KDE Lab at ImageCLEFmedical GANs 2024

Shota Fukuyama[1], Tetsuya Asakawa[1], Kazuki Shimizu[2], Kei Nomura[2] and Masaki Aono[1]

[1]Toyohashi University of Technology, 1-1 Hibarigaokam Tempakucho, Toyohashi, Aichi, 441-8580, Japan
[2]Toyohashi Heart Center, 21-1 Gobutori, Oyama-cho, Toyohashi, Aichi, Japan, 441-8530

Abstract

This paper describes KDE Lab approach in ImageCLEF GANs 2024. ImageCLEFmedical GANs 2024 task consists of two sub tasks, Identify training data fingerprints task and Detect Generative models'fingerprints task. Identify training data fingerprints task aims to examine the existing hypothesis that GANs are generating medical images that contain certain "fingerprints" of the real images used for generative network training. Detect Generative models'fingerprints task aims to investigate the hypothesis that generative models imprint distinctive "fingerprints" onto the generated images. We studied Identify training data fingerprints task in this research. For experiment, we attempted a methods that is approach using Multiple pre-training models and the results were compared. Also, we attempted a methods that is approach using super-resolution technique. Finally, the experiment was a failure and the results could not be compared. The competition result had an accuracy of 0.484.

Keywords

Medical Images, GANs, CNN, super-resolution

1. Introduction

In recent years, the emergence of various generative models has made it possible to construct highly accurate models based on a small amount of image data. In the medical field, it has been difficult to prepare an even and large amount of images with specific attributes, such as image data of rare diseases, resulting in biased data, but medical image generation AI is solving this problem. However, the problem arises when medical images can be falsified using these technologies. Generated images must be useful data for training models, and it must be possible to determine whether they are generated images or not.

This paper describes KDE Lab approach in Identify training data fingerprints task of ImageCLEF GANs 2024 [1]. ImageCLEF has been held as part of CLEF since 2003. ImageCLEF2024 [2] focus on multiple applications between different tasks,

and ImageCLEFmedical GANs Task is one of them. We have taken an approach to determine whether a generated image is a real image or a generated image. In the architecture, MobileNetv2 [3], DenseNet-121 [4] , and ResNet-50 [5], which were pre-trained on ImageNet [6] , were used for the CNN. In the following, we will describe dataset, methods, experiment and results.

2. Dataset

In this research, we use some of the images given in the 'Identify training data "fingerprints" Tasks' of the ImageCLEFmedical GANs 2024. The image data are axial slices of 3D CT images of about 8000 pulmonary TB patients. These images are stored in the form of 256x256 pixel 8-bit/pixel PNG images. From that image data, we used a total of 10,200 images, 10,000 images generated from GAN model 1, which is not revealed, and 100 images each annotated as having been used/not used for training the image generation for that model. The 10,000 images generated from GAN model 2 that were not

CLEF 2024: Conference and Labs of the Evaluation Forum, September 09–12, 2024, Grenoble, France

fukuyama.shota.jf@tut.jp (S. Fukuyama); asakawa.tetsuya.um@tut.jp (T. Asakawa); shimizu@heart-center.or.jp (K. Shimizu); kein312@gmail.com (K. Nomura); masaki.aono.ss@tut.jp (M. Aono)
https://www.kde.cs.tut.ac.jp/~aono/ (M. Aono)
0009-0005-1269-039X (S. Fukuyama); 0000-0002-8345-7094 (T. Asakawa); 0009-0000-3448-7986 (K. Shimizu); 0000-0003-2838-7844 (K. Nomura); 0000-0003-1383-1076 (M. Aono)

Table 1
CNN approach

approach	model	super-resolution
Method1	MobileNetV2	without
Method2	DenseNet-121	without
Method3	ResNet-50	without
Method4	MobileNetV2	with

revealed and the 100 images each annotated as used/not used for training the image generation of that model were not used this time.

3. Methods

3.1. Preprocessing

The images in the dataset are grayscale. Therefore, we attempted to colorize them using (tensorflow.image.grayscale_to_rgb). To augment the data, horizontal inversion and ImageGenerator were applied as follows.

