CustomMark: Customization of Diffusion Models for Proactive Attribution

Supplementary Material

1. Additional Experiments

Components Ablation. Tab. 1 presents a comprehensive ablation study to analyze the contribution of individual components in CustomMark to its overall performance. The complete implementation of CustomMark achieves the highest performance across all metrics, with bit accuracy at 96.10%, attribution accuracy at 91.83%, clip score at 0.80, and csd score at 0.77. These results highlight the framework's ability to maintain robust attribution while preserving image quality. The performance drop observed when specific components are removed demonstrates the critical role each plays in the model's functionality.

The removal of the concept encoder results in a significant drop in performance, with bit accuracy and attribution accuracy reduced to 81.21% and 65.19%, respectively. This highlights the encoder's essential role in embedding bi secret information effectively. Similarly, disabling the mapper reduces bit accuracy to 93.10% and attribution accuracy to 87.11%, indicating its importance in maintaining precise attribution. The absence of attention finetuning from LDM moderately impacts the bit accuracy and attribution accuracy. However, qualitative performance is greatly reduced with csd score falling to 0.65, showcasing its role in style matching of clean and watermarked generated images during training.

The removal of regularization loss leads to minor performance degradation for attribution, but it impacts the qualitative metrics like the csd score, which drops to 0.71, demonstrating its role in ensuring consistency during watermark embedding, even tough it's only for initial iterations. Notably, the exclusion of style loss has the most detrimental effect on attribution accuracy, which falls dramatically to 40.16%, emphasizing its importance in preserving stylistic fidelity during the watermarking process. These results collectively validate the carefully designed architecture of CustomMark, where each component contributes significantly in achieving both robust attribution and high-quality image generation.

Changed	Bit Acc. (%)↑	Attribution Acc. (%) ↑	CLIP Score ↑	CSD Score ↑
CustomMark	**96.10**	**91.83**	**0.80**	**0.77**
− Concept Encoder	81.21	65.19	0.65	0.61
− Mapper	93.10	87.11	0.79	0.78
− Att. Finetune	95.16	90.88	0.71	0.65
− Reg. Loss	95.31	90.12	0.77	0.71
− Style Loss	75.10	40.16	0.66	0.62

Table 1. Ablation study of various components of CustomMark for 10 concepts in training. [KEYS: att.:attention, reg. Regularization]

Sequential Learning Analysis. Fig. 1 demonstrates the performance of individual concepts during sequential learning with CustomMark, evaluated through CSD score deviation and attribution accuracy as new concepts are added. The graphs illustrate how CustomMark maintains robust performance while adapting to an increasing number of concepts, showcasing its scalability and efficiency.

In the CSD score deviation plot(Fig. 1(a)), the deviation remains minimal across most concepts, even as the number of concepts increases from 3 to 10. For instance, Hopper and Raphael exhibit only slight increases in deviation (+0.08 and +0.10, respectively) when additional concepts are introduced. This indicates that CustomMark effectively preserves stylistic fidelity for previously learned concepts while integrating new ones. Further, the CSD score before and after attribution remains almost similar. It decreases a little bit in start when the concept is introduced, but it gradually recovers to the original score. Notably, the deviation remains consistently low for concepts like Picasso and Monet, further validating the robustness of the model.

The attribution accuracy plot (Fig. 1(b)) highlights CustomMark's strong adaptability, with consistent attribution for new concepts added to training while maintaining high performance for earlier ones. This demonstrates that CustomMark's sequential learning approach effectively balances the retention of previously learned attributions with the incorporation of new ones, keeping in mind that CustomMark requires only about 10% additional training iterations per concept. These results underline the practical viability of CustomMark in dynamic, real-world scenarios where the set of concepts evolves over time.

Complex Prompts. Fig. 2 demonstrates the effectiveness of using complex and detailed prompts to generate images that accurately match the artistic styles of renowned painters. Each pair of images—one clean and one watermarked—illustrates that even though a long and complex prompt, CustomMark was able to insert the corresponding watermark onto the generated images as long as the concept token was perturbed.

Despite the complexity of the prompts, the generated images successfully capture the signature style of artists such as Dali, Monet, Van Gogh, Picasso, and Warhol. The results showcase precise interpretations of surreal, impressionistic, cubist, and other artistic movements, reinforcing the ability of GenAI model to replicate stylistic nuances into the watermarked images.

