AFS-Net: Adaptive Fusion Skip Network for Pediatric Brain Tumor Segmentation

Hao Li[1[0009−0000−7268−5510]], Yanjun Peng[1[0000−0002−8444−0622]⋆]

[1]Shandong University of Science and Technology
lihao@sdust.edu.cn
yjpeng@sdust.edu.cn

Abstract. Accurate pediatric brain tumor segmentation from multi-parametric magnetic resonance imaging is essential for quantitative diagnosis, treatment planning, and longitudinal follow-up. However, the BraTS 2026 Pediatric Brain Tumor Segmentation (PED) task remains challenging because pediatric tumors present heterogeneous appearance, variable anatomical locations, small lesion subregions, and strong class imbalance. In this work, we propose AFS-Net, an Adaptive Fusion Skip network built on nnU-Net v2 for reliable pediatric brain tumor segmentation. Rather than introducing a separate prior-guided model, we focus on improving the nnU-Net baseline through channel recalibration (SE-style attention), a widened encoder, specificity-driven regularization, and adaptive skip fusion. These design choices are motivated by prior BraTS-PED nnU-Net improvements and recent work on dynamic skip connections in U-like architectures. AFS-Net achieves DSC scores of 0.538, 0.913, 0.247, 0.000, 0.941, and 0.941 for ET, NET, CC, ED, TC, and WT, respectively, and NSD scores of 0.448, 0.638, 0.250, 0.000, 0.657, and 0.658. These results suggest that carefully designed nnU-Net enhancements remain a strong and reliable strategy for the BraTS 2026 PED task.

Keywords: Pediatric Brain Tumor Segmentation · Channel Recalibration · Specificity Regularization · Adaptive Fusion Skip

1 Introduction

Pediatric brain tumor segmentation is clinically important because precise delineation of tumor subregions supports diagnosis, surgical planning, radiotherapy planning, and follow-up assessment. Compared with adult gliomas, pediatric brain tumors often show diverse imaging appearance, variable locations, and complex subregion morphology. These properties make automatic segmentation particularly difficult for small and low-contrast tumor components.

The BraTS 2026 Pediatric Brain Tumor Segmentation (PED) task provides multi-parametric MRI scans and expert annotations for pediatric tumor cases [8,

⋆ Corresponding authors.

7]. We use nnU-Net v2 [1] as the baseline framework because of its strong self-configuration ability and robust performance across biomedical segmentation tasks. Therefore, this paper focuses on the more reliable nnU-Net-based modifications that produced stronger validation performance. The benchmark implementation and evaluation workflow follow the challenge infrastructure described by the organizers and the MedPerf-based evaluation platform [9].

Our method is mainly inspired by two recent lines of work. First, the BraTS-PED solution of Li et al. [3] showed that task-specific nnU-Net modifications, including channel attention, widened encoders, and specificity-aware regularization, can improve pediatric tumor segmentation. Second, Cao et al. [4] demonstrated that dynamic skip connections can enhance U-like networks by adaptively improving feature fusion between encoder and decoder. Following these insights, we develop AFS-Net, an Adaptive Fusion Skip network that strengthens nnU-Net with task-specific feature extraction and feature fusion enhancements. We refer to the final model family as AFS-Net throughout the paper. The public GitHub repository is available at `https://github.com/lihaohao-w/AFS-Net`.

The main contributions are:

- We establish a nnU-Net v2 baseline for BraTS 2026 PED and analyze the effect of several task-specific enhancements.
- We propose channel recalibration and a widened encoder to strengthen feature representation.
- We propose specificity-driven regularization to reduce false-positive predictions for absent or sparse lesion classes.
- We propose adaptive fusion skip enhancement for improving encoder–decoder feature fusion and report the best validation performance among the tested variants.

