

AI ART - LESSON 5 - APRIL 27, 2023

GETTING FAMILIAR WITH GAN

Benjamin Meadows | @TheBenMeadows





WHY AM I HERE?

- **MIDJOURNEY & STABLE DIFFUSION - NEWS**
- **HISTORY OF GENERATIVE ADVERSARIAL NETWORKS**
- **CURRENT GAN USE CASES & EXAMPLES**
 - Image Generation
 - Face Restoration
 - Colorization
 - Upscaling
 - Style Transfer

WHAT'S NEXT?

- **LESSON 1 - INTRODUCTION TO AI ART**
- **LESSON 2 - BASIC TOOLS FOR CREATING AI ART**
- **LESSON 3 - TOOLS & WORKFLOWS FOR AI ART (PART 1)**
- **LESSON 4 - TOOLS & WORKFLOWS FOR AI ART (PART 2)**
- **LESSON 5 - GETTING FAMILIAR WITH GAN**
- **LESSON 6+ - TO BE DETERMINED...**



MIDJOURNEY & STABLE DIFFUSION - NEWS

- NIJI VERSION 5 - 3 NEW STYLES
- MULTIPROMPT IMAGE BLENDING
- REAL-TIME DRAWING, V6/V7 & MORE
- DIFFUSIONBEE - ADDING CONTROLNET

New Niji V5 Styles

We added two new styles of NijiJourney (our anime themed Midjourney model)

To use them first enable Niji v5 under with `/settings` and then chose your style with `--style cute` or `--style expressive`

The cute style is quite cute! The expressive one is slightly more 3D and westernized. Feel free to share as always in `#niji-showcase`

#2 New NijiJourney Style

We added a new style to NijiJourney (our anime themed Midjourney model), this style is called 'scenic' and is meant for beautiful backgrounds and characters in beautiful environments.

To use it first enable Niji v5 under with `/settings` and then choose your style with `--style scenic`.

a lion looking out over the savannah --upbeta --q 2



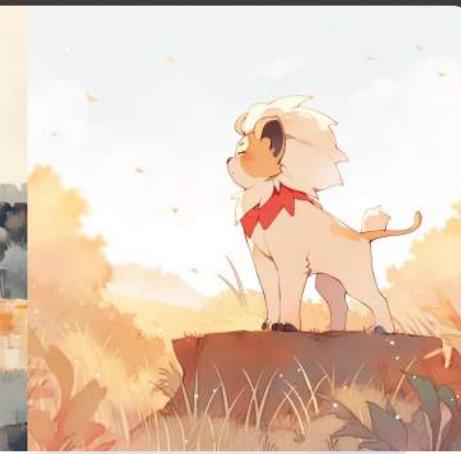
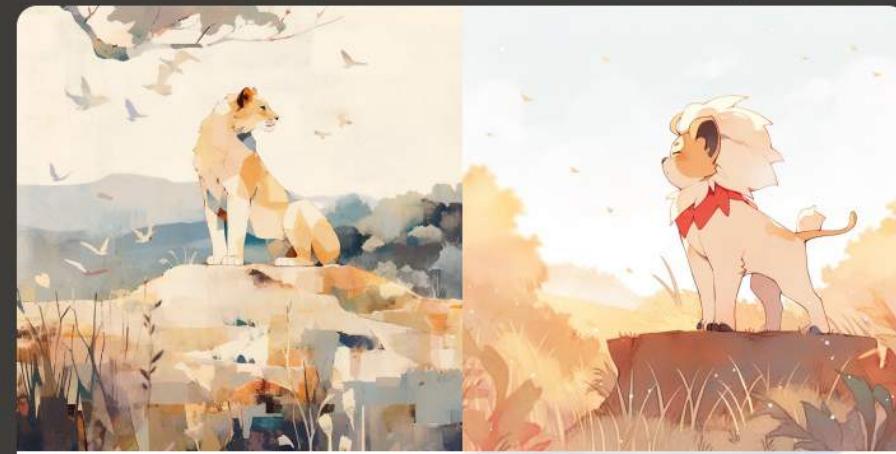
a lion looking out over the savannah --upbeta --q 2 --v 5



a lion looking out over the savannah --upbeta --q 2 --niji 5



a lion looking out over the savannah --style cute --upbeta --q 2 --niji 5



a lion looking out over the savannah --style expressive --upbeta --q 2 --niji 5



a lion looking out over the savannah --style scenic --upbeta --q 2 --niji 5



Christian Heidorn ✅
@ChrisHeidorn

...

