electronics

MDPI

Article

UTAC-Net: A Semantic Segmentation Model for Computer-Aided Diagnosis for Ischemic Region Based on Nuclear Medicine Cerebral Perfusion Imaging

Wangxiao Li and Wei Zhang *

School of Microelectronics, Tianjin University, Tianjin 300072, China; lwangxiao@tju.edu.cn
* Correspondence: tjuzhangwei@tju.edu.cn

Abstract: Cerebral ischemia has a high morbidity and disability rate. Clinical diagnosis is mainly made by radiologists manually reviewing cerebral perfusion images to determine whether cerebral ischemia is present. The number of patients with cerebral ischemia has risen dramatically in recent years, which has brought a huge workload for radiologists. In order to improve the efficiency of diagnosis, we develop a neural network for segmenting cerebral ischemia regions in perfusion images. Combining deep learning with medical imaging technology, we propose a segmentation network, UTAC-Net, based on U-Net and Transformer, which includes a contour-aware module and an attention branching fusion module, to achieve accurate segmentation of cerebral ischemic regions and correct identification of ischemic locations. Cerebral ischemia datasets are scarce, so we built a relevant dataset. The results on the self-built dataset show that UTAC-Net is superior to other networks, with the mDice of UTAC-Net increasing by 9.16% and mIoU increasing by 14.06% compared with U-Net. The output results meet the needs of aided diagnosis as judged by radiologists. Experiments have demonstrated that our algorithm has higher segmentation accuracy than other algorithms and better assists radiologists in the initial diagnosis, thereby reducing radiologists' workload and improving diagnostic efficiency.

Keywords: cerebral ischemia; segmentation; transformer; U-Net; computer-aided diagnosis

Citation: Li, W.; Zhang, W. UTAC-Net: A Semantic Segmentation Model for Computer-Aided Diagnosis for Ischemic Region Based on Nuclear Medicine Cerebral Perfusion Imaging. *Electronics* **2024**, *13*, 1466. https://doi.org/10.3390/electronics13081466

Academic Editors: Jawahar (Jay) Kalra and Patrick Seitzinger

Received: 14 March 2024
Revised: 3 April 2024
Accepted: 9 April 2024
Published: 12 April 2024

1. Introduction

Cerebral ischemia refers to a syndrome with a range of symptoms caused by inadequate blood supply to the brain, making it difficult to meet the metabolic needs of the brain [1]. "Time is the brain" [2]. Brain injury after cerebral ischemia can lead to irreversible neuronal damage, memory loss, cognitive decline, severe brain dysfunction, and long-term motor and cognitive impairment [3–6]. Cerebral perfusion imaging (CPI) with single-photon emission computed tomography (SPECT) is a perfusion imaging of cerebral blood flow after intravenous injection of radiopharmaceuticals, which is a well-established noninvasive test in terms of the evaluation of ischemia by directly reflecting the tracer uptake of the cerebrum. The analysis of medical images is the most important technical aspect [7].

Semantic segmentation plays a central role in medical image diagnosis [8]. In the initial diagnosis, radiologists have to segment the ischemic region from the complex background, which is subjective, time-consuming, and labor-intensive. The large number of people suffering from cerebral ischemia places a heavy burden on radiologists during the initial diagnosis, so they expect computer-aided diagnosis (CAD) to complete the task of image screening. Semantic segmentation has been widely used in a variety of clinical images, which greatly assists in medical diagnosis [9]. In a variety of important scenarios, such as image-guided interactions, radiological diagnostics, and radiology [10], it can help medical staff to effectively extract lesion areas, greatly reduce work intensity, and accurately perform disease analysis. Therefore, we design a CAD network for segmenting cerebral ischemia

regions, which can construct models to complete the image screening task, achieve the accurate positioning and segmentation of ischemic sites, improve the efficiency of film reading, and carry out a meaningful attempt and exploration for the wide application of intelligent medicine.

CAD, known as the "third eye" of the doctor, improves the diagnosis efficiency by locating the lesions, extracting the effective features of the lesions, and building an auxiliary diagnosis model [11,12]. At present, CAD has attracted great attention from medical researchers and clinical practitioners. Ayus et al. [13] proposed CAD for Alzheimer's identification, which may assist doctors and healthcare workers in improving AD diagnosis. Jasphin et al. [14] proposed a CAD network that can differentiate and categorize gastric cancers from intestinal disorders. Zhong et al. [15] proposed a CAD network for mammographic breast cancer screening. The majority of CAD studies have been designed based on magnetic resonance images and computed tomography scans, and only a small proportion of CAD studies have been designed on the basis of SPECT images. For instance, CAD is employed in research on the myocardial coronary artery, lumbar lesions, lung cancer, thyroid, bone, and brain. Papandrianos et al. [16] proposed CAD of the myocardial coronary artery in SPECT myocardial perfusion imaging for classifying SPECT as normal or abnormal. Petibon et al. [17] proposed CAD of lumbar lesions in Tc-99m-MDP SPECT to achieve effective detection of lumbar lesions. Xing et al. [18] proposed CAD of lung cancer in chest SPECT to correct the attenuation of chest SPECT images for the early diagnosis and evaluation of the treatment effects of lung cancer. Kwon et al. [19] proposed CAD of the thyroid in thyroid SPECT to generate an attenuation map and automatically segment the thyroid for the automatic evaluation of thyroid uptake. Lin et al. [20] proposed CAD of bone scan in SPECT to identify whether a SPECT image includes lesions by classifying the image into categories. Ni et al. [21] proposed CAD of Alzheimer in SPECT brain perfusion images to automatically evaluate Alzheimer's disease by extracting image features of SPECT BPI. The above studies have amply demonstrated that CAD can be an effective tool to aid in diagnosis using image processing techniques and SPECT image features.

However, most CAD that has been combined with deep learning has focused on image classification and image detection. While classification can determine whether medical images are normal or abnormal, it cannot provide the precise location of the lesion. Image detection detects abnormal locations on abnormal images by drawing frames, but it cannot provide the exact size and shape of the lesion. Image segmentation can segment the exact location and size of the lesion from the background to provide essential information for the doctor's diagnosis. In the case of SPECT CPI, there are various locations and shapes of ischemic regions. Therefore, we design UTAC-Net for CAD of cerebral ischemia based on image segmentation and SPECT CPI.

The development of image segmentation changes every day. Ronneberger et al. [22] introduced skip connections in CNN and proposed U-Net. U-Net combines low-resolution and high-resolution feature maps via skip connections. U-Net has now become the standard for most medical image segmentation tasks and has stimulated numerous improvements. Xiao et al. [23] proposed ResU-Net, which combined the advantages of ResNet and U-Net. Based on U-Net, by introducing multiple residual blocks and shortcuts in the encoder and decoder, it can better extract low-frequency information and alleviate the problem of missing semantic information and the gradient vanishing problem. Zhou et al. [24] proposed UNet++, which introduces a series of nested dense skip connections to effectively narrow the semantic gap between the encoder and decoder. Patel et al. [25] proposed EU-Net, which enhances semantic information by applying three parallel multiscale patches of nonlocal attention blocks and employs spatial cross-layer attention to focus on essential features and suppress unimportant and noisy features. Schlemper et al. [26] added the attention mechanism to the U-Net segmentation network and proposed A-UNet, which autonomously learns relevant features for segmentation tasks and suppresses irrelevant features. Zhang et al. [27] proposed SAU-Net. They added the self-attention module to the U-Net segmentation network to increase the global information. This module derives

an attention diagram along the spatial dimension to achieve adaptive feature refinement. Additionally, the original U-Net convolution block is replaced with a structured dropout convolution block to prevent network overfitting. Previous studies have demonstrated the high accuracy of U-Net with minimal data. Thus, we propose UTAC-Net as an improved version that employs U-Net as the backbone.

