Department of Computer Science
Hong Kong Baptist University

COST: Cloud-Oriented Style Transfer for Converting Natural Cloud Shapes to Animal Shapes

– COMP3076 Tech-Track Group Project

LI Jinchuan: 21252548
YU Fengfei: 21251215
ZHANG Kaicheng: 22221751

March 2025

1 Background

1.1 Generative Models and Their Applications

Generative models are a class of machine learning models designed to generate new data samples that resemble a given dataset. Unlike traditional discriminative models, which learn to classify or predict labels based on input data, generative models focus on understanding the underlying data distribution and creating novel instances from it. Therefore, generative models helps solve the data scarcity problem.

Several types of generative models have gained significant attention in recent years. One of the most prominent categories is Generative Adversarial Networks (GANs) (Goodfellow et al., 2020), which consist of a generator and a discriminator competing in a min-max game to improve the quality of generated samples. Another key class is Variational Autoencoders (VAEs) (Kingma, Welling et al., 2013), which learn a probabilistic latent space representation to generate new data samples. More recently, diffusion models, such as Denoising Diffusion Probabilistic Models (DDPMs) (Ho, Jain and Abbeel, 2020), have emerged as powerful approaches for high-quality image synthesis by progressively refining noisy data into structured outputs.

Generative models have numerous applications across various domains. In computer vision, they are used for image synthesis, style transfer, super-resolution, and data augmentation. In natural language processing, they power text generation models like GPT (Brown et al., 2020), enabling realistic text completion and dialogue systems. Additionally, they are widely applied in drug discovery, 3D modeling, and even music generation, demonstrating their versatility across creative and scientific fields.

1.2 Generative Art: Creativity Through AI

Generative models have also revolutionized artistic creation by enabling new forms of digital art, often referred to as generative art. Generative art leverages algorithms to autonomously create images, animations, and other artistic works, often incorporating randomness and machine learning to achieve unique results.

One popular approach to generative art is using GANs and diffusion models to create visually compelling artworks. Models like StyleGAN (Karras, Laine and Aila, 2019), and Stable Diffusion (Rombach et al., 2022) allow artists and designers to generate novel compositions, blend artistic styles, and even synthesize photorealistic images from textual descriptions. Text-to-image models such as DALL·E (Ramesh et al., 2021) and CLIP (Radford et al., 2021) have enabled artists to transform textual prompts into detailed visual artworks, expanding creative possibilities.

Another significant application of generative models in art is style transfer, a technique that allows the artistic style of one image to be applied to another. Originally popularized by Neural Style Transfer (NST) (Gatys, Ecker and Bethge, 2015), this approach uses deep neural networks to extract content and style features separately and recombine them to produce images that retain the structure of the original content while adopting the artistic appearance of the reference style. More advanced methods integrate GANs and diffusion models to achieve higher-quality style transfer, enabling artists to experiment with transforming photographs into paintings, merging different artistic styles, or even reinterpreting classical artworks in modern aesthetics.

Beyond generating entirely new images, generative models are also applied in art restoration and enhancement. These techniques can be used to restore damaged artworks, inpaint missing regions, or recreate lost masterpieces by learning artistic styles and patterns from historical data.

With the continuous evolution of generative models, AI-driven artistic creation is becoming increasingly sophisticated, offering new opportunities for both artists and researchers to explore the intersection of technology and creativity. The combination of text-to-image synthesis, style transfer, and restoration techniques demonstrates the growing impact of AI in redefining artistic workflows and expanding creative boundaries.

2 Problem Formulation

The objective of this project is to process cloud regions in images and transform their shapes into recognizable animal forms while preserving their natural cloud-like texture. The primary challenge is to ensure that the generated animal-shaped clouds remain visually coherent, context-aware, and seamlessly integrated into the original scene. This involves addressing different scenarios where clouds appear in various environmental conditions and resolving challenges related to shape-texture compatibility, spatial distribution, and semantic preservation.

2.1 Identified Scenarios

Cloud transformation varies depending on the composition of the input image. Several key scenarios have been identified:

- **Scenario 1: Clear Sky with Clouds (Minimal Noise)**
  This is the ideal scenario where the image consists solely of sky and clouds, without any foreground elements. The transformation focuses purely on reshaping the clouds into animal forms while maintaining realistic cloud textures, such as soft edges and varying degrees of translucency.

- **Scenario 2: Cloud Background with Foreground Objects**
  In this scenario, clouds serve as the background, while objects such as trees, buildings, or mountains are present in the foreground. The transformation must ensure that only the cloud regions are modified while preserving the integrity of foreground objects and the depth relationships within the scene.

- **Scenario 3: Partial Occlusions**
  Some cloud regions may be partially covered by thin objects such as power lines, poles, or birds. These occlusions introduce additional complexity in identifying and transforming cloud regions while maintaining a natural appearance. The presence of such elements may affect the perceived shape of the cloud and influence the transformation process.

- **Scenario 4: Different Weather Conditions**
  The structure and density of clouds vary significantly under different weather conditions. On a clear day, clouds may be sparse and thin, whereas overcast conditions result in thick, dense cloud formations. The transformation method must adapt to these variations to ensure that generated animal-shaped clouds fit naturally into their respective cloud distributions. For example, a sparse, flat cloud formation should not be transformed into a large, dense, and voluminous animal shape.