ICCV
#6700

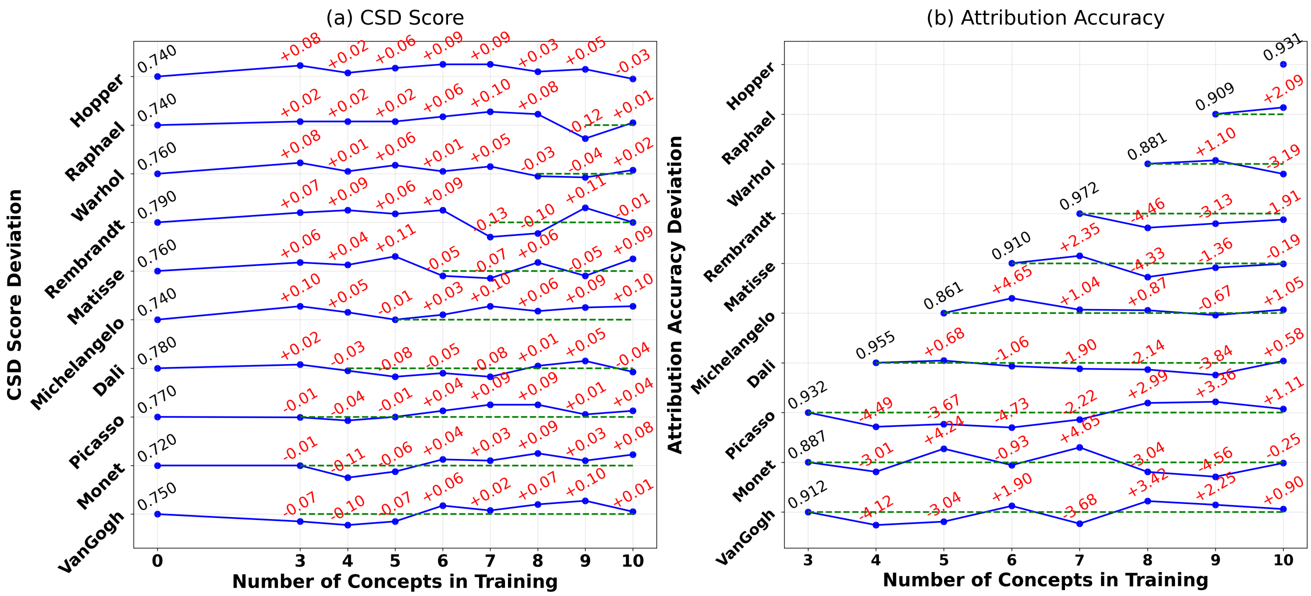

Figure 1. Performance variation of individual concepts during sequential learning.

Analysis of Token Embedding. Fig. 3 illustrates the analysis of original and perturbed tokens through t-SNE plots, norm distributions, and cosine similarity distributions. In the t-SNE plot (Fig. 3(a)), the original tokens (red) and perturbed tokens (blue) demonstrate a clear separation, signifying that the perturbed tokens effectively diverge from their original counterparts. This divergence is critical for embedding unique watermarks and facilitating robust attribution. The norm distributions (Fig. 3(b)) show that original tokens are centered very close to the norm 0 and exhibit a narrower range of vector norms, while perturbed tokens ahve high norms close to 100, and display a wider spread. This indicates that perturbations introduce divergence of the norm as compared to the original tokens and promotes controlled variability to the token space, contributing to their distinctiveness. The cosine similarity distribution (Fig. 3(c)) reveals that the similarity between original and perturbed tokens clusters around zero, highlighting that the perturbations maintain minimal overlap with the original token directions—a necessary condition for ensuring effective attribution.

In our proposed approach, we apply the regularization loss during the initial iterations of training. The regularization ensures that the perturbed tokens start with a meaningful deviation from the original tokens, setting a strong foundation for subsequent learning. To analyze this further, we don't switch off the regularization loss. We observe that continuing the regularization loss throughout the training process leads to the original and perturbed tokens becoming overly similar, undermining the ability to embed distinguishable watermarks and impairing attribution accuracy. With this approach, the model achieves a secret accuracy of 56.14% and an attribution accuracy of 1.54%, Therefore, we strategically switch off the regularization loss after the initial 200 iterations to allow the perturbed tokens to diverge as they want. This maintains the separation between original and perturbed tokens, ensuring that the model can generate robust watermarks while preserving the quality of attribution.

2. More Watermarked Samples

Fig. 4 provides a comparative analysis between clean images, ProMark [1], and CustomMark on the WikiArt dataset, showcasing their performance in attribution while preserving artistic styles across a range of renowned artists from WikiArt dataset. CustomMark demonstrates superior style adaptation compared to ProMark, consistently maintaining the unique stylistic elements and visual fidelity of the original artworks. For artists such as Degas, Picasso, and Van Gogh, CustomMark effectively replicates the signature brushstrokes, color palettes, and composition techniques, resulting in outputs that remain faithful to their distinctive styles. In contrast, ProMark introduces noticeable bubble like artifacts and style distortions that detract from the visual coherence of the images. Similarly, for detailed and intricate works by artists like Sargent and Dore, CustomMark preserves the depth and intricacy, while ProMark struggles with fidelity, leading to degradation in fine details.

Fig. 5 illustrates examples of clean and watermarked images for artists used as concepts, sampled from a model trained on 200 artists. Unlike Fig. 4, which focused on the