2 Method

Figure 1 summarizes the proposed AFS-Net framework. The overall pipeline follows nnU-Net preprocessing, patch-based training, and sliding-window inference, while architectural and loss-level modifications are evaluated cumulatively on top of the baseline. The final row in Table 2 corresponds to the full AFS-Net configuration, which combines channel recalibration, widened encoder design, specificity regularization, and adaptive fusion skip. The implementation details of the AFS module are intentionally conservative so that the final model remains close to a reproducible nnU-Net-based pipeline.

2.1 Baseline nnU-Net

nnU-Net v2 automatically configures preprocessing, patch size, batch size, network topology, data augmentation, and inference strategy according to dataset properties [1]. For the current dataset plan, the 3D full-resolution configuration uses a patch size of [96, 160, 160] and a batch size of 2, while the 2D auxiliary

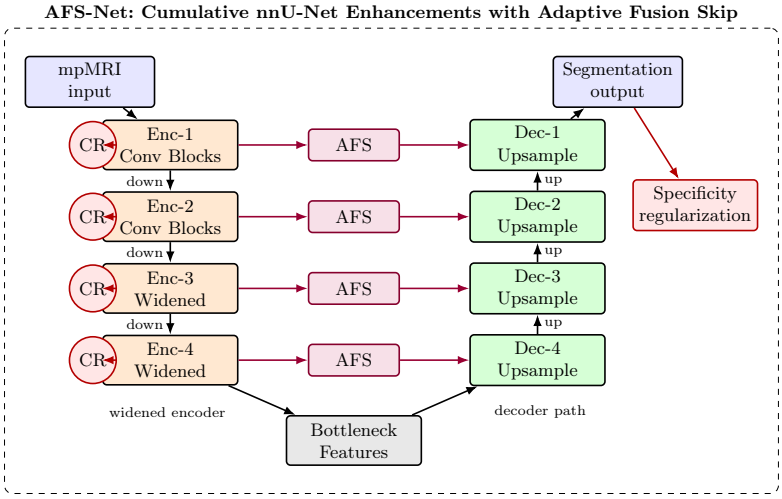

Fig. 1. Architecture of AFS-Net. The nnU-Net encoder–decoder is enhanced with channel recalibration, widened encoder stages, AFS modules for adaptive skip fusion, and specificity-driven regularization. The final AFS-Net model corresponds to the cumulative configuration shown in Table 2.

configuration uses [256, 224] and a batch size of 56. It uses an encoder–decoder U-Net architecture with skip connections, 3D convolutional blocks, instance normalization, and LeakyReLU activation. The default objective combines DSC and cross-entropy losses. This strong self-configuring behavior makes nnU-Net an appropriate baseline for pediatric brain tumor segmentation, where image spacing, tumor scale, and lesion composition vary substantially across patients.

2.2 Channel Recalibration

We evaluate channel recalibration (SE-style attention) in the nnU-Net architecture to emphasize tumor-relevant channels while suppressing less informative responses. Given a feature tensor F, global average pooling computes channel descriptors, which are then mapped to channel weights by a lightweight gating function. The recalibrated feature can be written as

$$\widetilde{F}_c = s_c F_c, \tag{1}$$

where s_c is the learned channel importance weight for channel c. We use this component as a direct plug-in to the encoder stages, following the BraTS-PED nnU-Net improvements in [3].

2.3 Widened Encoder

Pediatric tumor subregions may be small, heterogeneous, and difficult to distinguish from surrounding tissue. To increase the representational capacity

of the feature extractor, we evaluate a widened encoder variant. The encoder channels are increased relative to the baseline, allowing richer hierarchical representations before downsampling. The motivation is to strengthen low-level and mid-level feature extraction while retaining the nnU-Net decoding pathway for spatial reconstruction.

2.4 Specificity-Driven Regularization

Lesion-wise evaluation is sensitive to false positives, especially when a lesion class is absent in a case. This issue is common in pediatric brain tumor segmentation because not every case contains all annotated tumor subregions. Under lesion-wise DSC evaluation, even a small isolated false-positive region in an absent class can heavily penalize the score. Therefore, simply optimizing voxel-wise overlap may not sufficiently control false-positive predictions. The motivation for adding a specificity term is to reduce these spurious foreground activations without suppressing true-positive lesion voxels.

