α and β denote the weights of each loss. We empirically set α and β to 0.6 and 0.4.

Table 1. Acquisition details for the datasets. Acronyms TR -repetition time (ms); TE - echo time (ms); FA - Flip angle.

Type	Name	Num	TR	TE	FA	Resolution(mm)
T2*	RSS	34	45	31	13	0.49×0.49×0.8
	SABRE	11	1288	21	18	0.45×0.45×3.0
	ALFA	27	1300	23	15	1.0×1.0×3.0
SWI	HKCU	20	17	24	-	0.45×0.45×1.0
	AIBL	57	27	20	9	0.93×0.93×1.75

Table 2. Assessment indicators used in the experiment.

TPR	PPV	FP_{avg}	$F1$	$DICE$
$\frac{TP}{TP+FN}$	$\frac{TP}{TP+FP}$	$\frac{TP}{N}$	$\frac{2 \cdot TPR \cdot PPV}{TPR+PPV}$	$\frac{2 \cdot TP}{2 \cdot TP+FP+FN}$

3. EXPERIMENTS

3.1. Settings

Datasets. Our study assessed the effectiveness of the proposed methods across publicly accessible datasets, spanning two distinct modalities MRI. The T2* datasets are derived from the MICCAI VALDO challenge [15], specifically from three distinct institutions: RSS, SABRE, and ALFA. Besides, the SWI datasets originate from the HKCU [6] and AIBL[24]. We provide the details of these datasets in Table 1.

Data Preprocessing. Our pre-processing pipeline involved skull removal [25], cropping, standardization, and SR. Besides, we incorporated data augmentation techniques, including random inversion, voxel volume swapping, and noise augmentation. Each noise generation probability was consistently set to 0.2 to simulate potential distortions encountered during MRI acquisition. The SWI datasets, which initially only provided 3D center-of-mass coordinates, were manually annotated to segmentation-level label via ITK-SNAP.

Evaluation Protocol. To guarantee a fair and comprehensive comparison, we adopted 5-fold cross-validation and presented the results at both the detection and segmentation levels, shown in table 2. In the comparative analysis, we categorized existing research into two distinct groups: state-of-the-art results on the dataset and conventional segmentation networks. For the latter, we treated them as discriminative networks within the same two-stage framework, with the parameters of the screening network (FCN) held constant. All models were conducted on an NVIDIA 3090Ti GPU (24G). In addition, voxel volumes with dimensions of 20×20×16 were utilized in the first stage for training. Subsequently, in the second stage, the voxel volumes were enlarged to dimensions of 48×48×24 to incorporate a broader contextual perspective.

Table 4. Ablation experiment under two MRI modalities.

Modality	2D	3D	2D+3D	FCM	KD	PLKD	TPR	PPV	FP_{avg}	F1	Dice
T2* (SABRE+ALFA)	✓						51.3	52.4	2.3	51.8	52.6
		✓					52.7	56.3	1.9	54.4	53.6
			✓				56.8	56.5	1.9	56.8	54.8
			✓	✓			57.8	64.1	1.4	60.8	57.3
			✓	✓	✓		62.4	65.3	1.4	63.8	60.9
			✓	✓		✓	**63.6**	**68.1**	**1.3**	**65.8**	**61.2**
SWI (AIBL)	✓						86.5	43.2	5.3	57.6	50.2
		✓					87.3	45.2	4.6	59.6	52.7
			✓				93.1	44.4	5.1	59.9	51.6
			✓	✓			95.7	48.4	4.3	64.3	51.9
			✓	✓	✓		98.8	54.8	2.9	67.4	59.1
			✓	✓		✓	**98.8**	**61.9**	**2.1**	**76.1**	**64.8**

Table 3. Comparative experiment under two MRI modalities.

Mod	Method	TPR	PPV	FP_{avg}	F1	Dice
T2*	nn-unet [26]	49.3	44.7	4.7	46.9	53.1
	UNet [19]	38.9	31.3	10.3	34.7	38.9
	TransUNet [27]	**75.6**	41.0	6.6	53.2	42.9
	UNet++ [28]	52.3	48.6	3.6	50.4	48.6
	Ours	63.6	**68.1**	**1.3**	**65.8**	**61.2**
SWI	Momeni [24]	95.0	59.1	-	72.9	-
	UNet	87.4	44.1	5.3	58.6	50.4
	TransUNet	86.2	44.5	5.1	58.7	51.5
	UNet++	94.1	48.3	4.3	63.8	53.4
	Ours	**98.8**	**61.9**	**2.1**	**76.1**	**64.8**

Table 5. The effects of loss weights on network performance.