🖼️ Introducing Multiprompt Image Blending

Many of you have come to know and love the /blend command in [#midjourney](#).

But it doesn't always do what you want and often contains malformed objects or artifacts.

Luckily there's a better way to blend now!

🧵👉 Let me explain...



7:03 AM · Apr 15, 2023 · 330.2K Views

I found these 2 images on Unsplash and wanted to blend them in Midjourney.

1. The first one shows an elegant hotel lobby.
2. The second one is a photo of the Royal Greenhouse in Laeken (Brussels)

And of course, Midjourney will oblige.





By now you should all be familiar with Midjourney's /describe command, right?

Well, have you ever considered using it on a blended image?

Cause that's exactly what I did.

I took one of blended images, fed it into /describe and it gave me these 4 starter prompts.

Christian Heidorn used /describe

Midjourney Bot ✅ BOT Today at 09:14

1 the lobby of a modern building has several glass ceilings, in the style of tranquil gardenscapes, neoclassical simplicity, 32k uhd, indonesian art, neoclassical influences, rounded, green and bronze --ar 139:78

2 an enclosed lobby area with ceiling and tree trunks, in the style of fine lines, delicate curves, green and bronze, tang yau hoong, impressionist gardens, rounded, uhd image, neoclassical symmetry --ar 139:78

3 the hotel lobby at the victoria resort is decorated with tropical plants, in the style of chen zhen, transparent/translucent medium, 32k uhd, santiago calatrava, classical revival, camille pissarro, classical architecture --ar 139:78

4 the hall of the japanese palace in hokkaido, in the style of contemporary glass, enigmatic tropics, 32k uhd, rounded, green and bronze, commercial imagery, cartelcore --ar 139:78



1 2 3 4 C

I decided to go with the very 1st option on the list.

And what it gave me was already much crisper, more varied, and generally more interesting than any of the blending results.

So is that already it? Far from it.

This is just a very quick and easy fix for messed up blends.



You start off by feeding the original 2 images (the hotel lobby and the greenhouse) into the /describe command.

This gives you 2 sets of 4 starter prompts each.

If you want, you can double-check each prompt first, but you don't have to.

So what now?

Christian Heldorn used /describe
Midjourney Bot ✓BOT Yesterday at 15:38

1 an elegant lobby with lighting and white marble carpet, in the style of wood veneer mosaics, asian-inspired, uhd image, khmer art, rendered in maya, light aquamarine and light amber, intricate ceiling designs --ar 247:145

2 a beautiful lobby with white ceiling and wooden floors, in the style of light gold and light azure, dansaekhwa, precisionist lines and shapes, layered veneer panels, batik, realistic depiction of light, elegant, emotive faces --ar 247:145

3 the entrance to a hotel with a ceiling which is painted gold, in the style of realistic rendering, precisionist, zen buddhism influence, layered veneer panels, *dao trong* ie, light beige and beige, stock photo --ar 247:145

4 lobby design ideas and design for contemporary lobby, decorating ideas, in the style of eastern-inspired motifs, rendered in cinema4d, light amber and sky-blue, natural materials, uhd image, elegant, emotive faces, high detailed --ar 247:145



1 2 3 4 ⏪

Christian Heldorn used /describe
Midjourney Bot ✓BOT Yesterday at 15:40

1 an enclosed glass building inside the budapest gardens, in the style of tropical landscapes, gray and emerald, california impressionism, exotic flora and fauna, rounded, national geographic photo, stock photo --ar 96:145

2 a beautiful circular structure with plants and trees, in the style of enigmatic tropics, viennese actionism, grandeur of scale, emerald, impressionist atmospheric, *edogawa rampo*, *clodo meireles* --ar 96:145

3 the glass roof of the sydney conservatory, in the style of *fernando botero*, tropical landscapes, gray and emerald, *gustave moreau*, national geographic photo, high quality photo, nature-inspired --ar 96:145

