

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/395471935>

Diffusion-Based CT Image Segmentation for Intracerebral Hemorrhage

Conference Paper · September 2025

DOI: 10.1109/CIP55913.2025.11084347

CITATIONS

0

READS

9

5 authors, including:



Zhuangzhuang Gu

University of South Carolina

14 PUBLICATIONS 16 CITATIONS

SEE PROFILE

DIFFUSION-BASED CT IMAGE SEGMENTATION FOR INTRACEREBRAL HEMORRHAGE

Lili Wang, Pingping Cai, Zhuangzhuang Gu, Srihari Nelakuditi, Yan Tong

University of South Carolina
Molinaroli College of Engineering and Computing
South Carolina, USA

ABSTRACT

Intracerebral hemorrhage (ICH) is a life-threatening stroke that requires precise segmentation for effective treatment. To aid in diagnosing ICH, various deep learning-based methods have been proposed. However, these methods face challenges posed by irregular patterns and low-contrast boundaries in ICH images. In this paper, we introduce a novel conditional diffusion-based segmentation approach for ICH segmentation. Our framework leverages ResNet18 and a Transformer Block to enhance conditional feature extraction, as well as cross-attention mechanisms to align global and local features between the conditional and diffusion branches. Experiment results on the Instance2022 dataset show an improvement exceeding 10% in the Dice coefficient and IoU score, compared to the baseline diffusion-based model, MedSegDiffv2. More impressively, more than 20% improvement is achieved in terms of both the Dice coefficient and IoU score in cross-dataset evaluations on two benchmarks, i.e., the BHSD and PhysioNet, demonstrating excellent generalizability to unseen data.

Index Terms— Deep learning, Diffusion model, Medical image segmentation, Intracerebral hemorrhage (ICH)

1. INTRODUCTION

Intracerebral hemorrhage (ICH) is a type of disease that occurs when a blood vessel in the brain bursts. It affects more than 80,000 individuals annually in the United States, posing a high risk of mortality and causing significant functional disabilities [1]. Computed tomography (CT) imaging is a widely used tool for ICH diagnosis, but it still relies on subspecialty-trained neuroradiologists, which can delay timely treatment, especially in resource-limited settings [2]. Therefore, various segmentation algorithms have been proposed to automate ICH detection. However, they still face challenges due to CT variability, noise, artifacts, and hemorrhage heterogeneity [3].

Deep learning techniques have significantly advanced medical image segmentation by enabling pixel-level identification of complex patterns in CT images. Methods such as convolutional neural networks (CNNs) [4] have demonstrated success in bleeding area localization and boundary

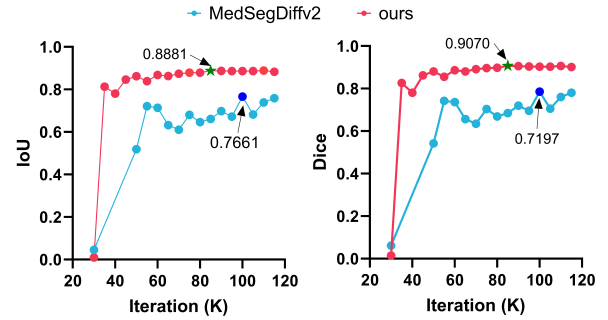


Fig. 1. Training curves comparing our method with MedSegDiffv2 in terms of IoU (left) and Dice (right). Our approach achieves faster convergence and more stable training dynamics on the **Instance2022** [7], reaching higher performance with fewer iterations.

delineation. However, manual annotations by radiologists are often prone to inter-observer variability, leading to training data uncertainty and inconsistent model performance [5]. To address these challenges, generative adversarial networks (GANs)[6] have been proposed as a solution for mitigating data limitations. By synthesizing realistic medical images and segmentation masks, GANs augment datasets with greater variability, simulate underrepresented classes, and reduce the reliance on extensive manual annotations. While these approaches have shown promise, challenges such as model collapse, synthetic-to-real gaps, and training instability limit their clinical utility.

More recently, diffusion-based segmentation frameworks [8] have emerged as an alternative for medical image segmentation. These methods utilize the input image as a condition for generating the segmentation mask in the diffusion framework, enabling precise boundary recovery and better alignment of model outputs with ground truth, making them particularly effective for medical image segmentation. However, for the ICH image segmentation task, as illustrated in Fig. 1, existing diffusion models often suffer from slow convergence. In addition, their generalizability remains limited, as evidenced by the consistently low performance observed in

Table 1 when evaluated on unseen datasets.

