Pattern Recognition Letters 200 (2026) 37–43

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

TriGAN-SiaMT: A triple-segmentor adversarial network with bounding box priors for semi-supervised brain lesion segmentation

Mohammad Alshurbaji [a], Maregu Assefa [b], Ahmad Obeid [a], Mohamed L. Seghier [c], Taimur Hassan [d], Kamal Taha [a], Naoufel Werghi [a,*]

[a] Department of Computer Science, Khalifa University of Science and Technology, Abu Dhabi, UAE
[b] Department of Computer Science, C2PS, Khalifa University of Science and Technology, Abu Dhabi, UAE
[c] Department of Biomedical Engineering and Biotechnology, Khalifa University of Science and Technology, Abu Dhabi, UAE
[d] Department of Electrical & Computer Engineering, Abu Dhabi University, Abu Dhabi, UAE

ARTICLE INFO

Keywords:
Brain lesion segmentation
Deep learning
Semi-supervised learning
Siamese
Mean-Teacher

ABSTRACT

Accurate brain lesion segmentation in MRI is critical for clinical decision-making, but pixel-wise annotations remain costly and time-consuming. We propose TriGAN-SiaMT, a novel semi-supervised segmentation framework that combines adversarial learning, consistency regularization, and bounding box priors. Our architecture comprises three segmentors (S_0, S_1, S_2) and two discriminators (D_0, D_1). It includes: (1) a supervised branch ($S_0 \leftrightarrow D_0$) trained on a small labeled subset; (2) a Siamese branch ($S_1 \leftrightarrow D_1$) with an identical architecture to $S_0 \leftrightarrow D_0$, but trained on unlabeled data; and (3) a teacher branch (S_2) updated via exponential moving average (EMA) from S_1, following the Mean Teacher (MT) paradigm. The teacher S_2 generates pseudo-labels to supervise S_1. It also provides soft segmentations to guide D_1, which does not see any labeled data. The model enforces consistency at multiple levels: between S_0 and S_1 (Siamese consistency), and between S_1 and S_2 (EMA consistency). Bounding box priors are incorporated as weak supervision for both labeled and unlabeled images, improving lesion localization. Evaluated on the ISLES 2022 and BraTS 2019 datasets, TriGAN-SiaMT achieves DSC scores of 84.80 % and 86.32 %, respectively, using only 5 % labeled data. These results demonstrate strong performance under limited supervision and robust generalization across brain lesions.

1. Introduction

Stroke and brain tumors are among the most common causes of neurological disability and mortality worldwide. Brain lesion segmentation plays a pivotal role in assessing damage severity and guiding timely clinical interventions [1–3]. Despite recent advances in deep learning for medical image segmentation, the development of robust models remains heavily constrained by the limited availability of high-quality annotated data and the costly pixel-level annotation. This implies that reducing labeling effort enhances the practicality of deep learning in clinical workflows [4–6].

Semi-supervised learning (SSL) has emerged as a promising alternative to fully supervised approaches by leveraging large amounts of unlabeled data alongside a small labeled subset [7]. SSL aims to transfer knowledge from labeled to unlabeled samples using pseudo-labels, consistency constraints, deep co-training, and entropy minimization [8]. Pseudo-labeling refers to a model trained on labeled data that generates predicted labels for unlabeled samples. These labels are treated as ground truth during training, although their quality strongly affects performance [9]. This approach is increasingly adopted in medical imaging tasks to alleviate the reliance on costly pixel-wise annotations.

Among SSL paradigms, the Mean-Teacher (MT) framework [10] has shown notable success. In this setup, a student model learns from both labeled and unlabeled data, while a teacher model-an exponential moving average (EMA) of the student-provides stable targets for unsupervised training. Numerous extensions of the MT framework [11–15] have shown strong performance across various medical imaging tasks, establishing MT as a solid foundation for brain lesion segmentation. Additionally, several other advanced SSL methods, such as [16–21], have achieved competitive results in other SSL domains.

Furthermore, Siamese strategies have been employed in SSL segmentation to enforce structural and feature-level consistency between supervised and unsupervised branches. By sharing weights between two network branches, they improve representation alignment, enhance

* Corresponding author.
E-mail addresses: 100063759@ku.ac.ae (M. Alshurbaji), maregu.habtie@ku.ac.ae (M. Assefa), ahmad.obeid@ku.ac.ae (A. Obeid), mohamed.seghier@ku.ac.ae (M.L. Seghier), taimur.hassan@adu.ac.ae (T. Hassan), kamal.taha@ku.ac.ae (K. Taha), naoufel.werghi@ku.ac.ae (N. Werghi).

https://doi.org/10.1016/j.patrec.2025.11.032
Received 12 July 2025; Received in revised form 17 October 2025; Accepted 20 November 2025
Available online 29 November 2025

training stability, and boost generalization under limited labeled data. Prior work has applied Siamese in medical image segmentation, for instance, Bortsova et al. [22] introduced a Siamese segmentation model trained to generate consistent outputs under elastic deformations using both labeled and unlabeled data. Also, jia et al. [23] used two identical segmentation networks with cross-supervision via pseudo-labels on unlabeled images; these networks teach each other in a Siamese manner, filtering out uncertain predictions to stabilize training. Furthermore, recent studies explored capsule-based architectures for medical image segmentation [24], but these models are computationally demanding and rarely used in semi-supervised settings. In contrast, our framework provides a lightweight and scalable solution for segmentation with sparse annotations.

Adversarial training and Generative Adversarial Networks (GANs) have demonstrated significant promise in SSL by using the discriminator to guide the generator toward producing realistic and accurate segmentation masks under limited labeled data. GAN-based frameworks can be used for pseudo-label refinement, for instance, Ou et al. [25] used adversarial training on both labeled and unlabeled data to improve stroke lesion segmentation. Similarly, the multi-scale consistency adversarial learning approach enforces feature-level consistency across scales while using a discriminator to evaluate the realism of segmentations in unlabeled data [26]. Lei et al. [27] proposed a model that uses adversarial consistency across MT branches. It employs dynamic convolution units to better exploit unlabeled medical data. These approaches collectively highlight how GANs provide strong regularization and structural realism for SSL segmentation.