4 a building with green building in the courtyard is jus, in the style of exotic landscapes, hall of mirrors, impressionist gardens, national geographic photo, atmospheric environment, rounded, stock photo --ar 96:145



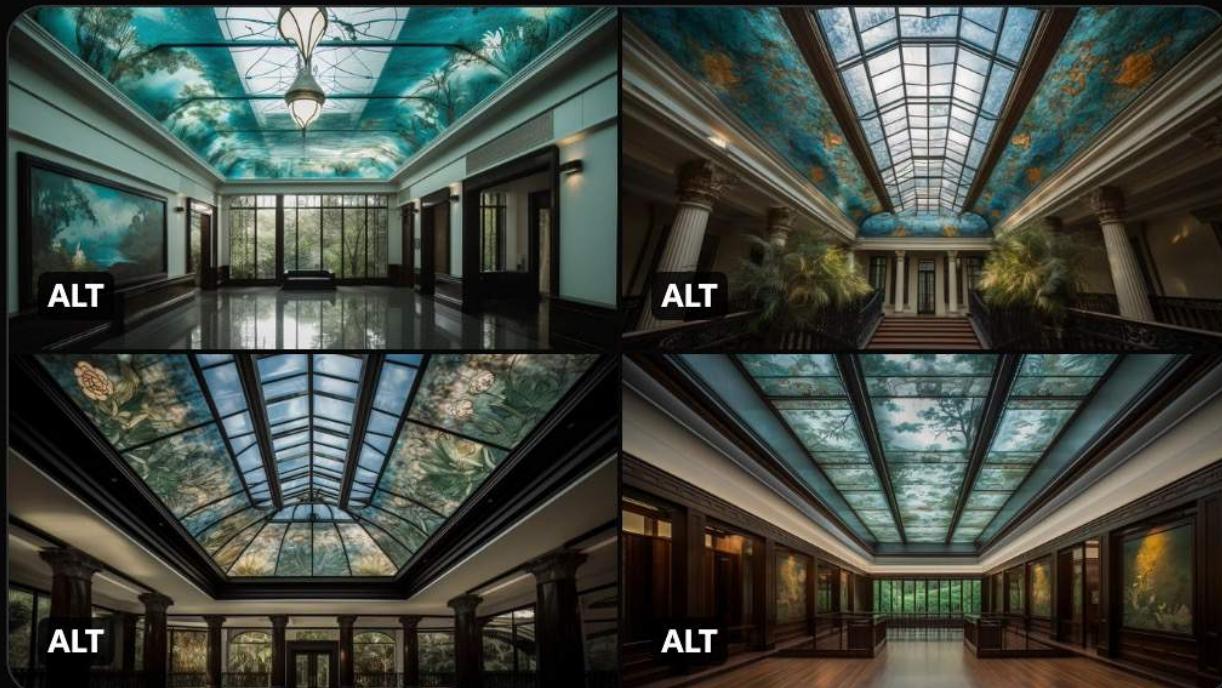
1 2 3 4 ⏪

Well, in this case, I decided to take No. 2 and No. 3 from each set.

I then combined them into a multiprompt with each starter prompt as its own segment.

