SV-UNet: Attention-based Fully Convolutional Network with Transfer Learning for Multimodal Infarct Segmentation

Han Xu

School of Computer Science, Queensland University of Technology, Brisbane, 4001, Australia
email: han.xu@connect.qut.edu.au

Abstract. Ischemic stroke has a devastating impact on global health, causing both death and disability. Automatic, accurate segmentation of these stoke areas, or infarctions, from Magnetic Resonance Imaging (MRI), can aid clinicians in personalized therapeutic strategies. Recent advances in merging fully convolutional networks with transfer learning show a promising outlook, but they rarely focus on multi-modalities analysis and leverage channel-wise anatomical information to improve segmented performance. The research introduces an attention-based SV-UNet model designed to identify infarctions in two MRI modalities: Diffusion-Weighted Imaging (DWI) and T1-Weighted (T1w) images. This model derives from the UNet architecture as the backbone, employing a pre-trained VGG16 model as a shared encoder connecting to two decoders with identical architecture. In each up-convolution operation, a Squeeze-and-Excitation Network is integrated to enhance feature restoration by analyzing global information. For comparison, a VGG16-Dual-UNet is established as the benchmark. This architecture is identical to the SV-UNet, except for the removal of the SENet module. The research evaluates the two networks using two datasets: Anatomical Tracings of Lesions After Stroke 2.0R and Ischemic Stroke Lesion Segmentation 2022. The study demonstrates that SV-UNet outperforms the baseline model in detecting small stroke lesions (minority pixels) within DWI data. While performance on T1w data remains comparable, the superior sensitivity in DWI data suggests promise for improved clinical applications.

Keywords: Ischemic Stroke, MRI Segmentation, Squeeze and Excitation Network, Transfer Learning, Fully Convolutional Network

1 Introduction

Acute cerebral vascular disorders, known as strokes, can result in mortality and long-term disability [1]. The World Stroke Organization (WSO) estimates a staggering 14 million new strokes and 5.8 million stroke deaths annually [2]. Ischemic strokes are dominant, accounting for 60–80% of all stroke patients [1]. It results from blood artery obstruction that restricts blood supply to the brain. In neuroimaging technologies, Magnetic Resonance Imaging (MRI) is a typical method to diagnose brain infarction [3]. In practice, neurologists need to read every scan to delineate infarcted regions to aid clinicians in deciding on treatment plans [4]. A potential problem indicates that artificial

Y. Wang (ed.), *Proceedings of the 2024 2nd International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI 2024)*, Advances in Computer Science Research 115,
https://doi.org/10.2991/978-94-6463-540-9_60

tracing is effort-intensive and subjective, which may cause delayed or biased diagnosis. It raises patients' indirect risk of permanent brain injury, which could result in death or disability. Therefore, investigating a reliable, objective infarct segmentation strategy is crucial to reducing this danger.

Most researchers believe that the Fully Convolutional Network (FCN) has the potential for automatic segmenting infarctions. This network evolved from the classical convolutional neural network (CNN), retaining convolution and pooling layers but replacing fully connected layers with convolutional layers [5]. It can capture low-level local characteristics (e.g., edges or texture) and high-level global features (e.g., brain structures), by down-sampling and up-sampling inputs sequentially. The mechanism avoids the loss of voxel-based spatial information and increases estimated accuracy. Moreover, the FCN lies in end-to-end learning, which removes the subjective intermediate interventions [5]. It speeds up the training process and discovers more effective features from the raw medical data. Many studies proposed variants of FCN-based models to experiment on MRI datasets, achieving effective segmentation [4, 6, 7]. However, exploring these models encounters obstacles, due to the expensive computational demand.

Transfer learning (TL) is identified as a viable solution for mitigating this demand and potentially attaining elevated segmentation accuracy [8]. This strategy enables pre-trained FCN-based models to be fine-tuned for solving a task-specific problem with prior knowledge. It significantly saves time and computational resources during model training. However, accessing these pre-trained models in sensitive medical domains might pose challenges due to privacy concerns and data scarcity [9]. As an alternative, Kora et al. demonstrated that publicly available, non-medical TL models, including GoogleNet, AlexNet, ResNet and VGGNet, were suited for medical image analysis after customized refinements [10]. Aside from these networks, Mohapatra et al. extra introduced IR (Inception-ResNet) V2, V3 and V4 networks to localize early infarctions on non-contrast CT images, with VGG16 achieving superior results [11]. In a similar task, Pravitasari et al. proposed VGG16-UNet achieved a 95.69% correct classification ratio on an MRI-based dataset in brain tumor segmentation [12]. This model employed a pre-trained VGG16 as an encoder, mirroring the down-sampling blocks of this encoder to build a decoder to form a symmetrical network. Existing transfer learning (TL)-based algorithms, while valuable references for the infarct segmentation task, lack reusability across multiple MRI modalities. This limitation requires separate models for each modality when dealing with the same tasks, potentially leading to outcome variability and hindering treatment decisions.