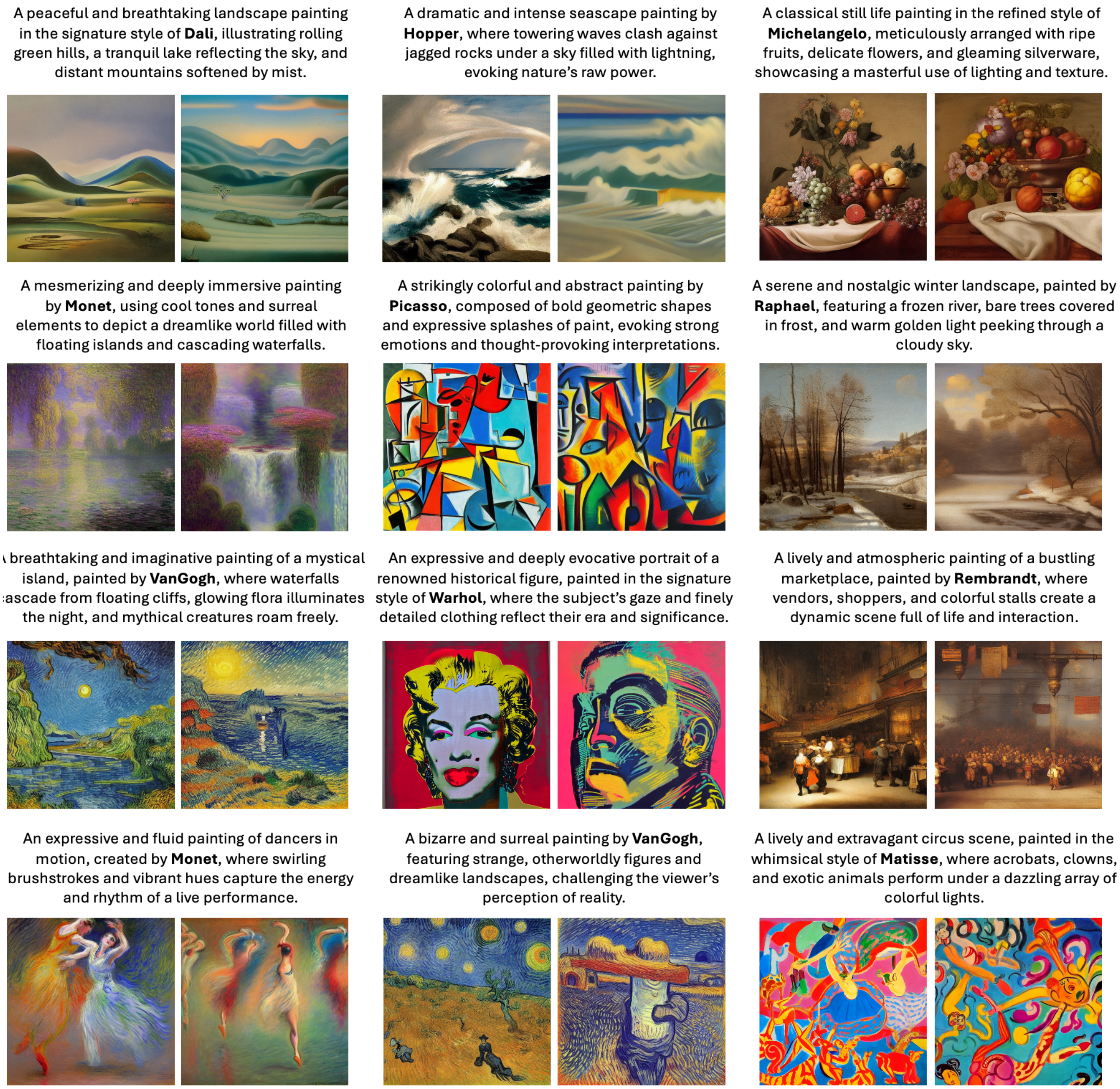

Figure 2. Generated clean (left) and watermarked (right) images pairs for artists as concepts sampled using big and complicated prompts.

WikiArt dataset and showcased the performance of CustomMark for 23 artists, this figure demonstrates the scalability of the method when extended to a much larger and diverse set of artistic concepts. Across a wide range of styles, from Bosch and Klimt's classic depictions to Koons and Haring's contemporary designs, the watermarked images retain the stylistic essence of the clean images while embedding imperceptible watermarks. Notably, the approach performs consistently well across different styles, capturing subtle details in works by artists such as Dürer, Toulouse, and Vermeer without introducing artifacts.

This comparison highlights CustomMark's ability to adapt seamlessly to various artistic styles, ensuring high-quality outputs that respect the original artistic intent, even when dealing with hundreds of distinct artistic styles. Its flexibility and fidelity make it a reliable solution for scenarios requiring robust watermarking without compromising on artistic integrity.

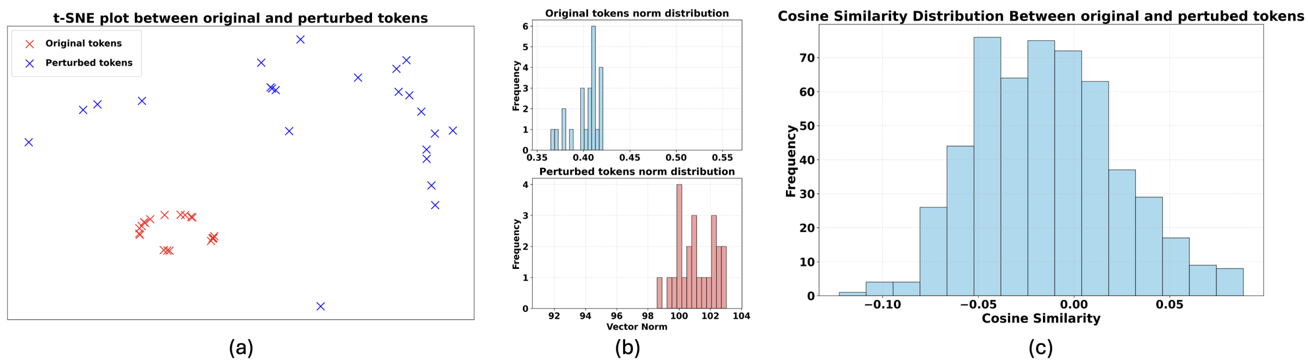

Figure 3. Analysis of original and perturbed tokens by (a) t-SNE plot, (b) norm distribution, and (c) distribution of cosine similarity between the two sets of embeddings.