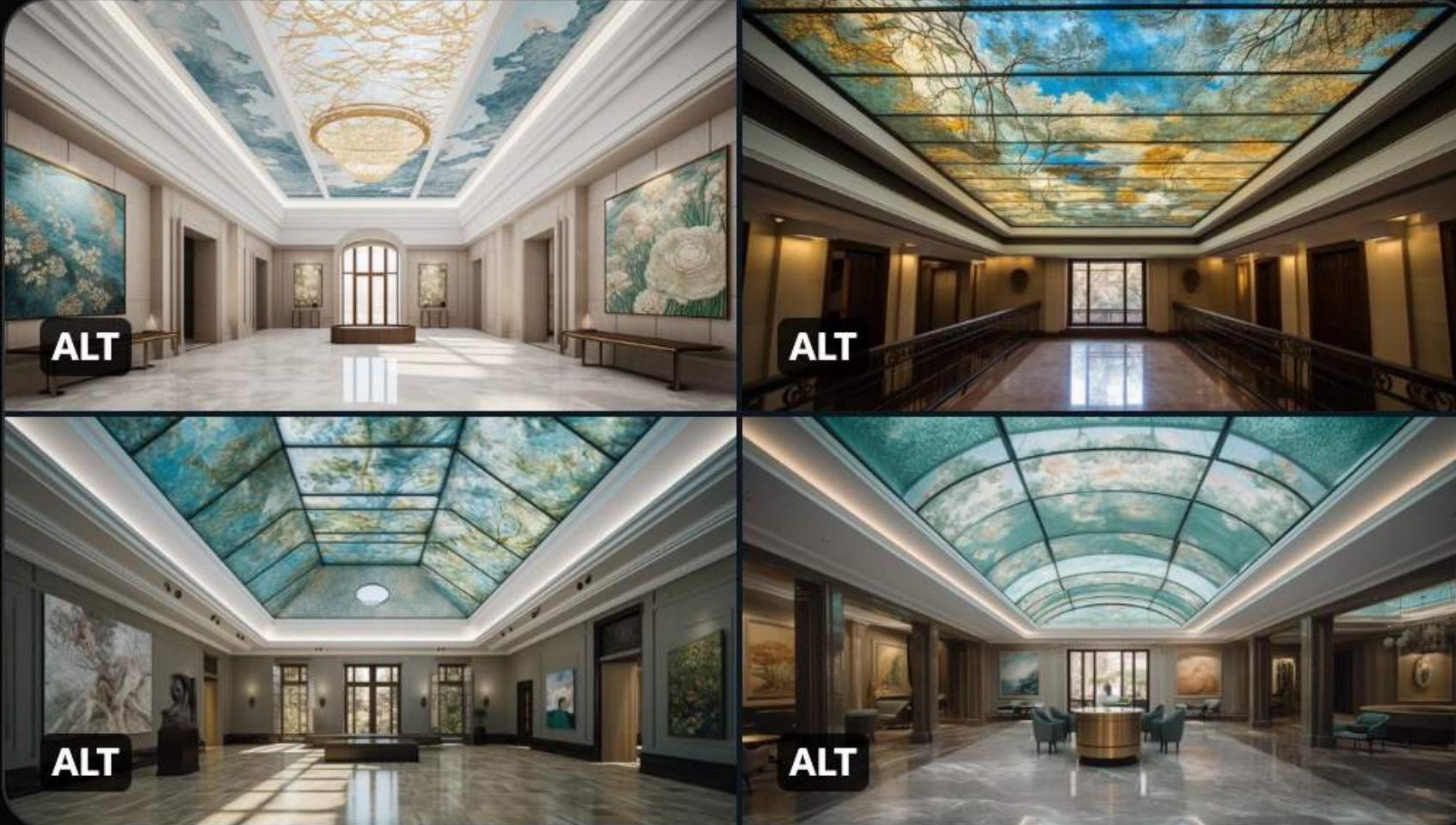
While this deviated a bit from the original blended concept, just look at these!

And best of all, we can adjust these 😊



By adjusting the weight of the first segment, we put more emphasis on the hotel lobby concept.

We get a more leveled camera angle and the greenhouse concept is less dominant.





Nick St. Pierre

@nickfloats

...

Midjourney's real-time drawing feature has only been mentioned a couple of times during office hours, but will be a game changer when it comes out.

