RESEARCH PROJECT

Kanokporn Pakdeenual

675020069-9

TITLE

Published In

Computer Vision – ECCV 2024: 18th European Conference, Milan, Italy, September 29 – October 4, 2024, Proceedings, Part XVIII

Sep 2024 | 578 pages

ISBN: 978-3-031-72648-4

DOI: 10.1007/978-3-031-72649-1

Editors: Aleš Leonardis, + 5

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025.

Publisher

Springer-Verlag
Berlin, Heidelberg

Publication History

Published: 30 September 2024

Zero-Shot Detection of AI-Generated Images

Davide Cozzolino[1], Giovanni Poggi[1], Matthias Nießner[2], and Luisa Verdoliva[1,2]

[1] University Federico II of Naples, 80125 Naples, Italy
[2] Technical University of Munich, 85748 Garching, Germany
{davide.cozzolino,poggi,verdoliv}@unina.it, niessner@tum.de

Conference Ranks

บทความ “Zero-Shot Detection of AI-Generated Images” เป็น**งานนำเสนอในการประชุม ECCV 2024 (European Conference on Computer Vision)**
ซึ่งถูกตีพิมพ์ใน **Proceedings (Lecture Notes in Computer Science, LNCS Vol. 15076)**
จาก **Springer-Verlag**

Conference Data

Search: ECCV

Name	Abbrv.	Rank	Source
European Conference on Computer Vision	ECCV	A	ERA
European Conference on Computer Vision	ECCV	A1	Qualis
Name	**Abbrv.**	**Rank**	**Source**

Zero-Shot Detection of AI-Generated Images

Davide Cozzolino[1], Giovanni Poggi[1], Matthias Nießner[2], and Luisa Verdoliva[1,2]

[1] University Federico II of Naples, 80125 Naples, Italy
[2] Technical University of Munich, 85748 Garching, Germany
{davide.cozzolino,poggi,verdoliv}@unina.it, niessner@tum.de

Abstract. Detecting AI-generated images has become an extraordinarily difficult challenge as new generative architectures emerge on a daily basis with more and more capabilities and unprecedented realism. New versions of many commercial tools, such as DALL·E, Midjourney, and Stable Diffusion, have been released recently, and it is impractical to continually update and retrain supervised forensic detectors to handle such a large variety of models. To address this challenge, we propose a zero-shot entropy-based detector (ZED) that neither needs AI-generated training data nor relies on knowledge of generative architectures to artificially synthesize their artifacts. Inspired by recent works on machine-generated text detection, our idea is to measure how *surprising* the image under analysis is compared to a model of real images. To this end, we rely on a lossless image encoder that estimates the probability distribution of each pixel given its context. To ensure computational efficiency, the encoder has a multi-resolution architecture and contexts comprise mostly pixels of the lower-resolution version of the image. Since only real images are needed to learn the model, the detector is independent of generator architectures and synthetic training data. Using a single discriminative feature, the proposed detector achieves state-of-the-art performance. On a wide variety of generative models it achieves an average improvement of more than 3% over the SoTA in terms of accuracy. Code is available at https://grip-unina.github.io/ZED/.

Abstract

Abstract

1. Problem – ปัญหา

- Nowadays, there are new AI image generation tools (Generative Models) such as DALL·E, Midjourney, and Stable Diffusion that are being rapidly developed.
- Traditional forensic detectors are no longer sufficient to detect images generated by AI, as they were trained on older patterns.
- However, new models keep evolving → causing delays and poor detection performance.
- Therefore, it is necessary to find methods that do not require image detection models or fake images for training.

Abstract

2. Proposed Solution – วิธีการที่นำเสนอ

Proposed Method: ZED (Zero-shot Entropy-based Detector)

- Propose a forensic detection tool called ZED, which operates in a "Zero-Shot" manner.
- **Key advantage:** Does not require any AI-generated image samples, not even a single one.
- Enables detection of images from newly developed AI models immediately, without needing prior examples.

Abstract

3. Method - ขั้นตอนการทำงาน

- **Main Concept:** Detect how “surprising” an image is when compared to real images.
- **Steps:**

a. Build a model that learns only the statistical characteristics of real images.

b. When testing, pass the image through a lossless image encoder to measure how unusual the pixel arrangement is.

c. If the tested image is significantly “unexpected” or “surprising,” and its properties deviate from the learned patterns of real images (i.e., it's very surprising), → it is likely to be an AI-generated image.

Abstract

4. Key result - ผลลัพธ์ที่น่าสนใจ

- Compatible with various AI image generation models
- Independent of fake image training data and model architecture details
- High performance: exceeds previous state-of-the-art methods by over 3% in accuracy
- Open-source code available: https://grip-unina.github.io/ZED/

THANK YOU