DragGAN: Interactive Point-Based Image Manipulation on Generative Adversarial Networks

Muran Wu

Data Science Institute, Shanghai Lida University, Shanghai, 201608, China

Abstract. Users have increasingly demanded greater control over generated images, including flexibility, precision, and versatility as a result of the development of Generative Adversarial Networks (GANs). This post introduces DragGAN, a picture-enhancing method that uses engaging dragging to obtain exact control over philosophical photos. DragGAN enables users to move, and adjust the position, style, location, and size of specific areas of a picture, and do so by integrating feature-operated motion supervision and point-tracking techniques. DragGAN demonstrates its ability to move key points more specifically to the precise positions in the framework of combined image restoration, for instance. Experimental results demonstrate that DragGAN outperforms conventional methods in terms of the generated graphics ' realism and objective-level accuracy. This method significantly enhances the flexibility and efficiency of picture editing, lowers the technical barrier, and enables quasi-expert users to easily accomplish higher-quality image editing, marking a significant advancement in the field of image synthesis. Future research will focus on curbing reliance on write-up-trained GAN patterns and increasing the person's steadiness and accuracy in complex scenes. This indicates that DragGAN's engineering is still developing, and future additions and changes may be made to improve the user experience and control results.

1 Introduction

In the field of image processing, Generative Adversarial Networks (GANs), and deep learning models have enormous potential and a wide range of application prospects [1]. The ability to control the created accessible information is a critical functionality requirement for these learning-based image synthesis techniques in actual-world applications. The three major requirements for people's analysis of designs are flexibility, accuracy, and flexibility. Models used to have photos that may provide the majority of people, but some models just managed to meet one or two of these conditions, with no great component able to fulfill all of them.

Most previous methods achieved controllability of GANs through the following two approaches: using prior 3D models or relying on supervised learning with manually annotated

Corresponding author: yuwangbao@ldy.edu.rs

data [2-4]. However, these methods had certain issues, such as being unable to generalize to new object categories, typically controlling only a limited range of spatial attributes, or providing minimal control over the editing process. Text-guided image synthesis gained attention recently, but this method lacked precision and flexibility in editing spatial attributes [5-7]. For instance, it achieved a certain level of text-guided synthesis but had an accuracy of only 77% [5].

Drag-and-cut processing has just attracted a lot of attention since the opened-resource release of DragGAN technology [8]. DragGAN, a strong pixel-level interactive editing technique, introduces an amazing drag-and-drop mechanism that enables users to alter and improve features like the position, shape, position, and size of the image using pull-and-drop operations. This pull-and-shoot method is generally appropriate and free of obvious distortion or relics. Moreover, its simple procedure drastically reduces technological obstacles to photo editing, making it simple for perhaps non-professional users to begin. To build a solid and user-friendly engaging photo editing approach, DragGAN is based on the StyleGAN2 design and combines feature-based motion control, point tracking, the efficiency of style information in StyleGAN, as well as user interaction and masking functions [9, 10].

The statement examines DragGAN's amazing features and capabilities. By examining its principal technology and actual-world applications, it is possible to understand how DragGAN's solid and user-friendly interesting image editing approach is aided by its combination of feature-based motion supervision, point tracking techniques, style information optimization, and user interaction functions. Based on the StyleGAN2 form, DragGAN represents a significant advance in image editing. Its potential and limitations will be revealed in the study.

2 Related work

2.1 Methods

A function-based movement direction component and a novel point-trapping technique make up DragGAN's vital components.

2.1.1 Feature-based motion power

DragGAN's motion control unit enables the exact action of fraudulent items from their starting jobs to designated certain areas. Using a characteristic-based decrease work, the GAN's latent code is tailored to achieve this accuracy. Features are extracted from a middle level of the GAN's algorithms, such as the second wall of StyleGAN2, and a feature patch loss is calculated between the specific layout and the handled image. While keeping the generated article's realism and regularity alive, this process ensures that the modifications are in line with the user's designed changes. It identifies motion supervision loss as, Given a manipulation point p_i and its target position t_i:

$$L_{motion} = \sum i \, ||F(qi) - F_0(pi) + d_i||_2^2 \quad (1)$$

Where F is the feature map extracted from GAN, F_0 is the original feature map, q_i is the current position of the manipulation point, and d_i is the direction vector pointing from p_i to t_i.

2.1.2 Point tracking

The point tracking unit changes the deception items ' opportunities to show their new areas in the altered photograph after each action is handled the action. It uses a function- based point- to- investing strategy because the motion supervision loss does not immediately give the new jobs.

