

Identifying Training Data "Fingerprints" Using Border Enhancing Image Processing Methods and Their Ensemble

Notebook for the inouekokiteam Lab at CLEF 2024

Koki Inoue^{1,*}, Tetsuya Asakawa¹, Kazuki Shimizu², Kei Nomura² and Masaki Aono¹

¹Toyohashi University of Technology, 1-1 Hibarigaoka, Tempaku-cho, Toyohashi, Aichi, 441-8580, Japan

²Toyohashi Heart Center, 21-1Gobutori, Ohyamacho, Toyohashi, Aichi, 441-8071, Japan

Abstract

This paper describes our approach to the Identify training data "fingerprints" task of ImageCLEFmedical GANs 2024. In Task 1, the goal 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 proposed method uses image processing as a preprocessing step, and a pre-trained model, Resnet-152, is used for training. We also integrated the predictions of each model. As a result, the model with histogram equalization was able to outperform the baseline model trained without preprocessing by 66.6%. The model with prediction integration achieved 63.1%.

Keywords

Image Processing, Integrated the Predictions, Histogram Equalization

1. Introduction

ImageCLEF has been held as part of CLEF since 2003, and ImageCLEF2024 [1] approaches different areas, including ImageCLEFmedical GANs 2024 [2]. In Task 1 (Task to identify training data "fingerprints"), the goal is to detect "fingerprints" within the synthetic biomedical image data to determine which real images were used in training to produce the generated images. We are participating as a member of the inouekokiteam and are challenging this task. This paper describes the approach used to determine which images were used in training to create the generative model.

2. ImageCLEF 2024 Dataset

This section describes the dataset for the Identify training data "fingerprints" task of ImageCLEFmedical GANs 2024 [2]. This task uses two generative models. The dataset contains images used to train each model, images not used for training, and images generated by the models.

2.1. Development Dataset

The first generative model consists of 200 images annotated as used/not used for training image generation and 10k generative images generated by model 1. The second generative model consists of 6k images annotated as used/not used for training image generation, and 10k generative images generated by model 2.

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

*Corresponding author.

✉ inoue.koki.we@tut.jp (K. Inoue); 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)

🆔 0009-0003-9600-1939 (K. Inoue); 0000-0002-8345-7094 (T. Asakawa); 0009-0000-3448-7986 (K. Shimizu);

0000-0003-2838-7844 (K. Nomura); 0000-0003-1383-1076 (M. Aono)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2.2. Test Dataset

The test dataset contains two CSV files and two folders, and does not specify which set of images was used to train the generative model. The ratio of generated to real images is not identical. The first folder contains 7200 generated images and 4000 real images. The second folder contains 5000 generated images and 4000 real images.

3. Proposed Method

In this section, we describe our approach to the task of identifying the training data "fingerprints" of ImageCLEFmed GANs 2024 [2]. We have observed that the color boundaries of the generated images are often unclear. Therefore, we propose a method that captures the boundary sharpness using a set of OpenCV [3] image processing functions as preprocessing for both training and prediction. We also propose a method to integrate the prediction results of each training model into a single result. The image processing methods used are shown below.

- Binarization
- Histogram Equalization
- Laplacian Process
- Contrast Adjustment

We also propose a method to integrate the predictions of each training model into a single prediction. A total of five models are used: one model trained without image processing and four models trained with the image processing described above. The integration procedure is described below.

- Take a majority vote of the five models' forecasts and make an integrated forecast.
- If the predictions of all five models are not in agreement, a negative result is assumed.

4. Preprocessing by Image Processing

This section describes the image processing preprocessing performed on the development and test datasets.

4.1. Binarization

Binarization was performed in preprocessing using OpenCV [3]. The image was loaded as grayscale, and Otsu binarization [4] was performed. It uses the threshold that maximizes the separation between classes.

4.2. Histogram Equalization

We describe the preprocessing histogram equalization performed using OpenCV [3]. The images were loaded as grayscale and subjected to histogram equalization. This is a process that transforms the density so that the histogram of pixel values is uniform throughout.

4.3. Laplacian Process

Laplacian processing was performed using OpenCV [3]. The images were loaded as grayscale and processed with a Laplacian filter. It detects edges where the difference in pixel values changes significantly.