Researchers have recently introduced the feature attention mechanism into FCN-based models for segmenting infarcted regions from MRI data [13]. This mechanism effectively suppresses background interference (0 pixels) and directs models to focus on infarct-specific characteristics, such as hyperintensity on diffusion-weighted imaging (DWI) scans. However, feature attention is computationally expensive due to the requirement to calculate all spatial relationships across all slices. Channel-wise attention, which possesses similar properties to feature attention in infarct detection, explores channel-level contributions by amplifying important information within channels to improve accuracy [14]. This mechanism offers a computationally more efficient alternative to feature attention by focusing on feature importance within each channel rather than calculating complex inter-slice relationships.

While integrating channel-wise attention mechanisms with transfer learning in FCNs presents a promising avenue to solve reusability issues and achieve exact infarct segmentation, this area remains largely unexplored. Addressing this gap, this paper introduces SV-UNet, a novel attention-based model with dual MRI modality branches, to identify infarctions in DWI and T1w images. Built on the UNet architecture, SV-UNet leverages a pre-trained VGG16 as a shared encoder, feeding two identical decoders. To enhance feature restoration, SENet modules are integrated after each up-convolution. The model performance is evaluated on Anatomical Tracings of Lesions After Stroke (ATLAS) 2.0R and The Ischemic Stroke LEsion Segmentation (ISLES) 2022 datasets.

This paper is structured as follows: Section 2 details the methodology employed to achieve the research goal. Section 3 presents the experimental results and discussions, while Section 4 concludes the work by outlining its key findings, limitations, and potential future directions.

2 Method

The experiment aims to validate a) the feasibility of using a pre-trained VGG16 to build an FCN-based network capable of analyzing T1w and DWI data for delineating regions of infarct; b) the impact of incorporating SENet into the proposed network. To achieve these objectives, two models were developed: VGG16-Dual-UNet, a baseline model with separate decoders for each modality, and SV-UNet, an improved version incorporating Squeeze-and-Excitation Networks (SENets) for feature map recalibration after up-convolutions. The following section elaborates on the experimental process, including dataset preparation, model construction, and implementation details.

2.1 Dataset Preparation

The ATLAS 2.0 dataset, provided by [15], was utilized for the infarct segmentation task. The original dataset consists of 955 T1-weighted (T1w) MRI images from multiple centers, 655 training images with manually segmented lesion marks and 300 as the testing set without marks. Since none of the marks are available in the testing set, the experiment selected sequentially 70 patients of T1w scans from the training images with corresponding lesion marks. Each patient has 189 slices with an individual shape of 197 x 233. In total, 13,230 slices were selected for the experiment. The ISLES 2022 contains 250 patients of skull-stripping Apparent Diffusion Coefficient (ADC), Diffusion-Weighted Imaging (DWI), FLuid Attenuated Inversion Recovery (FLAIR) and ground-truth scans [16]. The experiment chose 15,684 DWI and ground-truth slices from all patients. Fig. 1 depicted two gray-based, resized slices from T1w and DWI respectively.

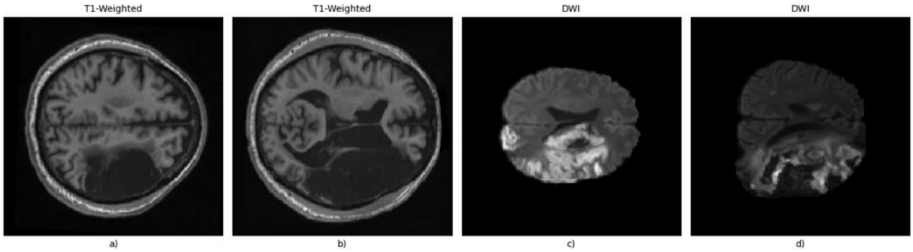

Fig. 1. a) and b) illustrate two T1-weighted slices resized to a shape of 256x256. c) and d) depict two skull-stripping diffusion-weighted images (DWI) with the same dimensions.

The ATLAS 2.0 dataset, provided by [15], was utilized for the infarct segmentation task. The original dataset consists of 955 T1-weighted (T1w) MRI images from multiple centers, 655 training images with manually segmented lesion marks and 300 as the testing set without marks. Since none of the marks are available in the testing set, the experiment selected sequentially 70 patients of T1w scans from the training images with corresponding lesion marks. Each patient has 189 slices with an individual shape of 197 x 233. In total, 13,230 slices were selected for the experiment. The ISLES 2022 contains 250 patients of skull-stripping Apparent Diffusion Coefficient (ADC), Diffusion-Weighted Imaging (DWI), FLuid Attenuated Inversion Recovery (FLAIR) and ground-truth scans [16]. The experiment chose 15,684 DWI and ground-truth slices from all patients. Fig. 1 depicted two gray-based, resized slices from T1w and DWI respectively.