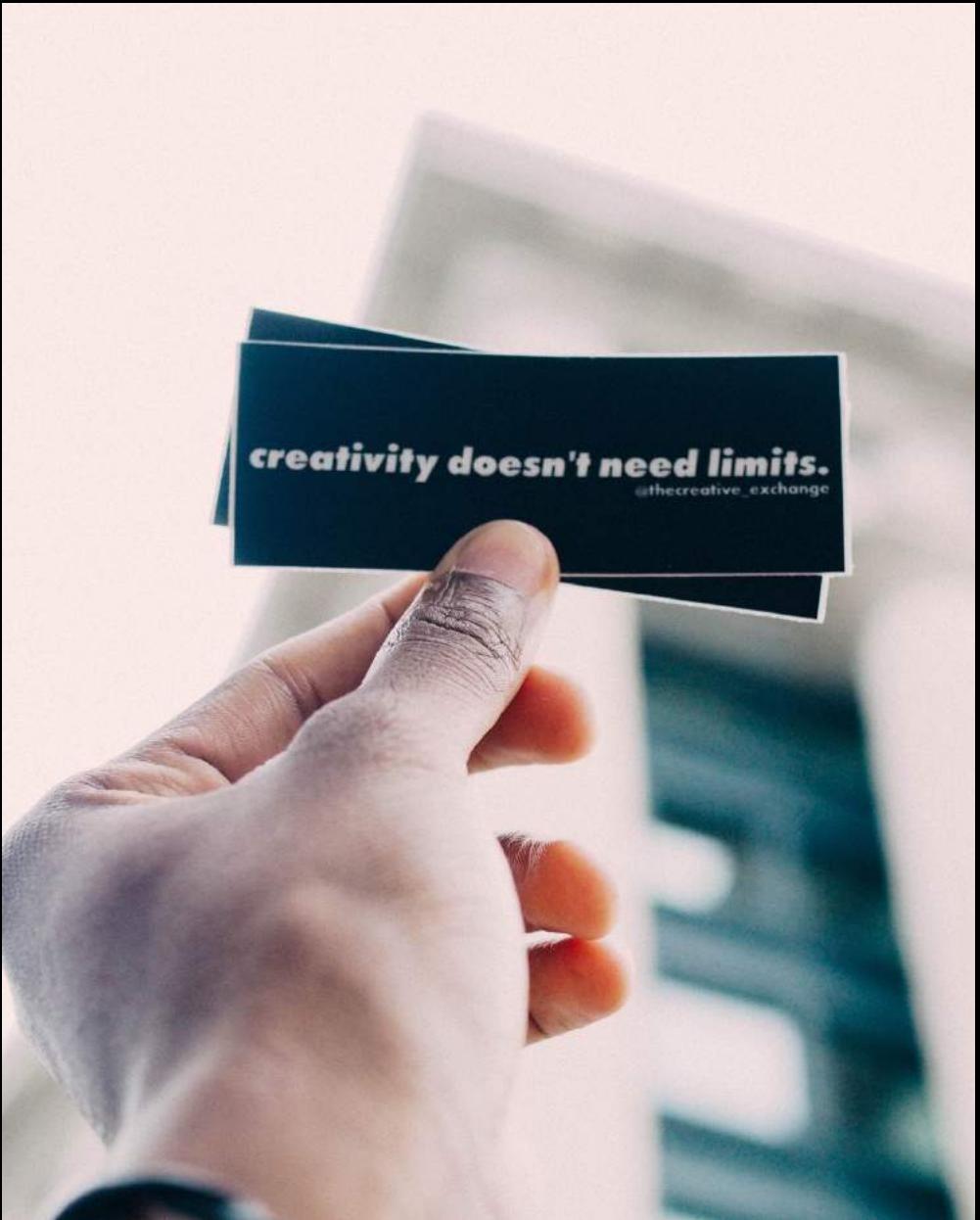
Imagine sketching on your iPad or Wacom tablet directly in midjourney, and watching your piece come to life in real-time.

Right now we have text and image prompts, with this feature your brushstrokes and sketches become prompts too. And combining them all together in midjourney is going to be revolutionary for many artists.

Nothing so far on how it will work, but I'd love to have the ability to prompt my canvas before I start drawing (with a general theme/style im looking for), layer control (plus automatic segmentation of layers, kinda like metas SAM), ability to prompt/reprompt layers (individual layers and multiple at once), ability to draw on top of images, and version history to name a few.

There's no timeline on when this will come out, but when it does it'll be big, and likely convince many more artists and illustrators to start using AI.

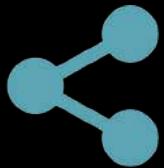
11:40 AM · Apr 19, 2023 · 489.3K Views



Overview of Generative Adversarial Networks (GANs)

GANs are composed of two neural networks, a generator and a discriminator, that work together to create new data.

