Preproceedings of the Federated Conference on Computer Science and Information Systems pp. 223–226
DOI: 10.15439/2021F145
ISSN 2300-5963 ACSIS, Vol. 26

Endoscopy Image Retrieval by Mixer Multi-Layer Perceptron

Quoc-Huy Trinh
University of Science, VNU-HCM, Vietnam
Vietnam National University, Ho Chi Minh City, Vietnam
Email: 20120013@student.hcmus.edu.vn

Minh-Van Nguyen
University of Science, VNU-HCM, Vietnam
Vietnam National University, Ho Chi Minh City, Vietnam
Email: 20127094@student.hcmus.edu.vn

***Abstract*—In Computer Vision, the Image Retrieval task is one of the interests of researchers, particularly medical image retrieval and endoscopy images. With the development of the Convolution Neural Network and Vision Transformer Technique, there are many proposals for using these techniques to make Image Retrieval Task and achieve a competitive result. In this paper, we propose a method that using Mixer Multi-Layer Perceptron architecture (Mixer-MLP) to build an Image Retrieval System with Medical images, particularly Endoscopic Images. This System base on the Classification process of Mixer-MLP architecture to generate vector representation for similarity calculation. The research result achieves competitively with efficient training time.**

I. Introduction

IMAGE RETRIEVAL TASK is the topic that using images to Retrieve the Image in the database[1]. In the Medical system, with the widespread in using digital imaging and storing, it causes difficult to query these large databases, this is the reason why there are high necessaries to use a content-based Image Retrieval system.[7]

An image retrieval system is a computer system for browsing, searching and retrieving images from a large database of digital images[17]. Most traditional and common methods of image retrieval utilize some method of adding metadata such as captioning, keywords, or descriptions to the images so that retrieval can be performed over the annotation words[18]. Manual image annotation is time-consuming, laborious and expensive; to address this, there has been a large amount of research done on automatic image annotation.

In recent years, the number of people that have been affected by colorectal cancer (CLC) is increasing. It is also on a third of the world for many years[12]. However, can we diagnose and prevent CLC is a crucial issue for the health organization[15]. Some studies illustrate that almost 95% of CLC is from the adenomatous polyp. The resection of Colorectal adenomatous polyps can reduce the CLC. On the other hand, the best way to deal with CLC is early diagnosis and have straight treatment. Nowadays, with the growth of the number of people that have CLC, digital imaging technique is applied to store the endoscopic images[13]. However, the doctor finds difficulty in querying the database because of the number of images in the database.[3]

Faculty of Information Technology, University of Science, Vietnam National University - Ho Chi Minh City

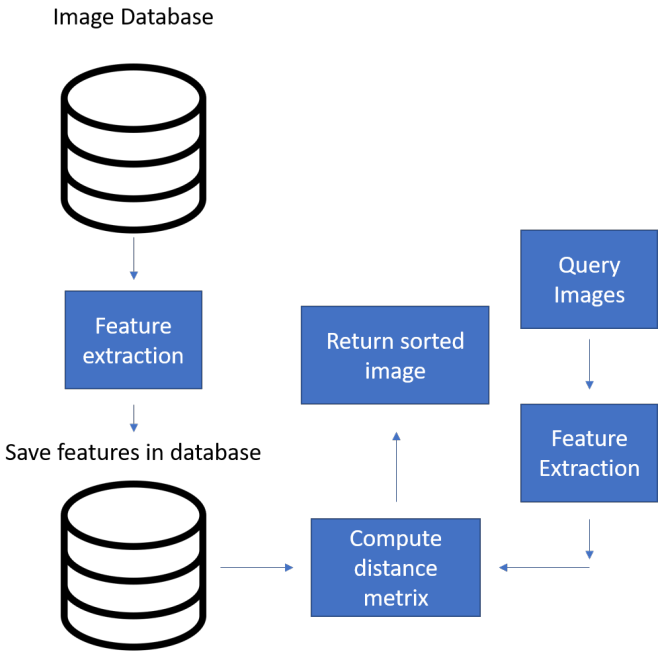

Fig. 1. Image Retrieval system

The figure above explain the image retrieval system. The method that we propose depend on this model.

