International Journal of Computational Intelligence Systems and Applications. Volume 11, Issue 1, (2024) 37-42
http://www.cayley-nielson.com/ijcisa/

TransUNet-CBAM: An Architecture Based on Deep Learning for Granular Layer Detection in Immunofluorescence Photographs

Zhixin Li[1], Tianyu Wu[1]
[1] North China University of Science and Technology, Tangshan, China
E-mail: *1034990898@qq.com*

Abstract

The nerve cells in the hippocampus are usually the focus in immunofluorescence experiments on animal neural centers, so it is often necessary to locate the hippocampus area when analyzing immunofluorescence photos with artificial intelligence technology. To address these challenges, we introduced an approach based on image segmentation technique to detect granular layers which are the border of hippocampus. In this paper, we proposed an architecture named TransUNet-CBAM to detect granular layers and trained the model on the dataset consisting of 1336 immunofluorescence photographs from multiple basic medical experiments. Through training and validation on the dataset, our model reached 88.85% and 10.93 on metrics of IoU and HD95. The results demonstrated that TransUNet-CBAM can accurately detect granular layers in immunofluorescence photos.

Keywords: Immunofluorescence; TransUNet; CBAM.

1. Introduction

In animal experiments to study neuroinflammation, the hippocampus bounded by granular layers are usually targeted for research. In this paper, we proposed TransUnet-CBAM network by integrating the CBAM (Convolutional Block Attention Module) attention module into the TransUNet framework to detect granular layers. Transformer has been widely applied in the field of NLP due to the excellent computational efficiency and scalability. In the field of image processing, Transformer has advantages in image segment and object detection of complex images due to its capability of capturing long-distance relationships between pixels through self-attention mechanism [1]. TransUNet architecture combines CNN and Transformer and is mainly composed of Multi-Head Attention mechanism and FFN (Feed-Forward Neural Network), which achieves superior performance of recording long-distance dependencies, capturing semantic information of images and generalization [2]. CBAM is a convolutional attention module that combines CAM (Channel Attention Module) and SAM (Spatial Attention Module). CBAM considers both channel and spatial dimensions of feature map separately to strengthen the attention on key information of images [3]. TransUNet-CBAM network improved the accuracy of granular layer edge detection by enlarging the receptive field of convolutional kernels while highlighting important features and suppressing unimportant feature.

2. Related work

In 2015, Olaf Ronneberger, Philipp Fischer, and Thomas Brox built upon UNET framework, a fully

convolutional network, for biomedical image segmentation [4]. Zongwei Zhou and Md Mahfuzur Rahman Siddiquee et al. presented UNET++, a new image segmentation network based on nested and dense skip connections, which can more effectively capture ne-grained details of the foreground objects [5]. Wang Shuang and He Xiaohai combined VGG16 network with UNET and CBAM module to classify fluorescence images [6]. Eric M. Christiansen, Samuel J. Yang showed a computational machine-learning approach to predict some fluorescent labels from transmitted-light images of unlabeled fixed or live biological samples reliably [7]. In 2021, Jieneng Chen and Yongyi Lu proposed TransUNet framework which achieved better performance than competing methods in medical image segmentation [8].

3. Method

In this paper, we designed TransUNet-CBAM network to detect the granular layers which are borders of the hippocampus. U-Net model has played a key role in the lesion segmentation such as tumor detection, pathological analysis and 3D reconstruction of medical image. To enhance the ability of model to capture global features, TransUNet framework combines U-Net and Transformer which encodes tokenized image patches from the feature map generated by CNN as the input sequence and achieves superior performances [9].

We selected TransUNet as the base framework to detect the granular layers and integrated CBAM module to take advantage of other properties of images besides spatial information, in which the input feature map is processed by the channel attention module, and then the spatial attention module to generate the output feature map [10]. As shown in Fig. 1, maxpool and avgpool are performed in the channel attention on the feature map in the spatial dimension to generate two vectors[c×1×1] which are processed by a shared MLP, add and sigmod, and then multiplied with the original feature map. The spatial attention module first performs maxpool and avgpool on the values of all channels at the same spatial position on the feature map[c×h×w] output by the channel attention module, and then concatenates the two vectors. After a convolution layer and sigmod , the result of vector[1×h×w] is multiplied with the original feature map to assign spatial weight for each position.