2.2 Key Challenges

To achieve a seamless transformation of clouds into animal shapes, the following challenges must be addressed:

1. **Shape-Texture Compatibility**
   The generated animal shapes must retain cloud-like textures and physical plausibility. Features such as soft edges, transparency, and natural lighting should be preserved to avoid artificial or jarring appearances. Additionally, the generated shape must be consistent with the spatial distribution of clouds.

2. **Context-Aware Integration**
   The generated animal-shaped clouds must align with the surrounding environment to ensure a natural

appearance. For example, if the original clouds are sparse and wispy, the transformed shape should maintain a similar level of density and dispersion rather than appearing overly dense or solid. The transformation should respect the original atmospheric conditions to avoid visual inconsistencies.

3. **Boundary Handling and Semantic Preservation**
   Clouds often blend smoothly with the sky and foreground objects, making boundary detection and modification challenging. The transformation method must prevent artifacts at cloud boundaries and ensure that foreground objects remain unaffected. Additionally, the generated cloud shape should not obscure important elements of the original image, such as mountains or buildings, in a way that alters its original meaning and depth perception.

By addressing these scenarios and challenges, the project aims to develop a method that can effectively transform cloud regions into realistic and context-aware animal shapes while maintaining the visual and semantic integrity of the original image.

3 Proposed Method

The proposed method is divided into four main steps: selecting the region of interest (ROI), processing the selected region, repainting it into a new shape, and post-processing to ensure seamless integration with the original image.

3.1 Step 1: Select Region of Interest (ROI)

The first step is to identify and extract the cloud region from the input image, which will be used for further processing. There are two primary approaches: object detection and segmentation. Object detection models, such as YOLO, can generate bounding boxes around cloud regions, while segmentation models like SAM provide more precise region extraction.

However, open-vocabulary segmentation methods, such as Grounding-DINO with SAM, struggle to differentiate clouds from the sky. To overcome this limitation, we propose fine-tuning an object detection model with a specialized cloud dataset. If a suitable dataset is unavailable, we can create our own dataset by re-describing existing cloud images using the ComfyUI workflow or generating cloud images using a diffusion model. This ensures accurate region extraction for downstream processing.

3.2 Step 2: Region Processing

After selecting the ROI, the next step is to preprocess the cloud region to make it suitable for repainting. A key requirement is to preserve the edge details of the extracted cloud region while making its features more distinguishable for subsequent steps.

The proposed method enhances the selected cloud region using basic computer vision operations, such as contrast enhancement and smoothing operations. Additionally, we apply resolution enhancement techniques to make the cloud features more prominent. This is important as the processed cloud image will later be used as a reference image for style transfer. Following these enhancements, Canny edge detection is used to identify and extract the outer edges of the cloud. By focusing on the boundary, we ensure that the repainted image aligns well with the original mask, preventing unnatural artifacts in the final output.

3.3 Step 3: Repainting

In this step, the selected cloud region is transformed into an animal shape using a Stable Diffusion model, guided by a given animal image or prompt-generated animal image. The process begins with using Canny edge detection on the provided animal image to capture its outline. Then, a carefully crafted text prompt is used to guide the diffusion model, ensuring that the generated animal shape is aligned with the cloud texture. The model is also guided to incorporate the specific cloud style into the generated animal shape.

To introduce variability, a custom node randomly selects an animal shape from a predefined list. This allows each execution of the workflow to generate a different animal transformation while ensuring that the generated shape maintains boundary consistency with the original cloud edges.

3.4 Step 4: Post-Processing

The final step addresses the potential shape mismatches between the generated animal-shaped clouds and the original cloud region. These mismatches may occur due to differences in shape or texture between the generated animal shape and the original clouds with respect to the surroundings. To resolve this, we use the original cloud environment to fill in the mismatched areas, ensuring that the final transformation is coherent with the surrounding clouds.

In addition to this shape adjustment, we perform color correction and texture matching to ensure that the final generated image retains consistent visual properties, such as color, opacity, and texture, with the original clouds. This post-processing step ensures that the transformed cloud region blends seamlessly with the rest of the image, maintaining both visual integrity and realism.

4 References

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A. et al., 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33, pp.1877–1901.

Gatys, L.A., Ecker, A.S. and Bethge, M., 2015. A neural algorithm of artistic style. *arxiv preprint arxiv:1508.06576*.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., 2020. Generative adversarial networks. *Communications of the acm*, 63(11), pp.139–144.

Ho, J., Jain, A. and Abbeel, P., 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33, pp.6840–6851.

Karras, T., Laine, S. and Aila, T., 2019. A style-based generator architecture for generative adversarial networks. *Proceedings of the ieee/cvf conference on computer vision and pattern recognition*. pp.4401–4410.

Kingma, D.P., Welling, M. et al., 2013. Auto-encoding variational bayes.

Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J. et al., 2021. Learning transferable visual models from natural language supervision. *International conference on machine learning*. PmLR, pp.8748–8763.

Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M. and Sutskever, I., 2021. Zero-shot text-to-image generation. *International conference on machine learning*. Pmlr, pp.8821–8831.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P. and Ommer, B., 2022. High-resolution image synthesis with latent diffusion models. *Proceedings of the ieee/cvf conference on computer vision and pattern recognition*. pp.10684–10695.