Finding the Source Images From the Generated Images with Contrastive Learning Methods

Notebook for the ImageCLEFmedical GANs Lab at CLEF 2023

Shitong Cao[1], Xiaobing Zhou*

School of Information Science and Engineering, Yunnan University, Kunming 650504, Yunnan, China

Abstract

This paper provides an notebook for the ImageCLEFmedical GANs lab at CLEF 2023. Using generative adversarial networks for medical image generation is a standard data-expanding method. However, the quality of the generated data is not high, and finding the source based on the generated images is a new task worth investigating, as finding the source of the real data can ensure the reliability of the generated data. The GANs task is a completely new challenge in the ImageCLEFmedical track. The task is focused on examining the existing hypothesis that GANs are generating medical images that contain the "fingerprints" of the real images used for generative network training. In this paper, our team(one five one zero) use contrastive learning to find real images with high response values through similarity calculations based on the natural similarity between real images and generated images. We use a triplet loss function for optimization.

Keywords

GANs, Contrastive Learning, Pre-trained Model, Triplet Loss

1. Introduction

Generating medical images is a fundamental problem in medical imaging and can be used for applications such as data enhancement and model training[1]. However, the commonly used methods for medical image generation are all deep learning-based Generative Adversarial Networks (GANs), which require a large amount of real image data for training. In practical applications, the acquisition of real image data is often restricted by many aspects, such as data privacy and data protection. Therefore, false data generation has become a hot research direction in the field of medical imaging.

However, due to the unclear data sources in the process of false data generation, the quality and reliability of incorrect data have also been widely questioned[2]. Therefore, in the field of medical image generation, it is important to find the provenance of the real data to ensure the quality and reliability of the generated data. In addition, finding the provenance of real data can also help us scale up the dataset's size, improve the generalization ability of the model, and protect the privacy of the data.

CLEF 2023: Conference and Labs of the Evaluation Forum, September 18–21, 2023, Thessaloniki, Greece

*Corresponding author.

stcao@mail.ynu.edu.cn (S. Cao); zhouxb@ynu.edu.cn (X. Zhou)

0009-0003-1298-4166 (S. Cao); 0000-0003-1983-0971 (X. Zhou)

The GANs task is an entirely new task in the ImageCLEFmedical track[3]. The task is focused on examining the existing hypothesis that GANs are generating medical images that contain the "fingerprints" of the real images used for generative network training[4]. The results of this task, for better or worse, can tell us something valuable. If the hypothesis is correct, artificial biomedical images may be subject to the same sharing and usage limitations as real sensitive medical data. On the other hand, if the hypothesis is wrong, GANs may be potentially used to create rich datasets of biomedical images that are free of ethical and privacy regulations.

To address the requirements of this task, this paper uses similarity calculations with the generated data as the target, the images used for a generation as positive examples, and the real images unused for a generation as negative examples. The model learns the data distributions of the three by learning them through contrastive learning, and since the generated images are output through the positive examples, the nature of generative adversarial networks is to learn the data distribution, so the generated data has a higher similarity to the data distribution than the positive examples, so contrastive learning is used as a way to distinguish the distribution of the data.

2. Data Description and Task Analysis

Investigate the hypothesis that GANs are generating medical images that are in some way similar to the ones used for the GAN training. The task is related to the problem of the security of personal medical image data in the context of generating and using artificial images in different real-life scenarios.

The objective of the task is to detect “fingerprints” within the synthetic biomedical image data to determine which real images were used in training to produce the generated images. The task is to analyze test image datasets and assess the probability with which certain images of real patients were used for training image generators and which were not.

2.1. Data Description

The benchmarking image data are the axial slices of 3D CT images of about 8000 lung tuberculosis patients. This particularly means that some of them may appear pretty “normal,” whereas the others may contain certain lung lesions, including severe ones. These images are stored in the form of 8-bit/pixel PNG images with dimensions of 256x256 pixels.

The published development dataset for the task includes 500 artificial images, 80 real images which were unused for training generative neural networks as well as 80 real images taken from the image set which has been used for training the corresponding generative model. The test dataset was created in a similar way. The only difference is that the two subsets of real images are mixed, and no proportion of non-used and used ones has been disclosed. Thus, a total of 10,000 were generated, and 200 real images were provided.

