Quantifying Likeness: A Computer Vision Approach to Identifying Style and Copyright Infringement in AI-Generated Artwork

Anonymous Authors[1]

1. Introduction

The rapid development of generative AI models has raised concerns among creators about potential job disruption and copyright infringement. As these technologies advance, they challenge our understanding of legal concepts like "likeness." This study aims to quantitatively explore "likeness," contributing to the establishment of a definition that protects artists, fosters innovation, and provides clear guidelines for the use of images and ideas. We introduce Iwerks, an image classification model that estimates the probability of an image bearing substantial similarity to copyrighted material. Using computer vision, it identifies the extent to which an image replicates the likeness of copyrighted entities, as demonstrated with synthetic images generated by DALL·E resembling Steamboat Willie era Mickey Mouse (see Figure 2 in Appendix for original Steamboat Willie). Through its design, we will engage with legal concepts of likeness and substantial similarity, aiming to provide a quantitative interpretability framework for style while considering the threshold between "influence" and "copying."

2. Legal Context

Intellectual property law protects innovations across industries by preventing protected works from being used, reproduced, or distributed without the author's consent. Those IP protections most relevant to literary and artistic works are copyrights and trademarks. Both offer some protection against unauthorized use, but there is no clear legal standard regarding how much of a work can be used without infringement (as highlighted in the recent landmark case, Andersen et al. v. Stability AI Ltd. et al (2023)).

The American Supreme Court has ruled that transformative works are fair use (Cam, 1994; War, 2023), while Canadian jurisprudence does not have a transformative likeness doctrine but instead authorizes several fair dealing exceptions (CCH, 2004). The difference in the scope of fair dealing and fair use has serious implications for the use of protected works in generative AI algorithms, particularly given that commercial use is not necessarily barred under either doctrine. As generative AI evolves, it becomes increasingly possible to generate new works based on the likeness of protected materials, raising the question: when does inspiration become infringement?

3. Model

We fine-tuned the fully connected (FC) layer of a ResNet-18 model, chosen for its superior performance and computational efficiency compared to AlexNet and VGG.

The ResNet architecture's deep residual learning framework helps alleviate the gradient degradation problem encountered in earlier models like AlexNet and VGG (He et al., 2015). Fine-tuning only the FC layer while keeping the pre-trained convolutional layers frozen was based on previous literature which suggests that pre-trained classification layers are necessary for better optimization during the fine-tuning process and increasing network depth (Shermin et al., 2019). This was confirmed by trial runs, which yielded higher accuracy and fewer false positives compared to fine-tuning all layers.

3.1. Data

The training set consists of 1026 images across three classes: 565 images of Mickey Mouse, 278 of Donald Duck, and 183 of Winnie the Pooh with random resizing, random cropping, horizontal flip, random rotation, colour jitter, random perspective, and random erasing applied during preprocessing. The Mickey Mouse images were drawn from stills of Steamboat Willie (1928) and Gallopin' Gaucho (1928), two works that entered the public domain this year. For the model classification tasks, we created a dataset of 50 images of cartoon mice with varying degrees of likeness to Steamboat Willie using DALL-E and generated prompts from the Claude API.

3.2. Model Setup

The ResNet-18 model consists of 18 layers, including convolutional, max-pooling, and a fully connected layer with 512 features. It takes a 224x224 input image and outputs predictions for 3 classes.

The model was trained using the cross-entropy loss function and the RMSprop optimizer with a learning rate of 0.0001 and a weight decay of 0.01. L2 regularization was applied to prevent overfitting. Hyperparameters were set to a batch size of 10 and a base learning rate of 0.001.

The model was trained for 100 epochs with a 70% training and 30% validation split using a GPU environment with CUDA, PyTorch, and TensorFlow.

4. Results

The results of the model performance are summarized in Table 1. The model demonstrated strong performance metrics, achieving a training loss of 0.1152 with an accuracy of 96.09%, and a validation loss of 0.0558 with an accuracy of 98.70%. The model reached its best validation accuracy at 99.21%.

Metric	Train	Validation
Loss	0.1152	0.0558
Accuracy	0.9609	0.9870
Best Validation Accuracy	0.9921	

Table 1: Performance metrics of the model on the training and validation datasets.

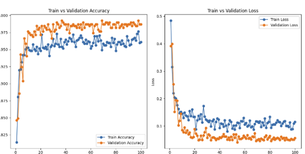


Figure 1: Model training accuracy plotted against validation accuracy

5. Likeness Prediction Scores

We used three visualization techniques to interpret our model's decision-making: feature mapping, saliency maps, and template matching (using Figure 2 as a reference image).

From our outputs, we selected Output 1 (see Figures 3a, 3b, and 3c in Appendix) and Output 2 (see Figures 4a, 4b, and 4c in Appendix) for this report as they vary greatly in their similarity to Steamboat Willie. The model identified Output 1 as having a 0.999 probability of belonging to the Mickey Mouse class, suggesting that this image is very likely to be reproducing the likeness of a copyrighted image. In Figure 3a and in Figure 3b, the importance is placed on the style of the ears, tail, and the shape of the cartoon's body. The pants are also a nearly exact match to Steamboat Willie's style of pants, which is highlighted by the saliency map.