- zoom_range : 0.97-1.03
  - Scaling is performed randomly within a scaling factor range of 0.97 1.03.
- rotation_range : 10
  - Randomly rotate the image in the range of -10 10 degrees.
- height_shift_range : 0.05
  - Vertical translation in the range of x 0.05 pixels of the original image height.
- width_shift_range : 0.05
  - Left-right translation within a range of x 0.05 pixels of the original image's width.

The default input channel is (224,224), so it was resized from (256,256) to (224,224). The horizontal inversion is performed only on the real image; 24 images are added in ImageGenerator for the image before and after horizontal inversion, respectively. The number of images can be reduced from 1 real image to 1 original image + 1 flipped image + 48 images converted by ImageGenerator = 50 images, which means that the data can be padded up to 10,000 images in total (1:1) when 200 real images are used. We thought that super-resolution technique would allow us to obtain more features that are not present in the real images but are prominent only in the generated images, so we performed super-resolution technique on the real images using ESRGAN [7] and then performed the same operation.

3.2. Architecture

We attempted CNN approach with three different CNN that employed ResNet50, MobileNetV2 and DenseNet-121. These CNNs used models pre-trained in imagenet(Method1-4). The input of the structure is set to (224,224,3), and the preprocess_input function is used before the CNN to perform preprocessing according to the weight data of each model. After the CNN, Dropout and Dense are installed. For Experimental details, Epoch is 50, Adam optimizer is used with learning rate 10-4. Additionally, Binary Cross Entropy is used as loss function. Figure 2 shows the architecture of our CNN approach. Method 4 uses images enhanced by ESRGAN. The shape of the image changed from (224,224) to (896,896) when the image was enlarged by ESRGAN. Therefore, the input of the architecture is adjusted to the shape. In addition, the pre-training of the model is "False".

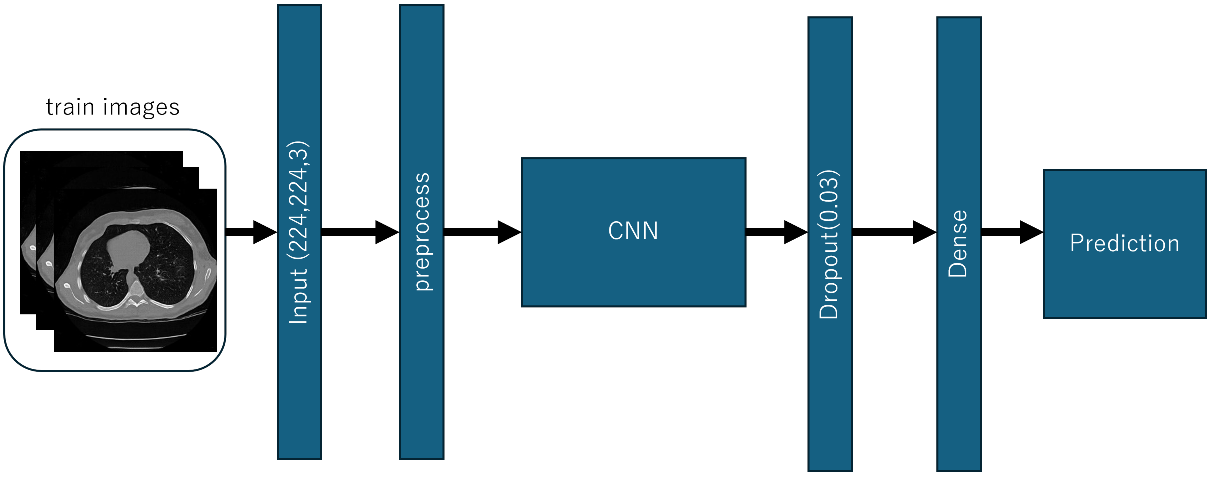

Figure 1: architecture

Table 2
confusion matrix

	prediction_Negative	prediction_Positive
Actual_Negative	TN(True Negative)	FP(False Positive)
Actual_Positive	FN(False Negative)	TP(True Positive)

3.3. Evaluation

Accuracy, Precision, Recall, and F1 values were used as evaluation indices. They can be calculated from the confusion_matrix of sklearn. The calculation formula is as follows(1)-(4).

