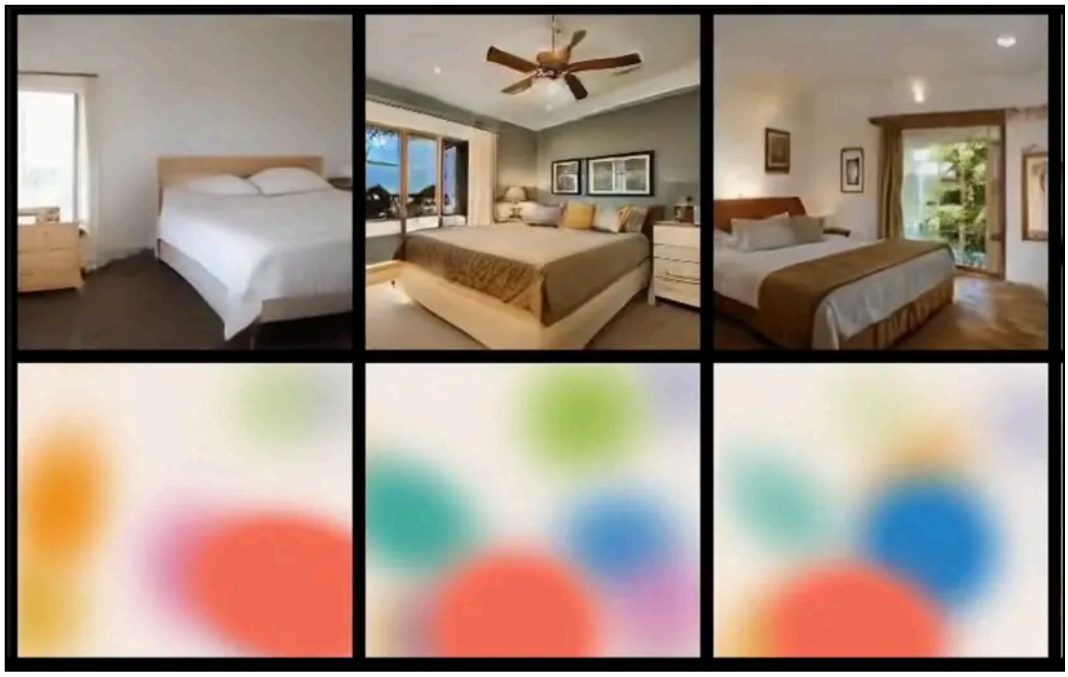
Editing a GAN's Latent Space With 'Blobs'

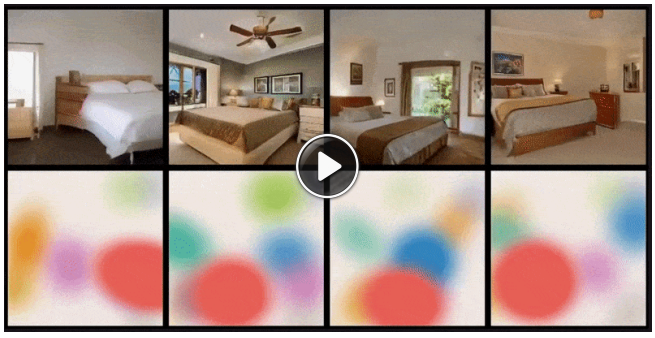
Originally published at https://www.unite.ai/editing-a-gans-latent-space-with-blobs/



New research from UC Berkeley and Adobe ADBE 0.82% offers a way to directly edit the hyperreal content that can be created by a Generative Adversarial Network (GAN), but which can't usually be controlled, animated, or freely manipulated in a manner long familiar to Photoshop users and CGI practitioners.

Titled *BlobGAN*, the method involves creating a grid of 'blobs' – mathematical constructs that map directly to content within the latent space of the GAN.

