Drag Your GAN with Adaptive GAN

Liav Eliyahu
Department of Electrical Engineering
Tel Aviv University
liav.eliyahu@gmail.com

Inbal Cohen
Department of Electrical Engineering
Tel Aviv University
inbalcohen282@gmail.com

Abstract

In 2023, a groundbreaking paper titled "Drag Your GAN: Interactive Point-based Manipulation on the Generative Image Manifold" [1] was published, showcasing advanced capabilities in image manipulation using Generative Adversarial Networks (GANs). The paper introduced a novel technology that allows for point-based manipulation of images on the generative manifold, utilizing reference points on the image to achieve precise and realistic results. This innovation opened new avenues in image processing and content creation.

However, this technology also brought challenges. The editing quality of images through point manipulation was highly dependent on the diversity of the training data. When attempts were made to create human poses that deviated from the training distribution, the results often included distortions and artifacts.

In our project, we aim to investigate and improve the performance of this technique, focusing on preserving object characteristics in images for more robust and realistic image manipulation. To address these challenges, we propose several enhancements to the model: training the model on each image multiple times before manipulation to better preserve the main characteristics of the object and incorporating a method similar to Adaptive GAN [2]. This involves quickly training the model on a specific image for a low number of iterations and at a significantly lower learning rate before manipulation. This allows the GAN to "remember" the key features of the image, enhancing performance.
$https://github.com/liaveliyahu/DragGAN\_with\_Adaptive\_GAN$

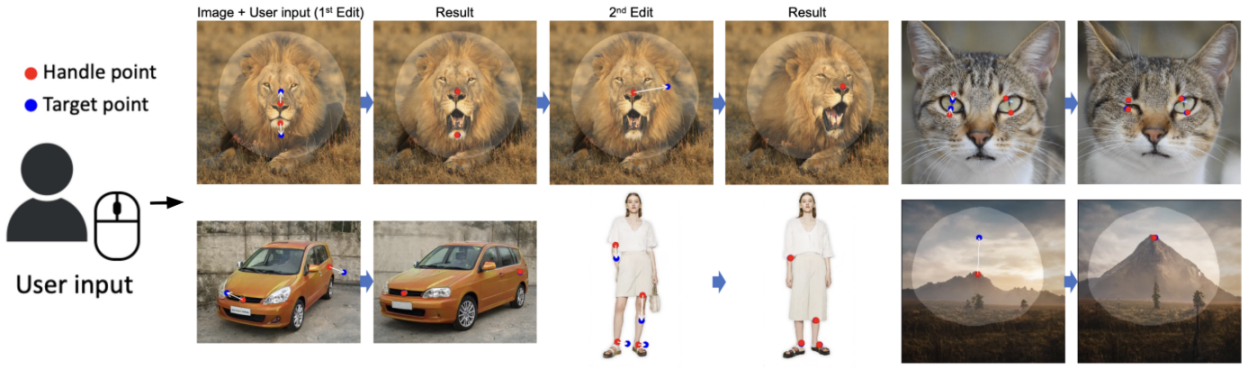

Figure 1: Demonstration of DragGAN interface. Image from the original paper.

1 Introduction

In this paper, we address the problem of preserving object characteristics in images during manipulation using Generative Adversarial Networks (GANs). Specifically, we investigate and aim to enhance the "Drag Your GAN: Interactive

Point-based Manipulation on the Generative Image Manifold" [1] approach. The current methodology struggles with maintaining realism and consistency, particularly when manipulating images into poses or configurations that deviate from the training data distribution, leading to noticeable artifacts. This challenge is significant as it impacts the quality and usability of image manipulation tools in various applications, from creative industries to medical imaging. Our approach involves training the GAN model on each image multiple times before manipulation, which helps in better preserving the main characteristics of the objects. Additionally, we integrate an Adaptive GAN [2] technique, where the model is quickly trained on a specific image for a low number of iterations and with a significantly lower learning rate before manipulation. This strategy allows the GAN to "remember" key features of the image, thus enhancing its performance. Our results demonstrate improved preservation of object characteristics, with the enhanced model showing more robust and realistic image manipulations. However, these improvements are relatively minor, and potential smearing in images indicates the need for further hyperparameter adjustments and additional research. Despite these limitations, our work contributes to advancing the field of image manipulation by highlighting areas for improvement and offering a foundation for more reliable and visually appealing techniques.