To address these challenges, we propose a diffusion-based segmentation framework tailored for ICH. Inspired by MedSegDiffv2 [9], we enhance ICH segmentation by introducing a conditional branch with ResNet18, leveraging residual connections and pre-trained weights for robust feature extraction. A Transformer Block is further included to capture long-range dependencies and improves contextual awareness. To enhance generalization, cross-attention mechanisms is introduced to align global and local features, improving accuracy and reducing false positives. These modifications address key limitations of diffusion models, leading to better boundary delineation, faster convergence, and greater generalizability. Experiment results on Instance2022 [7], BHSD [10], and PhysioNet [11] demonstrate the effectiveness our method. The main contributions of this paper are summarized as follows:

- We propose a new diffusion-based framework for ICH segmentation, integrating a newly-designed conditional feature extractor and cross-attention mechanisms to improve feature alignment and segmentation accuracy.
- Our method significantly accelerates convergence and enhances training stability compared to existing diffusion-based models, reducing computational costs while maintaining high performance.
- We evaluate our method on multiple datasets to demonstrate its segmentation accuracy and robust generalization capabilities.

2. RELATED WORKS

2.1. CNN-Transformer Hybrid Approaches for ICH Segmentation

Recent advances have seen the rise of hybrid architectures that integrate CNNs for local feature extraction with Transformers for global context modeling in ICH segmentation. For instance, TransHarDNet [12] enhances the U-Net architecture with Transformer blocks to effectively model long-range dependencies while maintaining computational efficiency, thereby improving both segmentation accuracy and inference speed. Similarly, STHarDNet [13] combines HarDNet with a Swin-Transformer, achieving superior performance on both CT and MRI datasets. In addition, another hybrid model [14] leverages CNNs enhanced with attention mechanisms and Atrous Spatial Pyramid Pooling (ASPP) to better delineate lesion boundaries. Building on these advancements, we incorporate a Transformer block into our diffusion-based framework, aiming to enhance global feature modeling without incurring significant additional computational overhead.

2.2. Diffusion-based Medical Image Segmentation

Diffusion models have recently emerged as a promising approach for medical image segmentation, particularly excelling in generating robust segmentation masks in noisy and low-contrast imaging scenarios. For example, brainSPADE [8] utilizes a latent diffusion model within a Variational Autoencoder (VAE) pipeline to generate synthetic brain MRI images and their corresponding segmentation masks, thereby enriching dataset diversity. Similarly, IISE [15] adapts the diffusion process for lesion segmentation by iteratively refining predictions and estimating uncertainty, which improves segmentation reliability. MedSegDiffv2 [9] further advances diffusion-based segmentation through innovations such as dynamic conditional encoding and frequency-based feature enhancements, demonstrating improved performance across various imaging modalities. However, despite these successes, diffusion models remain underexplored in the context of ICH segmentation. In this work, we address this gap by integrating Transformer-based enhancements into our diffusion framework, aiming to improve accuracy and generalization in challenging ICH segmentation tasks.

3. METHODS

Given an input CT image $I \in \mathbb{R}^{H \times W}$, the goal of ICH segmentation is to accurately identify the hemorrhage regions and produce a corresponding segmentation mask $X_0 \in \mathbb{R}^{H \times W}$. Diffusion-based methods iteratively refine an initial noisy image X_T through multiple denoising steps to approximate X_0 . This iterative refinement makes diffusion models particularly well-suited for capturing complex and heterogeneous structures often encountered in CT images.

To improve both the stability and generalizability of current diffusion-based methods, we propose a novel segmentation framework comprising two key components. A **Condition Branch** that leverages a ResNet18-based feature extractor and an integrated Transformer Block for enhanced multi-scale and global feature extraction. A **Diffusion Branch** that performs iterative denoising with a diffusion U-Net architecture, augmented by cross-attention mechanisms to effectively align features from the Condition Branch.

3.1. Overview

As illustrated in Fig. 2 the Condition Branch first extracts multi-scale hierarchical features using ResNet18, leveraging residual connections and pre-trained weights to ensure robust feature extraction, even with limited training samples. To further enhance contextual understanding, a Transformer Block is integrated to capture long-range dependencies, thus improving the global feature representation.