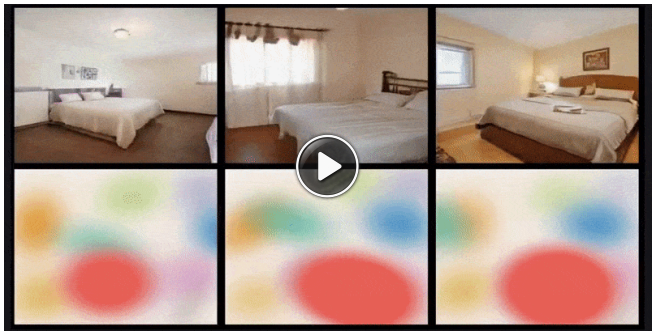
By moving the blobs, you can move the 'objects' in a scene representation, in an intuitive manner that's nearer to CGI and CAD methods than many of the current attempts to map and control the GAN's latent space:



CLICK TO VIEW ANIMATED GIF

Scene manipulation with BlobGAN: as the 'blobs' are moved by the user, the disposition of latent objects and styles in the GAN are correspondingly altered. For more examples, see the paper's accompanying video, embedded at the end of this article, or at https://www.youtube.com/watch?v=KpUv82VsU5k

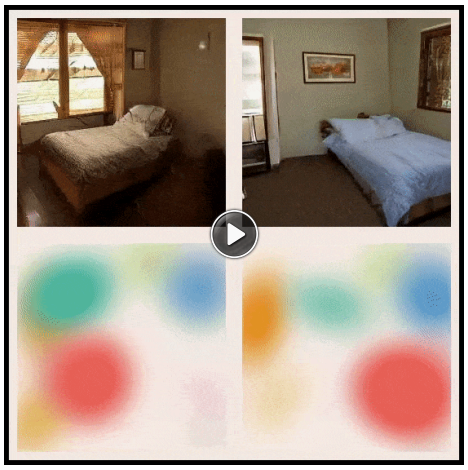
Since blobs correspond to 'objects' in the scene mapped out in the GAN's latent space, all the objects are disentangled *a priori*, making it possible to alter them individually:



CLICK TO VIEW ANIMATED GIF

Objects can be resized, shrunk, cloned, and removed, among other operations.

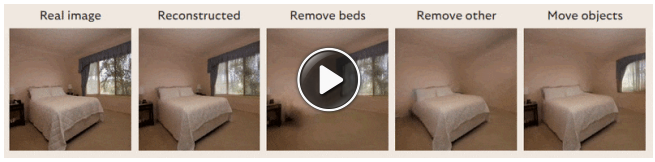
As with any object in photo editing (or even text-editing) software, a blob can be duplicated and subsequently manipulated:



CLICK TO VIEW ANIMATED GIF

Blobs can be duplicated in the interface, and their corresponding latent representations will also be 'copied and pasted'. Source: https://dave.ml/blobgan/#results

BlobGAN can also parse novel, user-selected images into its latent space:



CLICK TO VIEW ANIMATED GIF

With BlobGAN, you don't have to incorporate images that you wish to manipulate directly into the training data and then hunt out their latent codes, but can input select images at will and manipulate them. The photos being altered here are post-facto user input. Source: https://dave.ml/blobgan/#results

More results can be seen here, and in the accompanying YouTube video (embedded at the end of this article). There is also an interactive Colab demo*, and a GitHub repo**.

This kind of instrumentality and scope may seem naïve in the post-Photoshop age, and parametric software packages such as Cinema4D and Blender have been allowing users to create and customize 3D worlds for decades; but it represents a promising approach to taming the eccentricities and the arcane nature of the latent space in a Generative Adversarial Network, by the use of proxy entities that are mapped to latent codes.

The authors assert:

> *'On a challenging multi-category dataset of indoor scenes, BlobGAN outperforms Style-GAN2 in image quality as measured by FID.'*

The paper is titled *BlobGAN: Spatially Disentangled Scene Representations*, and is written by two researchers from UC Berkeley, together with three from Adobe Research.

Middle-man

BlobGAN brings a new paradigm to GAN image synthesis. Prior approaches to addressing discrete entities in the latent space, the new paper points out, have either been 'top-down' or 'bottom up'.

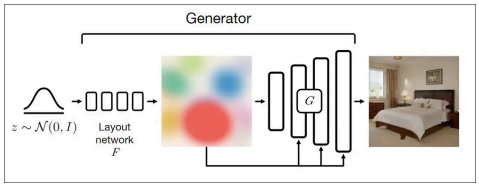
A top-down method in a GAN or image classifier treats images of scenes as classes, such as ‘bedroom’, ‘church’, ‘face’, etc. This kind of text/image pairing powers a new generation of multimodal image synthesis frameworks, such as the recent DALL-E 2 from OpenAI.

