THE MANY FACES OF ART

What techniques can we use to protect authentic artists from AI-generated art?

AUTHOR: SABINA-MALINA GRADINARIU

SUPERVISOR: DR. ANNA LUKINA, TU DELFT

1. BACKGROUND

- Generative models progressed in **mimicking creativity**;
- There is a knowledge gap in distinguishing between **AI-generated images** from **art**;
- Generative Adversarial Networks: a deep learning architecture that retains high-dimensional distribution over the dataset: **Midjourey**, **DALL-E 3**;
- **Diffusion Models**, the idea around them is destroying the training data by gradually adding Gaussian noise and recovering the data by reversing the processes: **Stable Diffusion XL**, **Adobe Firefly**;

2. RESEARCH QUESTION

What techniques can we use to protect authentic artists from AI-generated art?

Investigate:

- the effects of generated art on artists;
- current methods and tools for detection;
- art mimicry and tools that combat it;

3. METHODOLOGY

To address this hypothesis I employed several methods: data collection, sampling, investigating detection tools, using evaluation metrics, and conducting a literature review.

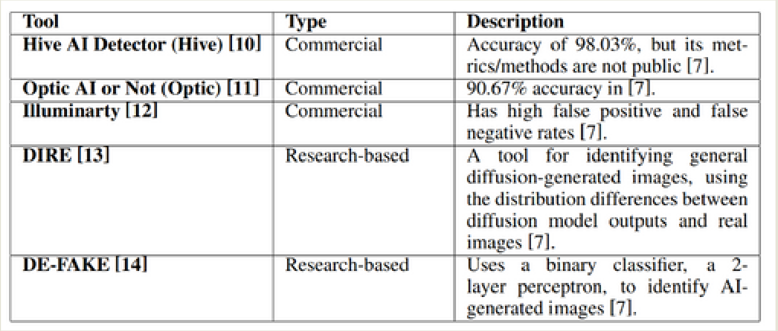
- Datasets used: **WIKIART** & **AI-ArtBench;**
- **Test** on **DE-FAKE** & commercial tools;
- Methodological choice to include many styles;
- The review emphasized the significance of employing detection tools and protective techniques to preserve integrity and **authenticity;**
- **End product:** create a pipeline that ensures that artists are shielded from the effects of AI in art;

Contact: s.gradinariu@tudelft.nl

4. CURRENT TOOLS

The state-of-the-art **detectors** are of two categories: **commercial** black-box detectors and **research-based** detectors. Better results are obtained in detection and minimising the false positives and negatives by the commercial detectors.

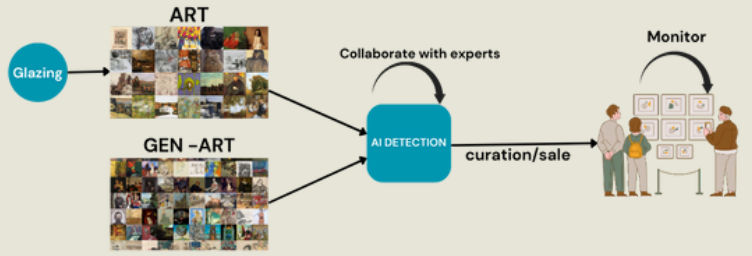
Artistic Style Protection:
Glaze is a system designed to disrupt style mimicry, by subtly altering an art piece's appearance, by adding **"style cloaks"** to an artist's work.



Tool	Type	Description
Hive AI Detector (Hive) [10]	Commercial	Accuracy of 98.03%, but its metrics/methods are not public [7].
Optic AI or Not (Optic) [11]	Commercial	90.67% accuracy in [7].
Illuminarty [12]	Commercial	Has high false positive and false negative rates [7].
DIRE [13]	Research-based	A tool for identifying general diffusion-generated images, using the distribution differences between diffusion model outputs and real images [7].
DE-FAKE [14]	Research-based	Uses a binary classifier, a 2-layer perceptron, to identify AI-generated images [7].

5. PIPELINE

Several crucial steps are included in the proposed pipeline to guarantee the authenticity and preservation of artwork.

