Attention Weighted Fully Convolutional Neural Networks for Dermatoscopic Image Segmentation

Michael Blackwell[1] and Qing Tian[1]

Bowling Green State University, Bowling Green OH 43403, USA
mblackw@bgsu.edu
qtian@bgsu.edu

Abstract. The goal of this project was to develop a fully convolutional neural network (FCNN) capable of identifying the region of interest (ROI) in dermatoscopic images. To achieve this goal, a U-Net style model was developed for this task and enhanced with an attention module which operated on the extracted features. The addition of this attention module improved our model's semantic segmentation performance and increased pixel-level precision and recall by 4.0% and 4.6% respectively. The code used in thie paper can be found on the project github page: https://github.com/Michael-Blackwell/CapstoneProject

Keywords: Attention · Semantic · Segmentation

1 Introduction

Melanoma is the deadliest form of skin cancer, resulting in more than 9,000 deaths each year in the US[1]. While early detection of melanoma is crucial for improving patient outcomes, the diagnostic accuracy of visual inspection from even trained professionals is estimated to be only 75%-84%[3]. This significant margin of error was the inspiration behind this project. If a model could be developed that is capable of identifying the region of interest (ROI) in dermatoscopic images, then it could be used to aid healthcare workers in detecting melanoma in earlier stages. In this paper, we focus on developing such a model by enhancing a U-Net style fully convolutional neural network (FCNN) using attention weighting.

The capability of FCNNs to produce state-of-the-art segmentations on medical images has been thoroughly demonstrated in recent years[4]. Yet, despite their success, FCNNs still have some inherent limitations; the fixed feature map structure prevents the model from effectively capturing relationships between non-local image features[2]. It has been proposed that applying attention weighting to extracted features can help to mitigate this limitation by aggregating non-local feature information[2][6]. This hypothesis will be tested by comparing the performance of our U-Net model with and without the attention module against a similar U-Net model developed by Intel AI[5].

2 Methods

2.1 Model Architecture

The U-Net model developed for this project consisted of 5 downscaling and 5 upscaling blocks. Each downscaling block consists of (in order) two convolution layers followed by a max pooling layer, batch normalization layer, and a dropout layer. Similarly, each upscaling block consists of (in order) a transposed convolution layer, a concatenation layer (for skip connections), two convolution layers, and a batch normalization layer. The attention module was added after the final downscaling layer, when the image features have been fully extracted. The structure and function of the attention module is discussed in the next section.

2.2 The Attention Module

Multiplicative attention weighting starts with the multiplication of two input tensors, *key* and *query*, which produce a tensor of attention weights. These weights are then multiplied with a third input tensor (*value*) to yield the final weighted output (figure 1). When the *key*, *query*, and *value* tensors are all identical, the process is known as self-attention and simply gives greater weight to more prominent features. For further details on attention weighting, refer to [7].

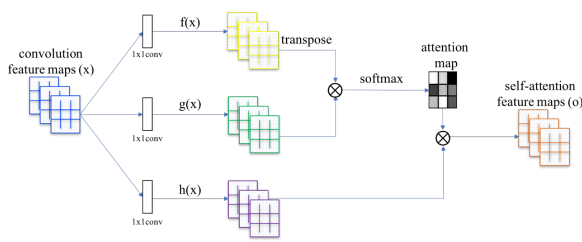

Fig. 1. Multiplicative Attention

The attention module used in this paper is inspired by the "Criss-Cross" method discussed in[2]. One of the primary differences here is our use of the Keras built-in *Attention* layer, which applies the multiplicative attention operation discussed above. Our module works by transposing the *key* before passing it into the *Attention* layer (where it will be transposed again). The result is a tensor of attention weights that compare rows to columns instead of the normal self-attention process which would compare rows to rows or columns to columns. Concatenating this "Criss-Cross" weighted tensor with the original input and passing it through a convolution layer helps to give the model more global contextual information than convolution alone can provide.

However, there is still contextual information to be gained by adding a second attention module in series with the first. This allows the contextual information gained in the first attention layer to further propagate through the model (figure 2) [2]. The result is features that are composed of contextual information from all pixels instead of only local ones.

