EMDA-Net: Earth Mover's Distance (EMD) influenced Attention-aided Neural Network for Medical Image Classification

Surya Majumder
Dept. of Computer Science & Engineering
Heritage Institute of Technology
Kolkata, India
suryamajumder0802@gmail.com
0009-0007-3272-7369

Kushaj Mallick
Dept. of Electrical Engineering
Jadavpur University
Kolkata, India
kushajmallick@gmail.com
0009-0009-0340-3052

Wrick Pal
Dept. of Electrical Engineering
Jadavpur University
Kolkata, India
wrick9122rkm@gmail.com
0009-0005-4064-3881

Somenath Chakraborty
Dept. of Computer Science & Information Systems
Leonard C. Nelson College of Engineering and Sciences
West Virginia, USA
somenath.chakraborty@mail.wvu.edu
0000-0002-3880-0308

Ram Sarkar
Dept. of Computer Science & Engineering
Jadavpur University
Kolkata, India
ram.sarkar@jadavpuruniversity.in
0000-0001-8813-4086

Dry-Run of the Proposed EMDA-Net Model for Medical Image Classification

In this section, we present an operational overview of the EMDA-Net model designed for medical image classification. As an illustration, we conduct a demonstration using the HAM10000 dataset, where the objective is to classify skin lesion images into seven distinct classes: Actinic Keratoses and Intraepithelial Carcinoma/Bowen's disease (AKIEC), Basal Cell Carcinoma (BCC), Benign Keratosis-like Lesions (BKL), Dermatofibroma (DF), Melanoma (MEL), Melanocytic Nevi (NV), and Vascular Lesions (VASC).

This demonstration primarily involves computing attention scores from the EMD_m and EMD_a layers at the 2^{nd} and 13^{th} blocks, respectively of the proposed model and analyzing their relationship with tensors derived from input feature maps to form the resulting feature matrices.

Initially, we select a sample image and proceed by resizing and normalizing it.

In the context of our proposed model, for the EMD_m layer at block_2, we extract 144 input feature maps. Subsequently, we compute the median of these 144 feature maps, as described by Equation 1 in our paper, represented as:

$$A = \begin{bmatrix} -9.3786709e-04 & -1.6929586e-04 & \cdots \\ -5.1927382e-05 & 5.4406666e-04 & \cdots \\ \vdots & \vdots & \vdots \\ -5.7209963e-05 & 4.8702920e-04 & \cdots \\ -4.2225551e-04 & 2.4250240e-04 & \cdots \end{bmatrix}$$

$$B = \begin{bmatrix} -1.22861375e-05 & -4.57077331e-05 & \cdots \\ 3.89492125e-05 & -9.63774946e-05 & \cdots \\ \vdots & \vdots & \vdots \\ 4.14363603e-05 & -9.80402547e-05 & \cdots \\ 8.99793522e-05 & -7.43575947e-05 & \cdots \end{bmatrix}$$

where A represents a sample input feature map and B denotes the median feature map.

Next, we compute the Cumulative Sum Feature Map (CSFM) as defined by Equation 2 in our paper and calculate the CSFM array for both the original input feature map and the median feature map.

The CSFM arrays corresponding to these feature maps are displayed below.

$$A' = \begin{bmatrix} -9.3786709e-04 & -1.1071625e-03 & \cdots \\ -6.9672260e-03 & -7.9532657e-03 & \cdots \\ \vdots & \vdots & \vdots \\ -4.6966157e+00 & -4.6976366e+00 & \cdots \\ -4.7745996e+00 & -4.7760496e+00 & \cdots \end{bmatrix}$$

$$B' = \begin{bmatrix} -1.22861375e-05 & -5.79938715e-05 & \cdots \\ -4.00437135e-03 & -4.10074880e-03 & \cdots \\ \vdots & \vdots & \vdots \\ -5.83244324e-01 & -5.83342373e-01 & \cdots \\ -5.92508435e-01 & -5.92582822e-01 & \cdots \end{bmatrix}$$

Here A' represents the CSFM array of a sample input feature map A and B' denotes the CSFM array of the median input feature map B.