To mitigate this problem, we evaluate a specificity-driven regularization term inspired by [3]. Let p_c denote the predicted foreground probability for class c, and let g_c denote the corresponding binary ground-truth mask. The false-positive volume for class c can be approximated by

$$N_{\mathrm{FP}}^{(c)} = \sum_{v} p_c(v)\,(1 - g_c(v))\,, \tag{2}$$

where v indexes voxels. The predicted and ground-truth foreground volumes are

$$N_{\mathrm{pred}}^{(c)} = \sum_{v} p_c(v), \qquad N_{\mathrm{gt}}^{(c)} = \sum_{v} g_c(v). \tag{3}$$

The specificity penalty is then formulated as

$$\mathcal{L}_{\mathrm{spec}} = \frac{1}{C} \sum_{c=1}^{C} \frac{N_{\mathrm{FP}}^{(c)}}{N_{\mathrm{pred}}^{(c)} + N_{\mathrm{gt}}^{(c)} + \epsilon}, \tag{4}$$

where C is the number of foreground classes and ϵ is a numerical stabilizer. The final training objective is

$$\mathcal{L} = \mathcal{L}_{\mathrm{DSC+CE}} + \theta \mathcal{L}_{\mathrm{spec}}, \tag{5}$$

where θ controls the regularization strength. This term encourages conservative predictions for sparse or absent lesion classes while retaining the standard DSC and cross-entropy supervision for region overlap. In the final experiments, we set $\theta = 0.1$ and $\epsilon = 10^{-5}$, and we use these values consistently across all ablation rows.

2.5 Adaptive Fusion Skip

Traditional U-Net skip connections transmit encoder features to decoder stages using fixed pathways. Although this design effectively restores spatial details, it assumes that the same feature fusion strategy is suitable for every image and every anatomical configuration. This assumption can be restrictive for pediatric tumor segmentation, where lesion appearance, size, location, and boundary clarity vary greatly across patients. Static skip connections may pass irrelevant low-level details to the decoder or fail to emphasize informative features needed for small subregions.

Inspired by dynamic skip connections [4], our Adaptive Fusion Skip (AFS) module addresses these issues using adaptive multi-scale feature refinement and input-dependent feature recalibration. AFS-Net adopts AFS enhancement to improve encoder–decoder feature fusion while preserving the nnU-Net backbone.

Let X_l be the encoder feature at level l, and let Y_{l+1} be the upsampled decoder feature from the deeper level. A conventional U-Net decoder stage can be written as

$$Y_l = \mathcal{F}_{\mathrm{dec}}(Y_{l+1}, X_l), \tag{6}$$

where the skip feature X_l is directly concatenated or fused with decoder features. In contrast, AFS refines the encoder feature before fusion:

$$Y_l = \mathcal{F}_{\mathrm{dec}}(Y_{l+1}, \mathcal{F}_{\mathrm{AFS}}(X_l)). \tag{7}$$

Here, $\mathcal{F}_{\mathrm{AFS}}$ denotes the adaptive fusion skip transformation. Conceptually, this transformation contains two complementary operations. The first is adaptive multi-scale filtering, which allows the skip pathway to capture both local boundary details and broader contextual patterns. The second is input-dependent feature recalibration, which makes the skip representation more compatible with the decoder feature of the current case.