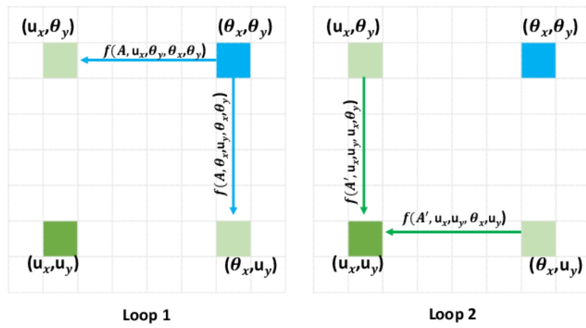

Fig. 2. Visualization of information propagation using the "Criss-Cross" method[2].

3 Data

The International Skin Imaging Collaboration (ISIC) 2017 Challenge dataset was used for the primary data source. This dataset was chosen for several reasons. First, it contains over 2,500 dermatoscopic images of potentially cancerous nevi which have already been divided into training, validation, and test sets[1]. Segmentation masks for these images are also included. Second, the segmentation masks have been drawn and reviewed by dermatology medical professionals[1]. The primary drawback to this dataset is that diversity in skin color is not sufficiently represented. While this is not problematic for an academic project, a more representative dataset would be required to train a model for clinical use.

4 Results

The attention module improved the precision, recall, and Dice score of our model by significant margins, as shown in table 1. While this performance boost comes at the cost of a large increase in trainable parameters (table 2), the model FLOPs were only modestly increased (table 3).

Table 1. Model metrics on the test dataset.

Model	Dice	mIoU	Recall	Precision
U-Net w/ Attention	81.1	80.3	76.7	88.9
U-Net w/o Attention	76.0	76.1	72.1	84.9
Intel Unet	77.5	76.2	68.8	89.5

Table 2. Model Parameters

Model	Trainable Params	Non-Trainable Params	Total Params
U-Net w/ Attention	3,521,901	2,500	3,524,401
U-Net w/o Attention	2,181,481	1,668	2,183,149
Intel Unet	1,941,105	0	1,941,105

Table 3. Model FLOPs

Model	FLOPs
U-Net w/ Attention	8,614,412,288
U-Net w/o Attention	8,413,184,000
Intel Unet	12,153,913,344

5 Conclusion

With a functional FCNN capable of accurate dermatoscopic image segmentation, the goal of this project has been met. The results indicate that attention mechanisms can effectively enhance the performance of encoder-decoder style models without significantly increasing computational resource requirements. The performance improvement from our attention layer warrants further investigation into applying similar methods to other encoder-decoder style models.

References

1. Codella, N.C.F., Gutman, D.A., Celebi, M.E., Helba, B., Marchetti, M.A., Dusza, S.W., Kalloo, A., Liopyris, K., Mishra, N.K., Kittler, H., Halpern, A.: Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (ISIC). CoRR **abs/1710.05006** (2017), http://arxiv.org/abs/1710.05006
2. Huang, Z., Wang, X., Wei, Y., Huang, L., Shi, H., Liu, W., Huang, T.S.: Ccnet: Criss-cross attention for semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence pp. 1–1 (2020). https://doi.org/10.1109/TPAMI.2020.3007032
3. Jerant, A.F., Johnson, J.T., Sheridan, C.D., Caffrey, T.J.: Early detection and treatment of skin cancer. Am Fam Physician **62**(2), 357–368 (Jul 2000)
4. Liu, X., Song, L., Liu, S., Zhang, Y.: A review of deep-learning-based medical image segmentation methods. Sustainability **13**(3) (2021), https://www.mdpi.com/2071-1050/13/3/1224
5. Ojika, D., Patel, B., Reina, G.A., Boyer, T., Martin, C., Shah, P.: Addressing the memory bottleneck in AI model training. CoRR **abs/2003.08732** (2020), https://arxiv.org/abs/2003.08732
6. Park J., Woo S., L.J.e.a.: A simple and light-weight attention module for convolutional neural networks. International Journal of Computer Vision **128**, 783–798 (2020), https://doi.org/10.1007/s11263-019-01283-0
7. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. CoRR **abs/1706.03762** (2017), http://arxiv.org/abs/1706.03762