Bounding-box (bbox) priors are a form of weak supervision in SSL segmentation. These coarse annotations are much faster and cheaper to obtain than pixel-level masks. For instance, BBox-Guided Segmentor [25] leverages bbox priors alongside fully labeled and unlabeled 2D images. A segmentation generator is trained on sparse full annotations and weakly-supervised bbox cases, while a discriminator encourages realistic mask shapes even when only bbox labels are available. In SSL, bbox priors offer an attractive compromise; they require only minimal effort from medical professionals yet provide strong cues for lesion localization. By involving clinicians directly in drawing these simple annotations, segmentation results become not only more accurate but also more trustworthy, as experts can validate the process with their own inputs. Motivated by this, our framework integrates bbox priors with consistency- and adversarial-based strategies to enhance both performance and clinical acceptance.

Despite recent advances, existing SSL frameworks often address only one aspect of SSL-for instance, focusing on either pseudo-labeling, consistency learning, or adversarial refinement in isolation. This gap motivates the need for a unified framework that integrates these paradigms. This aligns with recent findings that show hybrid paradigms combining adversarial learning with consistency regularization can improve segmentation performance in SSL [6,28–30].

In this paper, we present TriGAN-SiaMT, a novel SSL segmentation framework designed to address the challenge of sparse annotations in brain lesion imaging. To the best of our knowledge, it is the first framework to jointly unify (1) adversarial learning, (2) Siamese-based consistency, (3) MT-based guidance, and (4) bbox priors into one framework. The framework introduces an adversarial learning strategy comprising three segmentors (S_0, S_1, S_2) and two discriminators (D_0, D_1). The path $S_0 \leftrightarrow D_0$ forms the supervised branch trained on labeled data, while the path $S_1 \leftrightarrow D_1$ constitutes a Siamese counterpart of the supervised path that promotes consistency-driven learning on unlabeled data. S_2 serves as an EMA teacher model of its student model S_1, forming an MT pathway. Furthermore, S_2 acts as a pseudo-label predictor for D_1 that does not see any labeled images. Additionally, we incorporate bbox priors for both labeled and unlabeled data as a form of weak supervision to guide lesion localization. The main contributions of this work are as follows:

- Introduce a novel SSL segmentation framework (TriGAN-SiaMT) that integrates adversarial training, Siamese-based consistency, and MT-based training to enable robust brain lesion segmentation under sparse annotations.
- Multi-level consistency constraint losses are employed by combining both segmentor- and discriminator-level objectives, including Siamese consistency and EMA-based consistency, to promote stable and effective training.
- Incorporate bbox priors as a form of weak-supervision in both labeled and unlabeled branches, effectively guiding lesion localization
- Demonstrate state-of-the-art performance on ISLES 2022 [31] and BraTS 2019 [32–34] datasets using as little as 5 % labeled data, validating data-efficiency and generalization of the proposed approach for brain lesion segmentation.

2. Methodology

We propose TriGAN-SiaMT, a semi-supervised segmentation framework designed to leverage both labeled and unlabeled data for brain lesion segmentation. The architecture combines adversarial learning, MT-based, and Siamese-based consistency constraints with bbox priors. As illustrated in Fig. 1, our framework consists of three segmentors (S_0, S_1, S_2) and two discriminators (D_0, D_1), each serving distinct roles: S_0 and D_0 form the supervised branch, trained on labeled data with full annotations. S_1 and D_1 operate on unlabeled data, enforcing consistency-based learning. S_2 is an EMA teacher of S_1, forming the MT pathway, which also sees unlabeled data.

2.1. Adversarial semi-supervised training

At each iteration, the batch is divided into labeled and unlabeled subsets. The supervised branch ($S_0 \leftrightarrow D_0$) trains on full annotations, while the unlabeled branch ($S_1 \leftrightarrow D_1$) uses pseudo-labels generated by S_2. Adversarial consistency learning is applied through alternating updates: segmentors produce predicted masks, which are evaluated by discriminators together with their corresponding input images. During segmentor training, S_k learns to generate masks perceived as realistic by its discriminator, while the discriminators are optimized to better distinguish real from generated masks. This adversarial interaction forms a minimax game between segmentor and discriminator as defined in (1) [35], enabling robust learning under limited supervision.

$$\min_{S_k} \max_{D_k} \mathbb{E}_{y \sim \mathbb{P}_{\text{real}}}[\log D_k(y)] + \mathbb{E}_{\hat{y} \sim \mathbb{P}_{\text{fake}}}[\log(1 - D_k(\hat{y}))] \tag{1}$$

Here, $S_k(\cdot)$ denotes a segmentor and $D_k(\cdot)$ a discriminator, where applicable, and $\mathbb{P}_{\text{real}}$ and $\mathbb{P}_{\text{fake}}$ represent the real and predicted data distributions, respectively. The discriminator aims to assign high values to real masks $y \sim \mathbb{P}_{\text{real}}$ and low values to predicted masks $\hat{y} \sim \mathbb{P}_{\text{fake}}$ by the segmentor, while the latter seeks to segment images $\hat{y} = S_k(x)$ that maximize $D_k(\hat{y})$, where x is the input images.

2.2. Models architecture

All the segmentors follow U-Net [36] architecture that accepts input data with three channels: an MRI, a bbox prior, and a Gaussian noise map **N**. While the discriminators share the same structure of the discriminator in [15,37], which follows a five-stage convolutional (conv) architecture as illustrated in Fig. 1. Initially, each of the segmentation map and the image are processed by a separate conv layer, then their outputs are summed and passed through three conv layers. This design enables the discriminators to evaluate the joint consistency of the segmentation map and the original image.

2.3. Loss functions

This section outlines the loss used to train the proposed framework, which combines supervised, adversarial, and consistency-based components. The overall objective is designed to minimize the performance

