MCID: Multi-aspect Copyright Infringement Detection for Generated Images

Supplementary Material

The supplementary materials mainly include the following sections:

- In Sec. 6, we introduce additional cases involving copyright disputes of generated images.
- In Sec. 7, we present the main prompts used in the experiments.
- In Sec. 8, we provide more examples from LSCD and various infringement categories that we manually annotated.
- In Sec. 9, we present further examples of infringement detection using HIDM.
- In Sec. 10, we analyze the current limitations of our work and suggest future research directions.

6. Dispute Cases of Generated Image Copyright Infringement

In this section, we provide additional examples of copyright infringement disputes involving generated images. **1.** Getty Images has accused Stability AI of violating its intellectual property rights. The allegations pertain to Stability AI's purported use of Getty's images as data inputs for training and developing the Stable Diffusion model. Additionally, Getty Images asserts that the outputs generated by Stable Diffusion are synthetic images that significantly replicate its copyrighted works and/or display Getty's brand markings.

2. After the death of renowned South Korean artist Kim Jung Gi, a game developer released a tool that enabled users to create images similar to Kim's comics through text prompts. Although the developer claimed that the tool was meant to be a tribute, it quickly sparked intense criticism.

3. Japanese manga artist and politician Ken Akamatsu proposed that creators should have the right to exclude their work from datasets used to train AI programs. Alternatively, if they choose to include their work, they should receive appropriate compensation.

4. Greg Rutkowski is an artist renowned for his unique style, particularly in crafting fantasy scenes featuring dragons and epic battles, which have been utilized by fantasy games such as Dungeons and Dragons. His name has been associated with the generation of approximately 93,000 AI images on the platform Stable Diffusion. Rutkowski expressed his concerns, stating, "People are impersonating me, and I find this very troubling; it appears to be unethical."

These cases illustrate that the infringement of creators' rights by generated images has become a serious issue, which hinders the further development of generative models. There is an urgent need for methods to detect infringing images.

7. Detailed Prompts

7.1. Prompts for VLM in Infringement Detection

When only image pairs are input, the prompt guides the VLM to perform infringement detection:

We define infringement as the presence of a certain degree of similarity between two images in a specific evaluation dimension. Please adhere to this standard strictly. You're now a professional infringement detection expert, here the first image is a copyrighted artwork, while the second one is a generated image. Please determine whether the second image infringes on the first image. If there is any noticeable similarity between them, consider this as evidence of potential infringement. Additionally, do you think the second image similar to any famous IP protected character? If yes, also specify the name of the character. Do not use inferential conclusions. First, answer with a simple yes or no, then provide a brief analysis.

When image pairs and similarity scores are input, the prompt guides the VLM to perform infringement detection:

We define infringement as the presence of a certain degree of similarity between two images in a specific evaluation dimension. Please adhere to this standard strictly. You're now a professional infringement detection expert, here the first image is a copyrighted artwork, while the second one is a generated image. In addition, we provide you with similarity scores for the two images in three different evaluation dimensions: content, style, and structure. Each score ranges from 0 to 1, with higher scores indicating greater similarity between the two images in that evaluation dimension. Please determine whether the second image infringes on the first image by considering the image content and similarity scores. As long as the two images exhibit infringement in any one of the evaluation dimensions, it is believed there is infringement between them, since we believe infringement detection is a local attribute instead of a global judgment. Additionally, do you think the second image similar to any famous IP protected character? If yes, also specify the name of the character. First, answer whether there is an infringement in Yes or No. Then, provide a brief analysis explaining your reason about your judgment. The content similarity score is {content}, the style similarity score is {style}, the structural score is {structural}. Your inference process and conclusions should take into account both the scores and the images.

7.2. Prompts for Data Processing

Prompt for generating detailed caption in Fig. 3(a):

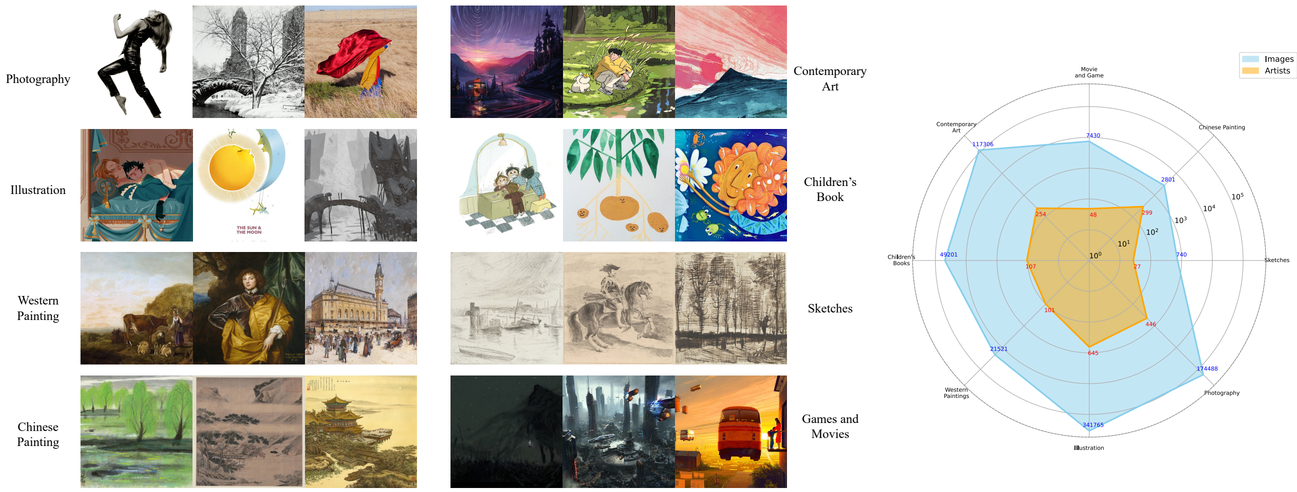


Figure 5. Left: Example images of different categories in LSCD. Right: Distribution of different data categories.