In our implementation, AFS is a plug-in modification rather than a redesign of the full nnU-Net architecture. This choice retains the strong preprocessing, training, and inference behavior of nnU-Net while specifically targeting the encoder–decoder fusion pathway, which is a critical information bottleneck in U-like networks. Let the encoder feature at stage l be $X_l \in R^{C_l \times H_l \times W_l \times D_l}$, and let the deepest bottleneck feature be X_b. The AFS block first compresses X_b with three parallel 3D convolution branches using kernel sizes 1^3, 3^3, and 5^3, respectively, followed by instance normalization and LeakyReLU. The three branch outputs are globally averaged to channel descriptors, concatenated, and converted into normalized gating weights by a lightweight $1 \times 1 \times 1$ convolution and softmax operation. The resulting weights re-scale the branch features before fusion, and the fused representation is projected back to the bottleneck channel width with a final $1 \times 1 \times 1$ convolution. The AFS module is inserted after each encoder stage before the corresponding decoder skip merge; in the current lightweight implementation, the bottleneck-only variant applies a single modulation block at the deepest feature level to reduce memory cost. The baseline feature widths follow the nnU-Net plan [32, 64, 128, 256, 320, 320], and

the widened-encoder variant increases the two deepest encoder stages to [32, 64, 128, 256, 384, 384]. The additional parameters are modest compared with the backbone because the AFS block operates only on the compressed bottleneck representation.

3 Training

All models are trained using the nnU-Net v2 pipeline. We use the same data split and preprocessing configuration for all variants to ensure fair comparison. The default nnU-Net augmentations are retained, including spatial transformations, intensity perturbations, Gaussian noise, Gaussian blur, gamma augmentation, low-resolution simulation, and mirroring, and these settings are applied identically to every ablation row. For the tested 3D full-resolution setting, nnU-Net uses a patch size of [96, 160, 160] and a batch size of 2, while the 2D auxiliary configuration uses [256, 224] and a batch size of 56 according to the dataset plan. The initial learning rate, optimizer, and scheduler follow the nnU-Net default setting unless otherwise specified; the default training schedule is 1000 epochs. We use a fixed hold-out validation split with 30 cases and report lesion-wise DSC on the official BraTS-PED regions ET, NET, CC, ED, TC, and WT.

Table 1. Training configuration used in our BraTS 2026 PED experiments.

Item	Setting
Framework	nnU-Net v2
Baseline	3D nnU-Net
Input	Multi-parametric pediatric brain MRI
Target regions	ET, NET, CC, ED, TC, WT
Training scheme	Patch-based 3D training
Optimizer	nnU-Net default SGD configuration
Initial learning rate	nnU-Net default setting
Loss	DSC + CE, optionally with specificity regularization
Specificity weight θ	Tuned on validation split
Stabilizer ϵ	Fixed small constant for numerical stability
Data augmentation	nnU-Net default 3D augmentation pipeline
Architectural variants	Channel recalibration, widened encoder, adaptive skip fusion
Deep supervision	nnU-Net default, except disabled when required by custom heads
Inference	Sliding-window inference with test-time mirroring enabled for all rows
Compile mode	Disabled for hardware compatibility
Submission	Docker container with selected model

4 Results

4.1 Quantitative Analysis

Table 2 reports the validation DSC and NSD scores of the evaluated variants. The ablation rows are cumulative: each successive row adds the listed modification on top of the previous configuration rather than evaluating the modification in isolation. The final row corresponds to the complete AFS-Net configuration, that is, the model obtained after adding adaptive fusion skip to the preceding cumulative setup. The baseline nnU-Net achieves strong TC and WT performance but lower ET and CC performance. Channel recalibration slightly improves ET and CC, while the widened encoder further improves ET. Specificity regularization gives a similar ET improvement. Among all tested variants, the final AFS-Net configuration achieves the best overall performance, with DSC scores of 0.538, 0.913, 0.247, 0.000, 0.941, and 0.941, respectively, and NSD scores of 0.448, 0.638, 0.250, 0.000, 0.657, and 0.658 for ET, NET, CC, ED, TC, and WT. The validation set contains 30 cases, and all rows are evaluated on the same split under the same preprocessing and augmentation pipeline.

Table 2. Validation DSC and NSD scores on BraTS 2026 PED. ET: enhancing tumor; NET: non-enhancing tumor; CC: cystic component; ED: edema; TC: tumor core; WT: whole tumor.