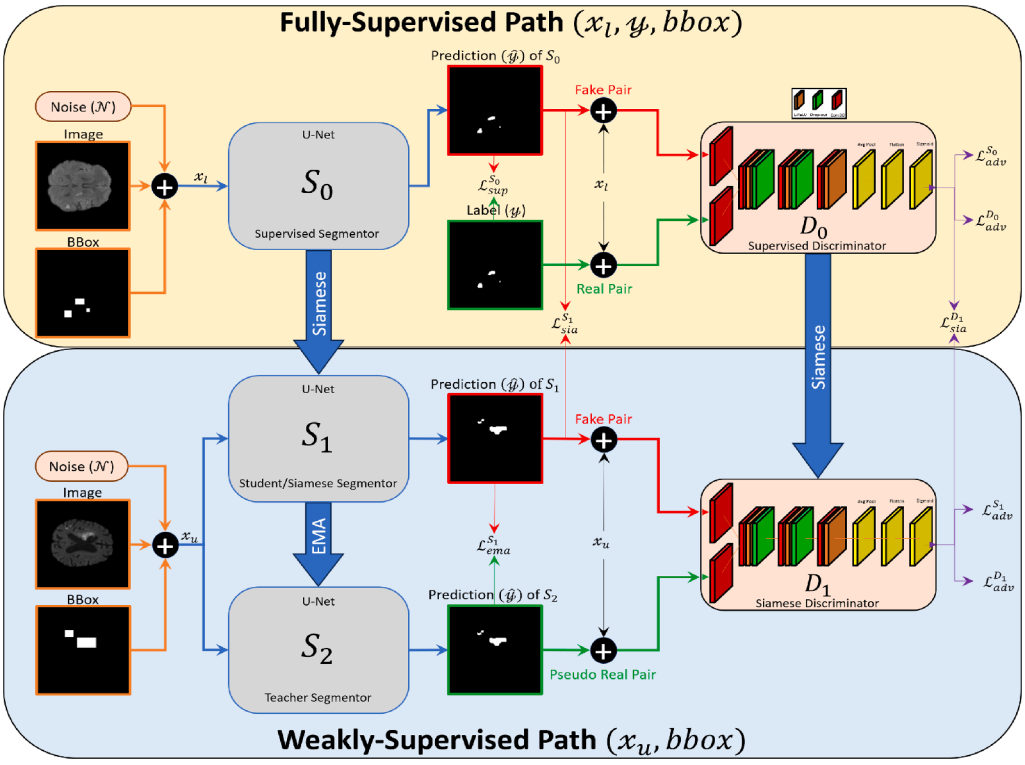

Fig. 1. An overview of our proposed semi-supervised TriGAN-SiaMT framework. Top panel: For fully labeled data x_l, the supervised segmentor S_0 is trained using real labels y, and predictions $\hat{y}$ are evaluated by the supervised discriminator D_0 through adversarial learning. Bottom panel: For unlabeled data x_u, the student segmentor S_1 is trained with adversarial feedback from D_1, guided by pseudo labels generated by the teacher S_2, and constrained by a Siamese consistency loss with S_0 and D_0. bbox priors are integrated as weak labels and are included as input along with the image and noise.

gap between supervised and unsupervised paths by having multiple consistency constraints.

2.3.1. Supervised loss

The supervised loss ($\mathcal{L}_{\text{sup}}$) is a combination of the widely defined Binary Cross-Entropy (BCE) loss ($\mathcal{L}_{\text{bce}}$) and the Dice loss ($\mathcal{L}_{\text{dice}}$), which is used to train the supervised segmentor S_0 as in Eq. (2).

$$\mathcal{L}_{\text{sup}}^{S_0}(x_l, y) = \mathcal{L}_{\text{dice}}[y, S_0(x_l)] + \mathcal{L}_{\text{bce}}[y, S_0(x_l)] \tag{2}$$

Where x_l is the labeled input image, P denotes the total number of pixels, and ϵ is a small constant (e.g., 10^{-6}) added for numerical stability.

2.3.2. Adversarial WGAN-GP loss

The discriminator adversarial loss consists of two parts: the critic loss $\mathcal{L}_{\text{critic}}$, which estimates the divergence between real and generated data distributions, and the gradient penalty loss $\mathcal{L}_{\text{gp}}$, which enforces the Lipschitz constraint by penalizing the gradient norm on interpolated inputs between real and fake masks. This formulation follows the Wasserstein GAN with gradient penalty (WGAN-GP) proposed by Gulrajani et al. [35].

$$\mathcal{L}_{critic}(x, \tilde{y}) = D_k(S_k(x)) - D_k(\tilde{y}) \tag{3}$$

$$\mathcal{L}_{\text{gp}}(y, \hat{y}) = \mathbb{E}_{\tilde{x}\sim\mathbb{P}_{\text{interp}}}\left[\left(\left\|\nabla_{\tilde{x}} D_k(\tilde{x})\right\|_2 - 1\right)^2\right] \tag{4}$$

$$\mathcal{L}_{\text{adv}}^{D_k} = \mathcal{L}_{critic}(x, \tilde{y}) + \lambda_{\text{gp}} \cdot \mathcal{L}_{gp}(y, \hat{y}) \tag{5}$$

Where x represents real samples (labeled or unlabeled). The term $\tilde{x}$ refers to samples interpolated between real and fake samples and equals: $\alpha y + (1-\alpha)\hat{y}$, where $\alpha \in [0, 1)$. Note that the term $\tilde{y}$ in Eq. (3) equals to y in the supervised path, and equals to $S_k(x_u)$ in the unsupervised path. The scalar λ_{gp} is a regularization coefficient controlling the weight of the gradient penalty term, and was set to 10 throughout all experiments.

On the other hand, the segmentor adversarial loss in Eq. (6) guides the segmentor to produce outputs that are indistinguishable from real annotations under the discriminator's evaluation.

$$\mathcal{L}_{\text{adv}}^{S_k}(x) = -D_k(S_k(x)) \tag{6}$$

2.3.3. Siamese consistency loss

To improve the reliability of the unlabeled branch, we employ a Siamese consistency strategy between S_1 and its supervised counterpart S_0, as well as between D_1 and its supervised counterpart D_0. The goal is to ensure that predictions from S_1 remain consistent with those from the supervised ones when applied to the same unlabeled input (x_u). Therefore, a consistency loss is applied between the prediction of S_1 and the pseudo-label generated by S_0, defined as:

$$\mathcal{L}_{\text{sia}}^{S_1}(x_u) = \mathcal{L}_{\text{dice}}[S_0(x_u), S_1(x_u)] \tag{7}$$

This loss encourages S_1 to maintain alignment with the behavior of the more stable supervised model S_0. Similarly, we enforce a consistency loss between D_1 and D_0 over both real and fake inputs as in:

$$\begin{aligned}\mathcal{L}_{\text{sia}}^{D_1}(x_u) = {} & \mathcal{L}_{\text{dice}}[D_0(S_2(x_u)), D_1(S_2(x_u))] \\ & + \mathcal{L}_{\text{dice}}[D_0(S_1(x_u)), D_1(S_1(x_u))]\end{aligned} \tag{8}$$

where $S_2(x_u)$ is the pseudo-label from the teacher, and $S_1(x_u)$ is the prediction from the student. The Siamese constraints act as a soft regularization, encouraging both S_1 and D_1 to emulate the behavior of their supervised counterparts.