However, the operation of CNN is subject to the receptive field, which restricts its ability to establish long-range dependencies. Chen et al. [28] introduced Transformer with the advantage of self-attention mechanism into the U-Net network and proposed TransUNet. It takes Transformer as a powerful encoder to extract global features. However, the lack of low-level detailed features in this model leads to issues with positioning. Wang et al. [29] put forward UCTransNet, which uses the CTrans module to replace skip connections in U-Net. It explores the global contextual information of a multiscale and narrows the semantic gap between low-level and high-level features. Liu et al. [30] proposed Swin Transformer, a hierarchical Transformer that represents computation by moving windows. This is performed by restricting the self-attentive computation to nonoverlapping local windows, while allowing cross-window connections. Zhang et al. [31] proposed TransFuse, a parallel branching architecture, which fuses Transformer and CNN in a parallel fashion to achieve a shallow network architecture to model global relationships and underlying details. The fusion of Transformer and CNN branches is performed by the newly proposed BiFusion module. Zhu et al. [32] proposed brain tumor segmentation based on the fusion of deep semantics and edge information in multimodal MRI, aiming to make fuller use of multimodal information for accurate segmentation. A shifted patch tokenization strategy is introduced into Swin Transformer to extract semantic features, and the edge spatial attention block is introduced into CNN to achieve edge detection. A multifeature inference block is designed to implement feature fusion. Ma et al. [33] proposed MedSAM, which is a generic medical image segmentation. The network architecture includes three components: an image encoder, a prompt encoder, and a mask decoder. To minimize the computational cost of adapting SAM to the medical image segmentation, the image encoder and prompt encoder are frozen and only the mask decoder is fine-tuned. MedSAM improves the ability to detect small objects and effectively reduces interference from objects around the segmentation target, thus reducing outliers. Fu et al. [34] proposed a novel multiscale network, HmsU-Net, based on CNN and Transformer, which facilitates the effective interaction of the extracted multiscale information. At the same time, they designed a lightweight attention mechanism to reduce the computational cost of the standard Transformer.

Previous studies indicate that U-Net is unable to fully extract global information due to its own receptive field, while Transformer benefits from establishing long-range dependence to compensate for the shortcomings of U-Net. Consequently, we proposed UTAC-Net based on U-Net and Transformer to extract rich features. It can combine the advantages of U-Net and Transformer to achieve accurate segmentation of SPECT CPI for cerebral ischemia.

The accuracy and reliability of diagnostic results can be directly influenced by the segmentation of medical images [35]. The shape and position of cerebral ischemic regions in CPI are highly variable, and they are essential for the identification of abnormal regions. The above studies for other diseases would be much less accurate if used in the diagnosis of cerebral ischemic diseases. Certain networks encounter difficulties extracting boundary information. In light of the characteristics of cerebral ischemia and the aforementioned issues with certain networks, we propose an improved network, UTAC-Net. It enables more precise segmentation of ischemic regions with distinct boundaries and accurate positions. UTAC-Net can achieve fast and objective diagnosis without the interference of radiologists' subjective factors and reduce manual processing time. CAD of cerebral ischemia improves the diagnostic efficiency of cerebral ischemia by assisting radiologists' diagnoses.

The CAD network of cerebral ischemia has several difficulties: the cerebral ischemia dataset is small due to difficulties in acquiring the dataset; the shape and position of cerebral ischemic regions in CPI are highly variable, and they are essential for the segmentation of

abnormal regions; and the boundary features of the abnormal regions in CPI are rich, so general networks cannot extract these irregular boundary features.

Aiming at the problems existing in the field of CAD for cerebral ischemia on a SPECT image, we propose a new algorithm with the main contributions as follows:

- We propose a novel segmentation network, called UTAC-Net, which adopts a branching code consisting of U-Net with Transformer, the attention branching fusion module (ABFM), and the contour-aware module (CAM).
- The dual-branch encoder combines the advantages of Transformer and U-Net. U-Net pays more attention to local features with complex details, and can effectively learn and recover the details of the image; Transformer pays more attention to global features and, through self-attention, deals with multiple elements in the sequence simultaneously to achieve highly parallel computation, which can capture long-range dependencies between different positions. The two branches form a complementary structure to extract more features on the SPECT image.
- ABFM selectively fuses global and local features to highlight those relevant to the segmentation task, and consists of the channel attention module, the spatial attention module, and the convolutional block attention module (CBAM) composition. The channel attention module filters the local features extracted by U-Net, the spatial attention module filters the global features extracted by Transformer, and the filtered features are fused and fed into CBAM to achieve full feature adaptation in both channel and spatial dimensions. This allows the network to highlight important features and suppress unimportant features.
- CAM fuses the contour features containing different scales from the decoding stage to further clarify the contour of the ischemic region. CAM performs an evolutionary deformation of the contour vertices of the ischemic region by the vertex iteration method designed in this paper so that the vertices keep approaching the contour of the ischemic region.

2. Materials and Methods

2.1. Overview

The structure of UTAC-Net is shown in Figure 1, adopting an architecture of encoder–decoder. The encoder uses a two-branch coding structure with U-Net and Transformer in parallel. The U-Net branch extracts the local features, while the Transformer branch extracts the global features. The local features and global features extracted by the dual encoder are then fused by the attention branching fusion module (ABFM). ABFM fuses the features extracted from the two branches with different weight values. The fused features are fed into a common decoder for upsampling to recover the image size. Different layers of features input the contour-aware model (CAM) to extract the contour information that may have been lost in the upsampling process. Finally, we refine the contour by making the extracted contour features complementary to the feature maps output from the backbone.

2.2. Transformer

Although U-Net has good performance, its limited receptive field means that it only works on local features and cannot capture long-range and global semantic information [8,36]. Transformer is able to establish long-range connections and has an advantage in capturing global information [37]. However, if the dataset is small, Transformer will overfit.

Inspired by TransUNet [28] and UCTransNet [29], we combine Transformer with U-Net to extract global and local information. Inspired by the article dual encoder [38,39], we consider using the Transformer encoder to form a dual encoder with the U-Net encoder. The global and local features extracted by the dual encoder are then fused by ABFM. The fused features are fed into a common decoder for upsampling to recover the image size. In order to extract the rich contour information in the abnormal regions, different layers of features decoded at different stages input the contour-aware model to extract the contour information that may be lost during the upsampling process.

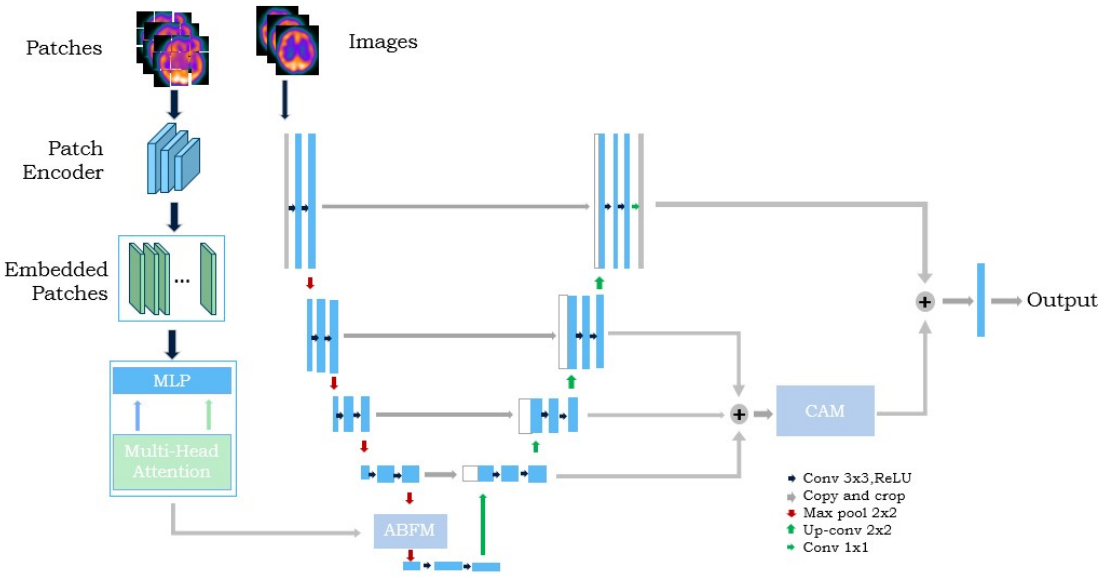

Figure 1. Overview of UTAC-Net.