The pre-processing steps included transforming 3D slices into 2D formats, resizing, normalizing, and filtering the slices. After pre-processing, T1w scans were reshaped from 70x197x233x189 (slices) to 3203x256x256x1 (the grey channel), with pixel values normalized into 0 and 1 by max-normalization. The infarct labels in the selected slices ranged from 110 to 6667. DWI slices were not uniform in shape, but the selected ones were resized to match the T1w dimensions, resulting in 3203x256x256x1. Their values were scaled between 0 and 1 using min-max normalization. The infarct pixels in filtered slices ranged from 100 to 9074. Finally, the T1w and DWI datasets were subsampled and split into training, validation, and testing sets with a 6:2:2 ratio.

2.2 Model Construction

Two FCN-based models with transfer learning were built for the experiment, named VGG16-Dual-UNet and SV-UNet. The proposed models were based on the U-Net architecture, a modification of classical convolutional neural networks (CNNs) [5]. Unlike standard CNNs, the symmetric U-Net architecture forgoes fully connected layers at the output and instead relies solely on convolutional and pooling layers. Its architecture constitutes a contracting path followed by an expanding path. A contracting pathway compresses pixel-based inputs into lower-dimensional features via downsampling, whereas an expanding pathway reconstructs these low-level features (e.g., texture or lines) through upsampling to restore spatial resolution lost during downsampling operations.

To mitigate the vanishing or exploding gradient problem, in the bi-decoders, the hidden convolutional layers with ReLU activation functions were initialized using Xavier

initialization [17]. This helped maintain a healthy distribution of gradients during training. For the final output convolutional layer with sigmoid activation, He initialization [17] was employed, which was suited for activated outputs with a non-zero lower bound. Furthermore, a masking decoder technique was employed during model training. It involved freezing the weights of one decoder while training the other. This prevented the decoders from interfering with each other's learning process.

VGG16-Dual-UNet. The baseline model was reliant on the UNet backbone, employing a shared VGG16 [18] encoder in the connection of two separate decoders, allowing the simultaneous analysis of T1w and DWI slices. VGG16 was pre-trained on the ImageNet dataset, capable of capturing general patterns from images. It was originally designed for RGB-channel images with an input shape of 224x224x3. To accommodate the T1w and DWI slices, its input layer was modified to 256x256, and the same input was triplicated to create a pseudo-RGB depth.

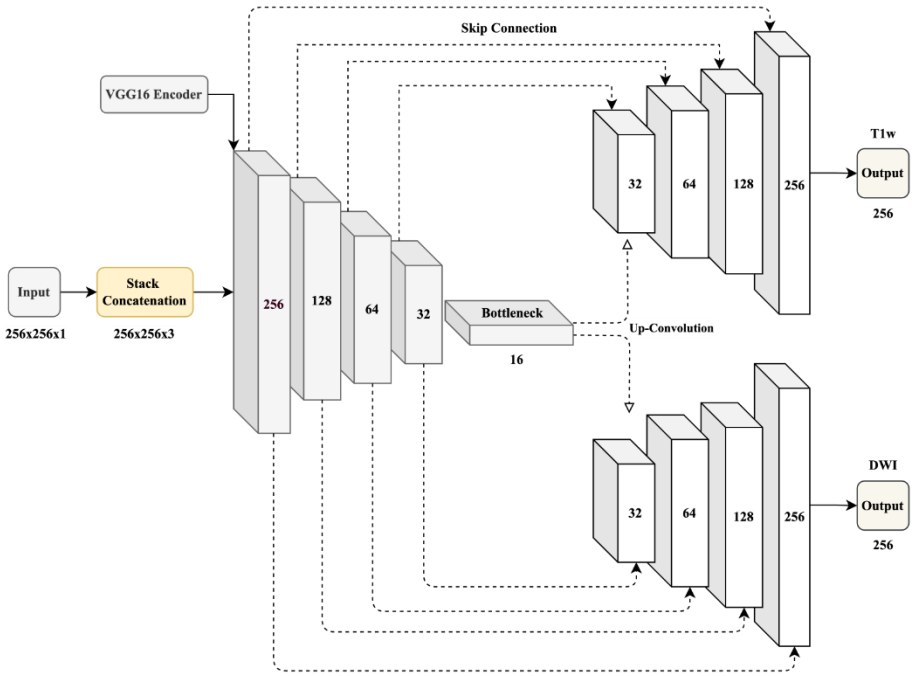

Fig. 2. VGG16-Dual-UNet architecture.