2.3.4. Consistency regularization

To improve the stability and reliability of pseudo-labels during training on unlabeled data, we adopt the EMA strategy to construct the teacher S_2, which maintains a smoothed version of the student S_1. The EMA teacher S_2 is used to generate pseudo-labels using x_u during training, while only the student S_1 is updated via backpropagation. Formally,

let θ_t denote the parameters of the student segmentor at the current epoch t, and θ'_t denote the EMA-updated parameters of the teacher. The update rule is defined as: $\theta'_t = \alpha\theta'_{t-1} + (1-\alpha)\theta_t$.

Here, $\alpha \in [0, 1)$ is a smoothing coefficient that controls the degree of temporal averaging. Previous studies suggest using a fixed value of $\alpha = 0.999$ [27]. However, we adopt a ramp-up strategy that adjusts α as a function of the training iteration [13], which is defined as: $\alpha_t = \min\left(1 - \frac{1}{t+1}, \alpha_{\max}\right)$. Here $\alpha_{\max} = 0.99$ is the target smoothing coefficient, which has been experimentally found to yield better results than $\alpha_{\max} = 0.999$. This formulation ensures that early updates incorporate more of the student model's parameters, thereby stabilizing the EMA updates and enabling faster adaptation during the initial training phase. As training progresses, the update becomes increasingly conservative, resembling a standard EMA update.

To further leverage the benefits of EMA, we introduce a consistency loss between the predictions of S_1 and S_2, as shown in Eq. (9) [13], which promotes temporal stability and regularization throughout training.

$$\mathcal{L}^{S_1}_{\text{ema}}(x_u) = \mathcal{L}_{\text{dice}}\left[S_2(x_u), S_1(x_u)\right] \tag{9}$$

To balance the influence of consistency-based losses over the course of training, we adopt a Gaussian ramp-up schedule for the consistency weight λ [27,38], which is defined as $\lambda = \delta \cdot \exp\left(-5 \cdot \left(1 - \frac{t}{t_c}\right)^2\right)$. Here, δ is the maximum weight (typically set to 1.0), and t_c is the ramp-up length in epochs, which experimentally was set to 40. At the early training stages, λ remains close to zero, making the model rely primarily on $\mathcal{L}_{\text{sup}}$. As training progresses and the model predictions become more reliable, λ increases, gradually incorporating consistency and adversarial losses on x_u. This ramp-up helps stabilize training and avoid early overfitting to noisy pseudo-labels [38].

2.3.5. Total loss

The training objective encompasses optimizing the five models: the supervised segmentor S_0, the student siamese segmentor S_1, the EMA teacher S_2, the supervised discriminator D_0, and the siamese discriminator D_1. This training procedure is illustrated in (Algorithm 1). Therefore, S_0 and D_0 are optimized using x_l, and their losses are defined as:

$$\mathcal{L}^{S_0} = \mathcal{L}^{S_0}_{\text{sup}}(x_l, y) + \mathcal{L}^{S_0}_{\text{adv}}(x_l) \tag{10}$$

$$\mathcal{L}^{D_0} = \mathcal{L}^{D_0}_{\text{adv}} = \mathcal{L}_{critic}(x, y) + \lambda_{\text{gp}} \cdot \mathcal{L}_{gp}(y, \hat{y}) \tag{11}$$

The student segmentor S_1 combines the adversarial objective, and two consistency regularizations: Siamese with S_0 and EMA with S_2 as in Eq. (12). The siamese discriminator D_1 is trained using the adversarial loss and a Siamese consistency loss with its supervised counterpart D_0 as in Eq. (13). The teacher segmentor S_2 is updated using the EMA and contributes through pseudo-label generation without direct optimization.

$$\mathcal{L}^{S_1} = \mathcal{L}^{S_1}_{\text{adv}}(x_u) + \lambda \cdot \left(\mathcal{L}^{S_1}_{\text{sia}}(x_u) + \mathcal{L}^{S_1}_{\text{ema}}(x_u)\right) \tag{12}$$

$$\mathcal{L}^{D_1} = \mathcal{L}_{critic}(x, \bar{y}) + \lambda_{\text{gp}} \cdot \mathcal{L}_{gp}(y, \hat{y}) + \lambda \cdot \mathcal{L}^{D_1}_{\text{sia}}(x_u) \tag{13}$$

Algorithm 1 TriGAN-SiaMT: training procedure.

Input: Labeled data x_l; Unlabeled data x_u
Init: Segmentors S_0, S_1, S_2; Discriminators D_0, D_1; EMA factor α; ramp-up epoch t_c
for epoch $t = 1$ to T **do**
  $\lambda_t \leftarrow$ ramp-up; $\alpha_t \leftarrow$ EMA update
  for minibatch (x_l, x_u) **do**
    Supervised: $\hat{y} \leftarrow S_0(x_l)$; compute $\mathcal{L}^{S_0}_{\text{sup}}$ and $\mathcal{L}^{S_0}_{\text{adv}}$; update S_0
    D_0: $\hat{y} \leftarrow S_0(x_l)$ (detach); $\mathcal{L}^{D_0}$; update D_0
    Pseudo-labels: $\hat{y}_0 \leftarrow S_0(x_u)$; $\hat{y}_2 \leftarrow S_2(x_u)$
    Unsupervised: $\hat{y}_1 \leftarrow S_1(x_u)$; compute $\mathcal{L}^{S_1}_{\text{adv}}$ and $\mathcal{L}^{S_1}_{\text{sia}}$; $\mathcal{L}^{S_1}_{\text{ema}}$; update S_1
    D_1: evaluate $D_1(\hat{y}_2), D_1(\hat{y}_1)$; compute $\mathcal{L}^{D_1}_{\text{adv}}$ and $\mathcal{L}^{D_1}_{\text{sia}}$; update D_1
    EMA Update: $S_2 \leftarrow \alpha_t\theta'_{t-1} + (1-\alpha_t)\theta_t$
  end for
end for

3. Experiments

3.1. Datasets and settings

Input data: In this study, each input slice is represented as $X = \text{concat}(I, B, N)$, where $I \in \mathbb{R}^{H\times W}$ is the MRI (first channel), $B \in \{0,1\}^{H\times W}$ is the bbox prior used as a form of weak supervision (second channel), and $N \sim \mathcal{N}(0, 0.1^2)$ is a Gaussian noise map serving as mild regularization (third channel). This zero-mean Gaussian noise with a standard deviation of 0.1 serves as a mild regularization signal, helping to improve generalization. The bbox priors were pre-generated algorithmically, extending one pixel beyond each lesion edge and intentionally kept loose-not tight-to mimic clinical practice. Overlapping boxes were recursively merged into a single box until no intersections remained, further reflecting the expert annotation process. Note that the bbox priors are provided only as an additional input channel and are not incorporated into the loss function. For computational efficiency and architectural compatibility, the images were resized to 128×128, converted to single-precision format, and their intensities were normalized to a [0, 1] range.