Mod	α	β	TPR	PPV	FP_{avg}	F1	Dice
T2*	1.0	0	57.8	64.1	1.4	60.8	57.3
	0.9	0.1	58.4	65.9	1.4	61.9	62.9
	0.8	0.2	61.6	**68.3**	**1.3**	64.8	57.4
	0.7	0.3	**63.6**	68.1	**1.3**	**65.8**	**61.2**
	0.6	0.4	59.3	61.1	1.7	60.2	58.8
	0.5	0.5	58.8	57.5	2.3	58.1	58.7
SWI	1.0	0	95.7	48.4	4.3	64.3	51.9
	0.9	0.1	96.2	51.2	3.9	66.8	58.7
	0.8	0.2	97.1	54.1	3.9	69.5	63.4
	0.7	0.3	97.1	59.0	2.6	73.4	62.3
	0.6	0.4	**98.8**	**61.9**	**2.1**	**76.1**	**64.8**
	0.5	0.5	98.4	61.1	2.3	75.4	63.1

3.2. Results

Comparative Experiment. Based on the results presented in Table 3, the proposed approach demonstrated significant performance improvements compared to the state-of-the-art method (nn-unet). Specifically, the F1 metric showed a remarkable 18.9 improvement, the Dice metric experienced an 8.1 boost, and the average false positive count decreased by 3.4. Furthermore, our method outperformed the findings reported by Momeni[22] for the SWI dataset, achieving a noteworthy 3.2 enhancement in the F1 metric. Consequently, proposed method represents a substantial breakthrough in enhancing the segmentation performance for LR.

Ablation Experiment. As detailed in Table 4, the results highlighted the substantial enhancement in the performance engendered by proposed module, particularly with the incorporation of the PLKD. Our method exhibited noteworthy improvements in F1, with increments of 2.0 and 8.7, along with an elevation in Dice scores by 0.3 and 5.7 when contrasted with the KD, as observed in both the T2* and SWI datasets.

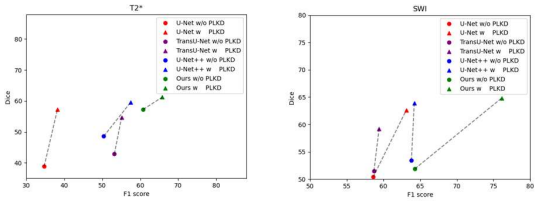

Fig. 6 Illustration of the performance improvement of the proposed PLKD for conventional segmentation networks.

Effect of Distillation Loss Weights on the Network. To examine the effect of distillation loss in Equ. (3), we systematically modulated the weight assigned to the loss function. The results shown in Table 5 reveal that the optimal performance is achieved within a range of 0.3 to 0.4 for two modalities.

Exploration of Progressive Distillation Across Other Networks. As shown in Fig. 6, we substantiated the versatility of the PLKD through a comparative analysis of the original results with those obtained after the application of PLKD. We had found that PLKD consistently improves the performance across the board, suggesting that PLKD serves as a beneficial augmentation to the existing learning method.

4. CONCLUSION

This paper has proposed a novel method of knowledge distillation based on progressively learning and an encoding-enhanced U-Net for automated segmentation of CMBs in LR MRI. The experimental results derived from publicly available datasets demonstrated the superior performance of our proposed method for the low-resolution segmentation task of small objects.

5. ACKNOWLEDGMENTS

This research work was partly supported by the Ningbo key research and development program (No.2023Z196), the Zhejiang Provincial Public Welfare Technology Research Project (No.LGF21F020008), and the Ningbo Municipal Public Welfare Technology Research Project (No.2022S134).

6. REFERENCES

[1] S. Geethanath and J. T. Vaughan Jr., "Accessible magnetic resonance imaging: a review," *Journal of Magnetic Resonance Imaging*, vol. 49, no. 7, pp. e65-e77, 2019.

[2] M. L. de Leeuw den Bouter, G. Ippolito, T. P. A. O'Reilly, et al., "Deep learning-based single image super-resolution for low-field MR brain images," *Scientific Reports*, vol. 12, no. 1, pp. 6362, 2022.

[3] Q. Lyu, H. Shan, C. Steber, et al., "Multi-contrast super-resolution MRI through a progressive network," *IEEE Transactions on Medical Imaging*, vol. 39, no. 9, pp. 2738-2749, 2020.

[4] S. Haller, M. W. Vernooij, J. P. A. Kuijer, et al., "Cerebral microbleeds: imaging and clinical significance," *Radiology*, vol. 287, no. 1, pp. 11-28, 2018.