The model identified Output 2 as having a 0.021 probability of belonging to the Mickey Mouse class. In Figure 3a, the only important feature that aligns with a Steamboat Willie reference image is the shape of the cartoon's ear. This inadvertently suggests a strong model performance — the mouse is surrounded by polygons very similar to the ear, and yet the model was able to distinguish between a polygon and a polygon that is part of an ear.

6. Discussion

The results demonstrate that our model was successful in separating style and content, a fundamental dichotomy in art criticism, when identifying the likeness of a cartoon figure in generative AI outputs. The model contextualized the polygons in Output 2 as belonging to an ear but did not classify the image as strongly similar to Steamboat Willie based on this content recognition alone. In Output 1, the style, remarkably similar to 1920s-era Mickey Mouse aesthetics, led to a very high probability score of reproducing likeness.

This study suggests that using AI to analyze the aesthetic influence and similarities of an image can effectively facilitate the discussion of determining when an image has crossed a threshold of likeness in the generation phase of the generative AI supply chain (Lee et al., 2023). Defining this threshold requires consultation with stakeholders, such as creators and legal domain experts. Quantitative methods force us to make our understandings of concepts explicit enough to be falsified and redefined. While this study suggests a method for making these concepts (likeness, style, shape) explicit, combining it with discourse in law and creators' interests can help develop clearer copyright guidelines in an age when authorship, creation, and inspiration are being renegotiated.

7. Future Work

To enhance our model, we propose exploring contrastive learning techniques, experimenting with distractor classes, class imbalance ratios, and data augmentation methods like AugMix (Hendrycks et al., 2020) to improve generalization, robustness, and performance. We will investigate adding layers beyond the FC layers and isolating certain layers for fine-tuning (Shermin et al., 2019). Furthermore, we aim to develop image classification models trained solely on human drawings.

We propose collaborating with artists whose works' likeness have allegedly been reproduced by generative AI models and training the model to classify between original work, generative work, and other works that artists identify as being influential on their own style. By comparing human, computational, art criticism and legal assessments of likeness, we aim to develop a framework for clearer guidelines on determining copyright infringement in AI-generated material, contributing to more robust and consistent standards for evaluating potential infringement.

A. Appendix

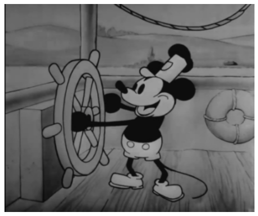
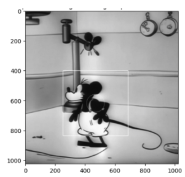
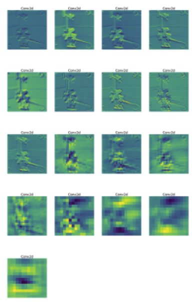


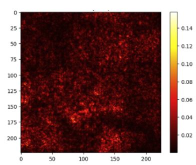
Figure 2: Steamboat Willie (1928)



(a) Template Matching applied to Output 1

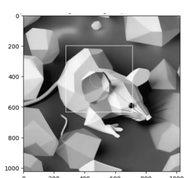


(b) Feature Map for Output 1

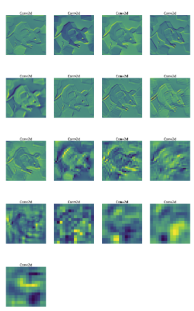


(c) Saliency Map for Output 1

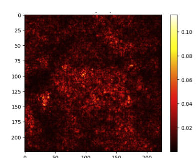
Figure 3: Analyses for Output 1



(a) Template Matching applied to Output 2



(b) Feature Map for Output 2



(c) Saliency Map for Output 2

Figure 4: Analyses for Output 2

References

Campbell v. acuff-rose music, inc. U.S. Supreme Court, 1994. 510 U.S. 569.

Cch canadian limited v. law society of upper canada. Supreme Court of Canada, 2004. [2004] 1 S.C.R. 339.

Andy warhol foundation for the visual arts, inc. v. goldsmith et al. U.S. Supreme Court, 2023. Important case on transformative use in copyright law.

Andersen, e. a. Andersen et al. v. stability ai ltd. et al. Bloomberg Law, 2023. URL `https://www.bloomberglaw.com/public/desktop/document/AndersenetalvStabilityAILtdetalDocketNo323cv00201` `1?doc_id=X7O8UTHBKAC83BA5ES7JLB5GFPL`. Case No. 3:23-cv-00201, N.D. Cal., filed Jan. 13, 2023.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. *arXiv*, abs/1512.03385v1, 2015. URL `https://arxiv.org/abs/1512.03385`.

Hendrycks, D., Mu, N., Cubuk, E. D., Zoph, B., Gilmer, J., and Lakshminarayanan, B. Augmix: A simple data processing method to improve robustness and uncertainty. In *International Conference on Learning Representations (ICLR)*, 2020.

Lee, K., Cooper, A. F., and Grimmelmann, J. Talkin' 'bout ai generation: Copyright and the generative-ai supply chain. *Journal of the Copyright Society*, July 2023. forthcoming.

Shermin, T., Teng, S. W., Murshed, M., Lu, G., Sohel, F., and Paul, M. Enhanced transfer learning with imagenet trained classification layer. *arXiv*, September 2019.