ISLES 2022: The ISLES 2022 dataset [31] consists of 250 annotated 3D patient cases of acute and sub-acute ischemic strokes, providing three MRI modalities for each patient. Here we use the diffusion-weighted imaging (DWI) modality as the first input channel. The 3D volumes were sliced into 2D axial slices to make 15,684 images in total. The data were divided into three subsets: training set with 187 cases (≈ 75 %), validation set with 25 cases (= 10 %), and testing set with 38 cases (≈ 15 %).

BraTS 2019: The BraTS 2019 dataset [32–34] comprises multimodal 3D brain MRI scans of 335 patients diagnosed with gliomas. Each case includes four MRI modalities, and three tumor sub-regions. In this work, we use the FLAIR modality as the first input channel, and all tumor sub-regions were merged into a single binary label. The 3D volumes were sliced into 73 2D axial slices for each patient, to make 24,455 images in total. The dataset was split into three subsets: training set with 252 cases (≈ 75 %), validation set with 33 cases (≈ 10 %), and testing set with 50 cases (≈ 15 %).

Implementation: All experiments were conducted on a server equipped with an NVIDIA A100 GPU with 80 GB VRAM, an AMD EPYC 7763 64-core processor, and 512 GB of RAM, running Ubuntu 22.04. The implementation was based on PyTorch 2.6 and CUDA 12.4. The model was trained using the Adam optimizer with a learning rate of 5×10^{-5}, a batch size of 1, and for a total of 200 epochs. Each segmentor has 1.81M parameters and 748.45M FLOPs, while each discriminator has 2.76M parameters and 411.04M FLOPs, resulting in a total of 9.14M parameters and 2.32G FLOPs for the complete TriGAN-SiaMT framework.

Evaluation Metrics: To assess the performance of brain lesion segmentation, we employ four widely used metrics: Dice Similarity Coefficient (DSC), Intersection over Union (IoU), 95th percentile Hausdorff Distance (HD95), and Average Surface Distance (ASD). These metrics are computed on 3D segmentations reconstructed by stacking 2D axial predictions for each patient [31].

3.2. Ablation study

To investigate the contribution of each component in the proposed framework, we conduct an ablation study using 5 % labeled setting on both datasets in Table 1. The ablation focuses on understanding the importance of adversarial training, Siamese consistency, and EMA components.

Table 1
Ablation study using 5 % labeled data on ISLES 2022 and BraTS 2019 datasets. X_l and X_u denote the number of labeled and unlabeled patient cases for training, respectively. The symbols ✓ and ✗ indicate the inclusion or exclusion of a module in the corresponding variant.

Variant	X_l/X_u	S_0	S_1	S_2	D_0	D_1	D_2	DSC(%)↑
— ISLES 2022 Dataset —								
Supervised-only	212/0	✓	✗	✗	✗	✗	✗	87.92
MT		✗	✓	✓	✗	✗	✗	81.37
GAN-MT		✗	✓	✓	✗	✓	✗	80.84
GAN-MT w D_2		✗	✓	✓	✗	✓	✓	81.74
w/o S_2		✓	✓	✗	✓	✓	✗	84.04
w/o GAN	11/201	✓	✓	✓	✗	✗	✗	83.93
w/o D_0		✓	✓	✓	✗	✓	✗	83.24
w D_2		✓	✓	✓	✓	✓	✓	84.22
w/o bbox priors		✓	✓	✓	✓	✓	✗	55.58
TriGAN-SiaMT (Ours)		✓	✓	✓	✓	✓	✗	**84.80**
— BraTS 2019 Dataset —								
Supervised-only	285/0	✓	✗	✗	✗	✗	✗	90.13
MT		✗	✓	✓	✗	✗	✗	84.81
GAN-MT		✗	✓	✓	✗	✓	✗	85.82
GAN-MT w D_2		✗	✓	✓	✗	✓	✓	84.53
w/o S_2		✓	✓	✗	✓	✓	✗	85.28
w/o GAN	15/270	✓	✓	✓	✗	✗	✗	84.98
w/o D_0		✓	✓	✓	✗	✓	✗	85.46
w D_2		✓	✓	✓	✓	✓	✓	84.74
w/o bbox priors		✓	✓	✓	✓	✓	✗	81.54
TriGAN-SiaMT (Ours)		✓	✓	✓	✓	✓	✗	**86.32**

We evaluate our framework against several variants: The **Supervised-only** baseline trains a single U-Net model with labeled data only. The **MT** variant uses just S_1 and S_2 without any discriminator, where S_1 sees labeled data. The **GAN-MT w D_2 EMA** variant is like its previous variant GAN-MT, but this one adds a third discriminator D_2 as an EMA of D_1, enforcing consistency between them. In **w/o S_2 EMA**, the teacher model S_2 is removed, which disables EMA consistency, and the results are predicted using S_1. The **w/o GAN** variant removes D_0 and D_1, which disables adversarial learning. The **w/o D_0 Siamese** variant disables the discriminator-level Siamese consistency loss between D_0 and D_1, without removing D_2 itself. The **w D_2 EMA** variant adds a third discriminator D_2 to TriGAN-SiaMT. Finally, the **w/o bbox priors** variant retains all modules of the main framework but excludes the third input channel containing the bbox priors.

The ablation results in Table 1 show that individual components of the TriGAN-SiaMT framework contribute differently across datasets. Certain variants outperform others on each dataset, indicating that the impact of each component is influenced by the dataset's inherent characteristics, including lesion type, size variability, or image distribution. Notably, while the performance impact of adding or removing a component is more pronounced on ISLES 2022, it appears less evident on BraTS 2019. This may be attributed to the higher homogeneity and structural consistency in BraTS 2019 tumor lesions compared to the more heterogeneous stroke lesions in ISLES 2022, making BraTS 2019 less sensitive to individual architectural modifications. Also, BraTS 2019 has a higher percentage of large lesions compared to ISLES 2022, which will be further discussed. These findings emphasize the complementary nature of TriGAN-SiaMT components. While each component brings unique advantages, their integration results in a robust and generalizable model.

