

GANs for All: Supporting Fun and Intuitive Exploration of GAN Latent Spaces

Wei Jiang

Richard Lee Davis

Kevin Gonyop Kim

Pierre Dillenbourg

*École Polytechnique Fédérale de Lausanne
Lausanne, Switzerland*

WEI.JIANG@EPFL.CH

RICHARD.DAVIS@EPFL.CH

KEVIN.KIM@EPFL.CH

PIERRE.DILLENBOURG@EPFL.CH

Editor: Douwe Kiela, Marco Ciccone, Barbara Caputo

Abstract

We have developed a new tool¹ that makes it possible for people with zero programming experience to intentionally and meaningfully explore the latent space of a GAN. We combine a number of methods from the literature into a single system that includes multiple functionalities: uploading and locating images in the latent space, image generation with text, visual style mixing, and intentional and intuitive latent space exploration. This tool was developed to provide a means for designers to explore the "design space" of their domains. Our goal was to create a system to support novices in gaining a more complete, expert understanding of their domain's design space by lowering the barrier of entry to using deep generative models in creative practice.

Keywords: GAN inversion, latent space exploration, latent space disentanglement, design space exploration, human centered artificial intelligence

1. Introduction

For design professionals, one of the key markers of expertise is a coherent understanding of the design space in which their work is situated. Novices lack this understanding, and as a result, they are likely to suffer from "design fixation." One way of supporting novices in building this expertise is to provide them with tools that allow them to explore the design space of their domains. Existing tools utilize methods such as merging designs using genetic algorithms and producing infinite variations of designs using parametric modeling. However, these tools lack a representation of the complete design space. This means a user can not freely explore the full design space, but is limited to the regions around or between examples.

Recently, deep generative models such as GANs (Radford et al., 2015) and VAEs (Kingma and Welling, 2014) have demonstrated the ability to encode large design spaces in their latent spaces. However, due to the high-dimensionality of these latent spaces, they are difficult to meaningfully explore. These spaces contain hundreds or thousands of dimensions which are typically entangled. This means interpolating a sample along any individual dimension will tend to change many of the characteristics of that image. Furthermore, while

1. A demonstration video can be viewed at <https://www.youtube.com/watch?v=dcC7G2zBuL8&t=42s>. A live demonstration can also be found at <https://generative.fashion>.

