Detecting AI-generated images with texture-based methods (*Computer Vision*)

Balashova Ekaterina
21ekaterina.b@vinuni.edu.vn
CS, VinUniversity
Vietnam

Nguyen Canh Huy
21huy.nc@vinuni.edu.vn
CS, VinUniversity
Vietnam

Abstract

This project aims to solve the problem of distinguishing real images from AI-generated ones. To achieve this, we combine PatchCraft, a texture-based detection method, with Error Level Analysis (ELA) to enhance forensic signals. Using the Tiny GenImage dataset, we trained a model on both PatchCraft-processed and PatchCraft+ELA-enhanced images. Our results show that the combined approach outperforms the PatchCraft baseline, suggesting that ELA may contribute useful information. However, further investigation is required to determine the practical significance of this improvement.

CCS Concepts

• **Do Not Use This Code → Generate the Correct Terms for Your Paper**; *Generate the Correct Terms for Your Paper*; Generate the Correct Terms for Your Paper; Generate the Correct Terms for Your Paper.

Keywords

AI-generated, Error Level Analysis, PatchCraft, Images, Convolutional Neural Networks, Generative Adversarial Networks

ACM Reference Format:
Balashova Ekaterina and Nguyen Canh Huy. 2024. Detecting AI-generated images with texture-based methods (*Computer Vision*). In *Proceedings of COMP4040/5110 Data Mining and Big Data Analytics (COMP4040/5110)*. VinUniversity, Vietnam, 3 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

The rapid progress of generative artificial intelligence (AI) has reached an unprecedented level, capable of generating extremely realistic images that are close to authentic photo-realism. This technological evolution poses significant challenges, particularly regarding identity integrity and the trustworthiness of information. For example, deepfake technology has been used in identity theft and disinformation operations affecting multiple industries from education to media and law enforcement. A recent study revealed that participants could only achieve an average accuracy of 61.3% in detecting state-of-the-art AI-generated images, with a misclassification rate of 38.7% [1]. Such figures emphasize the necessity of efficient tools to verify images and distinguish between authentic original material and machine-created content.

Existing approaches to detecting AI-generated images include feature extraction methods [2], machine learning techniques [3], GAN fingerprinting [4, 5, 6], and artifact analysis [7]. One common problem for AI-generated image detectors is the inability to generalize beyond the generators they were trained on, limiting their effectiveness as new models emerge. Moreover, solutions developed are usually competing against the advancement of GAN technologies. As generative AI tools evolve, distinguishing between deepfakes and genuine images becomes increasingly complex. Existing detection methods have been unable to keep up with how quickly new models like DALL-E or Midjourney are appearing, with one study showing that greater than 70% of AI-generated images can slip past existing forensic tools [8]. This ongoing "arms race" between generative technology developers and detection researchers necessitates innovative solutions.

Our project aims to address these challenges by developing novel approaches for verifying image authenticity. By enhancing the ability to classify images accurately, we seek to contribute towards building a safer world where users can browse images and content without fear of its nature.

2 Related Work

Of all the diverse methods for AI-generated image detection, we intend to focus on feature mining methods that spot textural inconsistencies.

One such method is PatchCraft, proposed by Zhong et al., which detects AI images by analyzing local texture inconsistencies [9]. It divides images into patches, classifies them as rich or poor in texture, and exploits the unnatural texture distribution common in AI-generated content.

We intend to combine it with Error Level Analysis (ELA), which detects compression inconsistencies by recompressing an image and highlighting regions where error levels differ unnaturally. This method has a history of being used to detect various kinds of image forgery [10]. Martin-Rodriguez et al. [11] showed the effectiveness of ELA for AI-generated image detection, pointing out that AI-generated images showed up as fully forged when subjected to it, and confirmed that this method works on AI-generated images obtained in the PNG format, as well.

Both techniques leverage pixel-level inconsistencies, but in different ways: PatchCraft focuses on spatial texture variations, while ELA highlights compression anomalies. Our approach combines these methods to assess whether their integration improves AI-image detection.

3 Method

3.1 Problem statement

With the advancements in generative AI, AI-generated images are becoming near-indistinguishable from real ones with the naked eye. We intend to use texture-based methods to provide a reliable method of AI-generated image detection.

3.2 Approach

We combine PatchCraft and Error Level Analysis (ELA) to improve AI-generated image detection. PatchCraft identifies inconsistencies in local texture distribution, while ELA highlights compression-level artifacts that may indicate synthetic content. Since both methods capture textural inconsistencies at different scales, we theorize that their integration could enhance detection performance.
Following the original papers, we apply these methods together with a CNN-based classifier. Our experiments compare PatchCraft alone vs. PatchCraft+ELA, assessing whether the latter provides a meaningful advantage in detecting AI-generated images.