3.2.1. Lesion size analysis

To assess model robustness across varying lesion sizes, we categorized the 3D patient cases according to the predominant lesion volume within each case, because multiple lesions can exist in each case. Lesions were classified as *large* (> 5 % of the brain volume), *medium* (1 %–5 %), or *small* (< 1 %), where the percentage is defined as $\left(\frac{\text{LesionPixels}}{\text{BrainPixels}}\right) \times 100\,\%$. Within the ISLES 2022 test set, 60.5 % of cases were dominated by small lesions, 31.6 % by medium lesions, and 7.9 % by large lesions. In contrast, the BraTS 2019 dataset included no small-lesion cases, with 12 % classified as medium and 88 % as large. This distribution highlights that BraTS 2019 is dominated by large lesion cases, while ISLES 2022 offers a more balanced distribution towards the smaller lesions.

Table 2
Performance comparison across lesion sizes.

Lesion Size	ISLES 2022		BraTS 2019	
	TriGAN-SiaMT	DyCON [21]	TriGAN-SiaMT	β-FFT [20]
Small	**86.05**	83.97	–	–
Medium	**84.11**	81.31	80.42	**84.51**
Large	**78.86**	77.84	87.29	**87.46**
Overall	**84.80**	82.70	86.32	**86.93**

Table 3
Impact of removing bbox priors on DSC (%).

Setting	ISLES 2022			BraTS 2019		
	TriGAN-SiaMT	DyCON [21]	β-FFT [20]	TriGAN-SiaMT	DyCON [21]	β-FFT [20]
w/o bbox	**55.58**	40.87	53.22	81.54	66.16	**83.58**
w bbox	**84.80**	82.70	79.77	86.32	85.63	**86.93**

Table 2 reports the DSC performance across different lesion sizes for the best-performing recent models under a 5 % labeled data setting. On ISLES 2022, TriGAN-SiaMT consistently outperforms DyCON across all lesion categories, with the largest margin observed for small lesions. This suggests that TriGAN-SiaMT is particularly effective in capturing small ischemic regions, which are often the most challenging to detect. Conversely, in BraTS 2019, TriGAN-SiaMT demonstrates strong results on large lesions but shows a notable drop in performance for medium lesions compared to β-FFT, indicating a sensitivity gap in handling medium-sized lesions. Therefore, medium-sized lesions in BraTS 2019 represent a failure mode for TriGAN-SiaMT.

3.2.2. The effect of BBox priors

The bbox priors were introduced to guide the model towards spatially relevant regions and reduce the search space for lesion localization. Their impact, however, differs significantly across datasets. As shown in Table 3, the removal of bbox priors causes a sharper decline in performance on ISLES 2022 compared to BraTS 2019. This indicates that bbox priors are especially valuable in cases dominated by small lesions, where precise localization is difficult without prior cues and limited labeled data supervision. By contrast, in BraTS 2019, which primarily consists of large lesions, the relative contribution of bbox priors is less pronounced. From a clinical perspective, this finding suggests that the annotation effort required to draw bbox priors is most justifiable for small ischemic lesions, where the gain in accuracy offsets the manual overhead, but may offer limited additional benefit for larger tumor cases.

Although TriGAN-SiaMT was specifically optimized to leverage bbox priors through its multi-level consistency constraints and adversarial training, its performance without bbox priors remains competitive. On ISLES 2022, it outperforms β-FFT [20] and DyCON [21], two of the latest SSL SOTA frameworks. However, on BraTS 2019 it fails to surpass β-FFT [20].

We further analyzed the effect of bbox margin size by extending lesion edges by 1 (our default), 2, and a random 1–3 pixels. On ISLES 2022 (5 % labeled), DSC scores were 84.80 %, 83.07 %, and 81.41 %, respectively, indicating that tighter (1-pixel) priors provide stronger localization cues, while larger margins reduce precision.

3.3. Comparison with state-of-the-art frameworks

Table 4 summarizes the performance of our proposed TriGAN-SiaMT framework under different labeling ratios (5 %, 10 %, and 40 %) on both

Table 4
Performance comparison with SOTA methods on ISLES 2022 and BraTS 2019 datasets under varying labeled data ratios. X_l and X_u denote the number of labeled and unlabeled patient cases for training, respectively.

SSL Method	X_l(%)	X_u(%)	DSC(%)↑	IoU(%)↑	HD95↓	ASD↓
— ISLES 2022 Dataset —						
CCT [16]			82.69	71.11	1.77	0.63
U^{2}PL [17]			70.70	57.81	–	–
URPC [18]			77.00	63.14	2.42	0.85
CT [19]	11(5%)	201(95%)	71.27	56.10	2.32	0.88
β-FFT [20]			79.77	66.99	2.00	0.69
DyCON [21]			82.70	70.93	2.12	0.70
TriGAN-SiaMT (Ours)			**84.80**	**74.04**	**1.57**	**0.52**
CCT [16]			83.14	71.67	1.75	0.60
U^{2}PL [17]			71.71	59.41	–	–
URPC [18]			80.11	66.50	1.92	0.64
CT [19]	22(10%)	190(90%)	72.10	56.95	2.69	0.87
β-FFT [20]			82.96	71.32	1.68	0.58
DyCON [21]			84.97	74.42	1.54	0.52
TriGAN-SiaMT (Ours)			**85.81**	**75.43**	**1.53**	**0.49**
CCT [16]			83.64	72.21	1.68	0.57
U^{2}PL [17]			73.94	62.44	–	–
URPC [18]			81.80	69.65	1.63	0.59
CT [19]	85(40%)	127(60%)	73.83	56.44	2.30	0.85
β-FFT [20]			84.89	74.10	1.40	0.46
DyCON [21]			82.54	70.68	2.06	0.68
TriGAN-SiaMT (Ours)			**87.66**	**78.41**	**1.33**	**0.42**
BBox-Guided Seg.[a] [25]	65(37%)	110(63%)	83.62	83.35	–	–
— BraTS 2019 Dataset —						
CCT [16]			85.16	77.16	3.55	1.15
U^{2}PL [17]			83.76	75.05	–	–
URPC [18]			85.45	76.95	3.57	1.28
CT [19]	15(5%)	370(95%)	85.95	77.50	3.45	1.22
β-FFT [20]			**86.93**	**77.95**	3.47	1.11
DyCON [21]			85.63	75.42	4.67	1.54
TriGAN-SiaMT (Ours)			86.32	77.91	**3.32**	**1.08**
CCT [16]			86.67	78.80	3.46	1.16
U^{2}PL [17]			85.64	77.71	–	–
URPC [18]			85.02	76.24	3.94	1.38
CT [19]	29(10%)	256(90%)	87.22	79.37	**3.31**	1.16
β-FFT [20]			**87.50**	**78.78**	3.37	**1.06**
DyCON [21]			86.29	76.65	4.18	1.37
TriGAN-SiaMT (Ours)			86.62	78.46	3.39	1.12
CCT [16]			87.52	79.87	3.22	1.10
U^{2}PL [17]			87.43	79.59	–	–
URPC [18]			85.88	77.46	3.69	1.29
CT [19]	114(40%)	171(60%)	87.14	79.27	3.01	1.07
β-FFT [20]			**89.06**	**81.51**	**2.74**	**0.94**
DyCON [21]			84.62	74.25	4.78	1.62
TriGAN-SiaMT (Ours)			87.95	79.91	3.05	1.04

[a] Reported from the original paper, as the code was not publicly available.

the ISLES 2022 and BraTS 2019 datasets. We benchmark our model against recent state-of-the-art (SOTA) SSL segmentation methods, highlighting its robustness across different supervision levels. To ensure a fair and consistent comparison, we provided all the benchmarking models with the same input configuration as our framework-namely, the raw MRI, the noise channel, and the corresponding bbox priors as the third channel.