The encoder comprised five VGG16 convolutional blocks (shown in Fig 2) for the feature extraction of input T1w or DWI slices. The initial two blocks each had two convolutional layers followed by max-pooling, which reduced spatial dimensions while increasing feature channels. The generated feature maps were shaped as 256x256x64 and 128x128x128. The following two blocks had three convolutional layers and one max-pooling, extracting more complex features, and yielded output shapes of 64x64x256 and 32x32x512. The fifth block acted as a bottleneck, using three

convolutional layers to compress spatial information into features with the shape of 16x16x512. These five blocks constituted the contracting path, which remained frozen during the training process. The layer hyperparameter settings of these blocks were consistent with the pre-trained VGG16.

The expanding path consisted of two separate decoders with identical architecture for processing the corresponding T1-weighted and DWI modalities. Each decoder comprised four up-convolutional blocks to reconstruct the compressed features. Each block followed a specific sequence: a 2x2 transposed convolution layer with 'same' padding and a stride of 2, followed by two sets of convolutional layers (3x3 filters and 'same' padding), batch normalization, and ReLU activation layers. Finally, a skip connection operation combines the relevant trimmed feature map from the contracting path, which is crucial for restoring missing pixels during the previous down-convolution process.

The decoder architecture mirrored the encoder's structure but with fewer parameters. It employed progressively smaller convolutional filters (32, 16, 8, and 4) during the up-convolution process. This resulted in feature maps with progressively increasing spatial dimensions (32x32x32, 64x64x16, 128x128x8, and 256x256x4). Finally, a 1x1 convolution layer convolved a feature map (256x256x1) into a binary class prediction using a sigmoid activation function.

SV-UNet. The SV-UNet emerged as an improved model, adopting a similar architecture to VGG16-Dual-UNet. However, it incorporated a SENet [14] embedded within each decoder after the up-convolution operations, shown in Fig. 3. This network excelled at capturing latent dependencies and relationships between feature maps (channels) generated by convolutional layers. The inclusion of SENet helped to recalibrate these feature maps by leveraging global information, ultimately enhancing the reconstruction process. The network comprised three modules: Squeeze, Excitation and Scale.

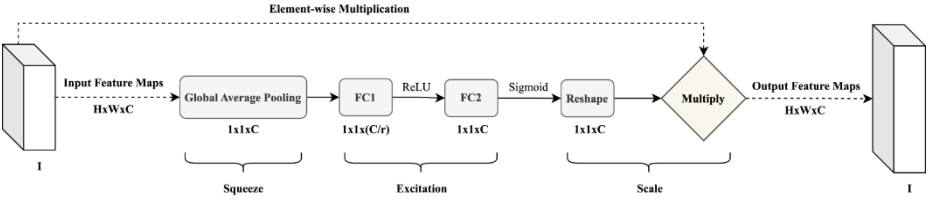

Fig. 3. Schematic flow for SENet embedded between two up-sampling blocks.

The Squeeze module leverages Global Average Pooling (GAP) to embed feature maps (I) into a global feature for reducing spatial complexity. These feature maps are generated by specific convolutional blocks, with dimensions of Height (H) x Width (W) x Channel (C). The GAP formula is defined as below:

$$V_c = \theta\left(I_c\right) = \frac{1}{H \times W} \sum_{n=1}^{H} \sum_{m=1}^{W} I_c(n, m) \tag{1}$$

The formula (θ) represents two mathematical computations. The first computation sums the value across all spatial locations (n, m) in H and W for a single channel (c).

The second computation divides this value by H x W for normalization. After repeating this process for the iteration of all channels in I_c, the feature maps are condensed into a squeezed vector of length c (1x1xC), denoted as V_c. In this vector, each value symbolizes compressed information of the corresponding channel, or feature map.

Excitation operates on the channel descriptor (V_c) with two fully connected (FC) layers to capture channel-wise dependencies. In the FC1, the dimensionality is reduced to C/r, where r is a reduction ratio that controls the capacity and computational cost. After ReLU activation, the FC2 restores the dimensionality back to C, followed by a sigmoid activation. In the experiment, this hyperparameter (r) is empirically set to 2. The mathematical equation for this process can be defined as:

$$g = \delta\big(W_2\sigma(W_1V_c)\big) \tag{2}$$

Where W_1 and W_2 are the learnable weights in the first and second fully connected layers (FC1 and FC2), respectively, σ and δ describe the ReLU and Sigmoid activations, and g denotes the excitation output. This output has the same dimensions (1 x 1 x C) as the squeezed output and contains the learned importance weights for each channel.

The scale operation performs element-wise multiplication between the original feature maps (I) and the excitation output (g). This multiplication selectively amplifies informative features and removes irrelevant ones, thereby recalibrating feature representation.