The Diffusion Branch then performs iterative denoising to reconstruct the segmentation map. To achieve optimal feature

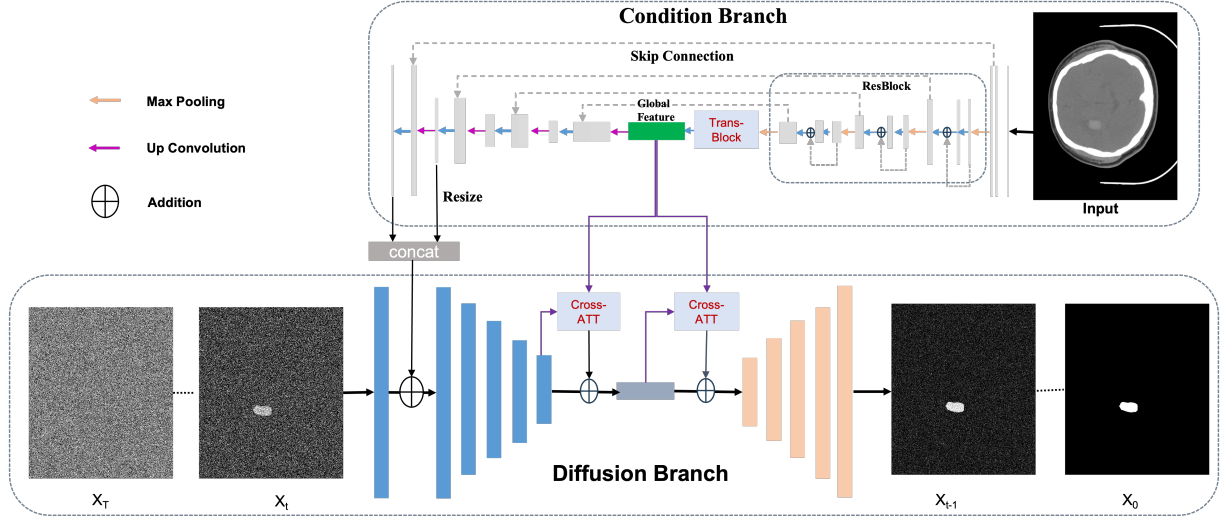


Fig. 2. The proposed diffusion-based segmentation framework for ICH. The architecture integrates a ResNet18 and a Transformer Block for conditional feature extraction, and a diffusion U-Net for iterative refinement, with cross-attention mechanisms enhancing feature alignment.

alignment between these two branches, we incorporate **cross-attention mechanisms** that facilitate the fusion of global and local information, leading to more precise segmentation.

3.2. Condition Branch

Unlike MedSegDiffv2 [9], which employs a U-Net trained from scratch, our framework uses a ResNet18-based feature extractor to improve efficiency and convergence, while leveraging pre-trained weights to alleviate the challenges of limited training data.

The Condition Branch starts with a 3×3 convolution that produces 32 low-level feature maps from the input CT image, followed by an encoder comprising the first three stages of ResNet18 (omitting its final classification stage). A Transformer Block—consisting of a multi-head self-attention layer and a feed-forward network with layer normalization and residual connections—is then applied to capture long-range dependencies. Notably, the pre-trained ResNet18 provides robust initialization in data-scarce scenarios, and the Transformer further enhances global feature modeling. Finally, up-convolution blocks upscale the features to the resolution required by the Diffusion Branch. The refined Transformer features serve as keys and values in the cross-attention mechanism, effectively fusing global and local information.

3.3. Diffusion Branch and Cross-Attention Integration

The diffusion process in our framework is composed of a forward and a reverse process.

Forward Process: Starting from the ground truth seg-

mentation X_0 , noise is progressively added over T steps:

$$q(X_t|X_{t-1}) = \mathcal{N}(X_t; \sqrt{\alpha_t}X_{t-1}, (1 - \alpha_t)\mathbf{I}), \quad (1)$$

where α_t controls the noise level at each step.