2.2. Task Analysis

The process of generating a model is essentially learning the distribution of accurate data. By analyzing the real training data, you can learn the distribution of the data, and having learned

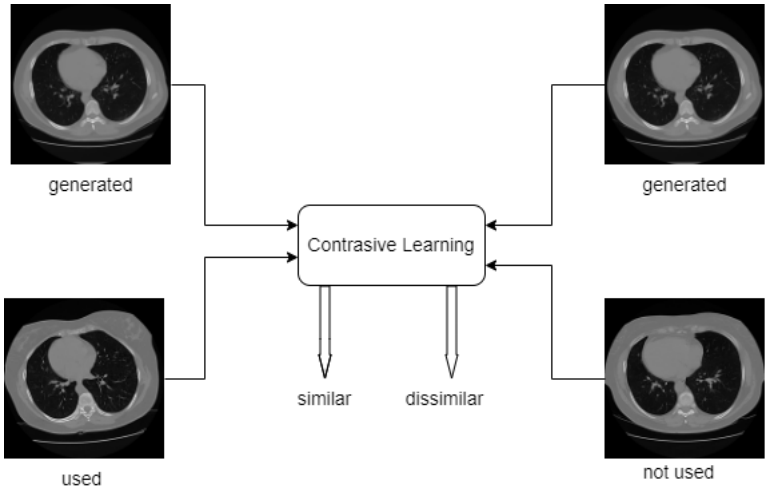

Figure 1: Comparative Learning Diagram. The generated image is obtained from the used image, so the two are similar in comparative learning. When comparing the generated image with the not used image for learning, it is dissimilar.

the distribution of the data, you can generate a lot of data that fits this data distribution based on the distribution. To find out the real images based on the generated fake images, the good idea is to fit the model, that is, to fit the generative ability of the model, because different models learn to learn different features; if you can know the network model and the parameters, you can achieve the inverse process, through the generated fake images reverse inference to the real images.

However, as the task did not tell us strictly what network was used, and the number of images given in the training set was very small, with only 80 of them used, it was difficult to train the generative adversarial network to achieve a simulation of the network model. Therefore, we can only focus on the distribution of the data, which is close to the distribution of the used images and differs significantly from the distribution of the unused data, so we can see the use of contrastive learning for the model's training.

3. System Description

3.1. Contrastive learning

Contrastive learning uses data that has been augmented as a positive sample and data that is not of the same category as a negative sample for comparison training[5, 6]. At the same time, the images generated by the generative network are not the data obtained by cropping, scaling, rotating, etc., in the traditional sense, but are also obtained from used images, so they can be understood as the same class of data with similar to the data distribution is identical. Hence, the features implied by both are similar. The unused image is not involved in generating the image, so the difference between its features and the data distribution of the generated image will be more pronounced, so a comparative learning approach is used in this paper to build it[7].

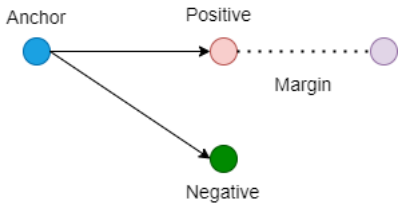

Figure 2: Schematic diagram of triplet loss. The generated image serves as the anchor, the used image serves as positive sample, and the not used image serves as negative sample.

3.2. Pre-training Models

This paper adopts a pre-trained model, which can be pre-trained on large-scale data to learn universal and robust feature representations[8]. These features can be migrated to various specific tasks, thereby improving the generalization ability and performance of the model. By utilizing pre-trained models, knowledge from large-scale data can be transferred to specific tasks, thereby reducing training time and data volume. This can greatly improve the efficiency and accuracy of the model. In some application scenarios, insufficient or incomplete data annotation makes it difficult to train models directly.

Pre-trained models can utilize large-scale unlabeled data to learn the potential structure and patterns of the data, thus addressing the problem of insufficient or incomplete data annotation. This paper uses various pre-training models, including Inception V3, ResNet, and EfficientNet.[9].

3.3. Triplet Loss Function

The triplet loss function describes the requirements of this task well[10]. In the design of the above model, the original image of the target is found by similarity by calculating the distance between the target and the positive and negative examples. This is consistent with the ternary loss function, as shown in Figure 2.

In this paper, our goal is to make the distance between the target and the positive example smaller than the distance between the target and the negative examples[11]. Therefore, we need to calculate the distance between the target, positive and negative examples, and use it as an input to the loss function. Specifically, we use the Euclidean distance to calculate the distance, as shown in Equation (1).

$$\text{loss} = \max(0, ||f(x) - f(x^+)||^2 - ||f(x) - f(x^-)||^2 + margin) \tag{1}$$

Where $f(x)$ denotes the generated image, $f(x^+)$ denotes the positive example, the used image, and $f(x^-)$ denotes the negative example, the unused image, and $margin$ is the hyperparameter.

Table 1
Quantity after data preprocessing

generated(train+test)	used	unused
11000	11000	11000

4. Experiments

4.1. Experimental Design

In existing experiments, training data was used for training and reasoning was done in test data. However, by analyzing the data, it can be seen that 500 images were generated in the training data, which were generated from 80 images. The data generated in the test data provided 10000 images. At this time, using training data for training and testing data for reasoning seemed inappropriate[12]. Due to the connection between the training data and the test data, it can be inferred that the 10000 images provided in the test have the same data distribution as the images in the training. Therefore, the generated images from the test data are also included in the training.