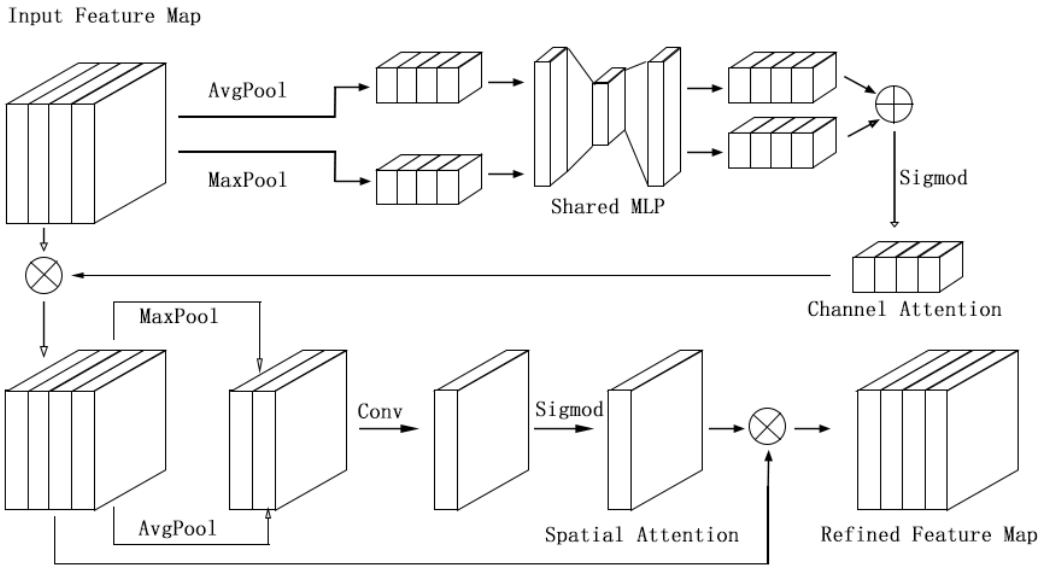


Fig. 1. Architecture of CBAM

In TransUNet-CBAM network, CBAM module was integrated into TransUNet framework before Transformer layer and each concatenation to ensure the feature map of each layer to be processed by CBAM during upsampling. The network architecture is illustrated in Fig. 2. The encoder consists of three convolution layers, a CBAM module and transformers. Three feature maps, t1[64×128×128], t2[128×64×64] and t3[256×32×32], were generated by convolution layers and t3 was passed into Transformer after CBAM and Linear Projection. The vector output by Transformer was reshaped and restored to feature map [64×128×128] by three rounds of upsampling, concatenation and convolution operation, and then additional upsampling and convolution were performed on the feature map and the output was passed to segmentation head to generate the area of the granular layers. We ignored area outside the hippocampus of immunofluorescence staining photographs in subsequent analysis because the astrocytes in the hippocampus were the research object of medical experiments providing dataset for this paper.