4 Experiments

4.1 Datasets

We use the Tiny GenImage dataset - a smaller version of the GenImage dataset by Zhu et al. [12] adapted by SanTai Yang on Kaggle [13]. It contains:

- Images from 8 different popular image-generating models
- 5,000 images per model, 40,000 in total
- Equal number of real and AI-generated images
- Equal number of JPEG and PNG images

The original dataset was created in 2023 and the models used in it are quite recent. Both the original dataset and the tiny version are freely available. We chose the tiny version due to resource limitations.
The dataset comes split into train/test 80/20. We additionally split the train set into train/val 90/10.

4.2 Metrics

To evaluate our model, we use Accuracy and Precision and Recall with respect to AI-generated images, as well as analyze the confusion matrix.

4.3 Baselines

We evaluate the proposed combination of PatchCraft and ELA against a PatchCraft-only baseline. The dataset is processed using both methods separately and used to train the same CNN architecture under identical conditions. We then compare the performance of both models to assess whether incorporating ELA provides a meaningful improvement.
While an ELA-only baseline would be useful for comparison, implementing it requires a different input format and model structure, making a direct comparison with PatchCraft-based methods complicated.

4.4 Implementation Details

Our process was as follows:

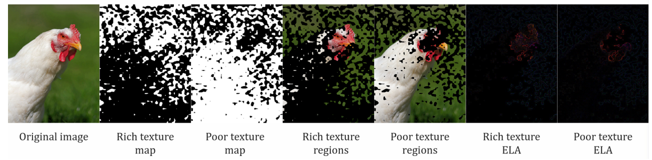

Figure 1: Preprocessing Results

(1) Data Preprocessing: train and test folders are separately passed through a preprocessing algorithm
  (a) Separate each image in a folder into rich- and poor-texture regions
  (b) Apply ELA separately
  (c) Save results into 2 separate folders: one for combined preprocessing, one for PatchCraft only
  (d) The functions for processing individual images are in the patchcraft-ela.py file, the script for processing the whole data folder is in the preprocessor.ibynb
(2) Model Training: two identical models are trained on the 2 sets of data created in the previous step
  (a) Use the pre-trained Resnet-18 in PyTorch
  (b) Adjust for 6-channel input (2 RGB images) and binary classification output
  (c) Fine-tune for 10 epochs with early stopping (we encountered an overfitting issue when testing the process on a smaller subset of data)
  (d) Save best models
  (e) The training scripts are in main.ipynb and baseline-patchcraft.ipynb respectively
(3) Model Evaluation: the resulting models are evaluated on identical evaluation scripts
  (a) Make predictions for (preprocessed) test data
  (b) Assess accuracy, precision and recall with respect to AI-generated image class, and confusion matrix
  (c) The evaluations are in the eval.ipynb and eval-baseline-patchcraft.ipynb

4.5 Results

4.5.1 Metrics. The model trained on the data preprocessed with the PatchCraft+ELA method showed the following results:

- Accuracy: 98.3
- Precision: 96.9
- Recall: 99.7

The model trained on the data preprocessed with only the PatchCraft method showed the following results:

- Accuracy: 97.6
- Precision: 96.6
- Recall: 98.7

4.5.2 Discussion. Our results suggest that combining PatchCraft and ELA for AI-generated image detection may be a promising approach. The PatchCraft+ELA model achieved slightly higher accuracy, precision, and recall compared to the PatchCraft-only baseline. While these improvements indicate that ELA may provide useful additional information, the differences are relatively small.

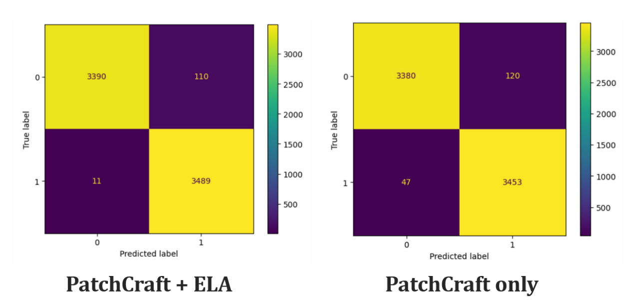

Figure 2: Confusion matrices

Further evaluation on larger and more diverse datasets is needed to determine whether this advantage holds practical meaning.

5 Conclusion

This project explores the combination of PatchCraft and Error Level Analysis (ELA) for AI-generated image detection. Our approach achieves slightly better performance than a PatchCraft-only baseline, suggesting that ELA might be able to supplement meaningful additional information for PatchCraft. While the results are encouraging, future work would be necessary to investigate the significance of this improvement across varied datasets and real-world AI-generated content.