Bottom-up approaches, instead, map each pixel in an image into a class, label, or category. Such approaches use diverse techniques, though semantic segmentation is a popular current research strand.

The authors comment:

> *‘Both paths seem unsatisfactory because neither can provide easy ways of reasoning about parts of the scene as entities. The scene parts are either baked into a single entangled latent vector (top-down), or need to be grouped together from individual pixel labels (bottom-up).’*

Rather, BlobGAN proffers an *unsupervised mid-level representation*, or proxy framework for generative models.



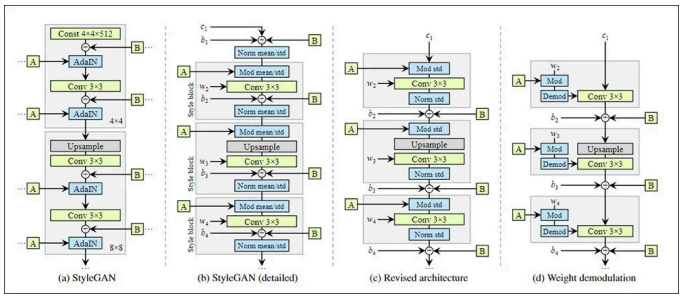
The layout network maps local (and controllable) ‘blob’ entities to latent codes. The colored circles in the center comprise a ‘blob map’. Source: https://arxiv.org/pdf/2205.02837.pdf

The Gaussian (i.e. noise-based) blobs are depth-ordered, and represent a bottleneck in the architecture that assigns a mapping to each entity, solving the biggest hurdle there is to GAN content manipulation: disentanglement (also a problem for autoencoder-based architectures). The resulting ‘blob map’ is used to manipulate BlobGAN’s decoder.

The authors note with some surprise that the system learns to decompose scenes into layouts and entities through an off-the-shelf discriminator which does not use explicit labels.

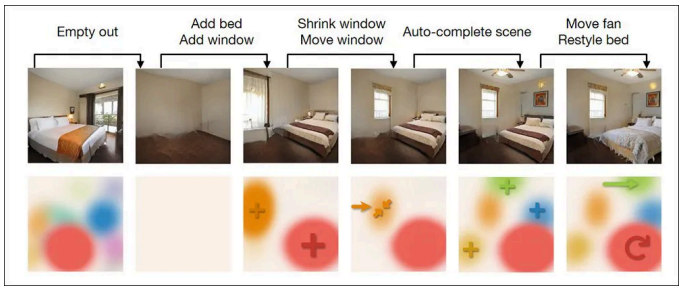
Architecture and Data

Entities in the blob map are converted into images via a revised StyleGAN2-derived network, in an approach that takes inspiration from prior NVIDIA NVDA -4.57% research.



A revised StyleGAN 2 derivative from NVIDIA Research. Some of the principles in this work were adopted or adapted for BlobGAN. Source: https://arxiv.org/pdf/1912.04958.pdf

StyleGAN 2 is modified in BlobGAN to accept input from the blob map instead of a single global vector, as is usually the case.



A series of manipulations made possible by BlobGAN, including the ‘autocompletion’ of an empty bedroom scene, and the resizing and relocation of the elements in the room. In the row below, we see the user-accessible instrumentality that enables this – the blob map.

By analogy, instead of bringing a vast and complex building (the latent space) into existence all at once, and then having to explore its endless byways, BlobGAN sends in the component blocks at the start, and always knows where they are. This disentanglement of content and location is the major innovation of the work.

BlobGAN: Spatially Disentangled Scene Representations

** Not functional at the time of writing*
*** Code not yet published at the time of writing*

First published 8th May 2022.

Writer on machine learning, domain specialist in human image synthesis. Former head of research content at Metaphysic.ai, until its dissolution into DNEG's Brahma.ai.
Portfolio site: martinanderson.ai
Contact: martin@martinanderson.ai

Discover More