TRANSFORMER-BASED DENOISING METHOD FOR MEDICAL IMAGES

Zhao Di, Gourinovitch A.B.
Department of Information Technologies in Automated Systems,
Belarussian State University of Informatics and Radioelectronics
Minsk, Republic of Belarus
E-mail: 3189124246@qq.com, gurinovich@bsuir.by

Biomedical image segmentation is essential for accurate disease diagnosis. However, issues like noise and artifacts in medical images can hinder effective diagnosis. This paper presents a Transformer-based method for medical image segmentation denoising, which uses self-attention to remove noise and retain image details, thus enhancing diagnostic accuracy.

Introduction

Medical imaging is a cornerstone in contemporary diagnostics, enabling radiologists to utilize non-invasive methods like magnetic resonance imaging (MRI), X-ray radiography, computed tomography (CT), and positron emission tomography (PET) to visualize internal tissues. These procedures are not only accessible but also designed to minimize patient discomfort and risk. Despite their benefits, the acquisition of medical images can introduce various forms of noise and streak artifacts, potentially obscuring diagnostic details. Consequently, the pursuit of effective medical image denoising and the enhancement of denoising precision are of paramount importance, holding significant value for both scientific inquiry and clinical practice.

I. Challenges and problems

Over the past decades, a plethora of medical image denoising algorithms has been developed, broadly classifying into traditional and deep learning-based approaches. Traditional methods exploit the pixel correlations within the original image block to estimate the gray values for noise reduction. While these traditional algorithms demonstrate effective denoising capabilities, they often require manual parameter tuning, leading to high computational complexity and are typically designed for single denoising tasks.

With the continuous improvement of network architecture, deep learning algorithms are gradually emerging, and more and more researchers and scholars apply them to the field of image denoising. Deep learning-based denoising algorithms are characterized by their precision in noise removal without embedding it into the image space. Convolutional neural network (CNN)-based denoising methods have emerged as a dominant trend in medical imaging, capable of preserving image details such as edges and texture structures through convolution operations, thereby effectively suppressing noise.Despite the superior performance of CNN-based methods in denoising, challenges remain[1]. The convolution kernels used during training are often not tailored to the image content, that can result in the loss of image detail information. Additionally, the neglect of non-local correlations within the image can lead to a loss of global information when modeling long-term dependencies, highlighting the need for more sophisticated modeling techniques that can capture both local and global image features.

II. Description of existing methods

Addressing the limitations and challenges prevalent in contemporary image denoising tasks, the article introduces a Transformer-based denoising algorithm. Initially introduced by Google in 2017 for natural language processing, the Transformer architecture diverges from the conventional Recurrent Neural Networks (RNNs) commonly employed in natural language processing(NLP). The common Transformer structure is shown in Figure 1, the Transformer consists of an encoder and a decoder, the encoder generates the input codes and the decoder receives all the codes from the encoder and utilizes them to integrate the contextual information to produce the output sequence. Each Transformer module consists of the following structures:

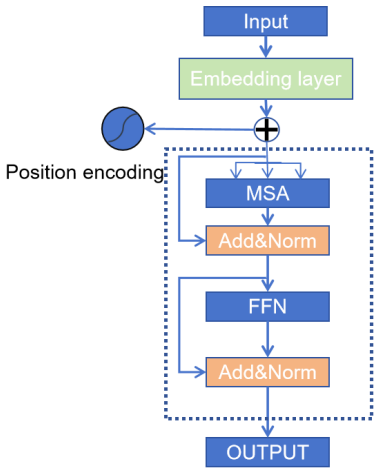

Рис. 1 – The structure of transformer

- Embedding layer : The medical image is segmented into small blocks (patches), each of which is considered as an element in the sequence. This step

is intended to reduce global dependencies and allow the model to focus on local features of the image, providing the basis for the denoising process. The embedding layer maps these localized image chunks into a high dimensional space to form embedding vectors. This process enhances the model's ability to capture local image features, providing rich information for the subsequent self-attention layer, which helps to recognize and remove noise.

- Self-Attention Layer : The self-attention mechanism is a core component of the Transformer architecture, the Self-Attention layer recognizes global patterns and noise distributions in an image by calculating the relationship between each image block and all other blocks in the sequence. This mechanism allows the model to take into account the contextual information of the entire image when processing each image block, thus effectively recognizing and removing noise while preserving important image features. The self-attention model uses the Query-Key-Value (QKV) model and the computation of the self-attention layer can be represented as:

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

where Q, K, and V are the query, key, and value matrices, respectively, and $d_k$is the dimension of the key vector.

- Multi-head self-attention : The multi-head self-attention mechanism further enhances the model's ability to capture features at different scales, and improves the denoising accuracy by processing information from multiple subspaces in parallel. This step allows the model to focus on both large-scale structures and small-scale details in the image, which helps to remove noise at different scales while recovering the detailed information of the image.

$$\text{MHA}(Q,K,V) = \text{Concat}(\text{head}_1, \dots \text{head}_n)W^o \quad (2)$$

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (3)$$

where MHA is the Multihead, n indicates the number of head,W_i^Q, W_i^K, W_i^V, and W^O are learnable weight matrices.

- Positional coding : Positional coding provides the model with information about the spatial location of each image block, which is crucial for understanding the structure and context of the image. During the denoising process, positional coding ensures that the model is able to correctly reconstruct the spatial structure of the image, avoiding structural distortions caused by noise.Positional coding provides the model with spatial location information for each patch.

- Layer Normalization : Layer normalization accelerates model convergence and improves training stability by stabilizing the input distribution. During the denoising process, layer normalization helps to reduce the internal covariate bias between different layers, which improves the robustness of the model to noise.

- Residual Connection : Residual connectivity transfers information from the previous layer to the subsequent layer through jump connections, which not only enriches the information transfer of the model, but also helps to reduce the problem of gradient vanishing in deep networks. During the denoising process, residual connections ensure that the original information of the image is preserved, while allowing the model to efficiently learn and adjust features in the deep network to remove noise.

- Feedforward Neural Network : The feedforward network is responsible for further processing the output of the self-attention layer to enhance the expressive power of the model through nonlinear transformations. This step helps the model to capture complex relationships in the image, especially in the presence of complex noise patterns, and the feed-forward network is able to adjust the features to better remove noise and recover image details.

Рис. 2 – The comparison between the image before denoising (a) and the image after denoising (b)

As shown in Figure 2.a is the medical image before denoising and Figure 2.b is the image after denoising, the noisy image before denoising contains a large number of random noise points, which seriously affects the visual clarity of the image. In the denoised image, artifacts and distortions in the image are significantly reduced and important structural and textural features are preserved and enhanced,. Through these steps, the Transformer model is able to effectively recover clear images from noisy images, providing a powerful tool in the field of medical image denoising.

III. CONCLUSION

Transformer-based medical image denoising method effectively removes the noise from the image through the self-attention mechanism while retaining the important features of the image. Future research can further explore the application of Transformer-based model in the field of medical image denoising.

1. Ilesanmi, A. E., & Ilesanmi, T. O. (2021). Methods for image denoising using convolutional neural network: a review. Complex & Intelligent Systems, 7(5), 2179-2198.
2. Sagheer, S. V. M., & George, S. N. (2020). A review on medical image denoising algorithms. Biomedical signal processing and control, 61, 102036.