Please generate a concise and complete image description (within 77 tokens) based on the following requirements based on the input image. The description should include the image's content, style, and structural information in a single, coherent sentence: 1. Content: Describe the main objects, people, or scenes in the image. 2. Style: Describe the artistic style, color palette, and overall mood of the image. 3. Structure: Describe the fine-grained geometric structure, including shapes, lines, and spatial relationships. Example: A golden retriever running on a grassy field, with bright colors and an impressionistic style, dynamic lines showing movement, and a background of blue sky and green grass. Please generate the description.

Prompt for rewriting caption in Fig. 3(d):

You are now an image caption restructuring expert. Please analyze the input text and image, identify all the entities within it, and randomly replace some entities' categories and attributes with similar entities without altering the relationships between the entities. For example: replace man with woman, replace cat with dog. Directly output the replaced text. Input text: {*caption*}

8. LSCD and Benchmark

8.1. Visualization of LSCD Dataset

We present a detailed artistic style distribution covered by the LSCD, along with corresponding examples, in Fig. 5.

8.2. Benchmark

In our MCID task, we subdivide the task of generating image infringements into four distinct categories and proposed a manually annotated test set to serve as a benchmark. It is worth noting that when constructing non-infringing pairs for the test set, we do not randomly select two non-infringing images, as such examples would be too easy and unable to evaluate performance effectively. Instead, we select sample pairs that have a certain degree of similarity but ultimately do not constitute infringement. Specifically, pairs with expert scores greater than 4.5 are considered infringing pairs, while those with scores between 3 and 3.5 are considered non-infringing pairs, forming the final test set. Due to the great difficulty of infringement-related annotations, which require experts with artistic appreciation abilities, the currently annotated benchmark data is still relatively limited (59 infringing pairs and 50 non-infringing pairs). We will continue to expand the scale of this benchmark in the future.

We visualize the results of different types of infringement samples of our proposed benchmark in Fig. 6. Content infringement, style infringement, and structure infringement images are generated using a method similar to that shown in Fig. 3. For IP infringement images, we referred to the generation strategy of CopyCat [17], which involves creating inducive captions to prompt the model to generate IP infringement images.

9. Case Study

In this section, we present additional examples of infringement detection using HIDM.

In case (a), the input consists of the original image of Snow White and a generated infringing image. Although both images have the same character, the three similarity scores are all relatively low, making it difficult to directly determine infringement. However, by combining VLM's understanding of the well-known IP image, it ultimately identifies the second image as Snow White, constituting IP infringement.

For case (b), the two images have extremely high structural similarity, thus they are successfully judged as infring-

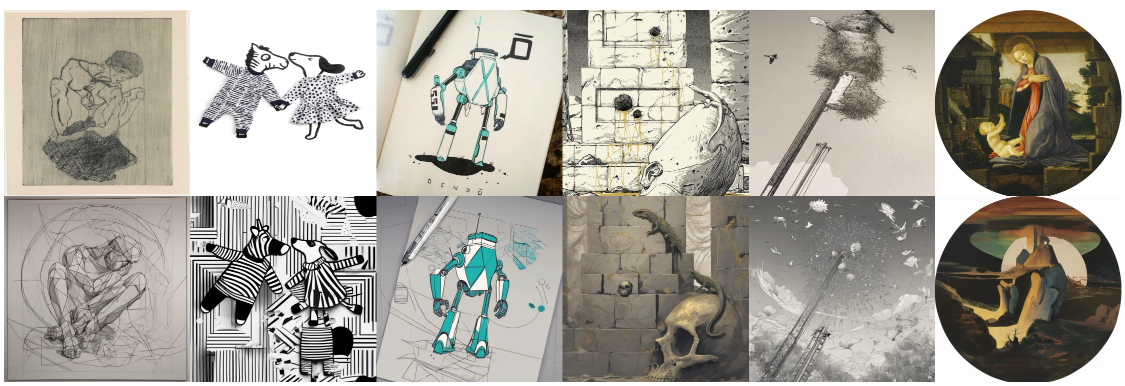
Content

(a)

Style

(b)

Structure



(c)

IP

(d)

Figure 6. Visualization of our manually annotated benchmark. Here, different types of infringement are primarily categorized based on the method of data generation.

ing. For case (c), the two images do not have significantly high scores in the three evaluation aspects, resulting in a judgment of non-infringement.

In summary, HIDM can integrate multiple aspects of similarity relationships to evaluate whether generated images constitute infringement. Additionally, it can leverage the understanding capabilities of VLM to detect more complex IP infringements and provide corresponding judgment criteria for inspectors.

10. Limitations

Despite the fact that our proposed HIDM is capable of detecting various types of infringement, there are still some limitations that need to be addressed. Firstly, the current approach requires three separate models to extract content, style, and structural features, and then calculate their similarities. This multi-model process results in relatively low efficiency. If a single model could perform all these tasks, it would significantly enhance the overall efficiency. Secondly, our method does not offer an effective solution for IP infringement detection and still relies heavily on Vision-Language Models (VLMs). This dependency on VLMs indicates a need for more robust and independent mechanisms to address IP infringement issues comprehensively.