3. Limitations

While CustomMark offers an efficient solution for concept attribution, it has some limitations. First, it relies on the explicit mention of concepts in prompts, making attribution challenging when an artist's style is indirectly referenced or subtly embedded in the generated image. CustomMark finds it challenging to embed large bit sequences due to the mapper network being too simple for mapping bit sequence to noise perturbation. A sophisticated mapper network might address this issue. Additionally, CustomMark has not been tested on multi-concept scenarios, such as prompts combining multiple artists or blending diverse styles, leaving its robustness in such cases unexplored. Another limitation of CustomMark is its reliance on generated data for training. If the original GenAI model fails to adequately capture an artist's unique style or nuances, the improved model with attribution capabilities may struggle to accurately reflect or attribute that style in the generated images. These limitations highlight areas for future improvement to enhance the system's versatility and robustness.

4. Potential Social Impact

The potential social impact of CustomMark lies in its ability to foster a collaborative and transparent relationship between AI model developers and the artists. By introducing attribution capabilities, this algorithm empowers artists to gain recognition for the influence of their styles on AI-generated content, promoting a sense of agency and fairness. Unlike adversarial strategies that often pit creators against AI systems, CustomMark provides a constructive mechanism to bridge this divide, offering a signal for transparency without compromising creativity. By focusing on attribution and transparency, CustomMark aims to support a harmonious integration of AI into the creative landscape, minimizing potential societal harm and building trust between artists and AI systems.

5. Implementation Details

Artist Lists. The list in Tab. 2 presents a comprehensive compilation of 200 artists, which serves as the foundation for our attribution experiments. For experiments requiring a specific number of artists (top-k), we systematically select the top-k artists based on their numerical ranking in the table. This approach ensures consistency and reproducibility across various experimental setups. An ablation study is conducted by varying k as discussed in the main paper, with artists chosen accordingly. The scalability and robustness of the attribution methodology are assessed under a range of configurations, from smaller subsets of artists to the full set of 200 artists. Furthermore, we extend our evaluation beyond 200 artists by leveraging $1,000$ classes from ImageNet as additional concepts, demonstrating the scalability and adaptability of our approach.

Distortion Applied for Robustness Evaluation. For robustness evaluation in Fig. 8 (main paper), we apply several post-processing distortions. These augmentations are applied by following [2]. Below are the details:

1. **Color Jitter**: For the color jitter augmentation, we modified several aspects of the images. The brightness factor, contrast factor, and saturation were adjusted to a value of 0.3, while the hue factor was set to 0.1 to introduce controlled variations in the image colors.
2. **Crop and Resize**: For the crop and resize augmentation, we randomly extracted 384×384 blocks from the original 512×512 images and resized these blocks to 256×256, simulating different framing conditions.
3. **Gaussian Blur**: We applied Gaussian blur with a kernel size of (3,3) and a sigma value of $(2.0, 2.0)$ to simulate soft-focus effects in the images.
4. **Gaussian Noise**: To introduce random noise, Gaussian noise was added to the images with a standard deviation of 0.05, creating a more realistic representation of noisy environments.