Specifically, it represents each manipulation point p_i by its feature descriptor $p_i = F_0(p_i)$ and search for the nearest neighbor of f_i in the feature map F of the manipulated image. This is achieved by computing the Euclidean distance between f_i and features within a local patch around each pixel in F:

$$\hat{p}_i = argmin_q ||F(q) - f_i||_2 \quad (2)$$

Where $\hat{p}_i$ is the updated position of the manipulation point.

2.1.3 Iterative selling

The motion control and point tracking methods are repeated incrementally until the most effective number of iterations is achieved before all adjusting opportunities are at their desired employment. To ensure safety and prevent major climbs in the latent space, just a little action is made in each era toward the target.

2.2 Datasets and implementation details

The method is implemented based on PyTorch [11]. The Adam optimizer is used to optimize the latent code w of the FFHQ (Flickr-Faces-High-Quality), AFHQCat (Animal Faces High Quality - Cats), and LSUN Car (Large-scale Scene Understanding Car Dataset) datasets with a step size of 2e – 3 [10, 12-14]. The hyperparameters are set as $\lambda = 20$, $r1$ = round(3/512 × $size$), and $r2$ = round(12/512 × $size$), where $size$ is the resolution of the generated image. In the implementation, the optimization process is stopped when the distance between all manipulation points and their corresponding target points is no more than d pixels. When there are no more than 5 manipulation points, d is set to 1; otherwise, d is set to 2.

2.3 Review and experimental design

Quantitative evaluation is carried out using the method of face landmark manipulation. Since face landmark detection using off-the-shelf tools is very reliable, its prediction results are taken as the true landmarks. Specifically, it randomly generates two face images using StyleGAN trained on FFHQ and detects their landmarks. The goal is to manipulate the landmarks of the first image to match those of the second image. After the manipulation, it detects the landmarks of the final image and calculates the mean distance (MD) to the target landmarks. The results are averaged over 1000 tests. The experiment is evaluated under two settings, including one and five different numbers of key points, to demonstrate the robustness of the method under different numbers of manipulation points. The Fréchet Inception Distance (FID) score between the edited image and the initial image is also reported as an indicator of image quality [15]. The same set of test samples is used to evaluate all methods. In this way, the final MD score reflects the ability of the method to move the landmarks to the target positions. The most crucial earlier job is mathematically compared to DragGAN and UserControllableLT [16]. The mean distance (MD) to the target landmarks and the Fréchet Inception Distance (FID) is used to measure the accuracy of reaching the target points and the realism of the generated images.

3 Results

The results, presented in Table 1, demonstrate that DragGAN outperforms UserControllableLT in both methods. In specific, DragGAN achieved MD distances of 2.44 and 3.18, indicating that its target-level efficiency is higher than that of UserControllableLT, which achieved MD distances of 11.64 and 10.41, which is higher. Also, DragGAN's FID tally of 9.28 is lower than UserControllableLT's 25.32, this indicates that DragGAN's photos are more appropriate.

The function-based point-to-point tracking technique employed in DragGAN is incredibly powerful at properly recording controlled positions, according to analysis. DragGAN effectively relieves the location of the revision points, but in complex and dynamic scenes, by representing each element point with feature descriptors and conducting a search for the nearest neighbors on the feature map of the governed image.

Table 1. Quantitative comparison with UserControllableLT.

Method	1 point	5 points	FID
No edit	12.93	11.66	-
UserControllableLT	11.64	10.41	25.32
Dragan	2.44	3.18	9.28

4 Discussion

DragGAN offers several advantages over ancient image manipulation methods:

Flexibility and Precision: With DragGAN, people can influence pictures by transferring specific features directly onto the image.

Generalizability: Without requiring more comments or prior knowledge, DragGAN may be applied to a range of photo categories and tough circumstances. It emphasizes its game skill in the ebook "DragGAN: Interactive Point-based Manipulation on the Generative Image Manifold" on distinct picture groups, including animals, cars, people, beauty, etc [8].

Efficiency: Because DragGAN operates in the GAN's latent space, quick and effective image manipulation is feasible.

Although DragGAN produces incredible results, there are still some limitations that need to be addressed in future work:

Dependence on post-trained GANs: This limits DragGAN's importance to fresh areas because it relies on post-trained GANs. Potential work was to discover domain adaptation strategies to increase DragGAN to mysterious regions. The study paper" Domain Generalization in Deep Learning: A Review" uses a variety of techniques to increase the universality of ideas across a variety of areas [17]. Some of these approaches and conversations do function as reference points for addressing DragGAN's limitations, particularly in terms of examining ways to lessen the reliance on post-trained models and improve the design to new domains.

Error Accumulation in Point Tracking: The point tracking system may produce errors over several versions, especially in difficult scenarios. To address this problem, more powerful point-to-place scanning techniques will be investigated in the future.