Due to the development of Convolution Neural Network (CNN), there are many deep architectures applied in the feature-vector generating process such as ResNet, DenseNet and EfficientNet, etc. [5]. In early 2021, there is a method of using Mixer Multi-Layer Perceptron (Mixer-MLP) is proposed to classify images. [2]

In this paper, we build an Image Retrieval system on Endoscopic Images with a training process on Mixer-MLP architecture and approach a method to generate image vector representation from the model trained before.

Different from the previous Mixer-MLP architecture that is proposed in May 2021. We propose an architecture that base on the previous architecture, by cutting the classify layers, we add an embedding layer to generate the feature vector to represent the images feature. This architecture has the merit that it is based exclusively on multi-layer perceptrons.

II. Related Work

In our methods, we use some previous work to build our system. To help our research briefly, we review some past research to our best knowledge.

A. Mixer-MLP

In half of 2021, Googe AI has published an architecture that uses a variety of Multi-Layer Perceptrons (MLPs). There are two types of layers in this architecture: MLPs apply independently to image patches, MLPs apply across patches. This model achieves the competitive score on the image classification benchmark. This result is quite positive when compared with state-of-the-art models.[2]

B. Content-based Image Retrieval

Content-based Image Retrieval is a well-studied problem in computer vision, with retrieval problems generally divided into two groups: category-level retrieval and instance-level retrieval[8]. Given a query image of the Sydney Harbour Bridge, for instance, category-level retrieval aims to find any bridge in a given dataset of images, whilst instance-level retrieval must find the Sydney Harbour bridge to be considered a match.[6]

C. Similarity Search

To have accurate retrieval, they propose diversity similarity methods such as cosine similarity, Euclid distance, Manhattan distance. In this paper, we use the traditional Similarity search method to retrieve the images in the database collections.[9]

D. Image classification by Deep Learning models

In recent years, there is a variety of Deep Convolution Neural Network architecture to classify images, almost the result of that architecture achieve competitive. In later years, methods use state-of-the-art architecture while they combine Convolution Neural Network with Transformer to get the better result[14]. However, we can have a new approach by using Multi-layer Perceptrons to each patch of the images instead of using Convolution layers, lead to the method of Mixer-MLP. [2]

III. Dataset

To evaluate our system, we experiment with Kvasir Dataset. The Kvasir Dataset is collected using endoscopic equipment at Vestre Viken Health Trust (VV) in Norway. The VV consists of 4 hospitals and provides health care to 470.000 people. One of these hospitals (the Bærum Hospital) has a large gastroenterology department from where training data have been collected and will be provided, making the dataset larger in the future. Furthermore, the images are carefully annotated by one or more medical experts from VV and the Cancer Registry of Norway (CRN).[3]
The dataset consists of 80000 images in 10 folds for cross-validation in the training and evaluating process. 80000 images are split into eight classes: dyed-lifted-polyps, dyed-resection-margins, esophagitis, normal-cecum, normal-pylorus, normal-z-line, polyps and ulcerative-colitis.[3]

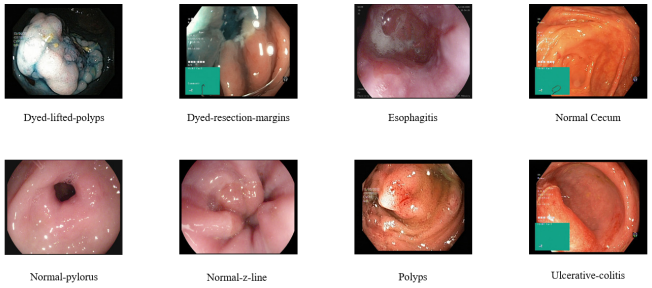

Fig. 2. The sample of Kvasir Dataset.

In our method, we propose a system with two parts: Collection generation and Retrieval. Meanwhile, we also approach to training classification model for the Kvasir dataset by using Mixer-MLP architecture.

IV. The Proposed Approach

A. Training Model with Mixers-MLP

1) Data Preparation: After loading data, we resize all the images to the size (150,150), then we split the dataset into the training set and validation set in the ratio of 0.75:0.25. After resizing and splitting the validation set, we rescale the data pixel down to be in the range [-1,1] by divide by 127.5.