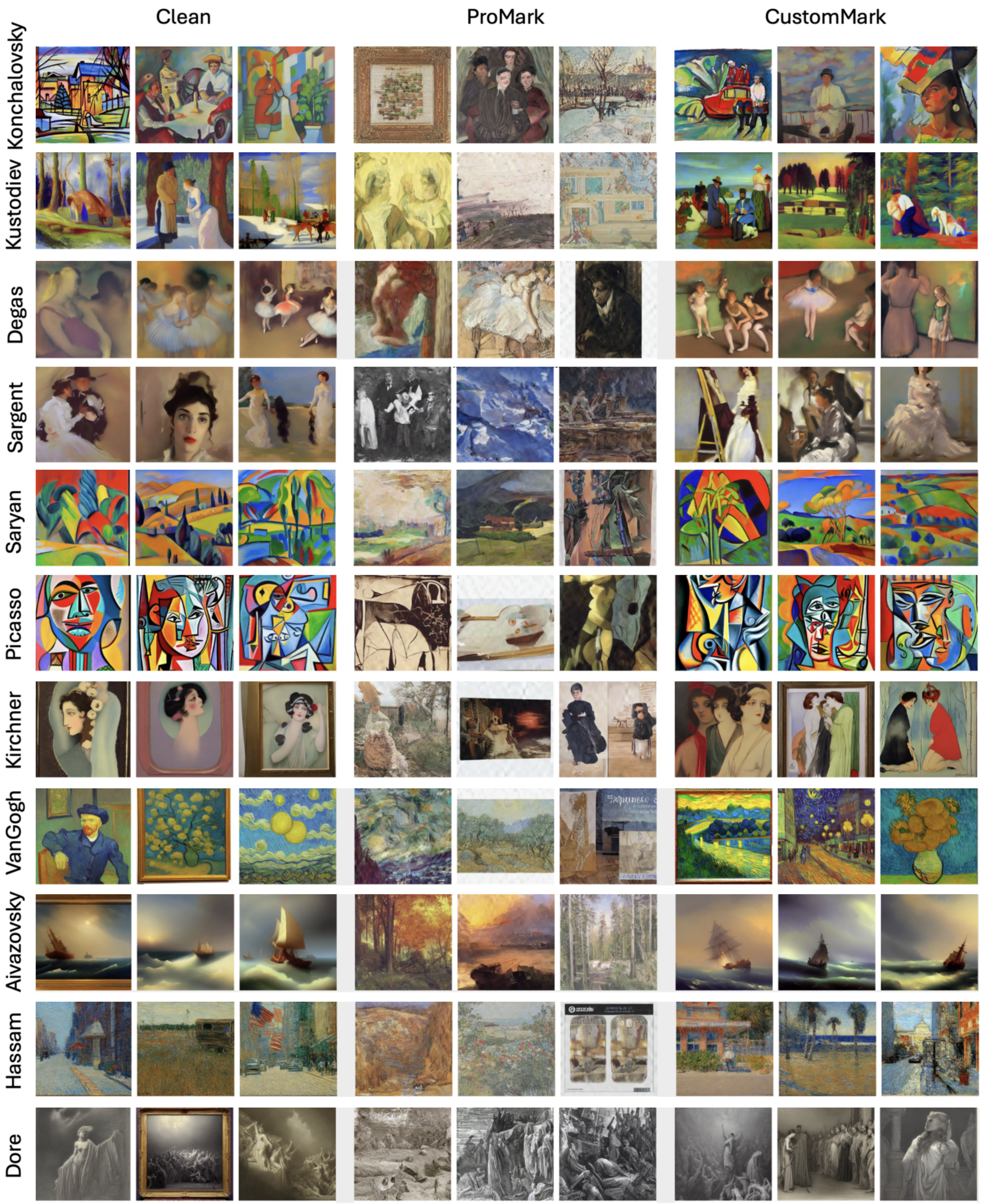

Figure 4. Comparison with ProMark on WikiArt dataset.