Reverse Process: In the reverse process, the model denoises X_T step by step to recover X_0 , conditioned on the features c extracted by the Condition Branch:

$$p_\theta(X_{t-1}|X_t, c) = \mathcal{N}(X_{t-1}; \mu_\theta(X_t, c, t), \Sigma_\theta(X_t, c, t)). \quad (2)$$

Cross-Attention Integration: To facilitate effective fusion between the Condition Branch and the Diffusion Branch, we employ cross-attention mechanisms. The attention operation is defined as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V, \quad (3)$$

where Q (queries) are derived from the Diffusion Branch, and K (keys) and V (values) come from the Transformer-enhanced features of the Condition Branch. Here, d_k denotes the dimensionality of the key vectors.

We apply the cross-attention mechanism at two critical stages to integrate global and local features, thereby enhancing segmentation accuracy:

Bottleneck Stage: The refined features from the Transformer Block, serving as keys and values, are aligned with the downsampled outputs (queries) from the Diffusion Branch before entering the bottleneck. This integration fuses global context with local details.

Upsampling Stage: During upsampling, the keys and values from the Transformer Block are further fused with the Diffusion Branch's features, enhancing high-resolution reconstruction and leading to more precise segmentation.

3.4. Loss Function

The proposed framework adopts the loss function from MedSegDiffv2 [9], combining the noise prediction loss L_n for training the diffusion model with an anchor loss L_{anc} to supervise the condition model. The anchor loss is defined as $L_{anc} = L_{dice} + \beta L_{ce}$, where L_{dice} and L_{ce} represent the soft Dice loss and cross-entropy loss, respectively, to enhance segmentation supervision. The total loss function is:

$$L_{total}^t = L_n^t + (t \equiv 0 \pmod{\alpha}) L_{anc}, \quad (4)$$

with $\alpha = 5$ controlling the frequency of supervision and $\beta = 10$ ensuring consistency with MedSegDiffv2 in our experiments.

4. EXPERIMENTS

4.1. Datasets and Evaluation Metrics

Datasets: The **Instance2022** [7] is a challenge dataset with annotated 3D CT volumes from 100 patients for ICH segmentation. The dataset is split into training (90%) and testing (10%) subsets at the patient level to prevent data leakage. Each 3D volume is converted into 2D slices along the axial plane, resized to 256×256 , and normalized. The training subset initially contained 2,687 slices with a significant imbalance between positive and negative samples. To address this, we downsampled the negative samples to achieve a 1:1 positive-to-negative ratio, resulting in 1,381 slices (691 positive and 690 negative) for training. The testing subset remains unchanged with 299 slices to simulate real-world conditions. The **BHSD** dataset [10] contains 192 labeled 3D volumes, from which 10% of patients were randomly sampled to yield 681 2D slices. Similarly, the **PhysioNet** dataset [11] comprises 82 labeled 3D volumes, with 10% sampling producing 346 2D slices. Both datasets are used solely for testing to evaluate our method’s generalizability. This 10% sampling balances computational efficiency with statistical representativeness, reducing inference time and memory usage.

Evaluation Metrics: Segmentation performance is measured using the **Dice similarity coefficient** and **Intersection-over-Union (IoU)** [3]. Binary masks are obtained by thresholding predicted probability maps at 0.1, 0.3, 0.5, 0.7, and 0.9. The final results are averaged across these thresholds for robust evaluation [9].

4.2. Implementation Details

The framework is implemented in PyTorch and trained on four NVIDIA RTX A6000 GPUs. The AdamW optimizer is used with an initial learning rate of 1×10^{-4} , weight decay for regularization, and the exponential moving average for stable training. The model is trained on the Instance2022 [7] training subset for 115,000 iterations with a batch size of 32, using a linearly decaying learning rate.

Inference: The trained model is evaluated on the **Instance2022** [7] testing subset, as well as the **BHSD** [10] and **PhysioNet** [11] datasets. Inference on both the **BHSD** [10] and **PhysioNet** [11] datasets is performed using the model trained solely on **Instance2022** [7] without additional fine-tuning, to assess its generalization to unseen data.

To evaluate the performance of our proposed method, we compare it with both previous and current state-of-the-art (SOTA) models, covering CNN-based and diffusion-based approaches. Specifically, we include **U-Net** [16], a widely used CNN-based segmentation model; **TransU-Net** [17], a hybrid model integrating Transformers with CNNs; **nnU-Net** [18], a self-adapting framework considered the SOTA among CNN-based models; **U-KAN** [19], a U-Net variant leveraging Kolmogorov-Arnold Networks (KANs) but prone to high false positives; and **MedSegDiffv2** [9], a diffusion-based framework representing the latest advances in generative segmentation.