2) Environment Setup: We prepare the training process on GPU Nvidia P100. The data is load with the batch size is 32, and preprocessing after loading.
With model initialization, we set the parameter for initializing the model in the following table:

B. Generating vectors for data representation

To generating the vector from the database and the vector for the query, we propose to replace the classification layers of Mixer-MLP with a Dense layer that enables to generate a vector for representation, then we will flatten the vector output to one dimension to facilitate similarity calculation.

TABLE I
Mixer-MLP initialization.

Parameter	Value
Number of MLPs block	8
Number of Patches	5
Input shape	(150,150,3)
Channel Mixing Unit	256
Token Mixing Unit	2048
Projection Unit	512
Patch size	5
Learning Rate	0.001
Optimizer	Adam

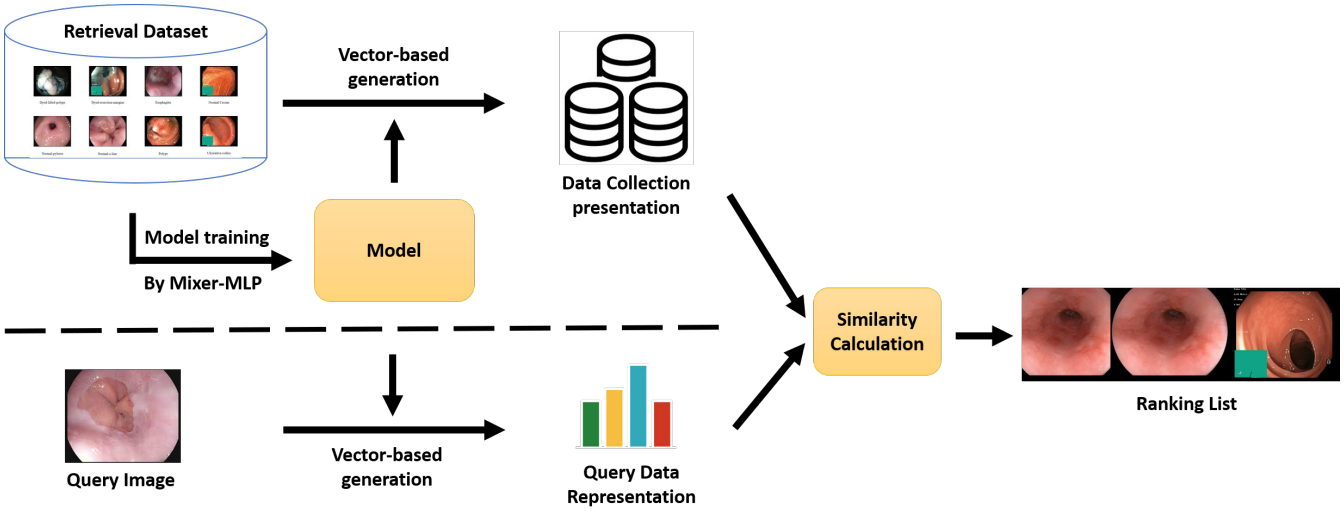

Fig. 3. Retrieval system by using Mixer MLP classification

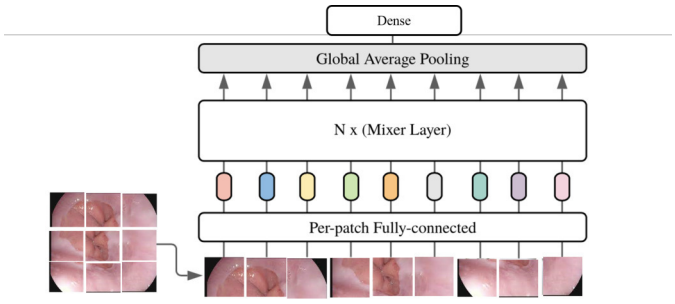

Fig. 4. Architecture of vector extraction for data representation

C. Similarity Calculation

To calculate between the query vector and database collection vector, we use cosine similarity with the following formula:[8]

$$Similarity(A,B) = cos(\phi) = \frac{\vec{A}\times\vec{B}}{|\vec{A}|\times|\vec{B}|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \times \sqrt{\sum_{i=1}^{n} B_i^2}}$$

Where:
A: is the vector A that has n elements
B: is the vector B that has n elements
The value of similarity is in the range from 0 to 1, depend on the value of cosine value. This can help cost efficient in computing.
After calculate the distance between the query vector and database collection vectors, the result is sorted to give back the rank of the result.