The key to the Transformer encoder is the attention, by which we mean finding the features in the input image that are relevant to the ischemic region and strengthening their relevance, while weakening irrelevant features. The Transformer encoder consists of a stack of N encoder layers, each encoder layer consists of two sublayer connection structures, the first sublayer connection structure includes a multi-self-attention (MSA) layer and a LayerNorm layer as well as a residual connection, where the multi-self-attention layer is obtained by parallel expansion of multiple self-attention layers. The second sublayer connection structure consists of a feed forward network (FFN) layer and a LayerNorm layer and a residual connection, where the feed forward network consists of two multi-layer perceptrons (MLPs). The image $x \in R_{H*W*C}$ is first cut into a sequence of patches $X = [X_1, X_2, \ldots, X_n] \in R^n * P^2 * C$, where (P, P) is the size of the patch, $n = (H * W)/P^2$ is the number of patches, and C is the number of channels. After each patch is tiled, it is converted into a patch embedding by linear projection to obtain a sequence of patch embeddings $X_0 = [E - X_1, \ldots, E - X_n] \in R^n * D$, where $E \in R^D * (P^2C)$. Since Transformer cannot recognize the position information, we add the position embedding $pos = [pos_1, \ldots, pos_n] \in R^n * D$ to the patch sequence to obtain the token $Z_0 = X_0 + pos$. We then exploit a stack of Transformer blocks encompassing an MSA and an MLP to learn the long-range contextual representation. As shown in Figure 2a, $Z_0 = [a_1, \ldots, a_n] \in R^n * D$ is input into three Transformation matrices in Transformer: W_q, W_k, W_v to obtain the corresponding values of q^i, k^i, v^i $(i = 1, 2, 3, \ldots, n)$, respectively. While q is the query, k is the key, v is the extracted information, and d is the length of vector k^i. The correlation between tokens is calculated by a dot product operation QK^T: the current token is dot-produced with the tokens in the whole image, and if the correlation of that token is large, the result of the dot product will be large. After obtaining the correlation, in order to highlight the tokens with high correlation, we will give different weights to the tokens; i.e., the greater the correlation, the greater the weight. The weights are obtained quantitatively using the weighted summed Softmax correlation, which measures the average correlation of the current token among all the tokens by solving for the expected, and this result is the weight value. Its calculation formula is shown in Equation (1). As shown in Figure 2b, b_i is obtained by weighting each v_i.

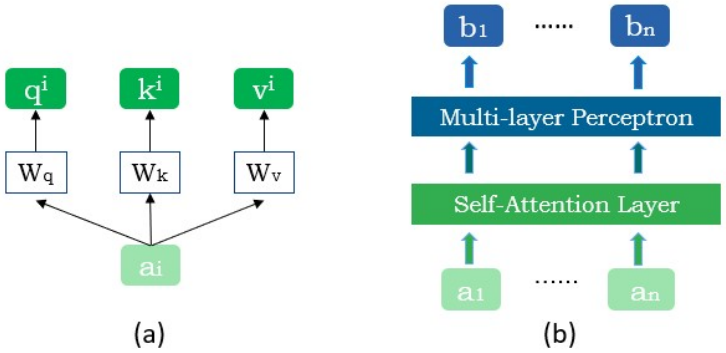

Figure 2. The transformation matrices (**a**); the process of weighting (**b**).

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{1}$$

The MSA consists of M parallel attention heads that accept multiple keys, queries, and values; generates multiple attention outputs from multiple scaled dot product attention blocks; and finally joins the multiple attentions to obtain a final attention t'_i. Its calculation formula is shown in Equation (2). Then t'_i inputs the LayerNorm, which takes the original embeddings, adds them to the embeddings of the multiple attention heads, and then normalizes them to a standard normal distribution with a mean of 0 and a variance of 1. The results of the LayerNorm are sent to the FFN to map and extract the input features. Its calculation formula is shown in Equation (3).

$$t'_i = MSA(Norm(t_{(i-1)}) + t_{(i-1)}) \tag{2}$$

$$t_i = MLP(Norm(t'_i) + t'_i) \tag{3}$$

2.3. Attention Branching Fusion Module

An attention module enhances important features and suppresses unimportant features to improve the representational ability of the network [27,28,40,41]. In order to effectively combine the global features extracted by Transformer with the local features extracted by the U-Net encoder, we design an attention mechanism to fuse the features of the two branches. Inspired by the convolutional block attention module (CBAM) [42], we propose the attention branching fusion module that mixes the combined channel information with spatial information. It is able to make the fusion process more focused on important features and suppress unnecessary features. CBAM is able to sequentially generate attentional feature maps in both channel and spatial dimensions. The input features are sequentially filtered by the channel attention module and the spatial attention module, and finally, the recalibrated features are obtained. The important features are emphasized, and the unimportant features are compressed. We propose a new attention module, ABFM, based on CBAM.

Since the channel attention is more focused on which features are of interest, we take the local features extracted by the U-Net encoder and first go through the channel attention to make the network more attentive to the local features in the channel information. The relative spatial attention is more focused on where the features of attention are, so we will take the global features extracted by Transformer and input the spatial attention to make the network pay more attention to the global features in the spatial information. The features adjusted by different attention modules are fused and input CBAM to perform all-round feature adjustment in both channel and spatial dimensions. This makes the network more selective to focus on the salient parts, emphasizing the important features and suppressing the unimportant ones.

Our proposed attention module is shown in Figure 3. X is the feature extracted from the U-Net branch, and Y is the feature extracted from the Transformer branch. X inputs the channel attention module to generate the channel attention weight X' by Equations (4) and (5), where σ is a sigmoid calculation, and $\otimes$ denotes pixel-by-pixel multiplication. Y inputs the spatial attention module to generate the spatial attention weight Y′ by Equations (6) and (7), and $f^{7\times7}$ denotes the convolution operation with a filter size of 7×7. Referring to Table II in [42], when comparing different convolutional kernel sizes, it is found that larger convolutional kernel sizes produce better accuracy. A large field of view (i.e., a large receptive field) is necessary to determine spatially important regions, and we used a convolutional layer with a large kernel size to compute spatial attention. X1 and Y1 are added at the same location to obtain F. F inputs the CBAM module. In the CBAM module, F is first input to the channel attention module to generate F′, by Equations (8) and (9), where σ is a sigmoid calculation, and $\otimes$ denotes pixel-by-pixel multiplication. F′ is then input to the spatial attention module to generate F″ by Equations (10) and (11), and $f^{7\times7}$ denotes the convolution operation with a filter size of 7×7. The channel attention and spatial attention are applied sequentially in CBAM so that the features of each branch generate the attention weights A independently and complementarily.

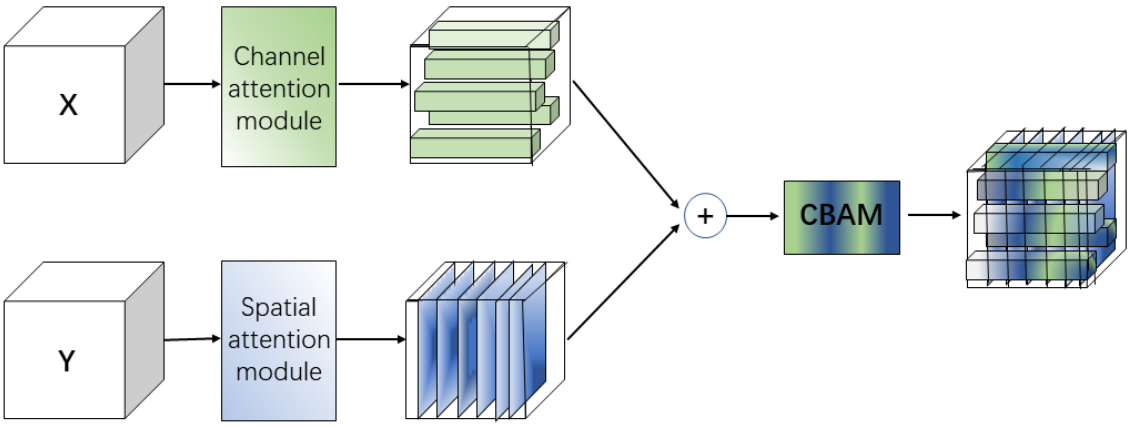

Figure 3. Attention branching fusion module.