According to the results of TriGAN-SiaMT, the performance gain is especially prominent in low-label settings, demonstrating the framework's ability to leverage unlabeled data effectively. Furthermore, Fig. 2 presents qualitative comparisons of TriGAN-SiaMT segmentation outputs against the four best performing SOTA frameworks. Notably, TriGAN-SiaMT generates masks that are visually close to the ground truth, effectively capturing lesion shapes and successfully identifying small lesions that are hard to segment.

It is well known that SSL methods often exhibit instability, especially when trained on limited labeled data. This instability can lead to

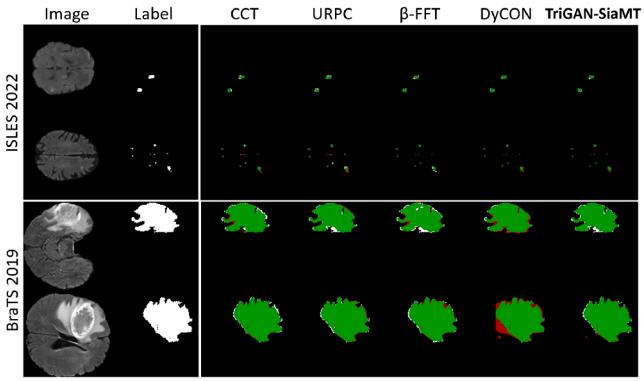

Fig. 2. Qualitative comparison of brain lesion segmentation results on the ISLES 2022 and BraTS 2019 datasets, under 5% labeled data setting. True positives are shown in green, false positives in red, and false negatives in white.

inconsistent performance across datasets with varying characteristics. For instance, while CT [19] and β-FFT [20] models underperform on ISLES 2022, they achieve strong results on BraTS 2019- with β-FFT even attaining the highest DSC scores in BraTS 2019. Moreover, all benchmarked SSL methods perform somehow similarly on BraTS 2019, with scores around those of TriGAN-SiaMT. This behavior can be attributed to the inherent nature of BraTS 2019, which contains structurally consistent and larger tumor lesions. In contrast, ISLES 2022 involves smaller, more heterogeneous stroke lesions with diffuse boundaries, making it more challenging for standard SSL models to generalize effectively.

In contrast to this variability, our TriGAN-SiaMT framework demonstrates strong training stability and robustness across both datasets. It reaches near-converged DSC performance under 15 epochs, underscoring its efficient optimization and reduced sensitivity to data distribution shifts. This stability stems from the integration of multi-level consistency constraints and adversarial learning, which together create a more resilient and generalizable segmentation framework.

4. Conclusion

We introduced TriGAN-SiaMT, a novel SSL segmentation framework that combines adversarial learning, multi-level consistency regularization, and bbox priors to address the challenge of sparse annotations in brain lesion segmentation. The architecture integrates three segmentors and two discriminators across supervised, Siamese, and Mean Teacher pathways. Extensive experiments on ISLES 2022 and BraTS 2019 datasets demonstrate the strong generalization ability of this framework, achieving state-of-the-art performance-particularly under low-label regimes (e.g., 5% labeled data)-along with stable and consistent convergence during training. The results highlight that bbox priors contribute most significantly to small lesions, establishing them as a valuable supervision signal when precise localization is essential. These findings highlight the effectiveness and reliability of our approach in real-world medical imaging scenarios where labeled data is limited.

CRediT authorship contribution statement

Mohammad Alshurbaji: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis; **Maregu Assefa:** Writing – review & editing, Supervision, Methodology, Investigation, Formal analysis, Conceptualization; **Ahmad Obeid:** Visualization, Validation; **Mohamed L. Seghier:** Writing – review & editing, Supervision, Project administration, Investigation, Data curation; **Taimur Hassan:** Supervision, Project administration; **Kamal Taha:** Supervision, Project administration; **Naoufel Werghi:** Writing – review & editing, Supervision, Software, Resources, Project administration, Investigation, Funding acquisition.

Data availability

This study relies exclusively on publicly accessible datasets. The ISLES 2022 dataset and the BraTS 2019 dataset can be accessed through their respective challenge websites. All implementation code developed for this work is openly accessible at: https://github.com/MAlshurbaji/TriGAN-SiaMT.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

During the preparation of this work, the authors used ChatGPT-4o (OpenAI, 2025) in order to enhance the readability and language of this manuscript. After using this tool, the authors reviewed and edited the content as needed and takes full responsibility for the content of the publication.