D. Data Augmentation

Data Augmentation is vital in data preparation process. Data Augmentation improve the number of data by adding sligtly modified copy of already exist data or newly created synthetic data from existing data to decrease the probability of the Overfitting problem, we use augmentation to generate the data randomly by random flip images and random rotation with an index of 0.2..[16]

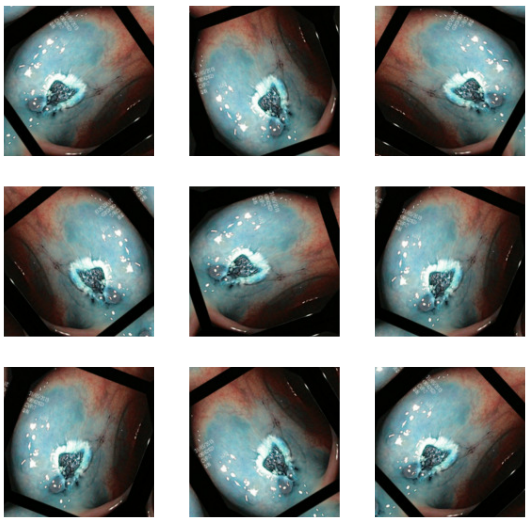

Fig. 5. Data after augmentation

V. Evaluation and Discussion

A. Model training

After 10 epochs, we got the following loss. The performance of the model on the Kvasir dataset is quite competitive, this can lead to the vector generation step can get a competitive score.

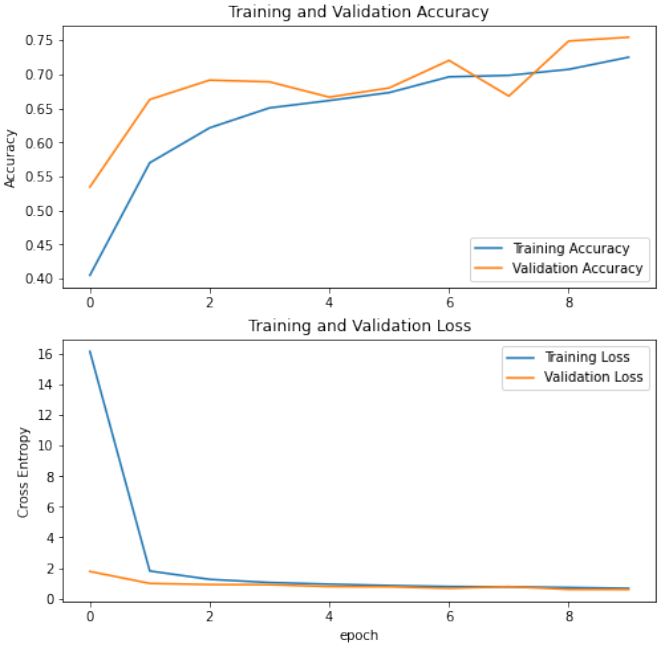

Fig. 6. Retrieval system by using Mixer MLP classification

The result on the test set achieves 0.8251 on the test set with 4000 images. This result can ensure the vector generation process of the system.

B. Discussion

Although our method gets a competitive score, there are some drawbacks in our methods: the training time gets long with 458ms/step, we can custom layers in the architecture to accelerate the computing cost. We can get more layers or can ensemble more backbones to achieve higher results.
The model can open a new approach to the Image Retrieval Task. However, there we have to deal with drawbacks by using this proposal:
- We have to improve the accuracy of the Mixer-MLP model to improve the quality of the vector generation process.
- We can improve the time for generating the vector representation for collection and query.

VI. Conclusion

In general, our research proposes the method that using Mixer Multilayers Perceptron to extract the feature of endoscopic images for image retrieval system. Our method achieves a competitive result on content-based retrieval. By using the classification model of Mixer-MLP and cutting the classification layer, we can generate a vector to represent the data feature. That will enable a new approach in the Content-based Image Retrieval task. Moreover, our method inspires the new approach to extract the feature for retrieval task instead of using the previous methods such as Deep Convolution Neural Network architecture or Vision Transformer.

VII. Acknowledgments

This research is supported by research funding from Faculty of Information Technology, University of Science, Vietnam National University - Ho Chi Minh City.

