Methods and Applications

Zero-Shot Medical Image Retrieval for Emerging Infectious Diseases Based on Meta-Transfer Learning — Worldwide, 2020

Yuying Zhao[1]; Hanjiang Lai[1,#]; Jian Yin[1]; Yewu Zhang[2]; Shigui Yang[3]; Zhongwei Jia[4]; Jiaqi Ma[2]

ABSTRACT

Introduction: Due to the increasing number of medical images, image retrieval has become an important technique for medical image analytics. Although many content-based image retrieval methods have been proposed, the retrieval of images in datasets related to emerging/new infectious diseases still remain a challenge–mostly due to the lack of historical data. As a result, the current retrieval models have limited functionality in helping doctors make accurate diagnoses of new diseases.

Methods: In this paper, we propose a zero-shot retrieval model based on meta-learning and ensemble learning, which can obtain a model with stronger generalizability without using any relevant training data, and thus performs well on new types of test data.

Results: The experimental results showed that the proposed method is 3% to 5% higher than the traditional method, which means that our model can retrieve relevant medical images more accurately for newly emerging data types and provide doctors with more effective assistance.

Discussion: When a new infectious disease occurs, doctors can use the proposed zero-shot retrieval model to retrieve all relevant cases, quickly find the common problems of patients, find the locations of the new infections, and determine its infectivity as soon as possible. The proposed method is a new computer-aided decision support technology for emerging infectious diseases.

INTRODUCTION

Recently, artificial intelligence (AI) technologies have been widely used in the medical industry. With the developments of digital imaging techniques, e.g., computed tomography (CT) and X-ray, millions or even billions of medical images have been generated. Image retrieval technology (*1*), which retrieves similar medical images from large-scale image datasets that contain patient physiological, pathological, and anatomical information, can be used as an important objective basis for assisting doctors in clinical diagnosis, disease tracking, and surgical research. With the help of image retrieval technology, it was possible to retrieve all similar cases in the database using the patient's medical pictures and to assist doctors in making more accurate and universal diagnoses.

In parallel, the newly emerging diseases, e.g., emerging infectious diseases, is a challenging problem for public health control. For example, coronavirus disease 2019 (COVID-19) emerged in Wuhan at the beginning of 2020 and caused numerous casualties and social losses. When a new infectious disease appears, it is difficult for the doctor, who can only rely on previous experiences, to quickly find the common patterns of the new disease without any historical data of the disease. In order to assist doctors in making a correct diagnosis quickly, a possible solution is computer-aided decision support technology, such as medical image retrieval. To analyze the new disease, the retrieval model can find all visually similar images of the new disease, which can be used to explore the common patterns of the disease, the therapeutic plan, etc. However, due to the lack of training data in new cases such as COVID-19, the performance of the existing retrieval models is greatly reduced.

The zero-shot retrieval model has been proposed to solve this problem. In the absence of relevant training data, the zero-shot retrieval model tries to find similar images from unseen image datasets. The current mainstream zero-shot learning models, such as SitNet (*2*) and AgNet (*3*), use text information as an aid to train the model. They map the image modality and the text modality into the same semantic space. In this way, semantic information of the label of unknown images can be used to learn the association between the unknown category and the known category despite the image data being unobtainable. However, for emerging infectious diseases, a new term is often used to name the disease, which has never appeared in the corpus of the text model. This will cause the text space itself to be unreliable, making it difficult to map the corresponding medical image to the same vector space.

In this paper, we propose a zero-shot hashing model with stronger generalizability, which can train the model without using text-based auxiliary information, so that the model can assist doctors in analyzing the emerging infectious diseases as soon as possible. Inspired from meta-learning (*4*) and ensemble learning (*5*), we combine the new training process (*6*) and model update method (*7*) to improve the generalizability of the final model, thereby improving its retrieval performance on new diseases. Our approach consists of two parts. First, we aimed to solve the shift problem that a model performs well in the training data domain while performing poorly in the other data domain with different statistics. To alleviate this problem, we introduced a virtual test domain, which was denoted as the meta-test domain in this paper, in the training process to simulate the domain shift between the training domain and the test domain (as shown in Figure 1). Without the need for semantic label information, the model's generalizability to the unknown domain was improved. Second, when the model was updated, the periodic learning rate was used to train the model, and the multipoint simple average of the Stochastic Gradient Descent (SGD) (*8*) trajectory was obtained as the final model by the moving average method, thereby obtaining a lower loss, globalized general solution, and further improving the generalizability.