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 = \frac{2 \cdot Recall \cdot Presision}{Recall \cdot Presision} \tag{4}$$

4. Results

The results obtained from the Method 1-4 approach were submitted and returned scores are shown in Table 3-5. Methods 2-4 failed to submit; Method 1 received a accuracy_score of 0.48425,precision_score of 0.480433, recall_score of 0.44825,f1_score of 0.4542. Failed of score was due to the fact that there were many of the same predicted results and that the method of submission was incorrect. The Method 1 values of Accuracy, Precision, Recall, and F1 were 0.484, 0.476, 0.317, 0.380 for Dataset1, 0.481, 0.484, 0.579, 0.527 for Dataset2.

5. Conclusion

This paper describes an approach to the training data fingerprint identification task in the ImageCLEF2024 GAN. We attempted to use multiple pre-training models and super-resolution techniques in our approach. However, this time, we were unable to produce correct values and could not compare the results for each model or image data with and without super-imaging. Therefore, it was not possible to

Table 3
Submission Result(Dataset1&Dataset2)

approach	accuracy	precision	recall	f1_score
Method1	0.484	0.480	0.448	0.454
Method2	Failed	Failed	Failed	Failed
Method3	Failed	Failed	Failed	Failed
Method4	Failed	Failed	Failed	Failed

Table 4
Submission Result(Dataset1)

approach	accuracy	precision	recall	f1_score
Method1	0.484	0.476	0.317	0.380

Table 5
Submission Result(Dataset2)

approach	accuracy	precision	recall	f1_score
Method1	0.481	0.484	0.579	0.527

identify features that are not present in the real images but are prominent only in the generated images. As a result of the competition, the CNN approach with EfficientNetV2 achieved an accuracy of 0.484.

Acknowledgments

A part of this research was carried out with the support of the Grant for Toyohashi Heart Center Smart Hospital Joint Research Course and the Grant-in-Aid for Scientific Research (C) (issue numbers 22K12149 and 22K12040).

References

[1] A. Andrei, A. Radzhabov, D. Karpenka, Y. Prokopchuk, V. Kovalev, B. Ionescu, H. Müller, Overview of 2024 ImageCLEFmedical GANs Task – Investigating Generative Models' Impact on Biomedical Synthetic Images, in: CLEF2024 Working Notes, CEUR Workshop Proceedings, CEUR-WS.org, Grenoble, France, 2024.

[2] B. Ionescu, H. Müller, A. Drăgulinescu, J. Rückert, A. Ben Abacha, A. García Seco de Herrera, L. Bloch, R. Brüngel, A. Idrissi-Yaghir, H. Schäfer, C. S. Schmidt, T. M. G. Pakull, H. Damm, B. Bracke, C. M. Friedrich, A. Andrei, Y. Prokopchuk, D. Karpenka, A. Radzhabov, V. Kovalev, C. Macaire, D. Schwab, B. Lecouteux, E. Esperança-Rodier, W. Yim, Y. Fu, Z. Sun, M. Yetisgen, F. Xia, S. A. Hicks, M. A. Riegler, V. Thambawita, A. Storås, P. Halvorsen, M. Heinrich, J. Kiesel, M. Potthast, B. Stein, Overview of ImageCLEF 2024: Multimedia retrieval in medical applications, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction, Proceedings of the 15th International Conference of the CLEF Association (CLEF 2024), Springer Lecture Notes in Computer Science LNCS, Grenoble, France, 2024.

[3] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen, Mobilenetv2: Inverted residuals and linear bottlenecks, 2019. arXiv:1801.04381.

[4] G. Huang, Z. Liu, L. van der Maaten, K. Q. Weinberger, Densely Connected Convolutional Networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 4700–4708.

[5] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.
[6] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, ImageNet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009), 2009, pp. 248–255. doi:10.1109/CVPR.2009.5206848.
[7] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, C. C. Loy, Y. Qiao, X. Tang, Esrgan: Enhanced super-resolution generative adversarial networks, 2018. arXiv:1809.00219.