References

[1] Fan Yang, Ryota Hinami, Yusuke Matsui, Steven Ly, Shin'ichi Satoh, Efficient Image Retrieval via Decoupling Diffusion into Online and Offline Processing.
[2] Ilya Tolstikhin, Neil Houlsby, Alexander Kolesnikov, Lucas Beyer, Xiaohua Zhai, Thomas Unterthiner, Jessica Yung, Andreas Steiner, Daniel Keysers, Jakob Uszkoreit, Mario Lucic, Alexey Dosovitskiy, Mixer-MLP: An all-MLP architecture for Vision.
[3] Konstantin Pogorelov, Kristin Ranheim Randel, Carsten Griwodz, Sigrun Losada Eskeland, Thomas de Lange, Dag Johansen, Concetto Spampinato, Duc-Tien Dang-Nguyen, Mathias Lux, Peter Thelin Schmidt, Michael Riegler, Pål Halvorsen, Kvasir: A Multi-Class Image Dataset for Computer Aided Gastrointestinal Disease Detection, In MMSys'17 Proceedings of the 8th ACM on Multimedia Systems Conference (MMSYS), Pages 164-169 Taipei, Taiwan, June 20-23, 2017.
[4] Filip Radenovic Giorgos Tolias Ond ´ ˇrej Chum, Fine-tuning CNN Image Retrieval with No Human Annotation
[5] Jinyun Lu , Image Retrieval Based on ResNet and KSH,Advances in Intelligent Systems Research, volume 147.
[6] Huiyi Hu, Wenfang Zheng, Xu Zhang, Xinsen Zhang, Jiquan Liu, Weiling Hu, Huilong Duan, Jianmin Si, Content-based gastric image retrieval using convolutional neural networks, Imaging Systems and Technology, pages 439-449.
[7] Sun Q, Yang Y, Sun J, Yang Z, Zhang J, eds. Using deep learning for content-based medical image retrieval. Paper presented at: Medical Imaging 2017: Imaging Informatics for Healthcare, Research, and Applications; 2017: International Society for Optics and Photonics.
[8] JC Felipe, AJ Traina, C Traina, eds. Retrieval by content of medical images using texture for tissue identification. Paper presented at: 16th IEEE Symposium Computer-Based Medical Systems, 2003 Proceedings; 2003: IEEE.
[9] A Rashno, S Sadri, eds. Content-based image retrieval with color and texture features in neutrosophic domain. Paper presented at: 2017 3rd International Conference on Pattern Recognition and Image Analysis (IPRIA); 2017 19–20 2017.
[10] Cai Y, Li Y, Qiu C, Ma J, Gao X. Medical image retrieval based on convolutional neural network and supervised hashing. IEEE Access. 2019; 7: 51877- 51885.
[11] Hasan MM, Islam N, Rahman MM. Gastrointestinal polyp detection through a fusion of contourlet transform and neural features. J King Saud Univ-Comput Info Sci. 2020.
[12] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei, "Line: Large-scale information network embedding," in Proceedings of the 24th International Conference on World Wide Web. ACM, 2015, pp. 1067–1077.
[13] F Sommen, S Zinger, EJ Schoon, eds. Computer-Aided Detection of Early Cancer in the Esophagus Using HD Endoscopy Images. Medical Imaging 2013: Computer-Aided Diagnosis. Vol. 8670. Florida: International Society for Optics and Photonics; 2013.
[14] Yu D, Seltzer ML, Li J, Huang J-T, Seide F. Feature learning in deep neural networks-studies on speech recognition tasks. arXiv. 2013;13013605.
[15] Nini Rao, Hongxiu Jiang, Chengsi Luo: Review on the Applications of Deep Learning in the Analysis of Gastrointestinal Endoscopy Images., Article in IEEE Access - September 2019
[16] Chung Y-A, Weng W-H. Learning deep representations of medical images using siamese cnns with application to content-based image retrieval. arXiv. 2017;171108490.
[17] Antoine Miech, Jean-Baptiste Alayrac, Ivan Laptev, Josef Sivic, Andrew Zisserman, Thinking Fast and Slow: Efficient Text-to-Visual Retrieval With Transformers, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 9826-9836
[18] JohnA.HForrestN.D.CFinlaysonD.J.CShearman, ENDOSCOPY IN GASTROINTESTINAL BLEEDING, the Lancet, Volume 304, Issue 7877, 17 August 1974, Pages 394-397.