Figure 2: Demonstration of the problem using the DragGAN tool. When moving the woman's hand, it changes the appearance of her. The woman got brighter hair and lighter jeans compared to the original image.

2 Related Work

The field of image manipulation using Generative Adversarial Networks (GANs) has seen significant advancements in recent years. One notable approach is the StyleGAN model, introduced by Karras et al [3], which allows for high-quality image synthesis and manipulation by controlling style at various levels of the network. StyleGAN has been widely adopted for tasks such as face editing and style transfer due to its ability to generate photorealistic images. However, it often struggles with maintaining consistency when applied to images outside the training distribution, leading to artifacts and unrealistic results.

The "Drag Your GAN" technique by Pan et al [1]. introduces a point-based manipulation method on the generative image manifold, enabling users to interactively adjust specific points in an image. This approach provides a more intuitive and user-friendly way to manipulate images compared to traditional methods. It also incorporates motion supervision and point tracking, allowing for more natural and controlled deformations in the manipulated images.

However, it is still limited by the diversity of the training data, which can result in artifacts when generating poses or configurations not seen during training.

In the "Drag Your GAN" technique, two key components: Feature-based Motion Supervision and Point Tracking, play crucial roles in enhancing the realism and control of image manipulations.

2.1 Motion Supervision

Let us denote the handle points as $\{p_i\}$ and the target points as $\{t_i\}$, with $i \in [n]$ if we have a total of n (handle, target) pairs. Let us further denote the points inside the circle centered at p_i with radius r_1 as $\Omega_1(p_i, r_1)$. Motion supervision is an iterative optimization process that manipulates the feature map F such that the region around each p_i is migrated to the region around t_i . Probably to enforce smoothness and continuity on the feature map, instead of a brutal cut-and-paste or copy-and-paste from $\Omega_1(p_i, r_1)$ to $\Omega_1(t_i, r_1)$, the DragGAN authors decided to use a gentler method. They first find the unit vector pointing from p_i to t_i:

$$d_i = \frac{t_i - p_i}{||t_i - pi||_2} \tag{1}$$

then the following loss is defined to guide the iterative update on the feature map:

$$\mathcal{L} = \sum_{i=0}^{n} \sum_{q_i \in \Omega_1(p_i, r_1)} ||F(q_i) - F(q_i + d_i)||_1 + \lambda||(F - F_0)(1 - M)||_1 \tag{2}$$

2.2 Point Tracking

The handle points are “updated” at each optimization step. More precisely, after each optimization iteration that manipulates the feature map, we need to re-localize the handle points in order to proceed to the next iteration. This re-localization job is performed by point tracking. The DragGAN authors decided to track the points in the feature space. Let us denote the points inside the square patch centered at p_i with side length of $2r_2$ as $\Omega_2(p_i, r_2)$. Let us additionally denote p_i at the k-th iteration as p_i^k, and the feature map at the k-th iteration as F_k. We can define the following rule for point tracking:

$$p_i^{k+1} = \underset{q_i \in \Omega_2(p_i, r_2)}{\operatorname{argmin}} ||F_{k+1}(q_i) - F_0(p_i^0)|| \tag{3}$$

This is a simple nearest-neighbor search over the square patch that looks for the point which is most similar to the initial feature map at the initial handle point. While we could potentially replace $F_0(p_i^0)$ with $F_k(p_i^k)$, we believe that the authors of DragGAN used the current formulation to mitigate drift error.