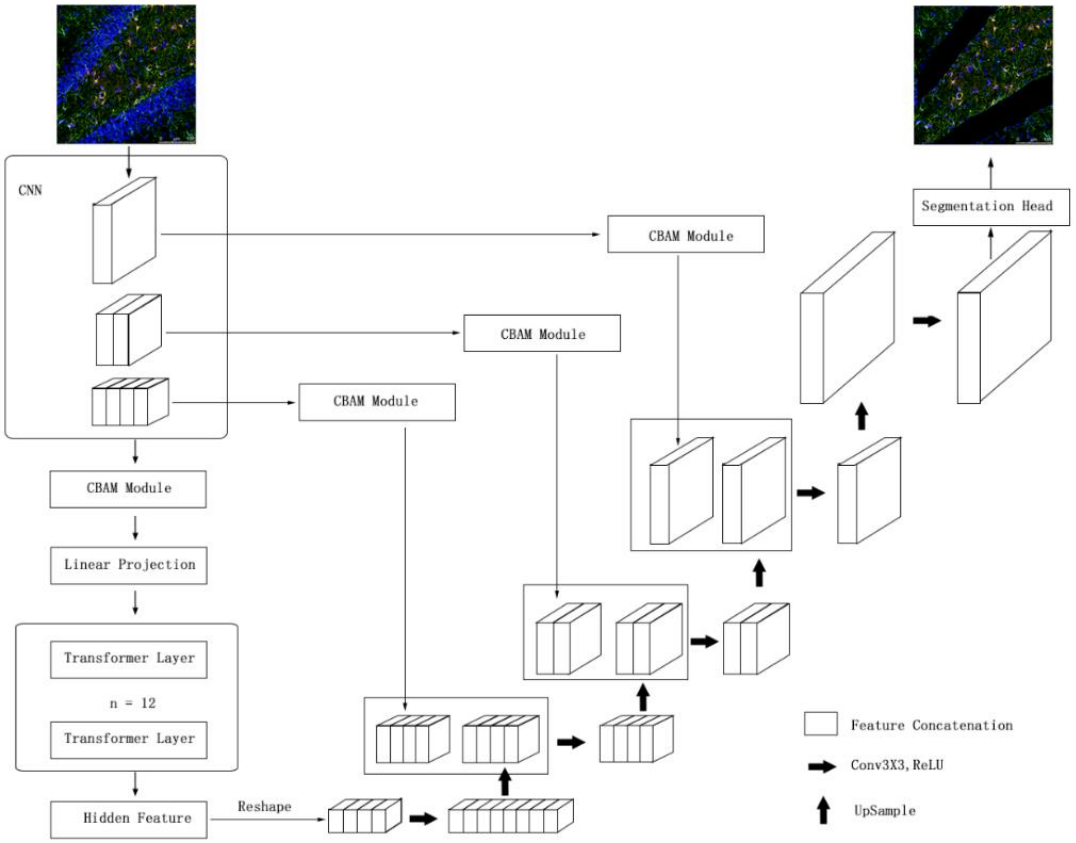


Fig. 2. Architecture of TransUNet-CBAM

4. Experiments

We trained and test the models in our experiment on a workstation with Intel i9-13900K processor, 64 GB of host memory and an NVIDIA GeForce RTX 3090 graphics card with 24 GB of video memory. The software environment was Ubuntu 22.04.5, Python 3.11 and the PyTorch 2.2.1 framework.

We evaluated TransUnet-CBAM model with PA(Pixel Accuracy), IoU(Intersection over Union)、DSC(Dice Similarity Coefficient) and HD9(Hausdorff Distance 95%). PA stands for the proportion of correctly classified pixels to the total pixels and it's computed as

$$PA = \frac{\sum_{i=0}^{n} p_{ii}}{\sum_{i=0}^{n} \sum_{j=0}^{n} p_{ij}} \quad (1)$$

where n is the count of categories, p_{ii} denotes the count of accurately classified pixels of the category i, p_{ij} indicates the count of pixels of category i that is classified incorrectly as category j.

IoU is the important evaluation matrix of image segmentation that computes the ratio of the intersection area and union area between the predicted box (A) and the ground truth (B) [11]. The higher the value of IoU, the higher the degree of overlap between A and B, indicates that the model's prediction is more accurate. The formular of IoU is

$$IoU = \frac{A \cap B}{A \cup B} \quad (2)$$

DSC stands for the similarity between A and B. In our experiment, it's computed by doubling the intersection area between the predicted box and the ground truth box, and then dividing it by the sum of count of the pixels in both boxes [12]. DSC is defined as follows:

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|} \quad (3)$$

HD95 is a definition of distance between two sets of points and mainly used to measure the accuracy of boundary of image segmentation. Assuming two sets of points A = $\{a_1,..., a_p\}$, B=$\{b_1,..., b_q\}$, the HD between A and B is defined as

$$H(A,B) = max(h(A,B), h(B,A)) \quad (4)$$

$$h(A,B) = \max_{a \in A} \{\min_{b \in B} \|a - b\|\} \quad (5)$$

$$h(A,B) = \max_{b \in B} \{\min_{a \in A} \|b - a\|\} \quad (6)$$

H(A,B) is Hausdorff distance of A and B，h(A，B) and h(B，A) are respectively one-way Hausdorff distances from set A to set B and from set B to set A. For each ai in set A, find minimum value in distances between ai and each bj in set B, and then take the maximum value of these minimum distances as h(A, B). h(B, A) can be similarly concluded and H(A,B) is the maximum between h(A,B) and h(B,A). HD95 is based on the calculation of the 95th percentile of the distances between boundary points in A and B. The purpose for using this metric is to eliminate the impact of a very small sub-set of the outliers.

5. Result

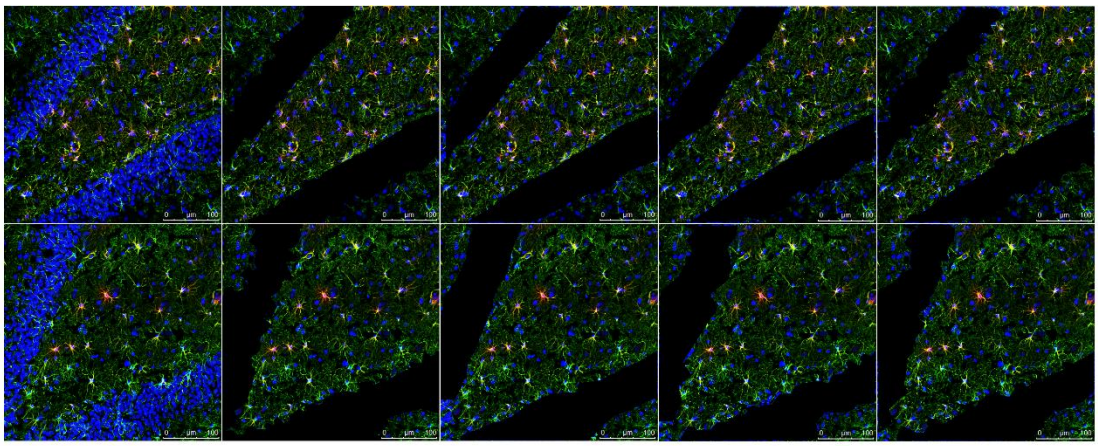
We trained TransUNet-CBAM model on 1336 immunofluorescence photographs, including 1236 photographs with granular layers and 100 photographs without granular layers, which were randomly divided into the training set with 1000 samples, the validation set with 300 samples and the testing set with 36 samples. UNet++ and TransUNet models were for comparison and learning rate is set to 0.01, batch size to 8 and epoch to 200. We compared the three models in terms of DSC, HD95, IoU and PA matrics, as shown in Table 1.