$$M_C(X) = \sigma(MLP(AvgPool(X) + MLP(MaxPool(X))) \tag{4}$$

$$X' = Mc(X) \otimes X \tag{5}$$

$$M_S(Y) = \sigma(f^{7\times7}([AvgPool(Y); (MaxPool(Y)])) \tag{6}$$

$$Y' = M_S(Y) \otimes Y \tag{7}$$

$$M_C(F) = \sigma(MLP(AvgPool(F) + MLP(MaxPool(F))) \tag{8}$$

$$F' = Mc(F) \otimes F \tag{9}$$

$$M_S(F') = \sigma(f^{7\times7}([AvgPool(F'); (MaxPool(F')])) \tag{10}$$

$$F'' = M_S(F') \otimes F' \tag{11}$$

2.4. Contour-Aware Module

The unclear boundary between ischemic and normal regions is a challenge for ischemic region segmentation. Inspired by the boundary-aware model [43], we propose the contour-aware module. Since the shallow features contain a large amount of boundary information, different scales of shallow features input to the contour-aware module are fused in the decoding stage of U-Net to further clarify the contours of the ischemic regions. The contours extracted by CAM are fused with the segmentation map output by the U-Net backbone network, and the two processes are kept independent and complementary.

Inspired by the boundary patch refinement in [44], we propose boundary point classifier refinement. Our method deforms the contour evolution process with a vertex classifier. As shown in Figure 4, we propose the vertex iteration method of CAM, where the vertices are continuously moved close to the boundaries of the ischemic regions. First, the initial contour $C^{(0)} : (x_i | i = 1, \ldots, N)$ is obtained by the decoder of the backbone, and then this series of vertices x_i are moved along their normal direction to continuously iterate towards the boundary. The vertex classifier $\phi(x_i)$ is used to predict the state of the vertex, determining its relative position (in or out of the region) to the ischemic region. The relative position is used to determine whether the vertex is moving along a positive or negative normal direction. For each vertex x_i, the vertex classifier $\phi(x_i)$ returns a scalar in the range $[0, 1]$, indicating whether the vertex is outside the region (e.g., $\phi(x_i) = 1$) or inside the region (e.g., $\phi(x_i) = 0$). If $\phi(x_i) = 0.5$, the vertex classifier is not sure whether the vertex is outside or inside; this means that the vertex may be on the boundary. Equation (12) uses $(\phi(x_i) - 0.5)$ to determine if that vertex is on the boundary. In Equation (13), d_i indicates the positive and negative direction in which the vertex moves along the normal. If $phi(x_i) = 1$, the vertex is outside the region, so let $d_i = -1$. If $phi(x_i) = 0$, the vertex is inside the region, so let $d_i = 1$. The vertex moves step by step along d_i, checking the value of $phi(x_i)$ at the current vertex after each move. If the vertex moves from the inside of the boundary to the outside (or from the outside to the inside), then the vertex stops to move. Then the vertex is called the position flip point and the number of moves from the original position to the flip point is s_i. We perform Equations (12) and (13) successively for each vertex of $C^{(0)}$ to obtain a series of contours $C^{(1)}, C^{(2)}, C^{(3)}, \ldots, C^{(n)}$, until all vertices stop moving; then the contour generated is the actual contour of the region.

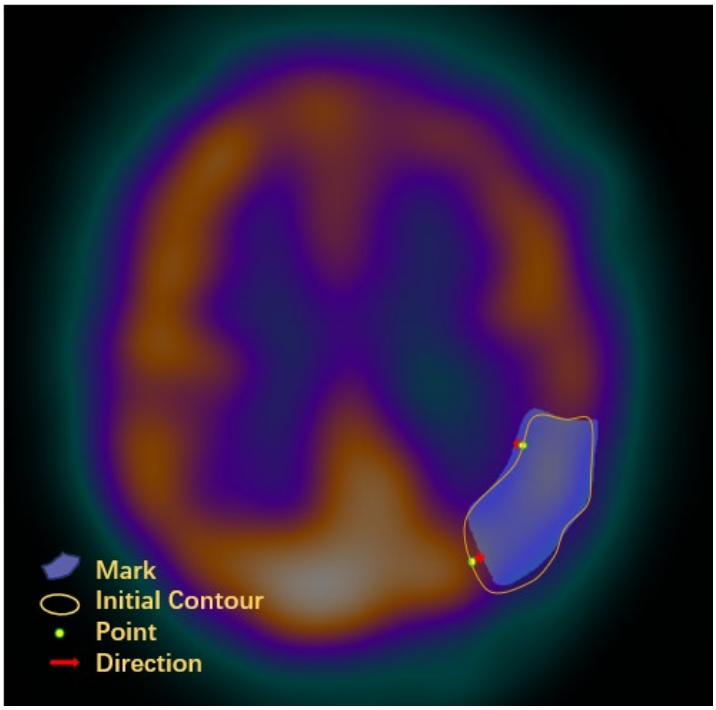

Figure 4. Contour-aware module.

$$s_i = \lambda|\phi(x_i) - 0.5| \tag{12}$$

$$x_i' = x_i + s_i d_i \tag{13}$$

where d_i is the direction of the movement, x_i' is the vertex corresponding to the next contour $C^{(')}$, and s_i is the moving step.

3. Results

3.1. Datasets

To evaluate the effectiveness of UTAC-Net, we test it on two medical datasets: the CPI dataset and the ISIC 2018 dataset.

The CPI dataset: it was collected from 97 patients, 60% male and 40% female, with a maximum age of 78 years and a minimum age of 33 years. All participants signed a written informed consent form. The imaging agent used in all patients was Tc-99m-ECD, a small molecule lipophilic imaging agent that freely crosses the blood–brain barrier and stably resides in the brain. The scanner is a Discovery NM/CT 670 CZT. The dataset contains 798 SPECT images, as shown in Figure 5, covering 12 annotated sites on both sides of the brain, including the frontal lobe, temporal lobe, parietal lobe, occipital lobe, thalamus lobe, and basal ganglia.

The CPI dataset is selected and annotated by three experienced radiologists. To ensure the true form of CPI, no preprocessing, such as data enhancement, is performed. Because the label type of the ischemic site is related to the location of the ischemic region, this study does not use methods such as rotation and folding to expand the dataset, but uses real cases, which has practical clinical significance. We separated these images randomly. Due to the small size of the dataset, ten-fold cross-validation is used to conduct experiments, randomly dividing the dataset into 10 parts, selecting 9 of them in turn to be used as the training set, and the remaining 1 to be used as the test set, and finally averaging the results over the 10 times.

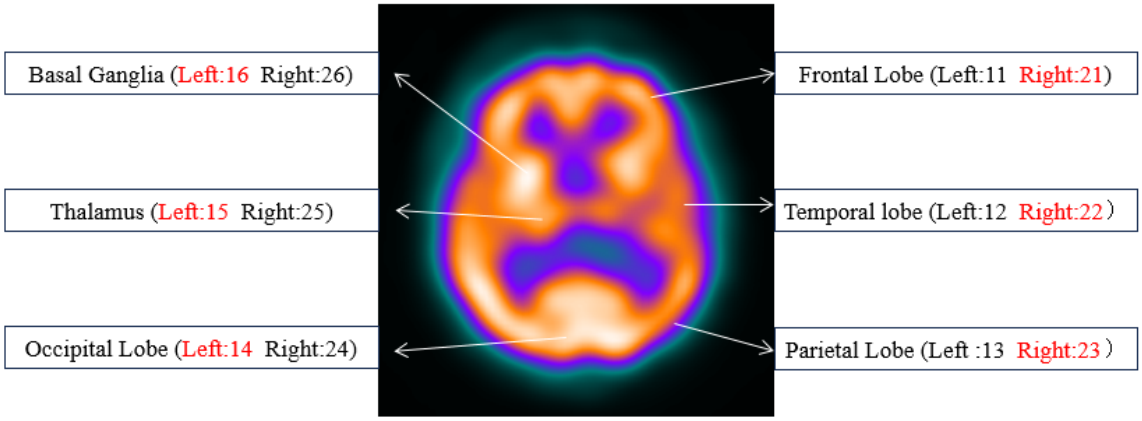

Figure 5. SPECT image with 12 annotated sites.