User software: The existing user interface is simple, but it could be made to support more hard adjustment things. Advanced users interface that let people discuss manipulation restrictions and boundaries may become explored in the future.

5 Conclusion

This file has uncovered several important conclusions through detailed DragGAN research. DragGAN is a groundbreaking engaging image-processing system that employs feature-controlled motion supervision and point-tracking techniques. This enables it to provide increased flexibility, accuracy, and efficiency in image manipulation.

More accurately, instinctive pull-and-cut software allows for the manipulation of image information. For example, just dragging information on an image does change unique features. This level of control makes great coarse changes to the generated pictures needed. Also, DragGAN makes the user system simpler, raising the technical barrier for non-skilled people. So, high-quality photo editing is simple to perform, even for those with little specific training.

DragGAN represents a significant advance in image synthesis. It opens up fresh avenues for creating customer-driven materials. For example, it is relevant to a range of applications, like digital art, images, and graphic design. DragGAN, a potent instrument for image processing, makes it easier to deliver innovative ideas to life with greater efficiency and comfort. However, DragGAN has the potential to revolutionize how image manipulation and material development are done.

DragGAN's successful actual-period processing capabilities are extremely important in promoting the development of image processing technology, and their successful application fills the gap in the accuracy and flexibility of existing photo processing technologies. It provides another expert with a strong program to study more difficult image manipulation issues. Also, DragGAN's simple layout makes it easier for non-expert users to enter the industry of photo editing, fostering a wider range of ideas and software.

Despite DragGAN's remarkable achievements in the field of creative control, there are still some limitations, such as the dominance on post- trained GAN models and the performance challenges in handling extremely complex scenes. Future research should look at ways to improve stability and accuracy in complex scenes, decrease the dependence on certain GAN designs, and increase the person's statement features. To build a more precise and powerful photo editing ecosystem, researchers may also look into how DragGAN can be integrated with other photo processing systems, such as heavy understanding and regular image editing tools.

References

1. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio. Generative adversarial nets. NeurIPS (2014).
2. Y. Deng, J. Yang, D. Chen, F. Wen, X. Tong. Disentangled and controllable face image generation via 3D imitative-contrastive learning. CVPR (2020).
3. P. Ghosh, P. Gupta, R. Uziel, A. Ranjan, M. Black, T. Bolkart. GIF: Generative interpretable faces. 3DV (2020).
4. A. Tewari, M. Elgharib, G. Bharaj, F. Bernard, H. Seidel, P. Pérez, M. Zollhofer, C. Theobalt. StyleRig: Rigging StyleGAN for 3D control over portrait images. CVPR (2020).
5. A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, M. Chen. Hierarchical text-conditional image generation with clip latents. arXiv preprint arXiv:2204.06125 (2022).
6. R. Rombach, A. Blattmann, D. Lorenz, P. Esser, B. Ommer. High-resolution image synthesis with latent diffusion models. arXiv:2112.10752 (2021).

7. C. Saharia, W. Chan, S. Saxena, L. Li, J. Whang, E. Denton, S. Seyed Ghasemipour, B. Ayan, S. Mahdavi, R. Lopes et al. Photorealistic text-to-image diffusion models with deep language understanding. arXiv preprint arXiv:2205.11487 (2022).
8. X. Pan, A. Tewari, T. Leimkühler, L. Liu, A. Meka, C. Theobalt. Drag Your GAN: Interactive point-based manipulation on the generative image manifold. In Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference (SIGGRAPH '23). ACM, New York, NY (2023).
9. T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, T. Aila. Analyzing and improving the image quality of StyleGAN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2020).
10. T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, T. Aila. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 4401-4410. IEEE, Los Angeles, CA (2019).
11. A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, A. Lerer. Automatic differentiation in PyTorch. (2017).
12. D. P. Kingma, J. Ba. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014).
13. Y. Choi, Y. Uh, J. Yoo, J. Ha. StarGAN v2: Diverse image synthesis for multiple domains. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2020).
14. F. Yu, A. Seff, Y. Zhang, S. Song, T. Funkhouser, J. Xiao. LSUN: Construction of a large-scale image dataset using deep learning with humans in the loop. arXiv preprint arXiv:1506.03365 (2015). DOI: 10.48550/arXiv.1506.03365
15. M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In 6th International Conference on Learning Representations, ICLR 2017, Toulon, France (2017).
16. Y. Endo. User-controllable latent transformer for StyleGAN image layout editing. Comput. Graph. Forum 41(7), 395–406 (2022).
17. K. Zhou, Z. Liu, Y. Qiao, T. Xiang, C. C. Loy. Domain generalization: A survey. IEEE Trans. Pattern Anal. Mach. Intell. (2022).