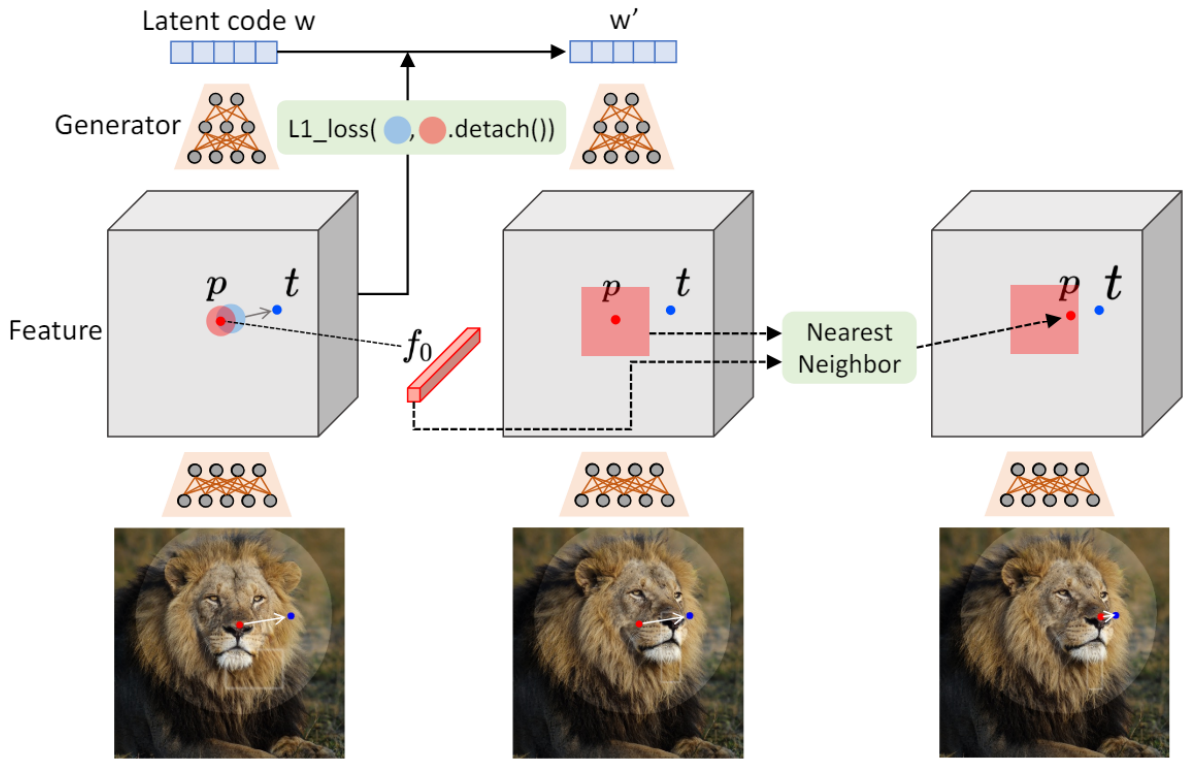

Figure 3: DragGAN method where motion supervision is achieved via a shifted patch loss on the feature maps of the generator. They perform point tracking on the same feature space via the nearest neighbor search. Taken from original paper.