An example of the dataset is shown in Figure 6. The symmetry and continuity of the distribution of the cortical contrast medium in the SPECT images, i.e., the contrast of the color grades in the bilateral images, were observed. If a decrease in radioactivity or a defect is detected, it is an abnormality. Normal and abnormal sites are nested and fused with each other, and their boundaries are blurred and difficult to distinguish. Moreover, the distribution of ischemic regions is random, with various shapes and rich boundary information. In this study, a new segmentation network is proposed to achieve accurate segmentation of ischemic regions based on these characteristics.

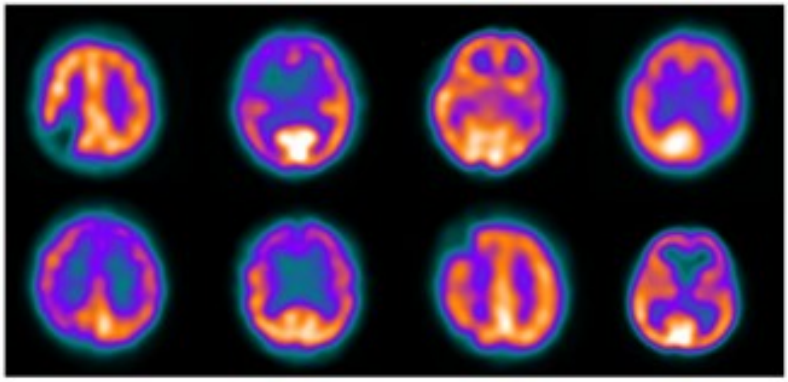

Figure 6. Example of the dataset.

The ISIC 2018 dataset: it is a large medical public dataset, which also features a random distribution of abnormal sites with different shapes, fuzzy boundaries, and rich features. The dataset contains dermoscopic images from different centers released by the ISIC 2018 challenge. The dataset has a total of 2594 images and corresponding annotations with 1868 images in the training set, 467 images in the validation set, and 259 images in the test set.

3.2. Experimental Configuration

The experimental platform is configured as follows: Intel(R) Xeon(R) Silver 4310 CPU @ 2.10 GHz, Tesla A40 GPU, and Ubuntu 18.04.6 LTS operating system. The experiments are implemented using PyTorch 1.10.1. To ensure fairness, all input images are resized to 512 $\times$ 512 using the same setup. The SGD optimizer is used. As an important superparameter in deep learning, the learning rate determines if and when the objective function converges to a local minimum. The higher the learning rate, the greater the magnitude of model parameter updates, leading to model dispersion. The smaller the learning rate, the smaller the magnitude of the model parameter update, resulting in slow model convergence. Therefore, in practice, the initial learning rate is usually 0.001. Momentum is another important parameter of the SGD algorithm, which helps the SGD algorithm to converge faster during optimization and prevents the model from stagnating at the local optimum. Momentum is normally set to 0.9. The decay rate is a parameter used to control the learning rate, allowing the learning rate to gradually decrease during the training process, making the model more stable as it approaches convergence. The decay rate is typically set to 0.0001. Batch_size is the size of the dataset for each training input in deep learning. When training neural networks, the dataset is usually divided into small batches for training, and each small batch contains multiple data samples, and setting the appropriate batch_size can speed up the training speed and improve the training effect. Therefore, the number of batches is 4.

3.3. Evaluation Metrics

In order to evaluate the comparative effectiveness of UTAC-Net and other segmentation networks, the Dice similarity coefficient and the intersection over union (IoU), which are commonly used in medical image segmentation, are used as an objective evaluation metric and combined with the subjective comparisons.

Dice: it is a set similarity measure function to calculate the similarity of two samples, taking the value [0, 1], as shown in Equation (14).

$$Dice = \frac{2 * (pred \bigcap true)}{pred \bigcup true} = \frac{2TP}{FP + 2TP + FN} \tag{14}$$

IoU: it is used to measure the degree of overlap between the predicted bounding box and the real bounding box, taking the value [0, 1], as shown in Equation (15).

$$IoU = \frac{(pred \bigcap true)}{(pred) + (true) - (pred \bigcap true)} = \frac{TP}{TP + FP + FN} \quad (15)$$

3.4. Comparison with Other Methods on the CPI Dataset

This section presents the training and testing results of UTAC-Net and other improved networks based on U-Net on the CPI dataset.

3.4.1. Quantitative Results

As shown in Tables 1 and 2, the objective evaluation of U-Net and other improved networks based on U-Net and UTAC-Net, all of the above networks are replicated in the same experimental environment and without using any data augmentation methods. The comparison shows that UTAC-Net has the best segmentation in 12 categories, with an mDice coefficient 9.16% higher than U-Net, 5.68% higher than SAU-Net, 3% higher than TransUNet, 1.71% higher than UCTransNet, and 3.52% higher than HmsU-Net. The mIoU of UTAC-Net is 14.06% higher than U-Net, 8.99% higher than SAU-Net, 4.86% higher than TransUNet, 2.8% higher than UCTransNet, and 5.23% higher than HmsU-Net.

Table 1. Dice (%) of different networks on the CPI dataset.

Categories	U-Net	SAU-Net	TransUNet	UCTransNet	HmsU-Net	UTAC-Net
background	99.64	99.64	99.67	99.68	99.65	99.69
11	86.14	84.09	92.25	92.51	90.15	92.72
21	86.90	88.59	82.84	91.57	89.22	92.31
12	81.94	82.24	88.07	84.87	84.42	86.84
22	73.90	85.70	85.31	85.46	84.86	85.81
13	76.08	87.59	87.73	88.50	85.18	88.65
23	81.53	82.01	89.92	89.30	86.50	90.92
14	77.51	80.02	80.20	86.39	78.61	87.13
24	76.82	78.96	79.62	83.76	78.82	86.67
15	75.66	76.31	83.77	85.00	85.46	86.50
25	77.88	77.65	85.70	85.29	85.52	87.92
16	86.01	87.79	86.88	88.03	90.50	94.48
26	84.24	86.01	90.42	89.87	90.32	93.03
mDice	81.19	84.67	87.35	88.64	86.83	90.35

Table 2. IoU (%) of different networks on the CPI dataset.

Categories	U-Net	SAU-Net	TransUNet	UCTransNet	HmsU-Net	UTAC-Net
background	99.28	99.28	99.34	99.36	99.30	99.38
11	65.65	72.55	85.61	86.06	82.07	86.43
21	66.83	79.52	70.71	84.45	80.54	85.72
12	69.40	69.83	78.68	73.71	73.04	76.74
22	58.61	74.98	74.38	74.61	73.70	75.15
13	61.39	77.92	78.14	79.37	74.19	79.61
23	68.82	69.51	81.69	80.66	76.21	83.36
14	63.28	66.70	66.96	76.04	64.76	77.20
24	62.37	65.23	66.15	72.06	65.04	76.48
15	60.85	61.69	72.07	73.91	74.61	76.21
25	63.78	63.46	74.98	74.36	74.70	78.45
16	75.45	78.23	76.81	78.62	82.65	89.53
26	72.77	75.45	82.51	81.60	82.35	86.96
mIoU	68.34	73.41	77.54	79.60	77.17	82.40

3.4.2. Qualitative Results

As shown in Figure 7, the subjective evaluation of U-Net and other improved networks based on U-Net and UTAC-Net, the images in the first column are the input images;

the images in the second column are images annotated under the guidance of clinically experienced radiologists; the images in the third, fourth, fifth, sixth, and seventh columns are the output images of the U-Net network and other improved networks based on U-Net; and the last column shows the output images of UTAC-Net. Comparing the six output images with the radiologists' annotated images, in the first row, when segmenting the ischemic region in the left frontal lobe, the boundaries of the output images of other improved networks based on U-Net in the blue box are rough, while the boundary of the output image of UTAC-Net is clear; in the second row, for the segmentation of the ischemic region in the right parietal lobe, the sizes of the output images of other improved networks based on U-Net in the yellow box are imprecise, while the size of the output image of UTAC-Net is precise. In the third row, for the segmentation of multicategory ischemic regions, when segmenting the ischemic region in the left frontal lobe, the output images of other improved networks based on U-Net in the green box are imprecise in size and have rough boundaries, while the output image of the UTAC-Net network is precise in size and has clear boundary; when segmenting the ischemic region in the left occipital lobe, the output images of other improved networks based on U-Net in the red box are imprecise in size and have rough boundaries, while the output image of the UTAC-Net network is precise in size and has clear boundary; and when identifying the ischemic region in the left occipital lobe, the output images of other improved networks based on U-Net in the red box are incorrectly segmented due to interference from other categories. Among them, U-Net annotates the ischemic region in the left occipital lobe as the left parietal lobe, and SAU-Net, TransUNet, UCTransNet, and HmsU-Net partially annotate the ischemic region in the left occipital lobe as the left parietal lobe. However, the output image of UTAC-Net is unaffected by interference, and the category is correct with clear boundaries, accurate size, and excellent performance.