Table 1 presents the segmentation performance on three benchmark datasets. On **Instance2022**, our model achieves 90.70% Dice and 88.81% IoU, outperforming MedSegDiffv2 (78.52% Dice, 76.60% IoU), confirming its superiority in diffusion-based segmentation. Furthermore, it performs comparably to the CNN-based SOTA, nnU-Net [18] (90.83% Dice, 88.98% IoU), while benefiting from the generative capabilities of the diffusion framework.

In cross-dataset evaluations, our model exhibits stronger generalization than competing methods. On **BHSD** [10], it achieved 75.20% Dice and 73.58% IoU, outperforming all baselines, including nnU-Net (71.96% Dice, 70.64% IoU) and MedSegDiffv2 (50.64% Dice, 49.35% IoU). Similarly, on **PhysioNet** [11], our model attained 83.67% Dice and 82.39% IoU, substantially exceeding MedSegDiffv2 (60.39% Dice, 60.10% IoU). Even compared to nnU-Net (88.82% Dice, 87.46% IoU), our approach demonstrates competitive performance with enhanced robustness across datasets. These results confirm our model’s state-of-the-art segmentation accuracy and superior generalization across both in-domain and out-of-distribution scenarios.

Fig. 3 illustrates representative segmentation results on the three benchmark datasets. Columns display original CT images, ground truth masks, predictions from our method and the SOTA methods in comparison. Note improvements in boundary accuracy and reduced false positives.

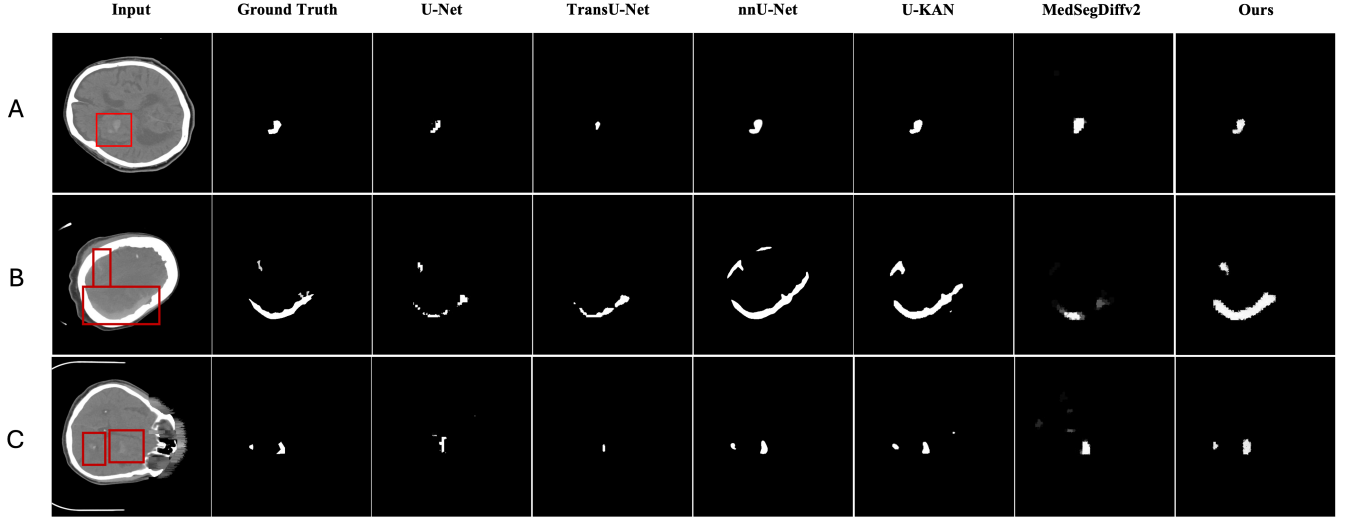
4.3. Ablation Study

Effectiveness of Architecture Design. We begin with the original MedSegDiffv2 [9] framework, which consists of a diffusion U-Net and a conditional U-Net. Our modifications enhance the conditional branch, as quantified in Table 2. Integrating cross-attention significantly improves Dice from 78.52% to 85.66%, demonstrating better feature alignment. Replacing the conditional U-Net with a ResNet18-based

Table 1. Comparison of ICH Segmentation Performance (Dice and IoU are reported as percentages %).