2.3 Implementation Details

In the experiment, the Adam optimizer (initial learning rate 0.0001) and focal loss (Gamma 0.5 and Alpha 0.75) [19] were applied to update the decoder's weights of VGG16-Dual-UNet and SV-UNet in the backpropagation process. Focal loss builds upon the standard binary cross-entropy loss function, with an introduced modulating factor $(1 - P_t)^{\gamma}$. The formula is defined as:

$$Focal\ Loss(p_t) = -a_t(1-p_t)^{\gamma}\log(p_t) \tag{3}$$

$P_t \in (0, 1)$ is the estimated probabilities for a pixel *t*, calculated by a sigmoid function. The α (alpha) $\in (0, 1)$ is a factor that harmonizes the contribution of well-classified and misclassified pixels. The former refers to how a model correctly predicts background and foreground pixels (0 and 1), while the latter indicates how a model assigns an incorrect pixel label. Gamma (γ) allocates higher loss weights to hard-to-class pixels, enabling a model to shift more attention to the minority infarct pixels.

The observation of validation curves depicted the model training process aimed to tweak hyperparameters. Early Termination monitored the validation loss with a patience of 30, returning the best weights during these patience periods. VGG16-Dual-UNet and SV-UNet models underwent training for 180 epochs, using a batch size of 32. The Area Under the Curve (AUC) was applied to examine model's capability to identify positive or negative labels. Recall and Precision were applied to evaluate how model correctly detects and predicts positive pixels (infarcts). The Dice Similarity Coefficient (DSC) harmonized Precision and Recall metrics, determining model

performance in infarct segmentation. The experiment was performed on V100 GPU in Google Colab.

3 Results and Discussion

The accuracy metric was not included in the experiment because it cannot truly reflect the model predictions due to the scarcity of positive pixels. The precision metric was chosen as an alternative due to its sensitivity to positive pixels. In the comparison of AUC, recall, precision, and DSC, the results reveal that SV-UNet outperforms VGG16-Dual-UNet on the DWI dataset, while demonstrating a similar performance to VGG16-Dual-UNet on the T1w dataset.

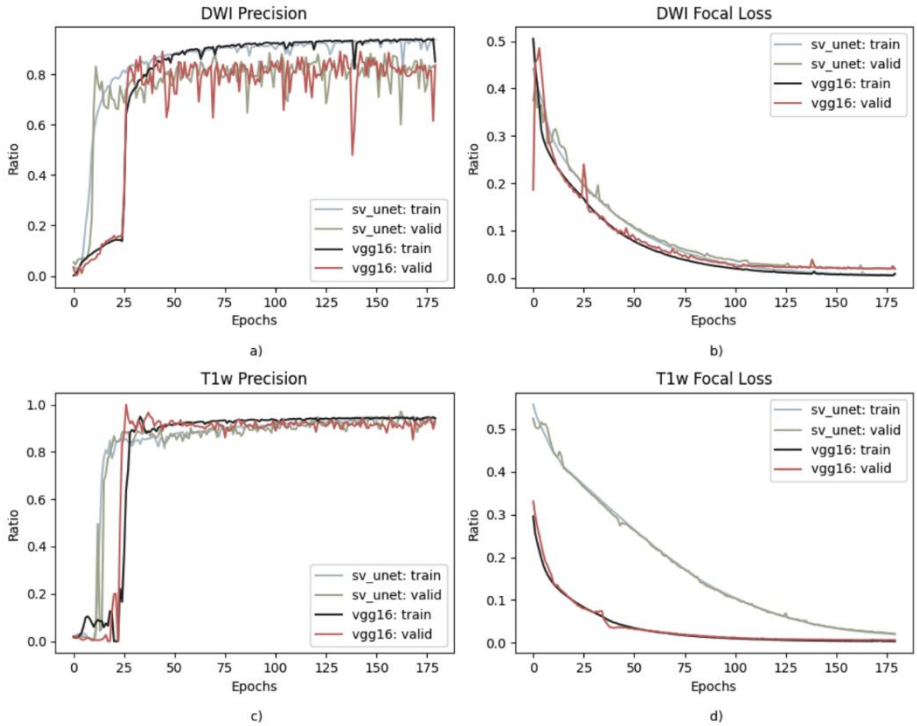

Fig. 4. Validation curves of Precisions and Losses for VGG16-Dual-UNet and SV-UNet training on T1w and DWI.

Fig. 4 demonstrates the concept of overfitting. In graph a), both models achieve high fitting to the training data but generalize poorly to the validation data. This is further evidenced by the high oscillation in the validation predictions. though the focal loss curves show a steady decrease with increasing epochs, shown as in the graph b). In contrast, graphs c) and d) highlight the effective learning behavior of the proposed models on the T1w dataset. This is evident in both training and validation predictions, accompanied by a smooth decrease in loss values. Notably, graphs (b) and (d) reveal an

advantage of the SENet module in the SV-NEet model. It improves sensitivity on minority pixels (infarcts) from the initial training epoch compared to VGG16-Dual-UNet, which requires 25 epochs for similar progress. Beyond this epoch, however, the two models exhibit a similar learning performance.