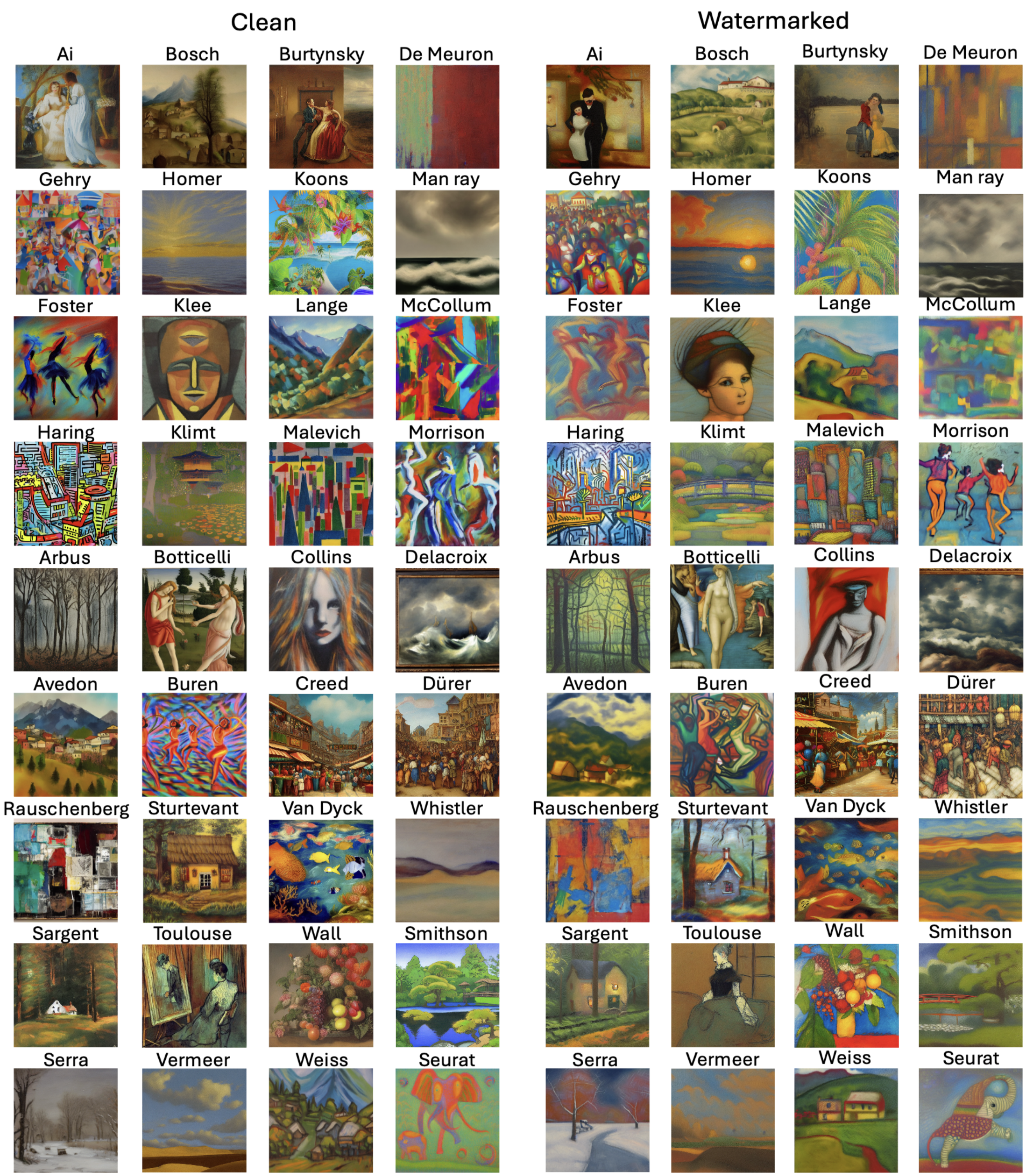

Figure 5. Generated clean and watermarked images for artists as concepts sample by model trained for attributing 200 artists.

5. **JPEG compression**: We used a quality setting of 50 to simulate compression artifacts often encountered in real-world image data.
6. **Rotation**: This augmentation was randomly applied to the images within a range of 0 to 180 degrees to account for changes in orientation during training.
7. **Sharpness**: For the sharpness augmentation, we set the intensity to 1, enhancing the clarity of certain features

1. Claude Monet	2. Pablo Picasso	3. Vincent van Gogh	4. Michelangelo Buonarroti	5. Raphael Sanzio
6. Rembrandt van Rijn	7. Salvador Dalí	8. Henri Matisse	9. Andy Warhol	10. Edward Hopper
11. Frida Kahlo	12. Edgar Degas	13. Paul Cézanne	14. Jackson Pollock	15. Edvard Munch
16. Gustav Klimt	17. Paul Gauguin	18. Pierre-Auguste Renoir	19. Johannes Vermeer	20. Caravaggio
21. Jan van Eyck	22. Édouard Manet	23. Georgia O'Keeffe	24. Francisco Goya	25. Albrecht Dürer
26. Sandro Botticelli	27. Titian	28. Diego Velázquez	29. Giotto di Bondone	30. El Greco
31. Peter Paul Rubens	32. Caspar David Friedrich	33. Wassily Kandinsky	34. Marc Chagall	35. Eugène Delacroix
36. Piet Mondrian	37. Roy Lichtenstein	38. Joan Miró	39. Hieronymus Bosch	40. Jean-Michel Basquiat
41. Gustave Courbet	42. Thomas Gainsborough	43. Jean-Auguste- Dominique Ingres	44. Élisabeth Vigée Le Brun	45. Artemisia Gentileschi
46. Camille Pissarro	47. Georges Seurat	48. Diego Rivera	49. Henri de Toulouse-Lautrec	50. Édouard Vuillard
51. Berthe Morisot	52. Mary Cassatt	53. James Abbott McNeill Whistler	54. John Singer Sargent	55. William Blake
56. David Hockney	57. Keith Haring	58. Jasper Johns	59. Alfred Sisley	60. Jean-Baptiste-Camille Corot
61. Winslow Homer	62. Grant Wood	63. Paul Klee	64. Yayoi Kusama	65. Egon Schiele
66. Amedeo Modigliani	67. Fernand Léger	68. Giorgio de Chirico	69. Henri Rousseau	70. Max Ernst
71. Kazimir Malevich	72. Mark Rothko	73. René Magritte	74. Alphonse Mucha	75. Francis Bacon
76. Marcel Duchamp	77. Leonardo da Vinci	78. Lucian Freud	79. Anselm Kiefer	80. Joseph Beuys
81. Bridget Riley	82. Anish Kapoor	83. Damien Hirst	84. Tracey Emin	85. Ai Weiwei
86. Gerhard Richter	87. Jeff Koons	88. Takashi Murakami	89. Zhang Xiaogang	90. Jenny Saville
91. Kara Walker	92. Yoko Ono	93. Cindy Sherman	94. Louise Bourgeois	95. Barbara Kruger
96. Richard Serra	97. Donald Judd	98. Sol LeWitt	99. Frank Stella	100. Ellsworth Kelly
101. Robert Rauschenberg	102. Claes Oldenburg	103. Paolo Veronese	104. Pieter Bruegel	105. Anthony van Dyck
106. J.M.W. Turner	107. John Constable	108. John Everett Millais	109. Dante Gabriel Rossetti	110. Edward Burne-Jones
111. David Alfaro Siqueiros	112. Rufino Tamayo	113. Victor Vasarely	114. Kurt Schwitters	115. Andy Goldsworthy
116. Richard Long	117. Robert Smithson	118. Christo Javacheff	119. Walter Gropius	120. Robert Venturi
121. Jean Nouvel	122. Daniel Libeskind	123. Richard Rogers	124. Renzo Piano	125. Norman Foster
126. Bjarke Ingels	127. Frank Gehry	128. Santiago Calatrava	129. Toyo Ito	130. Frank Lloyd Wright
131. Alvar Aalto	132. Dominique Perrault	133. Luis Barragán	134. James Stirling	135. Peter Zumthor
136. Kazuyo Sejima	137. Kengo Kuma	138. Jacques Herzog	139. Pierre de Meuron	140. César Pelli
141. Christian de Portzamparc	142. Stefano Boeri	143. Wang Shu	144. Olafur Eliasson	145. Thomas Hirschhorn
146. Felix Gonzalez-Torres	147. Gilbert	148. Ugo Rondinone	149. Paul McCarthy	150. Cory Arcangel
151. Elaine Sturtevant	152. Marcel Broodthaers	153. Maurizio Cattelan	154. Rirkrit Tiravanija	155. Allan McCollum
156. Glenn Ligon	157. Peter Fischli	158. David Weiss	159. Peter Doig	160. Thomas Schütte
161. Neo Rauch	162. Marlene Dumas	163. Felix Gonzalez-Torres	164. Lorna Simpson	165. Byrne Morrison
166. Glenn Martin	167. Dan Collins	168. Matthew Barney	169. Peter Hujar	170. Shirin Neshat
171. Thomas Demand	172. Alexander McQueen	173. Catherine Opie	174. Wolfgang Tillmans	175. Martin Creed
176. Olafur Eliasson	177. James Turrell	178. Bill Viola	179. Andreas Gursky	180. Lewis Baltz
181. Cindy Sherman	182. Man Ray	183. Bruce Nauman	184. Sol LeWitt	185. Richard Hamilton
186. James Rosenquist	187. Nam June Paik	188. Vito Acconci	189. Susan Rothenberg	190. Lawrence Weiner
191. Daniel Buren	192. Robert Gober	193. Adrian Piper	194. Katharina Fritsch	195. Christian Marclay
196. Richard Avedon	197. Jeff Wall	198. Edward Burtynsky	199. Julius Lange	200. Diane Arbus