References

[1] A. Bousselham, O. Bouattane, M. Youssfi, A. Raihani, Towards an accurate MRI acute ischemic stroke lesion segmentation based on bioheat equation and U-Net model, Int. J. Biomed. Imaging 2022 (2022) 1–12.
[2] M.S. Sheela, G. Amirthayogam, J.J. Hephzipah, R. Suganthi, T. Karthikeyan, M. Gopianand, Advanced brain tumor classification using DEEPBELEIF-CNN method, Babylonian J. Mach. Learn. 2024 (2024) 89–101.
[3] C. Sokea, S. Marina, Improving diagnostic accuracy of brain tumor MRI classification using generative AI and deep learning techniques, Babylonian J. Artif. Intell. 2025 (2025) 55–63.
[4] M. Al-Ayyoub, A. Alsmirat, A. Jararweh, Efficient annotation strategies in medical imaging: reducing the labeling burden for AI applications, Edraak J. AI Data Sci. 2 (1) (2024) 15–27.
[5] M. Soni, M.A. Shnan, Scalable neural network algorithms for high dimensional data, Mesopotamian J. Big Data 2023 (2023) 1–11.
[6] A. Marab, A. Bhadrashetty, A novel approach using deep convolutional neural networks for automated dementia detection and classification, EDRAAK 2023 (2023) 16–20.
[7] M.A. Almaiah, F.A. El-Qirem, An analytical comparison of transfer learning techniques for brain tumor detection and classification, Babylonian J. Mach. Learn. 2025 (2025) 106–115.
[8] Y. Lu, Z. Fan, M. Xu, Multi-dimensional fusion and consistency for semi-supervised medical image segmentation, 2023. arXiv:2309.06618
[9] C. Wang, C. Xiong, B. Zhao, S. Ding, CycleMatch: cyclic pseudo-labeling distillation in semi-supervised medical image segmentation, Pattern Recognit. Lett. 193 (2025) 135–141.
[10] A. Tarvainen, H. Valpola, Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results, Adv. Neural Inf. Process. Syst. 30 (2017) pp. 1195–1204.
[11] L. Yu, S. Wang, X. Li, C.-W. Fu, P.-A. Heng, Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation, *MICCAI 2019 - Medical Image Computing and Computer Assisted Intervention*, 2, Springer, 2019.
[12] Y. Zhang, J. Zhang, Dual-task mutual learning for semi-supervised medical image segmentation, 4 of PRCV 2021 - Pattern Recognition and Computer Vision, Springer, 2021.
[13] W. Cui, Y. Liu, Y. Li, M. Guo, Y. Li, X. Li, T. Wang, X. Zeng, C. Ye, Semi-supervised brain lesion segmentation with an adapted mean teacher model, 2019. arXiv:1903.01248
[14] K. Li, S. Wang, L. Yu, P.-A. Heng, Dual-teacher: integrating intra-domain and inter-domain teachers for annotation-efficient cardiac segmentation,2020 arXiv:2007.06279 .
[15] S. Li, C. Zhang, X. He, Shape-aware semi-supervised 3D semantic segmentation for medical images, 2020 arXiv:2007.10732.
[16] Y. Ouali, C. Hudelot, M. Tami, Semi-supervised semantic segmentation with cross-consistency training, in: Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition, 2020, pp. 12674–12684.
[17] Y. Wang, H. Wang, Y. Shen, J. Fei, W. Li, G. Jin, L. Wu, R. Zhao, X. Le, Semi-supervised semantic segmentation using unreliable pseudo-labels, 2022. arxiv:2203.03884
[18] X. Luo, G. Wang, W. Liao, J. Chen, T. Song, Y. Chen, S. Zhang, D.N. Metaxas, S. Zhang, Semi-supervised medical image segmentation via uncertainty rectified pyramid consistency, Med. Image Anal. 80 (2022) 102517.
[19] X. Luo, M. Hu, T. Song, G. Wang, S. Zhang, Semi-supervised medical image segmentation via cross teaching between CNN and transformer, in: Proc. Medical Imaging with Deep Learning (MIDL), PMLR, 2022, pp. 820–833.
[20] M. Hu, J. Yin, Z. Ma, J. Ma, F. Zhu, B. Wu, Y. Wen, M. Wu, C. Hu, B. Hu, Q. Wang, β-FFT: nonlinear interpolation and differentiated training strategies for semi-supervised medical image segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2025.
[21] M. Assefa, M. Naseer, I.I. Ganapathi, S.S. Ali, M.L. Seghier, N. Werghi, DyCON: dynamic uncertainty-aware consistency and contrastive learning for semi-supervised medical image segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2025.
[22] G. Bortsova, F. Dubost, L. Hogeweg, I. Katramados, M.D. Bruijne, Semi-supervised medical image segmentation via learning consistency under transformations, 2019. arXiv:1911.01218
[23] D. Jia, Semi-supervised multi-organ segmentation with cross supervision using Siamese network, in MICCAI Challenge on Fast and Low-Resource Semi-supervised Abdominal Organ Segmentation, Springer, 2022.
[24] E.T. Mahdi, S.M. Al-Barzinji, W.K. Awad, Object detection using capsule neural network: an overview, Babylonian J. Mach. Learn. 2024 (2024) 157–165.
[25] Y. Ou, S.X. Huang, K.K. Wong, J. Cummock, J. Volpi, J.Z. Wang, S.T.C. Wong, BBox-guided segmentor: leveraging expert knowledge for accurate stroke lesion segmentation using weakly supervised bounding box prior, Comput. Med. Imaging Graphics 107 (2023) 102236.
[26] X. Guo, K. Sun, Y. Zheng, Multi-scale consistency adversarial learning for semi-supervised 3D medical image segmentation, Biomed. Signal Process. Contr. 103 (2025) 107475.
[27] T. Lei, D. Zhang, X. Du, X. Wang, Y. Wan, A. Nandi, Semi-supervised medical image segmentation using adversarial consistency learning and dynamic convolution network, IEEE Trans. Med. Imaging 42 (5) (2023) 1265–1277.
[28] Y. Tang, S. Wang, Y. Qu, Z. Cui, W. Zhang, Consistency and adversarial semi-supervised learning for medical image segmentation, Comput. Biol. Med. 161 (2023) 107018.
[29] Y. Feng, X. Tang, Q. Sun, W. Li, S. Zheng, Semi-supervised medical image segmentation method using multi-scale consistency adversarial learning, Biomed. Signal Process. Contr. 111 (2026) 108250.
[30] T. Sakirin, S. Kusuma, A survey of generative artificial intelligence techniques, Babylonian J. Artif. Intell. 2023 (2023) 10–14.
[31] M.R.H. Petzsche, et al., ISLES 2022: a multi-center magnetic resonance imaging stroke lesion segmentation dataset, Sci. Data 9 (1) (2022) 762.
[32] B.H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, The multimodal brain tumor image segmentation benchmark (BRATS), IEEE Trans. Med. Imaging 34 (10) (2014) 1993–2024.
[33] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J.S. Kirby, J.B. Freymann, K. Farahani, C. Davatzikos, Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features, Sci. Data 4 (1) (2017) 1–13.
[34] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, R.T. Shinohara, C. Berger, S.M. Ha, M. Rozycki, Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge, 2018. edition, arXiv:1811.02629
[35] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A.C. Courville, Improved Training of Wasserstein GANs, 2017. arXiv:1704.00028
[36] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, 3 of MICCAI 2015 - Medical Image Computing, Springer, 2015.
[37] X. Luo, SSL4MIS, 2020. https://github.com/HiLab-git/SSL4MIS.
[38] S. Laine, T. Aila, Temporal ensembling for semi-supervised learning, 2016. arXiv:1610.02242