[5] S. M. Greenberg, M. W. Vernooij, C. Cordonnier, et al., "Cerebral microbleeds: a guide to detection and interpretation," *The Lancet Neurology*, vol. 8, no. 2, pp. 165-174, 2009.

[6] Q. Dou, H. Chen, L. Yu, et al., "Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1182-1195, 2016.

[7] Z. Fang, R. Zhang, L. Guo, et al., "Knowledge-guided 2.5D CNN for cerebral microbleeds detection," *Biomedical Signal Processing and Control*, vol. 86, 105078, 2023.

[8] S. Liu, D. Utriainen, C. Chai, et al., "Cerebral microbleed detection using Susceptibility Weighted Imaging and deep learning," *NeuroImage*, vol. 198, pp. 271-282, 2019.

[9] K. Koschmieder, M. M. Paul, T. L. A. van den Heuvel, et al., "Automated detection of cerebral microbleeds via segmentation in susceptibility-weighted images of patients with traumatic brain injury," *NeuroImage*, vol. 35, 103027, 2022.

[10] Y. D. Zh ang, X. X. Hou, Y. D. Lv, et al., "Sparse Autoencoder based deep neural network for voxelwise detection of cerebral microbleed," *in ICPADS*, pp. 1229-1232, 2016.

[11] M. A. Al-masni, W.-R. Kim, E. Y. Kim, et al., "A Two Cascaded Network Integrating Regional-based YOLO and 3D-CNN for Cerebral Microbleeds Detection," *in EMBC*, pp. 1055-1058, 2020.

[12] Z. Chen, C. Li, J. He, et al., "A novel hybrid convolutional neural network for accurate organ segmentation in 3D head and neck CT images," *in MICCAI*, pp. 569-578, 2021.

[13] S. Farsiu, D. Robinson, M. Elad, et al., "Advances and challenges in super-resolution," *International Journal of Imaging Systems and Technology*, vol. 14, no. 2, pp. 47-57, 2004.

[14] B. Baeßler, K. Weiss, D.P. Dos Santos, "Robustness and reproducibility of radiomics in magnetic resonance imaging: a phantom study," *Investigative Radiology*, vol. 54, no. 4, pp. 221-228, 2019.

[15] C. H. Sudre, K. Van Wijnen, F. Dubost, et al., "Where is VALDO? Vascular lesions detection and segmentation challenge at MICCAI 2021," *arXiv preprint* arXiv:2208.07167, 2022.

[16] G. Hinton, O. Vinyals, J. Dean, "Distilling the knowledge in a neural network," *arXiv preprint* arXiv:1503.02531, 2015.

[17] M. Zhu, K. Han, C. Zhang, J. Lin, Y. Wang, "Low-resolution Visual Recognition via Deep Feature Distillation," *in ICASSP*, pp. 3762-3766, 2019.

[18] S. Ge, K. Zhang, H. Liu, Y. Hua, S. Zhao, X. Jin, H. Wen, "Look One and More: Distilling Hybrid Order Relational Knowledge for Cross-Resolution Image Recognition," *in AAAI*, vol. 34, no. 07, pp. 10845-10852, 2020.

[19] O. Ronneberger, P. Fischer, T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *in MICCAI*, pp. 234-241, 2015.

[20] S. Shin, J. Lee, J. Lee, et al., "Teaching where to look: Attention similarity knowledge distillation for low resolution face recognition," *in ECCV*, pp. 631-647, 2022.

[21] H. Wang, Y. Wang, Z. Zhou, et al., "Cosface: Large margin cosine loss for deep face recognition," *in CVPR*, pp. 5265-5274, 2018.

[22] L. Sun, S. Cheng, Y. Zheng, et al., "SPANet: Successive pooling attention network for semantic segmentation of remote sensing images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 4045-4057, 2022.

[23] C. H. Sudre, W. Li, T. Vercauteren, et al., "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," *in MICCAI*, pp. 240-248, 2017.

[24] S. Momeni, A. Fazlollahi, P. Yates, et al., "Synthetic microbleeds generation for classifier training without ground truth," *Computer Methods and Programs in Biomedicine,* vol. 207, pp. 106127, 2021.

[25] M. Jenkinson, C. F. Beckmann, T. E. J. Behrens, et al., "Fsl," *Neuroimage*, vol. 62, no. 2, pp. 782-790, 2012.

[26] F. Isensee, P. F. Jaeger, S. A. A. Kohl, et al., "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203-211, 2021.

[27] J. Chen, Y. Lu, Q. Yu, et al., "Transunet: Transformers make strong encoders for medical image segmentation," *arXiv preprint* arXiv:2102.04306, 2021.

[28] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, et al., "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1856-1867, 2019.