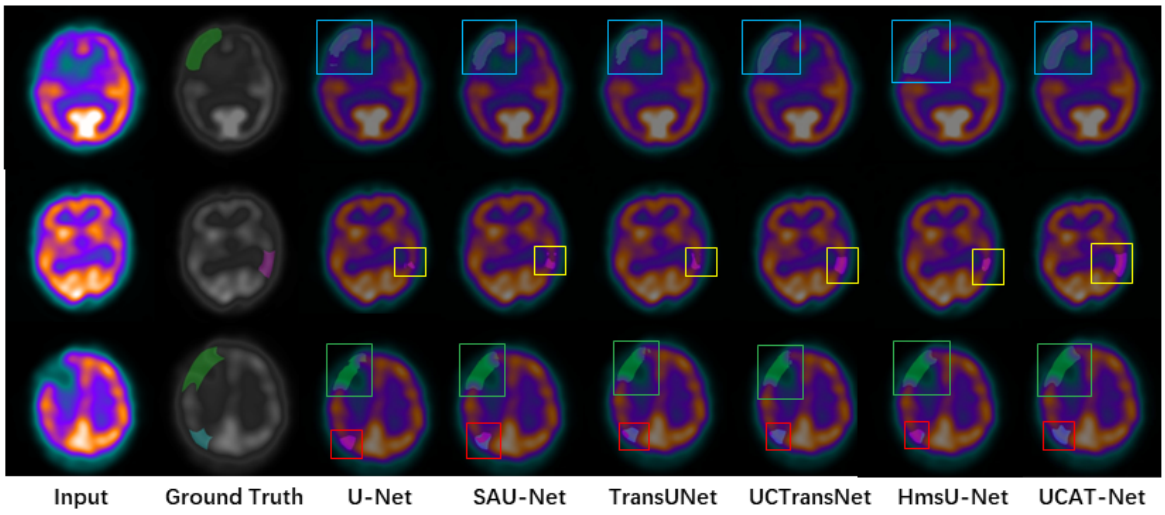

Figure 7. Subjective evaluation of other networks and UTAC-Net on the CPI dataset. Use color boxes to highlight where there are significant differences in the comparative results.

An objective evaluation and subjective judgment of the results of UTAC-Net compared with other networks shows that UTAC-Net can not only accurately segment ischemic regions with clear boundaries and precise sizes from the complex nested background, but also correctly identify the type of ischemia. UTAC-Net is resistant to interference and meets the requirements for lesion recognition in the medical field. The importance of the dual encoder and ABFM can be fully illustrated by the correct identification and accurate segmentation of different ischemic regions by excluding the interference of a similar degree; the smooth boundaries also fully illustrate the key role of CAM.

3.5. Comparison with Other Methods on the ISIC 2018 Dataset

To further evaluate the effectiveness of UTAC-Net in feature segmentation, we performed a comparative experiment on ISIC 2018. This section presents the training and testing results of UTAC-Net and other improved networks based on U-Net on ISIC 2018.

The comparative results are shown in Table 3. The Dice coefficient of UTAC-Net is 2.51% higher than that of U-Net, 1.14% higher than that of U-Net++, 1.32% higher than that of ResU-Net, 5.85% higher than that of Swin-Unet, 1.34% higher than that of DCSAU-Net, 1.94% higher than that of GAU-Net, 1.79% higher than that of MCNMF-Unet, 2.86% higher than that of LeaNet, and 0.17% higher than that of MDU-Net. The mIoU of UTAC-Net is 4.19% higher than that of U-Net, 1.93% higher than that of U-Net++, 2.23% higher than that of ResU-Net, 9.48% higher than that of Swin-Unet, 3.26% higher than that of GAU-Net, 3.01% higher than that of MCNMF-Unet, and 5.83% higher than that of LeaNet.

Table 3. Comparison of different networks on the ISIC 2018 dataset.

Network	Dice (%)	IoU (%)
U-Net [22]	89.24	80.57
U-Net++ [24]	90.61	82.83
ResU-Net [23]	90.43	82.53
Swin-Unet [45]	85.90	75.28
DCSAU-Net [46]	90.41	84.10
GA-UNet [47]	89.81	81.50
HmsU-Net [34]	91.85	84.93
MCNMF-Unet [48]	89.96	81.75
LeaNet [49]	88.89	78.93
MDU-Net [50]	91.58	84.81
UTAC-Net(Ours)	91.75	84.76

Comparisons show that UTAC-Net has excellent segmentation performance for features with varied shapes and fuzzy boundaries, with a Dice value of 91.75% and an IoU value of 84.76%, which is better than most networks.

3.6. Ablation Study

We propose a segmentation network, UTAC-Net, based on U-Net and Transformer, which includes CAM and ABFM. In order to show more clearly the effect of each improvement module on the segmentation effect of the network and to evaluate the role of each module in the network, ablation experiments were conducted based on U-Net. The experiments include U-Net and Transformer dual-branch coding, the ABFM module, and CAM. The ablation experiments add the above modules sequentially to the base network, and the results of the ablation experiments are shown in Table 4, where "√" means that the module is added and "-" means that the module is not added. The results of the ablation experiment for each module are analyzed in detail in the following section.

Table 4. Ablation experiments.

Network	Module			Metric	
	Transformer	ABFM	CAM	mDice (%)	mIoU (%)
U-Net	-	-	-	81.19	68.34
Ours	✓	-	-	87.35	77.54
Ours	✓	✓	-	88.15	78.81
Ours	✓	✓	✓	90.35	82.40

3.6.1. Transformer

The Transformer branch is added to U-Net to form a dual-branch coding structure to establish an effective joint CNN and Transformer mechanism. As shown in Table 4,

compared with the U-Net network, mDice is improved by 6.16% and mIoU is improved by 9.2% after adopting the dual-branch coding structure of CNN and Transformer. This shows that the improved network has significantly improved the segmentation effect of the whole network. U-Net focuses on local details and extracts local feature information, while the Transformer branch establishes long-distance dependency and extracts global feature information. The two parts form a dual-branch coding structure, which enables the network to extract more features, thus improving the segmentation accuracy of the network.

3.6.2. Attention Branching Fusion Module

On this basis, ABFM is added to fuse the multiscale features extracted from the two-branch coding structure. As shown in Table 4, after adding ABFM, mDice is improved by 0.8% and mIoU is improved by 1.27%, and the improved network has further improved the segmentation effect of the network. The experimental data show that ABFM fuses the multiscale features extracted from the two-branch coding structure. The parallel fusion of local feature information and global feature information strengthens the attention of the network to features relevant to the segmentation task, suppresses the interference of irrelevant features, and improves the segmentation accuracy of the network.

3.6.3. Contour-Aware Module

CAM is then added on this basis to further refine the contour information of the segmented image. As shown in Table 4, after adding CAM, mDice is improved by 2.2% and mIoU is improved by 3.59%, and the improved network has improved the segmentation effect of the network. This indicates that CAM helps the network to extract richer contour features, and complements the extracted contour features with the feature maps output from the backbone network to achieve precise outlining of the boundaries of the ischemic region, so that the output of the network is closer to the real value, thus improving the segmentation accuracy of the network.