Table 2. Comprehensive List of 200 Artists

within the images.

Architecture Details. We use several networks for designing CustomMark, which include concept encoder, secret mapper, and secret decoder. For concept encoder, a U-Net-inspired network designed for processing and transforming 1D sequential data is adopted. Initially, a fully-connected layer is maps the bit sequence to a feature vector which is concatenated with the token embedding. This is given as input to the encoder-decoder framework of U-Net to output the perturbed token embedding.

The mapper network is a feature transformation module designed to encode input indices into high-dimensional representations using an embedding-based approach. It employs a learnable embedding layer that maps input indices (*e.g.* 16) to vectors in a higher-dimensional space (*e.g.* 64). The embeddings are initialized orthogonally and scaled to maintain a unit standard deviation. During the forward pass, the network retrieves embeddings for all possible input indices, weights them element-wise based on the input tensor, and sums these weighted embeddings along the input dimension. The result is normalized by the square root of batch size and biased by adding 1, producing a robust high-dimensional representation for each input bit sequence.

Finally, we use the EfficientNet-B3 [3] architecture as its core backbone for secret decoder. The network is initialized with pre-trained weights from the ImageNet dataset for robust feature extraction. The final classifier layer of EfficientNet is replaced with a fully connected layer that outputs the predicted bit sequence.

Prompt Details. Following [2], we use various prompts for sampling clean and watermarked images which are used to train CustomMark. The collection of prompts is different, depending on the concept we attribute. We replace the "[name]" with the corresponding concept token. Below are the details:

1. Artists as concepts:
   - *"a painting, art by [name]"*
   - *"a rendering, art by [name]"*
   - *"a cropped painting, art by [name]"*
   - *"the painting, art by [name]"*
   - *"a clean painting, art by [name]"*
   - *"a dirty painting, art by [name]"*
   - *"a dark painting, art by [name]"*
   - *"a picture, art by [name]"*
   - *"a cool painting, art by [name]"*
   - *"a close-up painting, art by [name]"*
   - *"a bright painting, art by [name]"*
   - *"a cropped painting, art by [name]"*
   - *"a good painting, art by [name]"*
   - *"a close-up painting, art by [name]"*
   - *"a rendition, art by [name]"*
   - *"a nice painting, art by [name]"*
   - *"a small painting, art by [name]"*
   - *"a weird painting, art by [name]"*
   - *"a large painting, art by [name]"*
   - *"A serene landscape painting in the style of [name]"*
   - *"A bustling cityscape in the style of [name]"*
   - *"A painting of a cozy cottage in the woods in the style of [name]"*
   - *"A vibrant underwater scene in the style of [name]"*
   - *"A whimsical painting of a flying elephant in the style of [name]"*
   - *"A still life painting featuring fruit and flowers in the style of [name]"*
   - *"A portrait of a famous historical figure in the style of [name]"*
   - *"A painting of a dreamy night sky in the style of [name]"*
   - *"A colorful abstract painting in the style of [name]"*
   - *"A street scene from Paris in the style of [name]"*
   - *"A depiction of a beautiful sunset over the ocean in the style of [name]"*
   - *"A painting of a peaceful mountain village in the style of [name]"*
   - *"An energetic painting of dancers in motion in the style of [name]"*
   - *"A painting of a snow-covered winter scene in the style of [name]"*
   - *"A painting of a tropical paradise in the style of [name]"*
   - *"A painting of a magical forest filled with fantastical creatures in the style of [name]"*
   - *"A painting of a dramatic stormy seascape in the style of [name]"*
   - *"A portrait of a majestic lion in the style of [name]"*
   - *"A painting of a romantic scene between two lovers in the style of [name]"*
   - *"A painting of a serene Japanese garden in the style of [name]"*
   - *"A painting of a bustling marketplace in the style of [name]"*
   - *"A painting of a tranquil river scene in the style of [name]"*
   - *"A painting of a fiery volcano eruption in the style of [name]"*
   - *"A painting of a futuristic cityscape in the style of [name]"*
   - *"A painting of a whimsical circus scene in the style of [name]"*
   - *"A painting of a mysterious moonlit forest in the style of [name]"*
   - *"A painting of a dramatic desert landscape in the style of [name]"*
   - *"A portrait of a regal peacock in the style of [name]"*
   - *"A painting of a mystical island in the style of [name]"*
   - *"A painting of a lively carnival scene in the style of [name]"*
2. ImageNet classes as concepts:
   - *"a photo of a [name]"*
   - *"a rendering of a [name]"*
   - *"a cropped photo of the [name]"*
   - *"the photo of a [name]"*
   - *"a photo of a clean [name]"*
   - *"a photo of a dirty [name]"*
   - *"a dark photo of the [name]"*
   - *"a photo of my [name]"*
   - *"a photo of the cool [name]"*
   - *"a close-up photo of a [name]"*
   - *"a bright photo of the [name]"*
   - *"a cropped photo of a [name]"*
   - *"a photo of the [name]"*
   - *"a good photo of the [name]"*
   - *"a photo of one [name]"*
   - *"a close-up photo of the [name]"*
   - *"a rendition of the [name]"*
   - *"a photo of the clean [name]"*
   - *"a rendition of a [name]"*
   - *"a photo of a nice [name]"*
   - *"a good photo of a [name]"*
   - *"a photo of the nice [name]"*
   - *"a photo of the small [name]"*
   - *"a photo of the weird [name]"*
   - *"a photo of the large [name]"*
   - *"a photo of a cool [name]"*
   - *"a photo of a small [name]"*
   - *"a photo of a [name] playing sports"*
   - *"a rendering of a [name] at a concert"*
   - *"a cropped photo of the [name] cooking dinner"*
   - *"the photo of a [name] at the beach"*
   - *"a photo of a clean [name] participating in a marathon"*
   - *"a photo of a dirty [name] after a mud run"*
   - *"a dark photo of the [name] exploring a cave"*
   - *"a photo of my [name] at graduation"*
   - *"a photo of the cool [name] performing on stage"*
   - *"a close-up photo of a [name] reading a book"*
   - *"a bright photo of the [name] at a theme park"*
   - *"a cropped photo of a [name] hiking in the mountains"*
   - *"a photo of the [name] painting a mural"*

- *"a good photo of the [name] at a party"*
- *"a photo of one [name] playing an instrument"*
- *"a close-up photo of the [name] giving a speech"*
- *"a rendition of the [name] during a workout"*
- *"a photo of the clean [name] gardening"*
- *"a rendition of a [name] dancing in the rain"*
- *"a photo of a nice [name] volunteering at a charity event"*
- *"a photo of a [name] surfing a giant wave"*
- *"a rendering of a [name] skydiving over a scenic landscape"*
- *"a cropped photo of the [name] riding a rollercoaster"*
- *"the photo of a [name] rock climbing a steep cliff"*
- *"a photo of a clean [name] practicing yoga in a peaceful garden"*
- *"a photo of a dirty [name] participating in a paintball match"*
- *"a dark photo of the [name] stargazing at a remote location"*
- *"a photo of my [name] crossing the finish line at a race"*
- *"a photo of the cool [name] breakdancing in a crowded street"*
- *"a close-up photo of a [name] blowing out candles on a birthday cake"*
- *"a bright photo of the [name] flying a kite on a sunny day"*
- *"a cropped photo of a [name] ice-skating in a winter wonderland"*
- *"a photo of the [name] directing a short film"*
- *"a good photo of the [name] participating in a flash mob"*
- *"a photo of one [name] skateboarding in an urban park"*
- *"a close-up photo of the [name] solving a Rubik's cube"*
- *"a rendition of the [name] fire dancing at a beach party"*
- *"a photo of the clean [name] planting a tree in a community park"*
- *"a rendition of a [name] performing a magic trick on stage"*
- *"a photo of a nice [name] rescuing a kitten from a tree"*

References

[1] Vishal Asnani, John Collomosse, Tu Bui, Xiaoming Liu, and Shruti Agarwal. Promark: Proactive diffusion watermarking for causal attribution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10802–10811, 2024. 2

[2] Weitao Feng, Jiyan He, Jie Zhang, Tianwei Zhang, Wenbo Zhou, Weiming Zhang, and Nenghai Yu. Catch you everything everywhere: Guarding textual inversion via concept watermarking. *arXiv preprint arXiv:2309.05940*, 2023. 4, 8

[3] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019. 7