Table 1. Evaluative results of AUC, Recall, and Precision for VGG16-Dual-UNet and SV-UNet for specific datasets, accompanied with trainable and non-trainable weights; TP stands for Trainable Parameters, while NTP refers to Non-Trainable Parameters.

Model	Dataset	AUC	Recall	Precision	DSC	TP	NTP
VGG16-Dual-UNet	T1w	**0.95**	**0.91**	0.91	**0.91**	**577,938**	14,715,168
	Dwi	0.86	0.72	0.76	0.74		
SV-UNet	T1w	0.95	0.90	**0.92**	0.91	580,838	14,715,168
	Dwi	**0.86**	**0.73**	0.76	**0.75**		

Table 1 presents the evaluative metrics of two custom-designed models for infarct segmentation after 180 training epochs. For the T1w dataset, these metrics (AUC, Recall, and DSC) reveal close competition between the models. While VGG16-Dual-UNet achieves a slight edge in Recall (0.91), SV-UNet gains a marginal advantage in precision (0.92). Both models perform similarly in identifying infarcted pixels, but SV-UNet comes at the cost of requiring 2,900 more trainable parameters.

For the DWI dataset, AUC, Recall, and DSC again indicate tight competition, with SV-UNet exhibiting a slight advantage over VGG16-Dual-UNet in Recall. Precision remains comparable. Including the SENet module in SV-UNet benefits its Recall capability in predicting testing samples. This suggests that the SENet module prioritizes channels containing crucial infarct information while suppressing irrelevant features, ultimately enhancing SV-UNet's ability to detect positive pixels. In conclusion, SV-UNet outperforms VGG16-Dual-UNet on the DWI dataset.

Both models share a VGG16 architecture, contributing 14.7 million non-trainable parameters, with each reaching approximately 580,000 trainable weights. While both achieve a high DSC of 0.91 on the T1w dataset, performance drops to around 0.75 for the DWI dataset. This highlights the need for further research to improve DWI infarct segmentation.

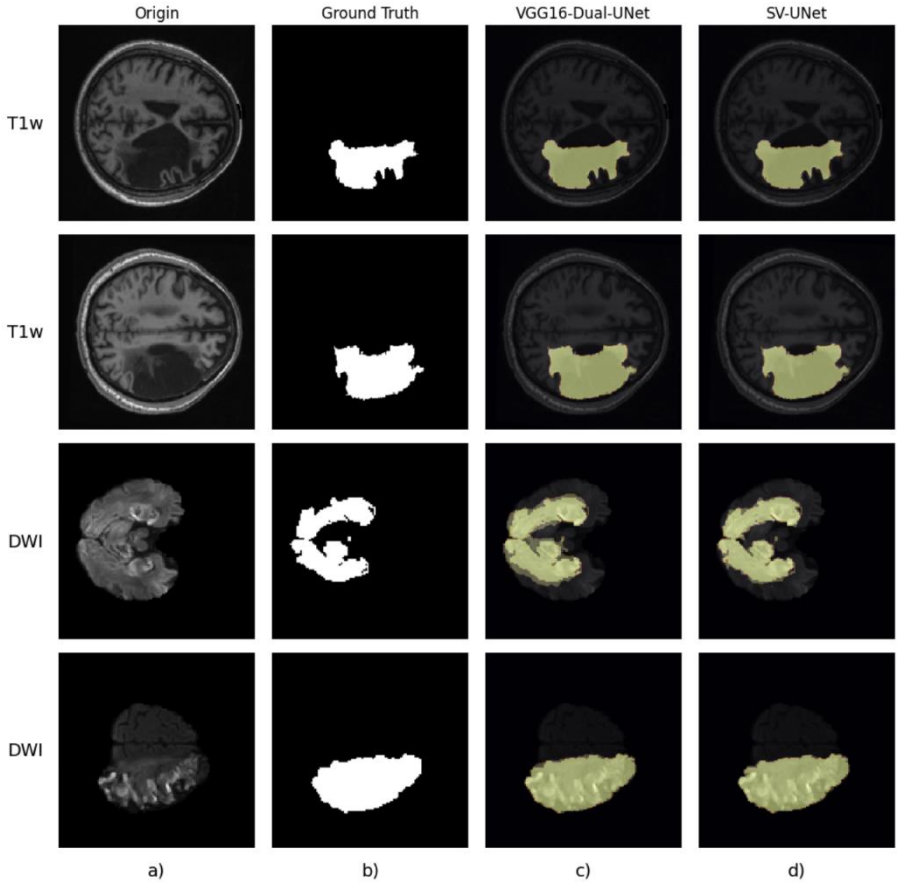

Fig. 5. Delineation of infarcted regions by VGG16-Dual-UNet and SV-UNet after 180 epochs training.