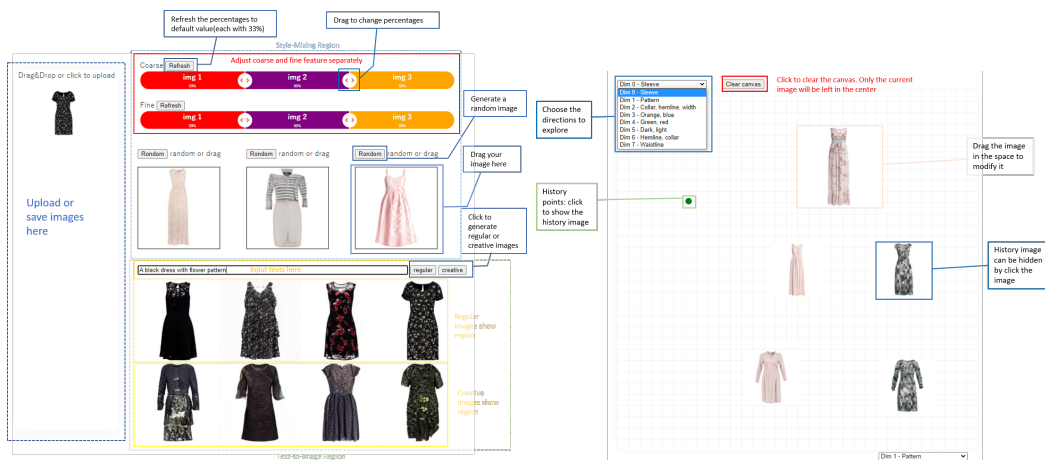


Figure 1: The interface of our neural design space exploration tool. Users can *upload* images in the workplace on the left or generate random image through *random* button. Also, they can generate examples via text descriptions using the *text box*. Users can *drag* these examples to the style-mixing region or save them in the workplace. Users can selectively combine elements from three designs using the visual style-mixing panel. The output image is shown in the center of the canvas on the right. The 2D-dimensional canvas represents the design space for two attributes in the horizontal and vertical axes, and these attributes can be changed by using a *drop-down* menu for each axis. *Dragging* the image within the canvas is equivalent to moving through the latent space of the GAN in semantically meaningful directions.

generating random samples is trivial, intentionally locating a starting point within the latent space based on a target image is not. Our tool provides technical solutions to these problems, and packages them in a web-based graphical interface that is designed for those with no programming experience (Figure 1). To the best of our knowledge, this makes our work the first example of a neural design-space exploration tool.

Our starting point was the StyleGAN2-ada model (Karras et al., 2020) trained on the Feidegger dataset (Lefakis et al., 2018). The next sections detail how we extended the model to achieve several features related to the search and exploration of the latent space.

2. Locating the Starting Point for Design Space Exploration

2.1. Searching the Latent Space with Text Prompts

It is important that a user is able to easily choose the starting point for their design space exploration. One intuitive way of achieving this is to describe a design and locate closely-matching images embedded in the latent space. We provided this capability in the form of a text box where users could write a short description of a design. We implemented two methods to locate the design. The first method was to randomly sample images from the latent space, then to pass these along with the text description through a CLIP (Radford

et al., 2021) model to find a small number of images which most closely matched the text. The second method was to fine-tune a DALL-E model (Ramesh et al., 2021) on the Feidegger dataset, and then to pass the text descriptions to DALL-E and let it generate designs. Surprisingly, the first method produced more satisfying results than the second method, though the DALL-E results were often more unexpected (see Appendix A).

2.2. Searching the Latent Space with Image Prompts

Another intuitive method for locating starting points in the latent space is to upload an example image and find the closest point in the latent space. To provide this, We employed the method described in (Abdal et al., 2019). Given a starting initialization w , we search for an optimized vector w^* that minimizes the objective loss function that measures the difference between the given image I and the image $G(w)$ generated from w . The optimal w^* is expressed as

$$w^* = \min_w L(w) = \min_w \|f(G(w)) - f(I)\|_2^2 + \lambda_{pix} \|G(w) - I\|_2^2$$

where the loss is calculated as the weighted sum of *perceptual loss* (Johnson et al., 2016) and *pixel-level loss* that is pixel-by-pixel MSE loss between two images. This loss noticeably improved our tool’s ability to embed out-of-sample examples in the latent space of the GAN, which made it possible for users to upload their own images and use them as starting points for their design space exploration. Results are shown in Appendix B.

3. Latent Space Exploration

3.1. Style Mixing of Multiple Images

Once starting images have been generated or uploaded, they can be dragged and dropped into the visual style mixing interface. Sliders are provided which allow the user to mix and combine features from each of the designs into a single example using the style-mixing method described in Karras et al. (2019). The coarse slider controls shape, and the fine slider controls pattern and color. Style mixing with multiple images is achieved by taking different percentages from each image in the aspect of the coarse and fine features. The output image is shown in the center of the latent-space exploration panel on the right in Figure 1.