	U-Net [16]		TransU-Net [17]		nnU-Net [18]		U-KAN [19]		MedSegDiffv2 [9]		Ours	
Dataset	IoU↑	Dice↑	IoU↑	Dice↑	IoU↑	Dice↑	IoU↑	Dice↑	IoU↑	Dice↑	IoU↑	Dice↑
Instance2022 [7]	76.92	79.01	79.73	81.82	88.98	90.83	69.79	71.99	76.60	78.52	88.81	90.70
BHSD* [10]	56.54	58.01	59.10	60.67	70.64	71.96	43.35	45.37	54.83	55.58	73.58	75.20
PhysioNet* [11]	58.85	60.21	58.75	59.74	87.46	88.82	50.46	52.15	60.10	60.39	82.39	83.67

* Inference on the BHSD [10] and PhysioNet [11] datasets is performed using the model trained exclusively on Instance2022 [7], without additional fine-tuning, to evaluate the generalization capability on unseen datasets.

**Fig. 3.** Segmentation results on (A) Instance2022 [7], (B) BHSD [10] and (C) PhysioNet [11]. Improvements in boundary accuracy and reduced false positives are highlighted in red boxes.**Table 2.** Ablation Study on Instance2022 [7] Dataset.

Model Variant	Dice (%)	IoU (%)	Best Iteration
Baseline (MedSegDiffv2 [9])	78.52	76.60	100K
+ Cross-Attention	85.66	83.84	55K
+ Reduce Layer	88.45	86.21	60K
+ ResNet18 Features	88.15	86.94	90K
+ Transformer Block (Full Model)	90.70	88.81	85K

extractor further boosts Dice to 88.15% by leveraging pre-trained weights and residual connections. Finally, incorporating a Transformer Block enhances long-range dependencies, achieving 90.70% Dice, confirming the effectiveness of our full model.

Stable Convergence and Efficiency. A key advantage of our method is its faster and more stable convergence. As shown in Fig. 1, our model reaches near-optimal performance by 35K iterations (82.59% Dice, 81.32% IoU), whereas MedSegDiffv2 [9] fluctuates significantly and stabilizes only after 55K iterations. By 85K iterations, our full model surpasses the baseline’s final performance (78.52% Dice, 76.60% IoU at 100K), highlighting its efficiency and robustness.

5. CONCLUSION

In this study, we present a novel diffusion-based segmentation framework for ICH detection. By integrating a ResNet18-based feature extractor, a Transformer block for enhanced conditional features, and cross-attention modules, our approach improves both training stability and segmentation accuracy. Extensive experiments on the Instance2022 [7] dataset, along with cross-dataset evaluations on BHSD [10] and PhysioNet [11], demonstrate significant improvements over the baseline MedSegDiffv2 [9] on unseen data.

Strengths and Limitations. Our method benefits from a streamlined architecture that enables efficient training and high accuracy, reducing false positives and enhancing recall—a crucial factor in clinical settings. However, processing solely on 2D slices limits the capture of 3D spatial context, which is important for segmenting complex structures. Future work will focus on integrating 3D spatial information and optimizing for high-resolution inputs, thereby balancing simplicity with comprehensive spatial awareness.