4.2. Experimental Data Processing

If the test data is placed in training, there will be imbalanced data. The generated data in the given test data has 10000 images, while the test data in the training data only has 500 pieces. Therefore, copying 500 pieces of the training data is equivalent to giving greater weight[13]. Thus achieving a balance between training and testing data. Secondly, there are only 80 used and unused images in the training, which means there are only 80 positive and negative images, respectively, which is a significant difference from the target's data volume. In comparative learning, it is necessary to take a target image, a positive image, and a negative image separately. After pre-processing, the number of data sets is shown in Table 1.

Due to the large size of the target data, the positive data was enhanced to 1000 images and replicated 10 times to maintain consistency with the target data. In this way, data can be taken from the target and then from positive and negative, respectively. The process of contrastive learning can be completed. The reverse update is completed through the triplet loss function, making the distance between the target and the positive closer and closer and the distance between the target and the negative farther and farther.

4.3. Experimental Results and Analysis

This task has three evaluation metrics: Accuracy, Recall, and F1 score. In this experiment, a total of three results were submitted, and three pre-trained models were used, i.e., Inception V3, ResNet and EfficientNet. The experimental equipment used in this experiment is RTX 3070, with a learning rate of 1e-4. The final submitted result scores are shown in Table 2.

As can be seen from Table 2, the best results were achieved by pre-training the model based on EfficientNet, which optimises the depth, width and resolution of the network at the same

Table 2
Presentation of experimental results

Pre-trained model	Accuracy	Recall	F1 score
Inception V3	0.48	0.57	0.522
ResNet50	0.515	0.5	0.507
EfficientNet	0.52	0.62	0.563

time and is highly versatile and scalable, and therefore performs better.

5. Conclusion

In this paper, to complete the task of finding the original image according to the generated image, from the perspective of the similar data distribution between the generated image and the original image, we use the contrastive learning architecture, combined with transfer learning, and use the pre-trained feature extraction module to find the target with large response value through the similarity calculation method, so as to find the original image. Future research can attempt more fine-grained networks, such as those in face detection, as the generated medical images are highly similar and differ mainly in some fine-grained features. This is in line with the networks in face detection, and further research can be considered from this perspective.

References

[1] T. Salimans, G. et al., Improved techniques for training gans, Advances in neural information processing systems 29 (2016).

[2] B. Dolhansky, B. et al., The deepfake detection challenge (dfdc) dataset, arXiv preprint arXiv:2006.07397 (2020).

[3] B. Ionescu, H. Müller, A. Drăgulinescu, W. Yim, A. Ben Abacha, N. Snider, G. Adams, M. Yetisgen, J. Rückert, A. Garcıa Seco de Herrera, C. M. Friedrich, L. Bloch, R. Brüngel, A. Idrissi-Yaghir, H. Schäfer, S. A. Hicks, M. A. Riegler, V. Thambawita, A. Storås, P. Halvorsen, D. J. A. A. A. R. I. C. V. K. A. S. G. I. Nikolaos Papachrysos, Johanna Schöler, H. Manguinhas, L. Ştefan, M. G. Constantin, M. Dogariu, J. Deshayes, A. Popescu, Overview of ImageCLEF 2023: Multimedia retrieval in medical, socialmedia and recommender systems applications, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction, Proceedings of the 14th International Conference of the CLEF Association (CLEF 2023), Springer Lecture Notes in Computer Science LNCS, Thessaloniki, Greece, 2023.

[4] A. Andrei, A. Radzhabov, I. Coman, V. Kovalev, B. Ionescu, H. Müller, Overview of ImageCLEFmedical GANs 2023 task – Identifying Training Data "Fingerprints" in Synthetic Biomedical Images Generated by GANs for Medical Image Security, in: CLEF2023 Working Notes, CEUR Workshop Proceedings, CEUR-WS.org, Thessaloniki, Greece, 2023.

[5] T. Chen, K. et al., A simple framework for contrastive learning of visual representations, in: International conference on machine learning, PMLR, 2020, pp. 1597–1607.

[6] T. Kipf, V. der Pol et al., Contrastive learning of structured world models, arXiv preprint arXiv:1911.12247 (2019).
[7] C.-Y. Chuang, R. et al., Debiased contrastive learning, Advances in neural information processing systems 33 (2020) 8765–8775.
[8] X. Qiu, S. et al., Pre-trained models for natural language processing: A survey, Science China Technological Sciences 63 (2020) 1872–1897.
[9] H. Chen, W. et al., Pre-trained image processing transformer, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 12299–12310.
[10] W. Deng, Z. et al., Rethinking triplet loss for domain adaptation, IEEE Transactions on Circuits and Systems for Video Technology 31 (2020) 29–37.
[11] B. Yu, L. et al., Correcting the triplet selection bias for triplet loss, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 71–87.
[12] S. Kotsiantis, K. et al., Handling imbalanced datasets: A review, GESTS international transactions on computer science and engineering 30 (2006) 25–36.
[13] N. V. Chawla, Data mining for imbalanced datasets: An overview, Data mining and knowledge discovery handbook (2010) 875–886.