How GANs Work and Train



GENERATIVE ADVERSARIAL NETWORKS (GANs) ARE A TYPE OF DEEP LEARNING ALGORITHM

GANs are composed of two neural networks, a generator and a discriminator, that compete against each other to generate new data



THE GENERATOR CREATES NEW DATA BASED ON A GIVEN INPUT

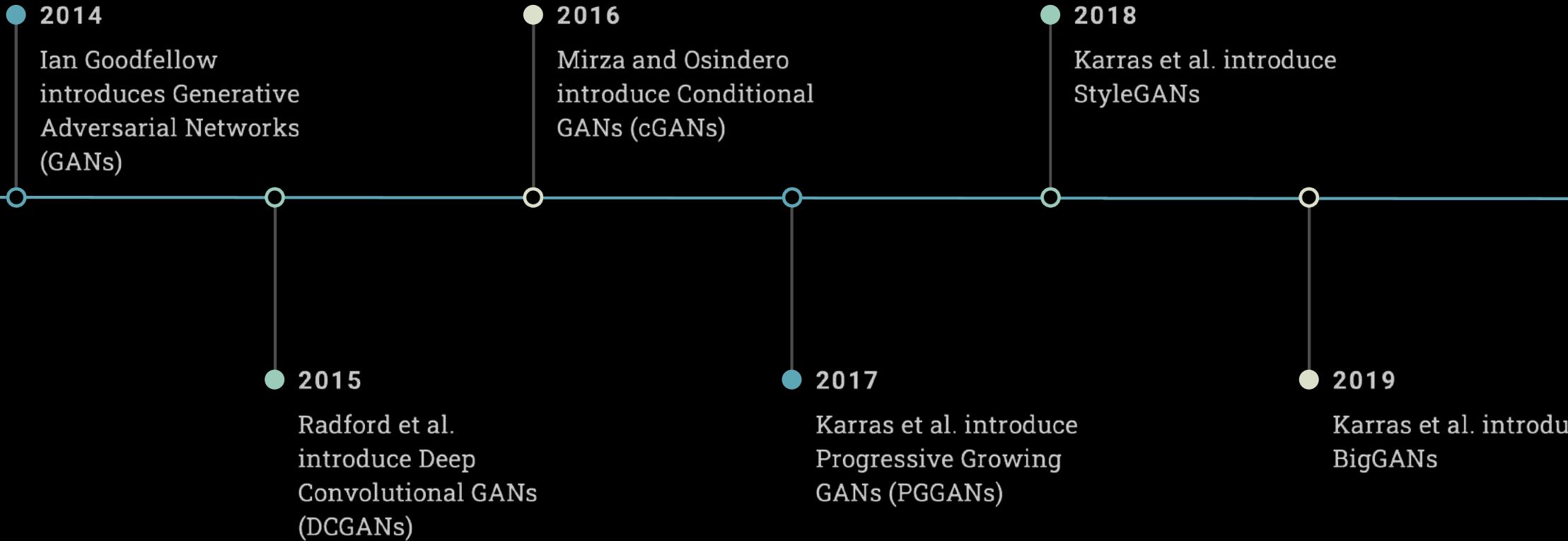
The generator is trained to create data that is indistinguishable from real data



THE DISCRIMINATOR EVALUATES THE GENERATED DATA

The discriminator is trained to distinguish between real and generated data

History and Timeline of GANs



General-Purpose GAN Models

- **STACKGAN**

Invented in 2017 by a group of Chinese researchers.

- **STYLEGAN**

Introduced in 2018 by Nvidia.

- **VQGAN+CLIP**

Introduced in 2020 (VQGAN) & 2021 (CLIP).

StackGAN

- **DETAILS**

Invented in 2017 by a group of Chinese researchers.

Type of PGGAN (Progressive Growing GAN).

StackGAN is a two-stage model. The first stage, called StackGAN-A, generates low-resolution images (64x64 pixels) from text descriptions. The second stage, called StackGAN-B, takes the low-resolution images from StackGAN-A and generates high-resolution images (256x256 pixels).

StackGAN uses a technique called progressive growing to generate high-resolution images. Progressive growing starts with a small image and gradually increases the size of the image in each iteration. This helps the model to learn the details of the image gradually, and it prevents the model from getting stuck in local minima.

This flower has long thin yellow petals and a lot of yellow anthers in the center



A small yellow bird with a black crown and a short black pointed beak



StyleGAN

- **DETAILS**

Introduced in 2018 by Nvidia.