6. REFERENCES

- [1] Connie W Tsao, Aaron W Aday, Zaid I Almarzooq, Cheryl AM Anderson, Pankaj Arora, et al., “Heart disease and stroke statistics—2023 update: a report from the american heart association,” *Circulation*, vol. 147, no. 8, pp. e93–e621, 2023.
- [2] Chelsea S Kidwell, Julio A Chalela, Jeffrey L Saver, Sidney Starkman, Michael D Hill, Andrew M Demchuk, John A Butman, et al., “Comparison of mri and ct for detection of acute intracerebral hemorrhage,” *Jama*, vol. 292, no. 15, pp. 1823–1830, 2004.
- [3] Xinyi Shi, Hanguang Xiao, Diyou Chen, and Yujia Wei, “Medical image segmentation of intracranial hemorrhage: A review,” in *Asian Conference on Artificial Intelligence Technology (ACAIT)*. IEEE, 2023, pp. 790–801.
- [4] Baris Kayalibay, Grady Jensen, and Patrick van der Smagt, “Cnn-based segmentation of medical imaging data,” *arXiv preprint arXiv:1701.03056*, 2017.
- [5] Kai Hu, Kai Chen, Xizhi He, Yuan Zhang, Zhineng Chen, Xuanya Li, and Xieping Gao, “Automatic segmentation of intracerebral hemorrhage in ct images using encoder–decoder convolutional neural network,” *Information Processing & Management*, vol. 57, no. 6, pp. 102352, 2020.
- [6] Siyi Xun, Dengwang Li, Hui Zhu, Min Chen, Jianbo Wang, Jie Li, Meirong Chen, Bing Wu, Hua Zhang, Xi-angfei Chai, et al., “Generative adversarial networks in medical image segmentation: A review,” *Computers in Biology and Medicine*, vol. 140, pp. 105063, 2022.
- [7] Xiangyu Li, Gongning Luo, Kuanquan Wang, Hongyu Wang, Jun Liu, Xinjie Liang, Jie Jiang, Zhenghao Song, Chunyue Zheng, Haokai Chi, et al., “The state-of-the-art 3d anisotropic intracranial hemorrhage segmentation on non-contrast head ct: The instance challenge,” *arXiv preprint arXiv:2301.03281*, 2023.
- [8] Amirhossein Kazerouni, Ehsan Khodapanah Aghdam, Moein Heidari, Reza Azad, Mohsen Fayyaz, Ilker Hacıhaliloglu, and Dorit Merhof, “Diffusion models in medical imaging: A comprehensive survey,” *Medical Image Analysis*, vol. 88, pp. 102846, 2023.
- [9] Junde Wu, Wei Ji, Huazhu Fu, Min Xu, Yueming Jin, and Yanwu Xu, “Medsegdiff-v2: Diffusion-based medical image segmentation with transformer,” in *AAAI*, 2024, vol. 38(6), pp. 6030–6038.
- [10] Biao Wu, Yutong Xie, Zeyu Zhang, Jinchao Ge, Kaspar Yaxley, Suzan Bahadır, Qi Wu, Yifan Liu, and Minh-Son To, “Bhsd: A 3d multi-class brain hemorrhage segmentation dataset,” in *International Workshop on Machine Learning in Medical Imaging*, 2023.
- [11] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, et al., “Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals,” *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [12] Zhigao Piao, Yeong Hyeon Gu, Hailin Jin, and Seong Joon Yoo, “Intracerebral hemorrhage ct scan image segmentation with hardnet based transformer,” *Scientific Reports*, vol. 13, no. 1, pp. 7208, 2023.
- [13] Zhigao Piao, Yeong Hyeon Gu, Seong Joon Yoo, and Myoung-ho Seong, “Segmentation of cerebral hemorrhage ct images using swin transformer and hardnet,” in *International Conference on Information Networking*. IEEE, 2023.
- [14] Pascal Spiegler, Amirhossein Rasoulzadeh, and Yiming Xiao, “Weakly supervised intracranial hemorrhage segmentation with yolo and an uncertainty rectified segment anything model,” *arXiv preprint arXiv:2407.20461*, 2024.
- [15] Julia Wolleb, Robin Sandkühler, Florentin Bieder, Philippe Valmaggia, and Philippe C Cattin, “Diffusion models for implicit image segmentation ensembles,” in *International Conference on Medical Imaging with Deep Learning*, 2022.
- [16] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *MICCAI*. Springer, 2015, pp. 234–241.
- [17] Jieneng Chen, Jieru Mei, Xianhang Li, Yongyi Lu, Qihang Yu, Qingyue Wei, Xiangde Luo, Yutong Xie, Ehsan Adeli, Yan Wang, et al., “Transunet: Rethinking the u-net architecture design for medical image segmentation through the lens of transformers,” *Medical Image Analysis*, vol. 97, pp. 103280, 2024.
- [18] Fabian Isensee, Tassilo Wald, Constantin Ulrich, Michael Baumgartner, Saikat Roy, Klaus Maier-Hein, and Paul F Jaeger, “Nnu-net revisited: A call for rigorous validation in 3d medical image segmentation,” in *MICCAI*, 2024.
- [19] Chenxin Li, Xinyu Liu, Wuyang Li, Cheng Wang, Hengyu Liu, Yifan Liu, Zhen Chen, and Yixuan Yuan, “U-kan makes strong backbone for medical image segmentation and generation,” *arXiv preprint arXiv:2406.02918*, 2024.