METHODS

The goal of zero-shot retrieval is to retrieve images of novel classes although there were no training samples of these categories in the training set. Let $S_{tr}=\{(x_i, y_i)\,|y_i \in Y_{train}\}$ be denoted as the training set, where x_i is the i-th image and y_i is the class label. We also denoted $S_{te}=\{(x_j, y_j)\,|y_j \in Y_{test}\}$ as the test set. Please note that none of the test classes occur in the training set. In this paper, we aimed to learn a retrieval model X->Y using only the training set, and the model can also perform well on the test set.

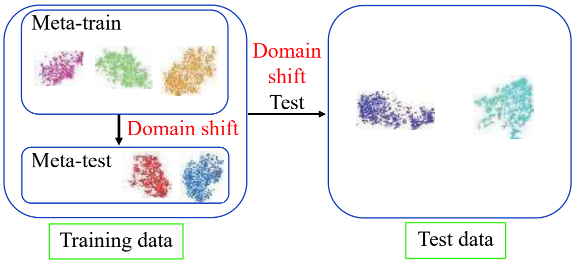


FIGURE 1. Meta-transfer learning for the domain shift problem where the distribution of different colors indicates different data domains.

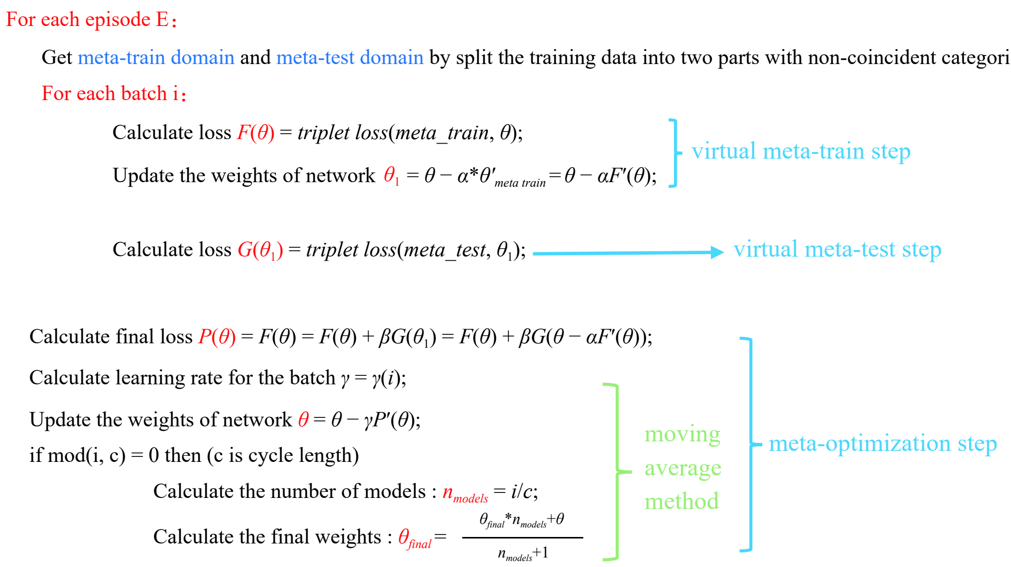
To establish good mapping, we needed to solve the domain shift problem. Our main idea is that the training process should be the same as the testing process. The main procedures are shown in Figure 2. To introduce the virtual test domain to simulate the real test process in the zero-shot task, we split the training data into two parts with non-coincident categories at the beginning of each round of training and get the meta-train domain and meta-test domain. For example, if we have one class as the test domain, and the last 9 classes as the training domain, we can in each round of training choose one class randomly from the training domain as the meta-test domain and the last 8 classes as the meta-train domain. Our training goal is to minimize the loss of the model on the virtual training domain, while also guaranteeing that the direction of the gradient update can reduce the loss on the virtual test domain. Our training goal is actually to train the model to generalize the unknown domain. The training process is divided into three steps.