The experiment randomly selects two slices from T1w and DWI respectively for further examining segmentation performance, shown as Fig. 5. Infarctions appear as a dark area on T1w slices and as a bright in DWI, easily distinguishable from surrounding brain tissues, shown as in column a). From column b), the Ground Truth illustrates the location and shape of infarcts. The segmented results are present in the graphs of c) and d). The highlighted areas indicate the intersection between the predicted regions and the ground truth.

The estimations of the two models on the T1w dataset are notably accurate, with only minor discrepancies observed. VGG16-Dual-UNet displays non-overlapping regions in its predictions for the DWI slice (row three), whereas SV-UNet predictions closely match the Ground Truth. These findings align with the evaluation metrics, indicating that post-training, both models demonstrate the ability to analyze multimodal data effectively and identify infarct areas within the experimental dataset.

Based on the evaluative results, three key findings are summarized in this section. The first finding concerns the quality of DWI data. Fig. 4(a) suggests that the proposed

models struggle to predict positive pixels (infarct pixels) in the validation set during training. Significant discrepancies in the DWI data patterns could hinder the models' ability to generalize to unseen validation data. In such an imbalanced dataset, a few incorrect predictions on positive pixels can lead to sharp variations in precision. This hypothesis is supported by the observation that both models exhibit stable learning and generalization on the T1w dataset, which contains a similar number of slices as the DWI dataset.

To further investigate how DWI data quality influences the models, future studies will involve visualizing the intensity distribution to explore signal intensity range and spread within the DWI data. Additionally, Signal-to-Noise Ratio (SNR) and Contrast-to-Noise Ratio (CNR) will be calculated for every slice to analyze signal and contrast intensities on a slice-by-slice basis [20]. By setting thresholds based on the intensity distribution analysis, high-quality slices will be selected from the pool of 3203 slices for model training. SNR serves as a metric for evaluating the signal's strength relative to background noise, with a higher value indicating well-delineated tissue signals. CNR, on the other hand, reveals the level of signal intensity differences between brain tissues. A higher CNR value signifies more pronounced and discernible contrasts within these tissues.

The second finding concerns overfitting in the DWI dataset. Despite having the same number of slices as the T1w data, both models exhibited overfitting. This suggests the models struggled to learn infarct features due to the dominance of negative pixels (healthy tissue) in each DWI slice compared to T1w data. While standard regularization techniques (e.g., Lasso, Ridge, or Dropout) were ineffective, future studies will explore Tomek Links [21]. This technique focuses on removing redundant healthy pixels from the DWI data, aiming to balance the dataset and improve model learning without losing valuable information.

The third finding reveals a complexity-interpretability trade-off with SENet. While SENet improves infarct segmentation accuracy for DWI data in the SV-UNet model, the inherent black-box nature of FCN-based models hinders interpretability. The current study lacks explanations for SENet's impact on T1w modality predictions and how it works for fine-grained DWI predictions. To address this, two techniques will be explored: Layer-wise Relevance Propagation (LRP) [22] and Local Interpretable Model-agnostic Explanations (LIME) [23]. LRP assigns importance scores to pixels during backpropagation, revealing key features for model decisions. LIME allows local analysis of the SV-UNet model, providing granular insights into its reasoning process for specific predictions.

4 Conclusion

This research introduces SV-UNet, a novel deep-learning architecture equipped with SENet attention for infarct segmentation in both T1w and DWI modalities. The SENet module refines feature maps after each up-convolution operation in the dual decoders of the network. While SV-UNet achieves comparative performance to the baseline VGG16-Dual-UNet on the T1w dataset, it demonstrates superior performance on DWI data.

This architecture is valuable considering the issues of data scarcity and class imbalance commonly faced in infarct segmentation. By combining SENet with a pre-trained VGG16 network, this approach allows the FCN-based model to achieve good performance even with limited medical data. This methodology offers a valuable reference for researchers and practitioners facing similar challenges.

The model encounters aleatoric and epistemic uncertainties. The former stems from the intrinsic variability in tissue properties within the DWI dataset. This variability can lead to unpredictable effects on model generalizability. The latter suggests that the current complex model produces unexplained predictions, hindering decisions for improvements. Additionally, SV-UNet exhibited overfitting on the DWI dataset even when the number of slices was identical to the T1w dataset. This suggests that the model may have difficulty capturing the latent patterns in the DWI data.

Future research will adopt a multi-pronged approach to solve these problems. First, a combined analysis of intensity distribution, SNR, and CNR identifies and filters high-quality slices from the DWI data for training. This will help mitigate the impact of noise and improve feature learning. Additionally, LRP and LIME will be employed to understand which image features are most critical for accurate class prediction, particularly in terms of spatial localization. This enhanced understanding will guide further model development. Finally, Tomek Links will be utilized to address the class imbalance by removing redundant pixels for the DWI dataset, mitigating overfitting issues.