Acknowledgments

We thank Professor Doan Dang Khoa, lecturer at College of Engineering and Computer Science, VinUniversity for his generous support.

Balashova Ekaterina

1. Synthesized research and literature review
2. Wrote sections 2d-2f of the preliminary proposal
3. Found the data
4. Made the slides for the proposal presentation
5. Wrote abstract and sections 2-4 of the final proposal
6. Developed, trained, and evaluated the model
7. Wrote the final report

Nguyen Canh Huy

1. Synthesized research and literature review
2. Wrote sections 1 and 2a-2c of the preliminary proposal
3. Formatted the preliminary proposal
4. Made the slides for the proposal presentation
5. Wrote abstract and section 1 of the final proposal
6. Formatted the final proposal
7. Coded PatchCraft and ELA and handled data preprocessing
8. Made the final slides

Impact Statement

As AI-generated images become increasingly realistic, improving detection methods alongside the evolving image-generating models is essential to prevent AI misuse, misinformation, and digital forgery. We hope that our project can contribute to building a future where real images remain trustworthy.

References

[1] S. S. H. A. R. A. M. Alshahrani, A. A. Alharthi, and M. M. Alzahrani. 2023. Seeing is not always believing: A quantitative study on human perception of AI-generated images. In *Proceedings of the International Conference on Artificial Intelligence and Data Science (ICAIDS)*. 2023, pp. 1-6, doi:10.1145/370262811.

[2] F. Zhang, Y. Zhang, and H. Wang. 2023. A survey of deep learning-based image forgery detection methods: Current trends and future directions. Sensors 23, 22 (2023), 9037, doi:10.3390/s23229037.

[3] J. Kwon and J.-H. Kim. 2023. An overview of AI-generated content detection: Current trends and challenges in the field of AI-generated media analysis. AI 5, 3 (2023), 76, doi:10.3390/ai5030076.

[4] Y. Chen, H.-Y. Lee, and S.-H. Chen. 2023. Detecting deepfake videos using a multi-modal approach based on convolutional neural networks and recurrent neural networks. In *2023 IEEE International Conference on Image Processing (ICIP)*, Taipei, Taiwan, 2023, pp. 102-106, doi:10.1109/ICIP46329.2023.10223798.

[5] G.-H. Lee and J.-H. Kim. 2022. Understanding the limitations of existing AI image detectors: An empirical study on deepfake detection accuracy and reliability in various contexts. arXiv preprint arXiv:2205.12543 (2022), doi:10.48550/arXiv.2205.12543.

[6] L.-Y. Yang, C.-Y. Chen, and Y.-C. Lin. 2024. Analyzing the effectiveness of AI-generated image detectors: Insights from a comprehensive evaluation study on detection performance across various generative models and datasets.*Digital Investigation* (2024), doi:10.1016/j.diinv.2024.100119.

[7] A. K. Jain, R. K. Gupta, and S. K. Singh. 2024. A comprehensive review of image forgery detection techniques: Current trends and future directions. *Digital Investigation 42* (2024), 100159, doi:10.1016/j.diinv.2024.10015

[8] Originality.ai Team,. 2024 Do AI image detectors work? An accuracy study on current detection tools and their performance against generative models in real-world scenarios.Originality.ai Blog. Retrieved from https://originality.ai/blog/do-ai-image-detectors-work-accuracy-study.

[9] N. Zhong, Y. Xu, S. Li, Z. Qian, X. Zhang. 2024. PatchCraft: Exploring Texture Patch for Efficient AI-generated Image Detection. arXiv preprint arXiv:2311.12397, 1-18, 2024

[10] N. B. A. Warif, M. Y. I. Idris, A. W. A. Wahab and R. Salleh. 2015. An evaluation of Error Level Analysis in image forensics. In *2015 5th IEEE International Conference on System Engineering and Technology (ICSET)*. Shah Alam, Malaysia, 2015, pp. 23-28, doi: 10.1109/ICSEngT.2015.7412439.

[11] F. Martin-Rodriguez, R. Garcia-Mojon, M. Fernandez-Barciela. 2023. Detection of AI-Created Images Using Pixel-Wise Feature Extraction and Convolutional Neural Networks. *Sensors (Basel, Switzerland)*, 23(22), 9037. https://doi.org/10.3390/s23229037

[12] M. Zhu, H. Chen, Q. Yan, X. Huang, G. Lin, W. Li, Z. Tu, H. Hu, J. Hu, Y. Wang. 2023. GenImage: A Million-Scale Benchmark for Detecting AI-Generated Image. arXiv preprint arXiv:2306.08571.

[13] SanTai Yang. 2024. tiny genimage. Kaggle. Available at: https://www.kaggle.com genimage

Received 20 January 2025