The first step is virtual meta-train. We calculate the loss of model θ on meta-train $F(\theta) = triplet\ loss(meta\_train,\ \theta)$ and backpropagate the obtained loss (θ) to update the network parameters so that we can obtain new weights $\theta_1 = \theta - \alpha F'(\theta)$. It is worth mentioning that the loss function we use is triplet loss, which can shorten the distance between the image's hash codes of the same category and increase the distance between the image's hash codes of different categories. In order to calculate the triplet loss, we first need to construct a tuple $<I, I_{pos}, I_{neg}>$ (where the origin I is a sample randomly selected from the training data, I_{pos} is the sample of the same category as I, and I_{neg} is a sample of a different category from I). The calculation formula of triple loss is as follows:

$$tripletloss = max(\|I - I_{pos}\|_2^2 - \|I - I_{neg}\|_2^2 + margin,\ 0)$$

The hyperparameter margin in the formula represents the minimum difference between $dis(I, I_{neg})$ and $dis(I, I_{pos})$.

The second step is virtual meta-test. Because our ultimate goal is not only to make the trained model perform well on the training domain, but also hope that the θ_1 model will also perform well on the test domain. We calculate the loss of model θ_1 on meta-test $G(\theta_1) = triplet\ loss(meta\_test,\ \theta_1)$.



```
For each episode E：
    Get meta-train domain and meta-test domain by split the training data into two parts with non-coincident categories.
    For each batch i：
        Calculate loss $F(\theta) = triplet\ loss(meta\_train, \theta)$;
        Update the weights of network $\theta_1 = \theta - \alpha*\theta'_{meta\ train} = \theta - \alpha F'(\theta)$;

        Calculate loss $G(\theta_1) = triplet\ loss(meta\_test, \theta_1)$;

    Calculate final loss $P(\theta) = F(\theta) = F(\theta) + \beta G(\theta_1) = F(\theta) + \beta G(\theta - \alpha F'(\theta))$;
    Calculate learning rate for the batch $\gamma = \gamma(i)$;
    Update the weights of network $\theta = \theta - \gamma P'(\theta)$;
    if mod(i, c) = 0 then (c is cycle length)
            Calculate the number of models : $n_{models} = i/c$;
            Calculate the final weights : $\theta_{final} = \frac{\theta_{final}*n_{models}+\theta}{n_{models}+1}$
```

FIGURE 2. Main procedures of the zero-shot medical image retrieval.

The third step is meta-optimization. We use the weighted sum of (θ) and $G(\theta_1)$, which is $P(\theta) = F(\theta) + \beta G(\theta_1) = F(\theta) + \beta G\left(\theta - \alpha F'(\theta)\right)$, as the final loss to update the model θ with gradient backhaul. Performing a first-order Taylor transformation on the second term (x), we can get

$$final\ loss = F(\theta) + \beta \times G(\theta) - \beta \times \alpha \times F'(\theta) \times G'(\theta).$$

From this formula, we can see that the *final loss* has two functions: 1) minimize the loss of the model in the two domains of virtual meta-train and virtual meta-test; 2) maximize the product of the loss gradient of the model in the virtual meta-train and virtual meta-test domains. The smaller the angle, the larger the vector product. Therefore, the gradient directions of these two fields can be made to be consistent. Because each round of training will re-divide meta-train and meta-test, the entire training process will make the gradient directions of any two domains to tend to be the same. Finding the direction in which the loss of two sub-problems decreases simultaneously each time to update the parameters can reduce overfitting to a single domain.

In addition, we were inspired by the idea that the random weight average can find a wider optimal range compared to SGD, so we used cyclic learning rate to train the model and used the moving average method to calculate the average of the multiple SGD trajectory as the final model.