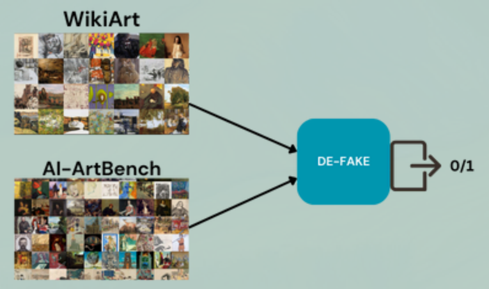


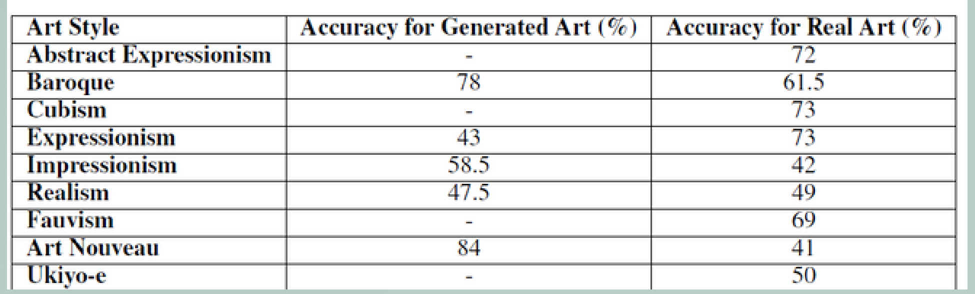
RELATED LITERATURE

1. A. Y. J. Ha, J. Passananti, R. Bhaskar, S. Shan, R. Southen, H. Zheng, and B. Y. Zhao, "Organic or diffused: Can we distinguish human art from AI-generated Images?" 2024.
2. Sha, Z. Li, N. Yu, and Y. Zhang, "De-fake: Detection and attribution of fake images generated by text-to-image generation models," 2023

6. EXPERIMENTAL FINDINGS

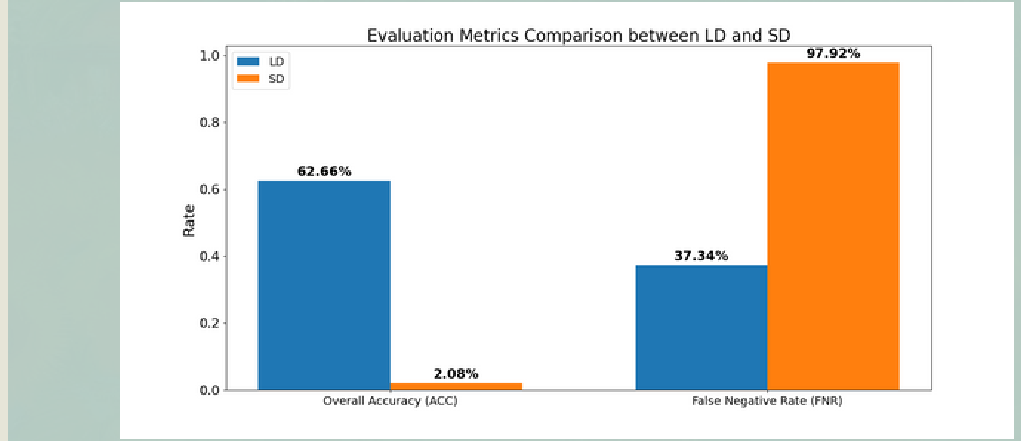
- **Hive**, the best at detecting AI Art is unfortunately unavailable, so **DE-FAKE** was used.
- **WikiArt**: over 80,000 images spanning 27 different artistic styles.
- **AI-ArtBench**: images generated by two generative models: Latent Diffusion (LD) and Stable Diffusion (SD), across ten styles.
- To evaluate the performance of the DE-FAKE algorithm **accuracy**, **false positives**, and **false negatives** were investigated.





Art Style	Accuracy for Generated Art (%)	Accuracy for Real Art (%)
Abstract Expressionism	-	72
Baroque	78	61.5
Cubism	-	73
Expressionism	43	73
Impressionism	58.5	42
Realism	47.5	49
Fauvism	-	69
Art Nouveau	84	41
Ukiyo-e	-	50

The results disclosed its proficiency in identifying images within **modern** styles scoring an accuracy between **70%** and **84%**. Its performance dropped when applied to more **realistic styles,** such as Baroque and Art Nouveau, with an accuracy rate of around a mere **40%**. Intuitively, the algorithm performed better with classes exhibiting distinctive pattern characteristics.



Results are very different depending on the generation type. The detection tool is unable to detect **SD** generated images.

7. ETHICAL IMPLICATIONS & DISCUSSION

- A significant risk to creators is the potential for AI systems to **mimic** distinctive artistic styles without permission or payment;
- We tried to mitigate the detrimental effects of AI on the art world and preserve the inherent value of true human creativity;
- This study is reproducible, transparent and everything is publicly available;

Future work

- More work should be done to improve detection algorithms until they can handle realistic art forms more effectively;
- Developing relationships with companies such as Hive, which have produced efficient AI detection tools, is one step forward;
- A solid and efficient framework is incorporating artists' insights;

8. CONCLUSION

The research findings and conclusions are meant to be the foundation for **further investigations**, guaranteeing that the line between AI-generated art and art will always be distinct and verifiable. The **pipeline** provides a tool for safeguarding the authenticity and **inherent worth** of human art in the age of **artificial intelligence**. Furthermore, exploring the integration of complementary detection methodologies and tools, drawing from **digital forensics** and **computational photography** should be investigated.