Three models performed comparably on PA. On DSC, TransUNet and TransUNet-CBAM exhibits improvement of 0.19% and 0.17% over UNet++ and TransUNet achieves the best performance. TransUNet-CBAM outperforms TransUNet on HD95 and IoU. The overall performance of TransUNet and TransUNet-CBAM is better than UNet++, which indicates that Transformer improves the accuracy of image segmentation. The best performance of TransUNet CBAM on the HD95 shows that the channel attention and spatial attention mechanisms improves the accuracy of boundary segmentation. Due to the integration

of channel attention and spatial attention mechanisms, the qualitative comparison of the results of the three models is shown in Figure 3.

Table 1. Granular layer detection results for UNet++, TransUNet and TransUNet-CBAM ++

Model	DSC	HD95	IoU	PA
UNet++	91.32%	12.36	88.64%	93.62%
TransUNet	91.51%	11.51	88.77%	93.54%
TransUNet-CBAM	91.49%	10.93	88.85%	93.57%



(a) (b) (c) (d) (e)

Fig. 3. Qualitative comparison of the results from UNet++, TransUNet and TransUNet-CBAM: (a) original image (b) ground truth (c) UNet++ result (d) TransUNet result (e) TransUNet-CBAM

6. Conclusion

This paper proposed the TransUNet-CBAM architecture by integrating CBAM attention module to TransUNet base framework to detect granular layers in immunofluorescence staining photographs automatically, efficiently and accurately. The proposed TransUNet-CBAM model achieved 91.49% on DSC metric, 10.93 on HD95 and 88.85% on IoU in granular layer detection. The result of granular layers detection represented that Transformer technology was able to improve accuracy of image segmentation and CBAM attention module was the effective mechanism to optimize the base framework.

References

[1] A. Dosovitskiy et al., “An image is worth 16X16 words: Transformers for image recognition at scale,” presented at Int. Conf. on Learning Representations, May 3-7, 2021.

[2] J. Chen et al., “Transformers Make Strong Encoders for Medical Image Segmentation,” ArXiv e-prints, 2021, 10.48550/arXiv.2102.04306.

[3] S. Woo et al., “CBAM: Convolutional Block Attention Module,” ArXiv e-prints, 2018, 10.1007/978-3-030-01234-2_1.

[4] O. Ronneberger et al., “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in Proc. MICCAI, Munich, Bavaria, Germany, 2015, pp. 234–241, 10.1007/978-3-662-54345-0_3.

[5] Z. Zhou et al., “UNet++: A Nested U-Net Architecture for Medical Image Segmentation,” in Proc. DLMIA 2018, 8th International Workshop, ML-CDS 2018, Granada, Spain, 2018, pp. 3-11.

[6] S. Wang et al., “Fluorescence image component classification based on deep learning,” Intelligent Computer and Applications, vol. 13, no. 3, pp. 175-181, Mar. 2023.

[7] E.M. Christiansen et al., “In Silico Labeling: Predicting Fluorescent Labels in Unlabeled Images,” Cell, vol. 173, no. 3, pp. 792-803, Mar. 2018, 10.1016/j.cell.2018.03.040.

[8] J. Chen et al., “Transformers Make Strong Encoders for Medical Image Segmentation,” ArXiv e-prints, 2021, 10.48550/arXiv.2102.04306.

[9] G. Sun et al., “DA-TransUNet: Integrating Spatial and Channel Dual Attention with Transformer U-Net for Medical Image Segmentation,” Frontiers in Bioengineering and Biotechnology, vol. 12, no. 5, May. 2024, 10.3389/fbioe.1398237.

[10] S. Woo et al., “CBAM: Convolutional Block Attention Module,” ArXiv e-prints, 2018, 10.1007/978-3-030-01234-2_1.

[11] H. J. Figuerola, “Attention Span For Personalisation,” ArXiv e-prints, 2016, abs/1608.00147.

[12] F, Milletarì, N. Navab and S. Ahmadi, “V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation.” 2016 Fourth International Conference on 3D Vision (3DV) (2016): 565-571.