4. Discussion

Based on U-Net and Transformer, we designed a new segmentation network, UTAC-Net. Experiments on a self-constructed dataset demonstrated that the network can segment ischemic regions in CPI SPECT images with excellent accuracy, consistency, and clinical utility. Experiments on publicly available datasets (ISIC 2018) demonstrate the excellent segmentation performance of UTAC-Net for features with varying shapes and fuzzy boundaries. It can therefore help doctors segment lesions in medical images as a diagnostic aid and assist them in their clinical practice. In the future, we will continue to develop new models for other diseases in medical images.

Author Contributions: Conceptualization, W.L.; data curation, W.L.; methodology, W.L.; resources, W.Z.; software, W.L.; supervision, W.Z.; validation, W.L.; visualization, W.L.; writing—original draft, W.L.; writing—review and editing, W.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are not available for privacy reasons.

Acknowledgments: We are grateful for the support of the Tianjin Medical University General Hospital.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CNN	convolutional neural network
CPI	cerebral perfusion imaging
SPECT	single-photon emission computed tomography
CAD	computer-aided diagnosis
ABFM	attention branching fusion module
CAM	contour-aware module
MSA	multihead self-attention
FFN	feed forward network
MLP	multilayer perceptron
CBAM	convolutional block attention module
Tc-99m-ECD	Technetium-99m-Ethyl Cysteinate Dimer
IoU	intersection over union
TP	true positives
TN	true negatives
FP	false positives
FN	false negatives

References

1. Wang, S.; Liu, F.; Ying, C.; Gao, C.; Zhang, Y. Action Mechanism of Traditional Chinese Medicine Combined with Bone Marrow Mesenchymal Stem Cells in Regulating Blood-brain Barrier after Cerebral Ischemia Reperfusion Injury. *Chin. J. Tissue Eng. Res.* **2023**, *27*, 5377.
2. Burns, M.M.J.; Pomager, L.; Young, B. Time is brain. *Am. Nurse J.* **2023**, *18*, 6–12. [CrossRef]
3. Zhang, H.Y.; Tian, Y.; Shi, H.Y.; Cai, Y.; Xu, Y. The Critical Role of the Endolysosomal System in Cerebral Ischemia. *Neural Regen. Res.* **2023**, *18*, 983. [CrossRef]
4. Shi, G.; Feng, J.; Jian, L.Y.; Fan, X.Y. DNA Hypomethylation Promotes Learning and Memory Recovery in A Rat Model of Cerebral Ischemia/Reperfusion Injury. *Neural Regen. Res.* **2023**, *18*, 863. [CrossRef] [PubMed]
5. Kahl, A.; Blanco, I.; Jackman, K.; Baskar, J.; Milaganur Mohan, H.; Rodney-Sandy, R.; Zhang, S.; Iadecola, C.; Hochrainer, K. Cerebral Ischemia Induces the Aggregation of Proteins Linked to Neurodegenerative Diseases. *Sci. Rep.* **2018**, *8*, 2701. [CrossRef] [PubMed]
6. Negredo, P.N.; Yeo, R.W.; Brunet, A. Aging and Rejuvenation of Neural Stem Cells and Their Niches. *Cell Stem. Cell* **2020**, *27*, 202–223. [CrossRef] [PubMed]
7. Haggenmüller, B.; Kreiser, K.; Sollmann, N.; Huber, M.; Vogele, D.; Schmidt, S.A.; Beer, M.; Schmitz, B.; Ozpeynirci, Y.; Rosskopf, J.; et al. Pictorial Review on Imaging Findings in Cerebral CTP in Patients with Acute Stroke and Its Mimics: A Primer for General Radiologists. *Diagnostics* **2023**, *13*, 447. [CrossRef] [PubMed]
8. Shamshad, F.; Khan, S.H.; Zamir, S.W.; Khan, M.H.; Hayat, M.; Khan, F.S.; Fu, H. Transformers in Medical Imaging: A Survey. *Med. Image. Anal.* **2022**, *88*, 102802. [CrossRef]
9. Wang, Y.; Yu, X.; Yang, Y.; Zeng, S.; Xu, Y.; Feng, R. FTUNet: A Feature-Enhanced Network for Medical Image Segmentation Based on the Combination of U-Shaped Network and Vision Transformer. *Neural Process. Lett.* **2024**, *56*, 84. [CrossRef]
10. Taghanaki, S.A.; Abhishek, K.; Cohen, J.P.; Cohen-Adad, J.; Hamarneh, G. Deep Semantic Segmentation of Natural and Medical Images: A Review. *Artif. Intell. Rev.* **2021**, *54*, 137–178. [CrossRef]
11. He, S.; Minn, K.T.; Solnica-Krezel, L.; Anastasio, M.A.; Li, H. Deeply-supervised Density Regression for Automatic Cell Counting in Microscopy Images. *Med. Image Anal.* **2021**, *68*, 101892. [CrossRef] [PubMed]
12. Zhang, M.; Wang, Y.; Ding, Q.; Li, H.; Chao, H. Application of Artificial Intelligence Image-Assisted Diagnosis System in Chest CT Examination of COVID-19. *Medicine* **2020**, *68*, 9. [CrossRef]
13. Ayus, I.; Gupta, D. A Novel Hybrid Ensemble Based Alzheimer's Ldentification System Using Deep Learning Technique. *Biomed. Signal Proces.* **2024**, *92*, 106079. [CrossRef]
14. Jasphin, C.; Geisa, J.M. Automated Identification of Gastric Cancer in Endoscopic Images by a Deep Learning Model. *Automatika* **2024**, *65*, 559–571. [CrossRef]
15. Zhong, Y.; Piao, Y.; Tan, B.; Liu, J. A Multi-Task Fusion Model Based on a Residual–Multi-Layer Perceptron Network for Mammographic Breast Cancer Screening. *Comput. Meth. Prog. Biomed.* **2024**, *247*, 108101. [CrossRef]
16. Papandrianos, N.; Papageorgiou, E. Automatic Diagnosis of Coronary Artery Disease in SPECT Myocardial Perfusion Imaging Employing Deep Learning. *Appl. Sci.* **2021**, *11*, 6362. [CrossRef]
17. Petibon, Y.; Fahey, F.; Cao, X.; Levin, Z.; Sexton-Stallone, B.; Falone, A.; Zukotynski, K.; Kwatra, N.; Lim, R.; Bar-Sever, Z.; et al. Detecting Lumbar Lesions in Tc-99m-MDP SPECT by Deep Learning: Comparison with Physicians. *Med. Phys.* **2021**, *48*, 4249–4261. [CrossRef] [PubMed]