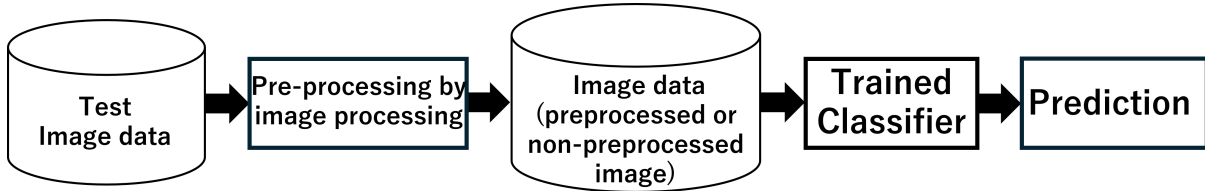


Figure 1: Flow of model Predictions

4.4. Contrast Adjustment

Contrast adjustment was performed in preprocessing using OpenCV [3]. The images were loaded as grayscale images and the contrast was adjusted. It was adjusted with $\alpha=1.5$ and $\beta=0$. v' is the output pixel value and v is the input pixel value.

$$v' = \alpha \times v + \beta \quad (1)$$

5. Train

In this section, we describe the training of the model. A pre-trained model from Resnet-152[5] was used for training. As training data, we used 3100 images each from generated_1 and generated_2 in the development dataset, for a total of 6200 images as generated, and all images from not_used_1, used_1, not_used_2, and used_2 as real. A total of 6200 images were considered REAL. In addition to preprocessing by image processing, random horizontal flipping was applied to the training images.

6. Prediction

In this section, we describe the prediction using the model described in the previous section and the integration of the prediction results. Test dataset preprocesses the models for prediction by image processing according to the model used. A total of five models are used for the prediction, one trained without image processing and four trained with different image processing methods.

6.1. Model Predictions

The prediction for each model is described in the following section. The detailed flow is shown in Figure 1. For the prediction of a trained model without image processing, no image processing is applied to test dataset. For the trained model with image processing, the same image processing was applied to test dataset to make predictions.

6.2. Integration of Prediction

We describe the integration of the predictions, using two methods: one with no image processing on test dataset, and the other with four different image processing methods. For the integration of the predictions, we used majority voting and perfect agreement. The integration flow is shown in Figure 2. For perfect agreement, the results were accepted only when all the results predicted by the five models were in agreement, and rejected when they were not.

7. Submission Results

In this section we describe the results of our team's submissions. The submissions included predictions for each of the five models (Run ID: 891-896) and the integration of the predictions (Run ID: 301, 890). The prediction for the model without added preprocessing (Run ID: 896) was 66.3%. The highest score