Our approach builds on these foundations by incorporating strategies aimed at enhancing the preservation of object characteristics during manipulation. We utilize techniques from the work on Adaptive GAN [2], which involve quickly training the model on a specific image with a low learning rate. This allows the network to "remember" key features and produce more consistent results. While this method has shown promise in improving the model's ability to handle variations in input data, it often requires careful tuning of hyperparameters. Additionally, we incorporated methods from Shady Abu Hussein's work on image-adaptive GANs. In this approach, the GAN is fine-tuned specifically on the input image before applying manipulation techniques. This fine-tuning process involves training the GAN on the target image for a low number of iterations with a significantly reduced learning rate. The goal is to enhance the model's ability to maintain consistency and preserve object characteristics even when the image undergoes significant manipulations. By combining this image-adaptive fine-tuning with the "Drag Your GAN" approach, we aim to improve the robustness and realism of the manipulation results. Our results show that the combined approach of multiple training iterations and image-adaptive GAN fine-tuning leads to improved preservation of object characteristics, with more robust and realistic manipulations. However, these improvements are relatively minor, and potential smearing in images indicates the need for further hyperparameter adjustments and additional research. In summary, our approach leverages and extends existing techniques in GAN-based image manipulation by integrating methods from both Adaptive GANs and image-adaptive GANs. While we have made some progress in enhancing the consistency and realism of manipulated images, our findings suggest that further research is necessary to fully address the challenges posed by diverse training data and complex image manipulations. Future work could focus on refining hyperparameter tuning and developing more sophisticated methods for retaining image features to achieve more significant improvements.

3 Data

In our project, we worked with several datasets generated by different StyleGAN2 models [3]. These datasets consist of high-quality images synthesized for various objects and categories, including animals and human images. The datasets we used are as follows:

- StyleGAN2-AFHQCat-512x512: This dataset contains images of cats at a resolution of 512x512 pixels, generated using the StyleGAN2 architecture trained on the AFHQCat dataset. The AFHQCat dataset comes from the "AFHQ" (Animal Faces-HQ) dataset, which was collected and curated for high-quality animal face images.
- StyleGAN2-Car-Config-F: This dataset comprises images of cars generated by the StyleGAN2 model, configured with the "F" setup for detailed car synthesis. The original car images were collected from various sources to create a comprehensive dataset for training the model.
- StyleGAN2-FFHQ-512x512: This dataset includes high-resolution images of human faces at 512x512 pixels, created using the StyleGAN2 model trained on the Flickr-Faces-HQ (FFHQ) dataset. The FFHQ dataset consists of high-quality images of human faces collected from Flickr and other public sources.
- StyleGAN2_Dogs_1024_PyTorch: This dataset features images of dogs at a resolution of 1024x1024 pixels, synthesized by the StyleGAN2 model implemented in PyTorch. The original dog images were sourced from various public datasets and curated for high quality.
- StyleGAN2_Elephants_512_PyTorch: This dataset consists of elephant images at 512x512 pixels, generated using the StyleGAN2 model implemented in PyTorch. The elephant images were collected from public image repositories and curated to ensure diversity and quality.
- StyleGAN2_Horses_256_PyTorch: This dataset includes images of horses at 256x256 pixels, created using the StyleGAN2 model implemented in PyTorch. The horse images were collected from public sources and curated for quality and diversity.
- StyleGAN2_Lions_512_PyTorch: This dataset features lion images at a resolution of 512x512 pixels, synthesized by the StyleGAN2 model implemented in PyTorch. The lion images were collected from various public image sources and curated to ensure high quality and diversity.

The data we worked with is synthetic and generated by StyleGAN2 models trained on various image datasets. The images span multiple categories and resolutions, providing a diverse set of data for our project. Both the original authors of the "Drag Your GAN" paper and we load pre-trained models for our experiments. Additionally, we train the models further using the mentioned datasets to enhance their performance. We did not have to perform any preprocessing, filtering, or other special treatment to use this data in our project. The images generated by the StyleGAN2 models are already in a suitable format for our experiments and analyses. This allows us to directly focus on enhancing the "Drag Your GAN" technique and evaluating its performance in preserving object characteristics during image manipulation.

4 Methods

In this section, we elaborate on our approach for preserving object characteristics during image manipulation using Generative Adversarial Networks (GANs), building upon the "Drag Your GAN" technique and integrating the Adaptive GAN method proposed by Shady Abu Hussein. Below, we detail the process as depicted in the included diagram.

