Modified U-Net for vessel centerline extraction and arteries veins classification

Bochra Ayed

10 juin 2023

Chapitre 1: Modified U-net proposed architecture

Goal

The proposed architecture has been used to extract the centerline of vessels and classify their types as arteries or veins.

1.1 State of the art

Several methods have been suggested, aiming to detect the skeleton of blood vessels. In this context, most papers have based their proposals on image processing. In[1] The authors propose to work with monochrome images obtained by extracting the green component from RGB fundus images, offering better visibility of blood vessels. Morphological operations such as opening have been applied to fill in vessel gaps that could cause subsequent errors. For example, the retina presents high-gradient structures that are not vessels (e.g., Optic Disc). The scientists applied a double "top-hat" operation to extract the entirety of vessels and eliminate unnecessary structures. Finally, the extraction of vessel centerlines was performed by combining the "Hit-or-miss transformation" morphological operation with the stochastic watershed transformation technique. To segment blood vessels, the authors proposed starting with the extraction of centerlines [2], [3]. In [2], the researchers utilized graph theory. They considered arteries/veins as edges and the optic nerve as the root of the tree. They applied a heuristic search algorithm to correct errors resulting from the graphical estimation by considering all possible trees. The importance-based search algorithm was adopted to find the closest tree to the ground truth. However, in [3], the researchers estimated the confusion matrix based on a group of points extracted from the image histogram. From this matrix, quantification was performed to obtain the initial vessel network. Subsequently, median filtering was applied to the initial vessel network to remove unnecessary lines.
In the literature, research in this field is oriented in two directions. The first approach focuses on extracting only the vessel centerlines [1], which provides valuable results that can be utilized as biomarkers for detecting other diseases. The second approach involves extracting the vessel centerlines for blood vessel segmentation [2], [3].

Based on the analysis of existing methods, it is observed that all these methods rely on image processing. However, image processing alone may not yield the best results, especially in critical tasks such as medical applications. Convolutional Neural Networks (CNN) have demonstrated their effectiveness in image processing and analysis in various domains, including agriculture [4], art [5], security [6], and even more critical and sensitive medical fields [7], among others.

Therefore, we propose to extract vessel centerlines using a deep learning architecture based on U-Net, which is built upon CNN[8]. U-Net is an improvement over the Fully Convolutional Neural Network (FCNN) architecture proposed by [45]. U-Net has the advantage of being trained on a reduced number of images while ensuring semantic segmentation. Furthermore, U-Net was initially proposed for the segmentation of biomedical images and has shown valuable results [8]. The architecture consists of two symmetrical pathways : the contracting pathway (encoder) and the expansive pathway (decoder), with a bridge connecting the two pathways to allow for concatenation. Figure 1 illustrates our proposed model.

1.2 Description of the proposed architecture

The encoder is composed of four downsampling blocks, with each block consisting of four 3x3 convolutions followed by the Rectified Linear Unit (ReLU) activation function and the MaxPooling layer. It's worth noting that after each block, the number of feature maps is doubled, and the spatial resolution of the image is reduced.

The decoder is composed of four upsampling blocks, with each block consisting of four3×3 convolutions, each followed by ReLU activation, and a 2×2 transposed convolution that duplicates the spatial resolution. As a result, the number of feature maps is halved. The output of each transposed convolution is concatenated with the feature maps of the same spatial resolution from the corresponding level of the encoder through a skip connection. This allows the encoder features to be transferred to the decoder at different resolution levels, resulting in more detailed output from the network.

The last part of the network includes four 3×3 convolutions followed by the ReLU activation function, applied to the output of the last upsampling block. Then, a final 1x1 convolution followed by the softmax activation function is applied to obtain the desired classes.

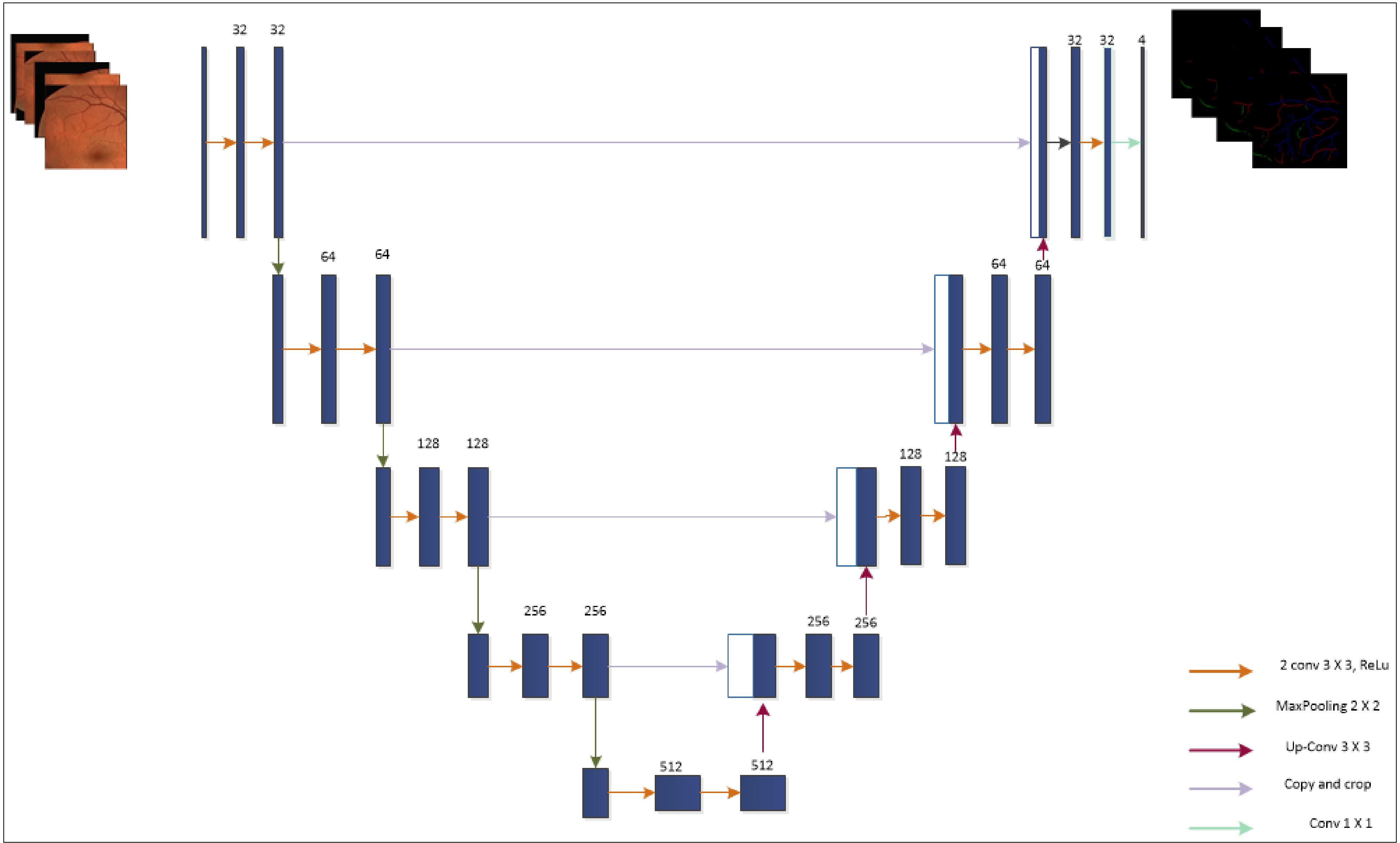

FIGURE 1. Proposed architecture