18. Xing, H.; Wang, T.; Jin, X.; Tian, J.; Ba, J.; Jing, H.; Li, F. Direct Attenuation Correction for Tc-99m-3PRGD(2) Chest SPECT Lung Cancer Images Using Deep Learning. *Front. Oncol.* **2023**, *13*, 1165664. [CrossRef] [PubMed]
19. Kwon, K.; Hwang, D.; Oh, D.; Kim, J.H.; Yoo, J.; Lee, J.S.; Lee, W.W. CT-Free Quantitative SPECT for Automatic Evaluation of% Thyroid Uptake Based on Deep-Learning. *Ejnmmt. Phys.* **2023**, *10*, 20. [CrossRef]
20. Lin, Q.; Man, Z.; Cao, Y.; Wang, H. Automated Classification of Whole-Body SPECT Bone Scan Images with VGG-Based Deep Networks. *Int. Arab. J. Inf. Technol.* **2023**, *20*, 1–8. [CrossRef]
21. Ni, Y.C.; Tseng, F.P.; Pai, M.C.; Hsiao, I.T.; Lin, K.J.; Lin, Z.K.; Lin, W.B.; Chiu, P.Y.; Hung, G.U.; Chang, C.C.; et al. Detection of Alzheimer's Disease Using ECD SPECT Images by Transfer Learning from FDG PET. *Ann. Nucl. Med.* **2021**, *35*, 889–899. [CrossRef] [PubMed]
22. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; Proceedings, Part III 18; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241. [CrossRef]
23. Xiao, X.; Lian, S.; Luo, Z.; Li, S. Weighted Res-UNet for High-Quality Retina Vessel Segmentation. In Proceedings of the 2018 Ninth International Conference on Information Technology in Medicine and Education(ITME 2018), Hangzhou, China, 19–21 October 2018; pp. 327–331. [CrossRef]
24. Zhou, Z.; Rahman Siddiquee, M.M.; Tajbakhsh, N.; Liang, J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In Proceedings of the Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision, Granada, Spain, 20 September 2018 ; Springer International Publishing: Cham, Switzerland, 2018; pp. 3–11. [CrossRef]
25. Patel, K.; Bur, A.M.; Wang, G. Enhanced U-Net: A Feature Enhancement Network for Polyp Segmentation. In Proceedings of the 2021 18th Conference on Robots and Vision (CRV), Burnaby, BC, Canada, 26–28 May 2021; pp. 181–188. [CrossRef]
26. Schlemper, J.; Oktay, O.; Schaap, M.; Heinrich, M.; Kainz, B.; Glocker, B.; Rueckert, D. Attention Gated Networks: Learning to Leverage Salient Regions in Medical Images. *Med. Image Anal.* **2019**, *53*, 197–207. [CrossRef] [PubMed]
27. Zhang, S.j.; Peng, Z.; Li, H. SAU-Net: Medical Image Segmentation Method Based on U-Net and Self-Attention. *Acta Electonica Sin.* **2022**, *50*, 1.
28. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. Transunet: Transformers Make Strong Encoders for Medical Image Segmentation. *arXiv* **2021**, arXiv:2102.04306. https://doi.org/10.48550/arXiv.2102.04306.
29. Wang, H.; Cao, P.; Wang, J.; Zaiane, O.R. UCTransNet: Rethinking the Skip Connections in U-Net from a Channel-wise Perspective with Transformer. In Proceedings of the Thirty-Sixty AAAI Conference on Artificial Intelligence, Online Conference, 22 February–1 March 2022; pp. 2441–2449. [CrossRef]
30. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV 2021), Montreal, QC, Canada, 10–17 October 2021; pp. 9992–10002. [CrossRef]
31. Zhang, Y.; Liu, H.; Hu, Q. TransFuse: Fusing Transformers and CNNs for Medical Image Segmentation. In Proceedings of the Medical Image Computing and Computer Assisted Intervention-Miccai, Strasbourg, France, 27 September–1 October 2021, PT I ; Lecture Notes in Computer Science; DeBruijne, M., Cattin, P., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., Essert, C., Eds.; Springer: Berlin/Heidelberg, Germany, 2021; Volume 12901, pp. 14–24. [CrossRef]
32. Zhu, Z.; He, X.; Qi, G.; Li, Y.; Cong, B.; Liu, Y. Brain Tumor Segmentation Based on the Fusion of Deep Semantics and Edge Information in Multimodal MRI. *Inform. Fusion* **2023**, *91*, 376–387. [CrossRef]
33. Ma, J.; He, Y.; Li, F.; Han, L.; You, C.; Wang, B. Segment Anything in Medical Images. *Nat. Commun.* **2024**, *15*, 654. [CrossRef] [PubMed]
34. Fu, B.; Peng, Y.; He, J.; Tian, C.; Sun, X.; Wang, R. HmsU-Net: A hybrid Multi-Scale U-net Based on a CNN and Transformer for Medical Image Segmentation. *Comput. Biol. Med.* **2024**, *170*, 108013. [CrossRef] [PubMed]
35. Lyu, Y.; Xu, Y.; Jiang, X.; Liu, J.; Zhao, X.; Zhu, X. AMS-PAN: Breast Ultrasound Image Segmentation Model Combining Attention Mechanism and Multi-Scale Features. *Biomed. Signal. Proces.* **2023**, *81*, 104425. [CrossRef]
36. Lin, A.; Chen, B.; Xu, J.; Zhang, Z.; Lu, G.; Zhang, D. Ds-Transunet: Dual Swin Transformer U-Net for Medical Image Segmentation. *IEEE T. Instrum. Meas.* **2022**, *71*, 4005615. [CrossRef]
37. Valanarasu, J.M.J.; Oza, P.; Hacihaliloglu, I.; Patel, V.M. Medical Transformer: Gated Axial-Attention for Medical Image Segmentation. In Proceedings of the Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, 27 September 27–1 October 2021; Proceedings, Part I 24; Springer: Berlin/Heidelberg, Germany, 2021; pp. 36–46. [CrossRef]
38. Wu, H.; Chen, S.; Chen, G.; Wang, W.; Lei, B.; Wen, Z. FAT-Net: Feature Adaptive Transformers for Automated Skin Lesion Segmentation. *Med. Image. Anal.* **2022**, *76*, 102327. [CrossRef] [PubMed]
39. Liu, B.; Wu, R.; Bi, X.; Xiao, B.; Li, W.; Wang, G.; Gao, X. D-Unet: A Dual-encoder U-Net for Image Splicing Forgery Detection and Localization. *arXiv* **2020**, arXiv:2012.01821.
40. Karri, M.; Annavarapu, C.S.R.; Acharya, U.R. Explainable Multi-module Semantic Guided Attention Based Network for Medical Image Segmentation. *Comput. Biol. Med.* **2022**, *151*, 106231. [CrossRef] [PubMed]
41. Wang, B.; Qin, J.; Lv, L.; Cheng, M.; Li, L.; Xia, D.; Wang, S. MLKCA-Unet: Multiscale large-kernel Convolution and Attention in Unet for Spine MRI Segmentation. *Optik* **2023**, *272*, 170277. [CrossRef]

42. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In *Proceedings of the European Conference on Computer Vision (ECCV)*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 3–19. [CrossRef]
43. Hu, S.; Zhang, J.; Xia, Y. Boundary-Aware Network for Kidney Tumor Segmentation. In Proceedings of the Machine Learning in Medical Imaging: 11th International Workshop, MLMI 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, 4 October 2020; Proceedings 11; Springer: Berlin/Heidelberg, Germany, 2020; pp. 189–198. [CrossRef]
44. Tang, C.; Chen, H.; Li, X.; Li, J.; Zhang, Z.; Hu, X. Look Closer to Segment Better: Boundary Patch Refinement for Instance Segmentation. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13926–13935. [CrossRef]
45. Cao, H.; Wang, Y.; Chen, J.; Jiang, D.; Zhang, X.; Tian, Q.; Wang, M. Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation. In *Proceedings of the ECCV Workshops*; Springer: Berlin/Heidelberg, Germany, 2021. [CrossRef]
46. Xu, Q.; Duan, W.; He, N. DCSAU-Net: A Deeper and More Compact Split-Attention U-Net for Medical Image Segmentation. *Comput. Biol. Med.* **2022**, *154*, 106626. [CrossRef] [PubMed]
47. Pang, B.; Chen, L.; Tao, Q.; Wang, E.; Yu, Y. GA-UNet: A Lightweight Ghost and Attention U-Net for Medical Image Segmentation. *J. Imaging Inform. Med.* **2024**, *(Early Access)*. [CrossRef]
48. Yuan, L.; Song, J.; Fan, Y. MCNMF-Unet: A mixture Conv-MLP Network With Multi-Scale Features Fusion Unet for Medical Image Segmentation. *PeerJ. Comput. Sci.* **2024**, *10*, e1798. [CrossRef]
49. Hu, B.; Zhou, P.; Yu, H.; Dai, Y.; Wang, M.; Tan, S.; Sun, Y. LeaNet: Lightweight U-Shaped Architecture for High-Performance Skin Cancer Image Segmentation. *Comput. Biol. Med.* **2024**, *169*, 107919. [CrossRef] [PubMed]
50. Liu, Y.; Yao, S.; Wang, X.; Chen, J.; Li, X. MD-UNet: A Medical Image Segmentation Network Based on Mixed Depthwise Convolution. *Med. Biol. Eng. Comput.* **2023**, *62*, 1201–1212. [CrossRef] [PubMed]