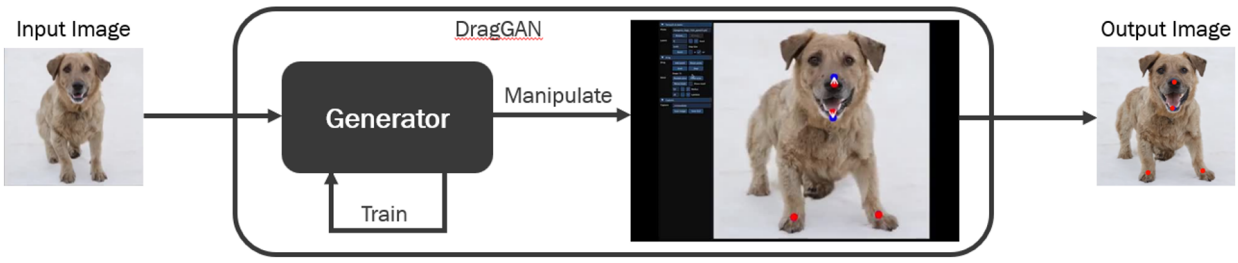

Figure 4: Diagram of our proposed method. Pick an input image, train the generator on the input image, continue to manipulate the image using the DragGAN tool and finally get an output image.

4.1 Input Image Processing

The workflow begins with an input image, which undergoes preprocessing to align it for manipulation.

4.2 DragGAN Generator Training

The core of our approach involves training the DragGAN generator. This training is tailored to each specific image, enhancing the generator's ability to adapt to unique image characteristics.

4.3 Manipulation Using DragGAN

Post-training, the image is manipulated using the trained DragGAN model. This step includes interactive point-based manipulations that allow for precise control over the image's attributes.

4.4 Adaptive GAN Integration

A significant enhancement in our methodology is the integration of the Adaptive GAN approach. Before manipulation, the model is quickly trained on the specific image for a few iterations at a low learning rate, which allows the GAN to "remember" the main features of the image, thereby significantly improving the manipulation quality.

4.5 Output Image Generation

The final step in the process is the generation of the output image, which reflects the manipulations applied, with a focus on maintaining the integrity and realism of the original object characteristics.

Comparison with Alternative Approaches: We also compared our method with traditional GANs and style transfer techniques, which were less effective in preserving object characteristics, especially for images that deviated from the training distribution. Our approach, leveraging the Adaptive GAN method, demonstrated superior performance in terms of preserving object characteristics and producing high-quality manipulated images.

Experimental Validation: The experimental setup involved manipulating several images, including animals and vehicles, to assess the effectiveness of our enhanced DragGAN model against traditional methods. The results, illustrated in the accompanying figures, show our method's ability to maintain object integrity and realism more effectively than existing techniques.

Unsuccessful Proposed Method: We attempted to incorporate self-attention [4] into our approach to enhance the generator's ability to focus on relevant image features during manipulation, potentially improving the preservation of object characteristics. Self-attention mechanisms have shown promising results in various computer vision tasks, particularly in capturing long-range dependencies and improving model performance. However, despite the potential benefits, the integration of self-attention in our GAN model did not yield the expected results. In some cases, it led to the degradation of images, possibly due to the complexity introduced by the additional attention mechanism, which may have interfered with the GAN's learning process. As a result, we decided to exclude self-attention from our final approach and focus on other techniques, such as the Adaptive GAN method, which demonstrated superior performance in preserving object characteristics and producing high-quality manipulated images.

5 Experiments

In our experimental analysis, we evaluated our image manipulation method against the original DragGAN approach using several examples. The focus of our experiments is on the ability to preserve the original characteristics of various subjects during manipulation, with both qualitative and quantitative assessments.

Preservation of Original Characteristics Our method demonstrates a slightly better ability to retain the inherent characteristics of subjects during image manipulation. For instance, when adjusting facial expressions, our approach maintains the nuances of facial features, such as the shape of the eyes and the curve of the mouth, more effectively than the original DragGAN. This subtle improvement is particularly evident in complex images where minor details are crucial.