Using Equation 3 from our main paper, we compute and construct a matrix containing scores for each input feature map, as illustrated below.

$$Scores\_EMD_m = \begin{bmatrix} 8696.4970703125 \\ 288.5041809082 \\ \vdots \\ 3873.7133789062 \\ 5914.3642578125 \end{bmatrix}$$

the dimension being (number of features $\times 1$)

Subsequently, we negate these scores and obtain the Hadamard product and perform linear interpolation across all values from 1 to 0.5, following equation 4 outlined in the main paper. The resulting array, MV_1, represents the effective attention scores derived from the EMD_m layer.

$$MV_1 = \begin{bmatrix} 0.857003629207611 \\ 0.997565746307373 \\ \vdots \\ 0.937629342079162 \\ 0.903514385223388 \end{bmatrix}$$

where the dimension is (number of features $\times 1$)

From each of the 144 input feature maps, we extract every element of an expanded tensor and multiply it with its corresponding MV_1 value to get the final result feature matrix, which is then forwarded to the next layer in the proposed model.

$$Result_1 = \begin{bmatrix} \begin{bmatrix} -8.0375548e-04 \\ -1.6888375e-04 \\ \vdots \\ -9.7467290e-04 \\ 3.5968557e-04 \end{bmatrix} \\ \vdots \\ \begin{bmatrix} -1.5388840e-03 \\ -4.1931336e-05 \\ \vdots \\ -6.6172704e-04 \\ 6.9439277e-04 \end{bmatrix} \end{bmatrix}$$

Result_1 as shown above denotes the resultant feature matrix.

For the EMD_a layer, we gather the input feature maps from the 13^{th} block, where the EMD_a functionality is applied and compute the average feature map. This layer comprises 576 input feature maps. X shows a sample feature map as follows:

$$X = \begin{bmatrix} -1.9922839e-09 & -9.0121448e-09 & \cdots \\ -7.2541444e-09 & -1.6929788e-09 & \cdots \\ \vdots & \vdots & \vdots \\ -2.6470788e-09 & -4.9096327e-09 & \cdots \\ -9.8750330e-10 & -3.6327894e-09 & \cdots \end{bmatrix}$$

Y denotes the mean feature map as follows:

$$Y = \begin{bmatrix} 1.7125739e-10 & 5.1911047e-10 & \cdots \\ -4.6496326e-12 & 2.6156599e-10 & \cdots \\ \vdots & \vdots & \vdots \\ -7.0361237e-11 & 3.0403313e-10 & \cdots \\ -3.9098910e-10 & -1.9523464e-10 & \cdots \end{bmatrix}$$

We then obtain the CSFM arrays for each input feature map and mean feature map and calculate scores according to Equation 2 and Equation 3, respectively as mentioned in the main paper.

X' denotes the CSFM array for the sample input feature map X and Y' represents CSFM array for the mean feature map Y, as follows:

$$X' = \begin{bmatrix} -1.9922838e-09 & -1.0934984e-10 & \cdots \\ -9.3061551e-08 & -9.2594959e-08 & \cdots \\ \vdots & \vdots & \vdots \\ 1.1881010e-06 & 1.2004004e-06 & \cdots \\ 1.3099764e-06 & 1.3227319e-06 & \cdots \end{bmatrix}$$

$$Y' = \begin{bmatrix} 1.7125739e-10 & 6.9036787e-10 & \cdots \\ 5.7084018e-09 & 5.9699676e-09 & \cdots \\ \vdots & \vdots & \vdots \\ 9.2276466e-08 & 9.2580499e-08 & \cdots \\ 1.0156344e-07 & 1.0136820e-07 & \cdots \end{bmatrix}$$

The score matrix is shown as:

$$Scores\_EMD_a = \begin{bmatrix} 0.00015478982822970 \\ 0.00042745511746034 \\ \vdots \\ 0.00070360308745876 \\ 0.00051849614828825 \end{bmatrix}$$

where the dimension is (number of features $\times 1$)

Furthermore, we negate these scores and obtain their Hadamard product and interpolate them linearly from 1 to 0.5, as specified in Equation 4 from the main paper:

$$MV_2 = \begin{bmatrix} 0.9444286823272705 \\ 0.8423951864242554 \\ \vdots \\ 0.7390584945678711 \\ 0.8083269596099854 \end{bmatrix}$$

the dimension being (number of features $\times 1$)

MV_2 provides us with the effective attention scores for the EMD_a layer.

We can obtain the resultant features array by following the same procedure as that of the EMD_m layer. $Result_2$ denotes this resultant feature matrix, as shown below:

$$Result_2 = \begin{bmatrix} \begin{bmatrix} -1.8815700e-09 \\ -7.5991787e-09 \\ \vdots \\ -2.4999771e-09 \\ -9.3262642e-10 \end{bmatrix} \\ \vdots \\ \begin{bmatrix} 6.8050197e-09 \\ 1.2046731e-08 \\ \vdots \\ 1.5541302e-08 \\ 1.4987041e-08 \end{bmatrix} \end{bmatrix}$$

These features are then forwarded to subsequent blocks of the proposed EMDA-Net model.