Model	DSC						NSD					
	ET	NET	CC	ED	TC	WT	ET	NET	CC	ED	TC	WT
nnU-Net baseline	0.509	0.909	0.207	0.000	0.937	0.937	0.419	0.642	0.211	0.000	**0.664**	**0.664**
+ CR attention	0.514	0.908	0.232	0.000	0.934	0.934	0.422	0.642	0.221	0.000	0.654	0.654
+ Widened encoder	0.527	0.899	0.232	0.000	0.927	0.927	0.426	**0.645**	0.227	0.000	0.661	0.662
+ Specificity regularization	0.528	0.904	0.212	0.000	0.931	0.932	0.426	0.638	0.200	**0.001**	0.653	0.654
AFS-Net (full model)	**0.538**	**0.913**	**0.247**	0.000	**0.941**	**0.941**	**0.448**	0.638	**0.250**	0.000	0.657	0.658

4.2 Qualitative Analysis

Figure 2 presents representative validation cases together with the corresponding AFS-Net predictions. The visual examples complement the quantitative DSC and NSD results in Table 2 by showing that AFS-Net produces plausible tumor segmentations across cases with different tumor appearances and spatial configurations. In particular, the predictions remain spatially coherent and visually consistent with the input anatomy, which supports the improvement observed in the quantitative results.

The final AFS-Net configuration corresponds to the complete cumulative model and therefore should be interpreted as the full method rather than as the isolated effect of the AFS block alone. Its strong ET performance and competitive results on the other regions make it the most suitable candidate for final model selection. The observed improvements are modest but consistent with

the intended effect of the proposed modifications. Further cross-validation and Docker validation remain necessary before final submission.

5 Discussion

The proposed nnU-Net-based framework provides a strong and reproducible baseline for the BraTS 2026 PED task. Across the validation split, the baseline already achieves competitive performance on the larger composite regions, especially TC and WT, while the smaller subregions ET and CC remain more difficult to segment. This pattern is consistent with the heterogeneous appearance and limited spatial extent of pediatric tumor subregions. Because all variants are evaluated on the same 30-case validation split with identical preprocessing and augmentation settings, the observed differences can be attributed to the proposed modifications rather than to changes in the experimental protocol.

Among the evaluated components, channel recalibration yields modest improvements on ET and CC, indicating that channel-wise reweighting can help emphasize tumor-relevant features. The widened encoder further improves ET, suggesting that additional representational capacity is beneficial for small and heterogeneous lesions. At the same time, these gains are accompanied by slight reductions on some composite regions, which reflects the usual trade-off between sensitivity to fine-grained structures and stability on larger tumor categories. This behavior underscores the importance of balancing local detail preservation with global region consistency in pediatric brain tumor segmentation.

Specificity-driven regularization is introduced to reduce false-positive predictions under lesion-wise evaluation. In the reported experiments, it improves ET but does not provide uniform gains across all classes. This indicates that the regularization term changes the sensitivity–specificity balance and should be tuned carefully with respect to the validation distribution. In particular, small structures such as ET and CC may benefit from reduced false positives, whereas overly aggressive suppression can also limit true-positive detections. The fixed values of θ and ϵ used in this study provide a stable setting for comparison, but class-specific tuning may be worth exploring in future work.

Adaptive fusion skip is the most effective component in the current study. By refining encoder features before skip fusion, AFS-Net achieves the best ET DSC and strong overall performance while remaining competitive on NET, TC, and WT. This is consistent with the general observation that fixed skip connections may be suboptimal when lesion appearance varies substantially across cases. In the present setting, adaptive skip refinement improves the compatibility between encoder and decoder features and appears to better preserve clinically relevant information in difficult regions.

The final AFS-Net configuration should therefore be interpreted as the complete cumulative model rather than as the isolated effect of a single block. Its performance indicates that combining moderate architectural changes with a specificity-aware objective can yield a useful improvement over the nnU-Net baseline without abandoning the stability of the underlying framework. Nev-

ertheless, the present study remains limited to a single validation split. Cross-validation, official leaderboard evaluation, and additional uncertainty analysis are still required to establish the robustness of the findings and to support the final challenge submission.