Figure 5: Results of a few samples from different datasets. With each sample we can see that the characteristics of the objects are more similar with our method compared to the suggested method of DragGAN.

5.1 Qualitative Evaluation

Although the improvements are minor, our method consistently handles intricate details more effectively. In images of animals, for example, our approach preserves the texture of fur and the shape of ears better than the original DragGAN, which occasionally results in smoother but less accurate details. This consistency in handling fine details is a significant advantage in high-fidelity image editing tasks.

5.2 Quantitative Evaluation

The tables present the results of a quantitative evaluation comparing our image manipulation method to the DragGAN approach. We utilized VGG, a deep convolutional neural network, to extract features from the images for this evaluation. Two similarity metrics were employed: Cosine Similarity and Mean Squared Error (MSE). Table 1 displays the Cosine Similarity results, where higher values indicate greater similarity. Our method consistently shows higher similarity scores compared to DragGAN, indicating better preservation of the original image characteristics. For instance, the Red Car sample achieved a Cosine Similarity of 0.93 with our method, compared to 0.90 with DragGAN. Table 2 presents the MSE results, where lower values signify better performance. Our method also outperforms DragGAN in this metric, demonstrating lower MSE values across all samples. For example, the Red Car sample had an MSE of 0.0099 with our method, compared to 0.0158 with DragGAN. Overall, these results highlight that our method maintains more fidelity to the original images than DragGAN does.

Table 1: Cosine Similarity Results

Cosine Similarity using VGG extraction		
Sample	Ours	DragGAN
Red Car	0.93	0.90
Black Car	0.95	0.87
Lion	0.86	0.84
Woman	0.96	0.95

Table 2: MSE Similarity Results

MSE Similarity using VGG extraction		
Sample	Ours	DragGAN
Red Car	0.0099	0.0158
Black Car	0.0061	0.0160
Lion	0.0295	0.0330
Woman	0.0075	0.0078

6 Conclusion

Our experimental analysis has shown that training the model on the image helps to better preserve the main characteristics of the object, leading to more faithful manipulations. While the improvements we achieved are notable, it's important to highlight that they are still considered minor. This suggests that there is room for further enhancement of our method.

One interesting observation from our experiments is that adjusting the weights of the generator can have a significant impact on the quality of the resulting images. We found that improper weighting can lead to artifacts or distortions, highlighting the importance of carefully tuning these parameters. Moreover, our experiments revealed the need for fine-tuning the training hyperparameters per class. Different classes of objects may require different settings for optimal performance, and this fine-tuning process could further improve the fidelity of our manipulations.

Looking ahead, there are several avenues for future research and application of our ideas. One direction could be to explore other types of image manipulation tasks, such as image inpainting or style transfer, where our method could be adapted and potentially yield even more impressive results. Additionally, further refinement of our approach, perhaps by incorporating more advanced deep learning techniques or by leveraging larger datasets, could lead to significant improvements in image manipulation fidelity.

References

[1] Xingang Pan. Drag your gan: Interactive point-based manipulation on the generative image manifold. *arXiv:2305.10973, 18 May 2023. http://arxiv.org/abs/2305.10973*.

[2] Shady Abu-Hussein, Tom Tirer, and Raja Giryes. Image-adaptive gan based reconstruction. *arXiv:2305.10973, 18 Nov 2019. https://arxiv.org/abs/1906.05284*.

[3] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. *arXiv:1912.04958, arXiv, 23 Mar. 2020. https://arxiv.org/abs/1912.04958*.

[4] Han Zhang, Ian Goodfellow, Dimitris Metaxas, and Augustus Odena. Self-attention generative adversarial networks. *arXiv:1805.08318, 14 June 2019. http://arxiv.org/abs/1805.08318*.

$https://github.com/liaveliyahu/DragGAN\_with\_Adaptive\_GAN$

$https://github.com/XingangPan/DragGAN$

$https://github.com/shadyabh/IAGAN$