References

1. Chugh, C.: Acute ischemic stroke: management approach. Indian Journal of Critical Care Medicine 23(Suppl 2), S140 (2019).
2. Kim, J., Thayabaranathan, T., Donnan, G.A., Howard, G., Howard, V.J., Rothwell, P.M., Thrift, A.G.: Global stroke statistics 2019. International Journal of Stroke 15(8), 819-838 (2020).
3. Kim, B.J., Kang, H.G., Kim, H.J., Ahn, S.H., Kim, N.Y., Warach, S., Kang, D.W.: Magnetic resonance imaging in acute ischemic stroke treatment. Journal of Stroke 16(3), 131 (2014).
4. Shin, H., Agyeman, R., Rafiq, M., Chang, M.C., Choi, G.S.: Automated segmentation of chronic stroke lesion using efficient U-Net architecture. Biocybernetics and Biomedical Engineering 42(1), 285-294 (2022).
5. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, LNCS, vol. 9351, pp. 234-241. Springer, Cham (2015).
6. Qi, K., Yang, H., Li, C., Liu, Z., Wang, M., Liu, Q., Wang, S.: X-net: Brain stroke lesion segmentation based on depthwise separable convolution and long-range dependencies. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2019, LNCS, vol. 11767, pp. 247-255. Springer, Cham (2019).
7. Yu, Y., Xie, Y., Thamm, T., Gong, E., Ouyang, J., Huang, C., Zaharchuk, G.: Use of deep learning to predict final ischemic stroke lesions from initial magnetic resonance imaging. JAMA Network Open 3(3), e200772 (2020).

8. Karimi, D., Warfield, S.K., Gholipour, A.: Critical assessment of transfer learning for medical image segmentation with fully convolutional neural networks. arXiv preprint arXiv:2006.00356 (2020).
9. Upadhyay, A.K., Bhandari, A.K.: Advances in Deep Learning Models for Resolving Medical Image Segmentation Data Scarcity Problem: A Topical Review. Archives of Computational Methods in Engineering 31(3), 1701-1719 (2024).
10. Kora, P., Ooi, C.P., Faust, O., Raghavendra, U., Gudigar, A., Chan, W.Y., Acharya, U.R.: Transfer learning techniques for medical image analysis: A review. Biocybernetics and Biomedical Engineering 42(1), 79-107 (2022).
11. Mohapatra, S., Lee, T.H., Sahoo, P.K., Wu, C.Y.: Localization of early infarction on non-contrast CT images in acute ischemic stroke with deep learning approach. Scientific Reports 13(1), 19442 (2023).
12. Pravitasari, A.A., Iriawan, N., Almuhayar, M., Azmi, T., Irhamah, I., Fithriasari, K., Ferriastuti, W.: UNet-VGG16 with transfer learning for MRI-based brain tumor segmentation. TELKOMNIKA 18(3), 1310-1318 (2020).
13. Liu, L., Kurgan, L., Wu, F.X., Wang, J.: Attention convolutional neural network for accurate segmentation and quantification of lesions in ischemic stroke disease. Medical Image Analysis 65, 101791 (2020).
14. 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).
15. Liew, S.L., Lo, B.P., Donnelly, M.R., Zavaliangos-Petropulu, A., Jeong, J.N., Barisano, G., Yu, C.: A large, curated, open-source stroke neuroimaging dataset to improve lesion segmentation algorithms. Scientific Data 9(1), 320 (2022).
16. Hernandez Petzsche, M.R., de la Rosa, E., Hanning, U., Wiest, R., Valenzuela, W., Reyes, M., Kirschke, J.S.: ISLES 2022: A multi-center magnetic resonance imaging stroke lesion segmentation dataset. Scientific Data 9(1), 762 (2022).
17. Datta, L.: A survey on activation functions and their relation with xavier and he normal initialization. arXiv preprint arXiv:2004.06632 (2020).
18. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014).
19. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980-2988 (2017).
20. Ueda, T., Ohno, Y., Yamamoto, K., Murayama, K., Ikedo, M., Yui, M., Toyama, H.: Deep learning reconstruction of diffusion-weighted MRI improves image quality for prostatic imaging. Radiology 303(2), 373-381 (2022).
21. Swana, E.F., Doorsamy, W., Bokoro, P.: Tomek link and SMOTE approaches for machine fault classification with an imbalanced dataset. Sensors 22(9), 3246 (2022).
22. Binder, A., Montavon, G., Lapuschkin, S., Müller, K.R., Samek, W.: Layer-wise relevance propagation for neural networks with local renormalization layers. In: Artificial Neural Networks and Machine Learning – ICANN 2016, LNCS, vol. 9886, pp. 63-71. Springer, Cham (2016).
23. Kumarakulasinghe, N.B., Blomberg, T., Liu, J., Leao, A.S., Papapetrou, P.: Evaluating local interpretable model-agnostic explanations on clinical machine learning classification

models. In: 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS), pp. 7-12. IEEE (2020).