3.2. Intentional and Intuitive Latent Space Exploration

The latent-space exploration panel allows users to take a starting design and move it along meaningful directions in the latent space. Each axis of this two-dimensional canvas corresponds to a semantically-meaningful direction in the latent space, and the direction corresponding to each axis can be changed using a drop-down menu. Dragging an image within the canvas is equivalent to interpolating a point in the latent space along the directions selected in the drop-down boxes. Each newly-generated point is shown on the canvas as a history point. For a simplified representation of this interface see Figure 3.2. Behind the scenes, we used a method described in (Härkönen et al., 2020) to identify semantically-meaningful directions using PCA on the latent w space corresponding to sleeve length,

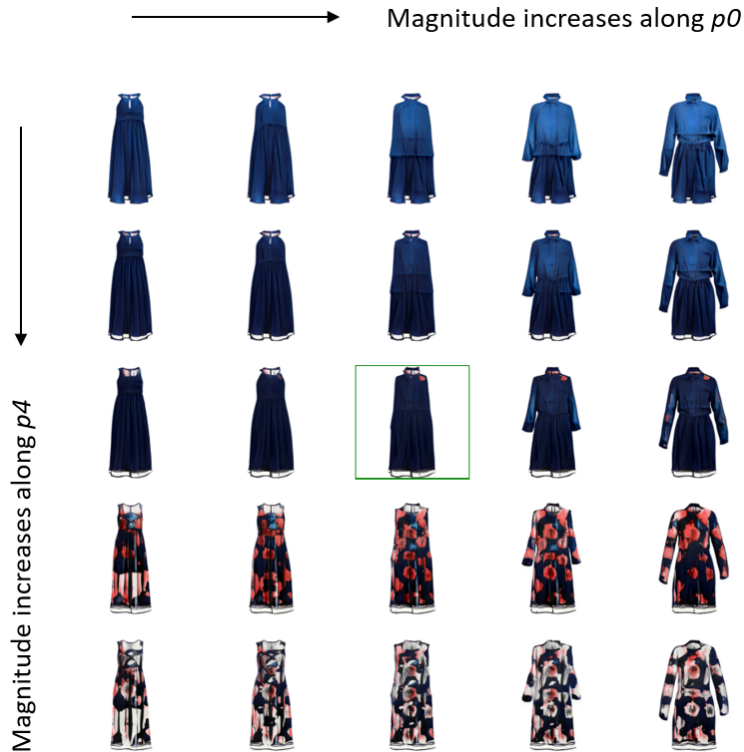


Figure 2: Examples of interpolating images simultaneously along two meaningful directions in the latent space (sleeve and pattern) found using PCA. The image in the green box shows the original image with 0 magnitude.

pattern, color, hemline, waistline, and more. For more information on this method and for more examples of the results of interpolating along the different principal components, see Appendix C.

4. Conclusion

Our work combines a number of methods from the literature into a single system. This combination provides a fun and intuitive way to meaningfully explore the latent space of a GAN. The power of our system is brought to the fore by the graphical interface, which is its primary novelty. Users can locate points in the latent space by writing text descriptions or uploading images. In the latter case, even uploaded images which are far out-of-sample can be located in the latent space. Users can then interpolate their starting images along semantically meaningful directions in the latent space using a drag-and-drop interface, providing an intuitive way to intentionally explore the latent space. This interface is what makes it possible for a novice designer with no programming experience to meaningfully explore the latent space of a GAN. To the degree that this latent space maps onto the design space of a domain, our system supports novices in expert-like exploration of the design space.

References

- Rameen Abdal, Yipeng Qin, and Peter Wonka. Image2stylegan: How to embed images into the stylegan latent space?, 2019.
- Erik Härkönen, Aaron Hertzmann, Jaakko Lehtinen, and Sylvain Paris. Ganspace: Discovering interpretable gan controls, 2020.
- Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution, 2016.
- Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019.
- Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training generative adversarial networks with limited data. *arXiv preprint arXiv:2006.06676*, 2020.
- Diederik P Kingma and Max Welling. Auto-encoding variational bayes, 2014.
- Leonidas Lefakis, Alan Akbik, and Roland Vollgraf. Feidegger: A multi-modal corpus of fashion images and descriptions in german. In *LREC 2018, 11th Language Resources and Evaluation Conference*, 2018.
- Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks, 2015.
- Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision, 2021.
- Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. Zero-shot text-to-image generation, 2021.