To verify the validity of this method, we do experiments on a widely-used medical dataset to evaluate the proposed method. We randomly sampled 5% of images from the NIH Chest X-Ray Dataset (*9*) and created a smaller dataset (*10*), which contains 5,606 images that were classified into 15 classes, including 14 common chest lesions (such as atelectasis, consolidation, infiltration, pneumothorax, edema, etc.) and one for "No findings." To simulate the situation of new diseases, such as an emerging infectious disease, we randomly selected in our experiment one disease (e.g., infiltration) as the new disease, and the other 14 types of diseases as the training set. All the samples were used as the database for retrieval. In our experiment, we trained a retrieval model on the 14 diseases and aimed to achieve a good retrieval performance on the new disease without using any data from the new disease.

In the experimental setting, all images were resized to 224×224 resolution, and we used the pretrained model Alexnet (*11*) to extract image features with 4,096 dimensions. The learning rate was set to 10^{-4} and the momentum was set to 0.9. The weight decay parameter was 0.0005. The mini batch was set to 64. We chose the conventional training method and network update method as the baseline, and mean Average Precision (mAP) based on Hamming ranking as the evaluation metric.

All the experimental settings of the traditional method we used for comparison experiments were the same as the above settings. The only difference was that meta-learning was not used in the training process, and ensemble learning was not used in the network update process. The traditional method only used Alexnet (*11*) to extract image features, and then obtained the hash code of the image and finally used triplet loss as the loss function and SGD as the network update method to train the network.

RESULTS

The comparison results were shown in Table 1. From Table 1, we can see that, in terms of retrieval index mAP, the proposed method is 3% to 5% higher than the traditional method, indicating that our method is effective. In addition, we can also see that we have tried 4 different lengths of hash codes, which are 8 bits, 16 bits, 32 bits, and 48 bits, which increase to 5.342%, 3.148%, 3.769%, and 4.527%, respectively. In the case of all hash code lengths, the proposed method has higher retrieval accuracy than traditional methods, which demonstrates the effectiveness of our proposed method.

DISCUSSION

In our study, we pointed out the importance of popularizing artificial intelligence applications in the diagnosis of emerging infectious diseases and analyzed the limitations of existing image retrieval models on this issue. In response to the lack of training samples and inaccurate cognition of new terms, we proceeded from the perspective of improving the model's generalizability for new categories, and finally proposed a zero-shot hashing model that can achieve good retrieval results without using additional text tags. We verified the effectiveness and feasibility of the proposed method on a widely used medical dataset and found that the simulation test process can indeed make the model accustomed to identifying new categories. The application flowchart of our image retrieval model was shown in Figure 3. Experimental results showed that our model could retrieve relevant pictures more accurately, so the proposed model could be used to assist doctors in making the correct diagnosis quickly when an emerging infectious disease occurs and improve public health.

TABLE 1. mAP of baseline and proposed method on Chest X-ray Dataset.

Method	8 bits	16 bits	32 bits	48 bits
Baseline	27.523	31.231	31.472	32.825
Ours	32.865	34.379	35.241	37.352

This study was subject to some limitations. First, the loss function used in this method was no different from the ordinary retrieval model. We can further explore better loss functions that can improve the generalizability of the model, such as adding regularization items or considering the relationship between different classes. Second, if supplementary information of the sample is added to this method, such as image attributes, the retrieval effect can be further improved. Next, we will consider and practice these ideas in more detail.

Fundings: This study was supported by grants from the Key Joint Project for Data Center of the National Natural Science Foundation of China and Guangdong Provincial Government (U1611264), and the Pearl River Nova Program of Guangzhou (201906010080).

doi: 10.46234/ccdcw2020.268

# Corresponding author: Hanjiang Lai, laihanj3@mail.sysu.edu.cn.

[1] School of Computer Science and Engineering, Sun Yat-Sen University, Guangzhou, Guangdong, China; [2] Center for Public Health Surveillance and Information Service, Chinese Center for Disease Control and Prevention, Beijing, China; [3] State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, Zhejiang, China; [4] School of Public Health, Peking University, Beijing, China.