Type of PGGAN (Progressive Growing GAN),
but with notable improvements.

One of the biggest improvements is the use of a style transfer technique called adaptive instance normalization (AdaIN). AdaIN allows StyleGAN to learn the style of an existing image and apply it to a new image. This makes it possible to create images with a wide variety of styles, from realistic to cartoony.

Another improvement is the use of a noise vector to add variation to the generated images. This prevents the images from becoming too repetitive or boring.

StyleGAN3 (2021)

- Project page: <https://nvlabs.github.io/stylegan3>
- ArXiv: <https://arxiv.org/abs/2106.12423>
- PyTorch implementation: <https://github.com/NVlabs/stylegan3>

StyleGAN2-ADA (2020)

- ArXiv: <https://arxiv.org/abs/2006.06676>
- PyTorch implementation: <https://github.com/NVlabs/stylegan2-ada-pytorch>
- TensorFlow implementation: <https://github.com/NVlabs/stylegan2-ada>
- MetFaces dataset: <https://github.com/NVlabs/metfaces-dataset>

StyleGAN2 (2019)

- ArXiv: <https://arxiv.org/abs/1912.04958>
- Video: <https://youtu.be/c-NJtV9Jvp0>
- TensorFlow implementation: <https://github.com/NVlabs/stylegan2>

StyleGAN (2018)

- ArXiv: <https://arxiv.org/abs/1812.04948>
- Video: <https://youtu.be/kSLJriaOumA>
- TensorFlow implementation: <https://github.com/NVlabs/stylegan>
- FFHQ dataset: <https://github.com/NVlabs/ffhq-dataset>

Progressive GAN (2017)



VQGAN+CLIP

- **DETAILS**

Introduced in 2020 (VQGAN) & 2021 (CLIP).

VQGAN+CLIP is a combination of two neural networks that can be used to generate images from text descriptions. VQGAN is a vector-quantized generative adversarial network, while CLIP is a contrastive language-image pre-training model.

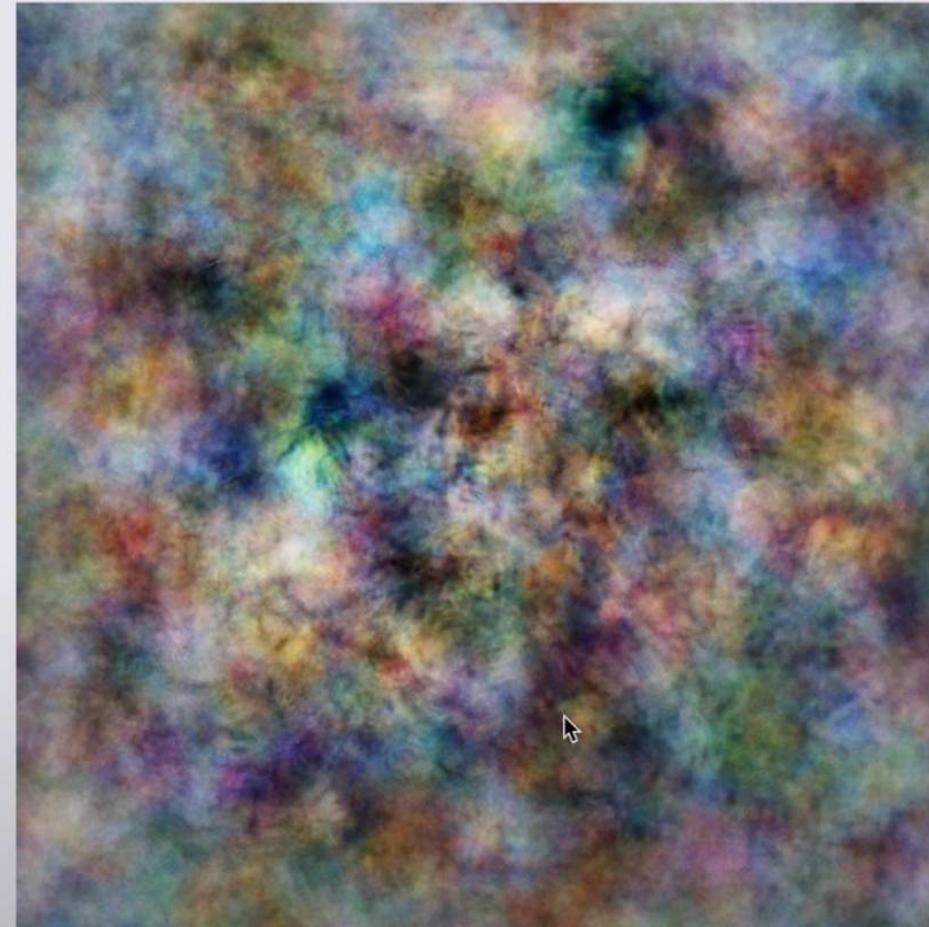
In other words, VQGAN is a generative adversarial neural network that is good at generating images that look similar to others (but not from a prompt), and CLIP is another neural network that is able to determine how well a caption (or prompt) matches an image.