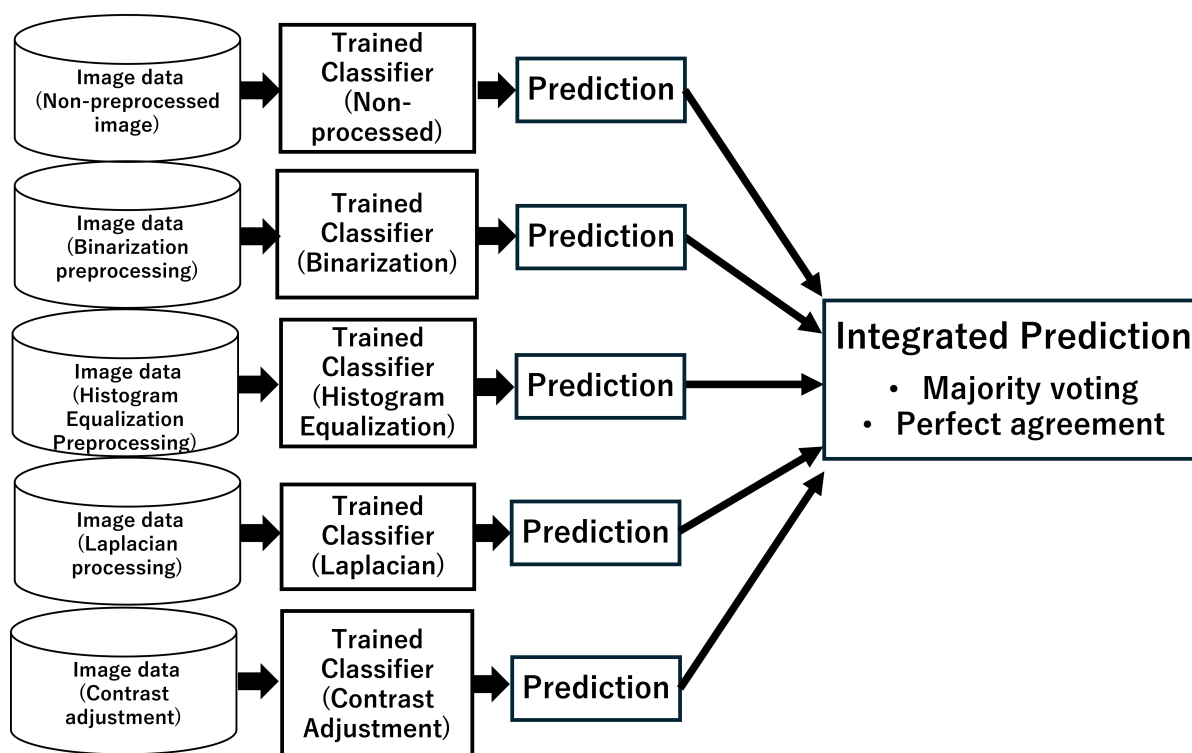


Figure 2: Integration of the five models' predictions

Table 1
Submission Results

Run ID	method name	Score	Acc	M1			M2			
				Prec	Recall	F1	Acc	Prec	Recall	F1
896	Non-Preprocessed	0.663	0.495	0.497	0.987	0.661	0.5	0.5	0.996	0.66
895	Binarization	0.638	0.484	0.49	0.838	0.619	0.503	0.501	0.951	0.656
894	Contrast Adjustment	0.660	0.491	0.495	0.973	0.656	0.499	0.499	0.993	0.664
892	Histogram Equalization	0.666	0.499	0.499	0.998	0.665	0.501	0.5	0.999	0.667
891	Laplacian Process	0.663	0.484	0.49	0.838	0.619	0.503	0.501	0.951	0.656
301	Majority Voting	Non								
890	Perfect Agreement	0.631	0.473	0.484	0.805	0.604	0.508	0.504	0.945	0.657

for the prediction using the model with histogram equalization (Run ID: 892) was 66.6%. No score was returned for majority voting (Run ID: 301), one of the proposed methods. The reason for not returning a score is believed to be that it produced the same prediction result for all test data. For perfect agreement (Run ID: 890) the score was 63.1%.

8. Discussion

In this section, we describe the submitted results. The model with histogram equalization and laplacian processing outperformed the baseline model with no preprocessing (Run ID: 896). Other models with additional preprocessing underperformed the baseline. This suggests that histogram equalization is an effective image processing method for detecting "fingerprints" within the synthetic biomedical image data to determine which real images were used in training to produce the generated images. We were not able to exceed the baseline for perfect agreement in predictive integration. One possible reason for this is that histogram equalization was effective, but other image processing methods were not. It is also possible that the acceptance method of rejecting all predictions if they did not match resulted

in the rejection of accurate predictions. No results were returned for majority voting for prediction integration. A possible reason for this could be that the prediction was not accepted because it was used for all images.

9. Conclusion

This paper describes an approach to the identify training data "fingerprints" task of ImageCLEFmedical GANs 2024[2]. We applied image processing as a preprocessing step, and attempted training and prediction. We also made predictions for each model, and attempted to integrate the predictions using majority voting and perfect agreement methods. The results showed that only the models with histogram equalization and laplacian processing were able to exceed the 66.3% of the models without image processing that were set as the baseline. Both predictions integration failed to exceed the baseline.

This paper describes an approach to the task of identifying training data "fingerprints" of ImageCLEFmedical GANs 2024 [2]. We applied image processing as a preprocessing step and attempted training and prediction. We also made predictions for each model and attempted to integrate the predictions using majority voting and perfect agreement methods. The results showed that only the models with histogram equalization and laplacian processing were able to exceed the 66.3% of the models without image processing, which was set as the baseline. Both prediction integrations failed to outperform the baseline.

10. 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] B. Ionescu, H. Müller, A. Drăgulescu, J. Rückert, A. Ben Abacha, A. Garcia Seco de Herrera, L. Bloch, R. Brüngel, A. Idrissi-Yaghir, H. Schäfer, C. S. Schmidt, T. M. 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.
- [2] 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.*
- [3] G. Bradski, *The OpenCV Library*, Dr. Dobb’s Journal of Software Tools (2000).
- [4] N. Otsu, A threshold selection method from gray-level histograms, *IEEE Transactions on Systems, Man, and Cybernetics* 9 (1979) 62–66. doi:10.1109/TSMC.1979.4310076.
- [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.