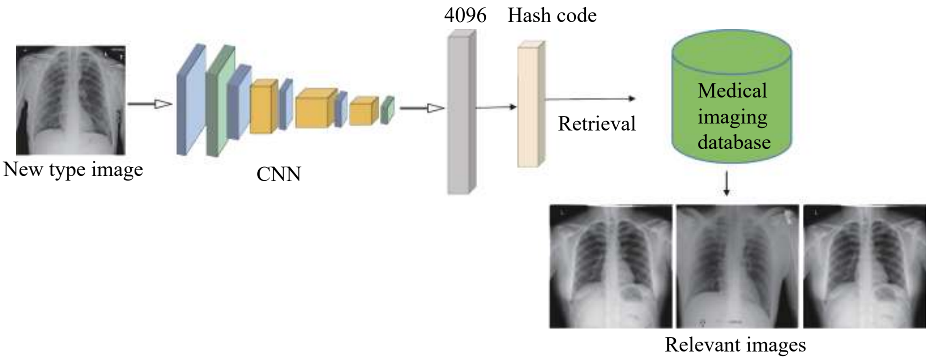


FIGURE 3. Schematic diagram of image retrieval for new type of lung disease image.

Submitted: December 03, 2020; Accepted: December 21, 2020

REFERENCES

1. Li ZY, Zhang XF, Müller H, Zhang ST. Large-scale retrieval for medical image analytics: a comprehensive review. Med Image Anal 2018;43:66 – 84. http://dx.doi.org/10.1016/j.media.2017.09.007.
2. Guo YC, Ding GG, Han J, Guo Y. SitNet: discrete similarity transfer network for zero-shot hashing. In: Proceedings of the 26th international joint conference on artificial intelligence. Melbourne, VIC, Australia: IJCAI. 2017;p.1767 – 73. http://dx.doi.org/10.24963/ijcai.2017/245.
3. Ji Z, Sun YX, Yu YL, Pang YW, Han JG. Attribute-guided network for cross-modal zero-shot hashing. IEEE Trans Neural Netw Learn Syst 2020;31(1):321 – 30. http://dx.doi.org/10.1109/TNNLS.2019.2904991.
4. Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. In: Proceedings of the 34th international conference on machine learning. Sydney, NSW, Australia: ACM. 2017;p.1126 – 35. https://dl.acm.org/doi/10.5555/3305381.3305498.
5. Dong YP, Liao FZ, Pang TY, Su H, Zhu J, Hu XL, et al. Boosting adversarial attacks with momentum. In: Proceedings of 2018 IEEE/CVF conference on computer vision and pattern recognition. Salt Lake City: IEEE. 2018;p.9185 – 93. http://dx.doi.org/10.1109/CVPR.2018.00957.
6. Li D, Yang YX, Song YZ, Hospedales TM. Learning to generalize: meta-learning for domain generalization. https://arxiv.org/abs/1710.03463. [2020-11-28].
7. Izmailov P, Podoprikhin D, Garipov T, Vetrov D, Wilson AG. Averaging weights leads to wider optima and better generalization. https://arxiv.org/abs/1803.05407. [2020-11-28].
8. Ruder S. An overview of gradient descent optimization algorithms. https://arxiv.org/abs/1609.04747. [2020-11-28].
9. Kaggle. NIH Chest X-rays. https://www.kaggle.com/nih-chest-xrays/data. [2020-11-28].
10. Kaggle. Random Sample of NIH Chest X-ray Dataset. https://www.kaggle.com/nih-chest-xrays/sample. [2020-11-28].
11. He KM, Zhang XY, Ren SQ, Sun J. Deep residual learning for image recognition. In: Proceedings of 2016 IEEE conference on computer vision and pattern recognition. Las Vegas: IEEE. 2016;p.770 – 8. http://dx.doi.org/10.1109/CVPR.2016.90.