6 Conclusion

We propose an enhanced nnU-Net framework for the BraTS 2026 Pediatric Brain Tumor Segmentation task. Starting from the nnU-Net baseline, we evaluate channel recalibration, a widened encoder, specificity-driven regularization, and adaptive fusion skip enhancement. Validation results show that adaptive fusion skip achieves the best ET DSC and competitive overall performance, while other modifications provide class-dependent gains. These findings guide our final model selection for Docker-based challenge submission. The selected submission model is the best-performing validated variant under the current training and inference budget.

Acknowledgements This work was supported by the BraTS 2026 Pediatric Brain Tumor Segmentation challenge resources. The authors thank the organizers for providing the benchmark and evaluation platform. Data used in this publication were obtained as part of the Challenge project through Synapse ID syn74274097. We also acknowledge the MedPerf evaluation platform and the related challenge infrastructure. We further thank the BraTS-PED organizers and the authors of the referenced challenge materials for making the benchmark and reporting requirements available.

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

References

1. 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**, 203–211 (2021)
2. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention, pp. 234–241. Springer (2015)
3. Li, X., Xu, Z.Q.J., Ren, Y., Qiu, T., Wang, X.: Towards reliable pediatric brain tumor segmentation: task-specific nnU-Net enhancements. arXiv preprint arXiv:2511.00449 (2025)
4. Cao, Y., He, Q., Wang, K., Xiong, J., Yi, Z., He, T.: Enhancing feature fusion of U-like networks with dynamic skip connections. Medical Image Analysis **110**, 104010 (2026)
5. Hu, J., Shen, L., Sun, G.: Squeeze-and-Excitation Networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018)

6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
7. Kazerooni, A.F., et al.: The Brain Tumor Segmentation (BraTS) Challenge 2023: Focus on Pediatrics (CBTN-CONNECT-DIPGR-ASNR-MICCAI BraTS-PEDs). arXiv preprint arXiv:2305.17033 (2023)
8. Kazerooni, A.F., et al.: The Brain Tumor Segmentation in Pediatrics Challenge: Focus on Pediatrics. arXiv preprint arXiv:2404.15009 (2024)
9. Karargyris, A., Umeton, R., Sheller, M.J., et al.: Medperf: open benchmarking platform for medical artificial intelligence using federated evaluation. arXiv preprint arXiv:2110.01406 (2021)
10. BraTS 2026 Challenge Organizers: BraTS 2026 Pediatric Brain Tumor Segmentation challenge instructions and data usage agreement. Challenge documentation (2026)
11. Sun, Y., Wang, X., Liu, Z., Miller, J., Efros, A., Hardt, M.: Test-time training with self-supervision for generalization under distribution shifts. In: International Conference on Machine Learning, pp. 9229–9248 (2020)
12. Sun, Y., Li, X., Dalal, K., Xu, J., Vikram, A., Zhang, G., Dubois, Y., Chen, X., Wang, X., Koyejo, S., et al.: Learning to (learn at test time): RNNs with expressive hidden states. arXiv preprint arXiv:2407.04620 (2024)
13. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B.: Attention U-Net: learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)

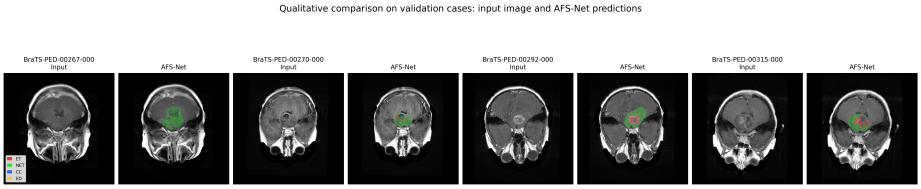

Fig. 2. Qualitative comparison on representative validation cases. For each case, the input MRI slice is shown together with the corresponding AFS-Net prediction. The visualizations provide visual support for the quantitative DSC and NSD results in Table 2.