CLIP + {AN IMAGE GENERATOR}

TEXT PROMPT

“A beautiful
painting of a +
dog riding a
dolphin”

GENERATE IMAGE



= CLIP
5.3%

CLIP + {AN IMAGE GENERATOR}

"A beautiful
painting of a +
dog riding a
dolphin"



CLIP
= 40.7%

STEP 50

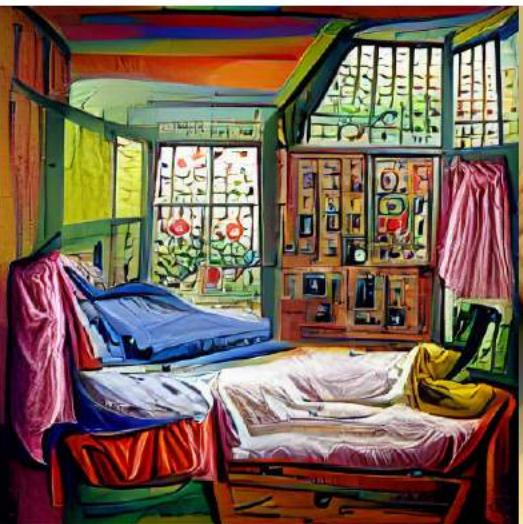
CLIP + {AN IMAGE GENERATOR}

"A beautiful
painting of a +
dog riding a
dolphin"



= CLIP
70.7%

STEP 500



How is VQGAN+CLIP different?

- **VQGAN+CLIP VS STYLEGAN**

VQGAN+CLIP can generate images from **open domain text prompts** without any prior training, while StyleGAN requires **pre-training** on a specific domain of images (such as faces, cars, etc.).

VQGAN+CLIP uses CLIP as a **guidance** for VQGAN to match the image and text, while StyleGAN uses a **mapping network** to map points in latent space to an intermediate latent space that controls the style of the image.

VQGAN+CLIP can generate images of **variable size**, while StyleGAN can only generate images of a **fixed size** that depends on the pre-trained model.

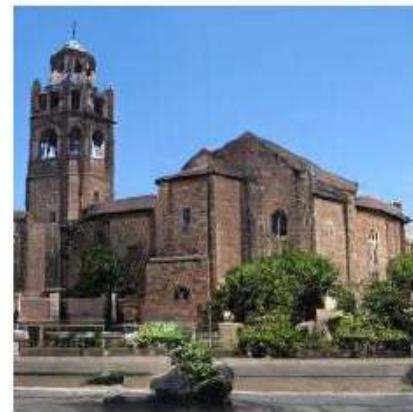
VQGAN+CLIP can also **edit** existing images by adding or removing text prompts, while StyleGAN can only **synthesize** new images.

StyleCLIP

- **DETAILS**

2021 Paper: "**StyleCLIP: Text-Driven Manipulation of StyleGAN Imagery**"

A method to manipulate images using a driving text. Our method uses the generative power of a pre-trained StyleGAN generator, and the visual-language power of CLIP.



“Emma Stone”

“Mohawk hairstyle”

“Without makeup”

“Cute cat”

“Lion”

“Gothic church”

Face Restoration

- **GPGAN (2021)**

GPGAN aims at developing a **Practical Algorithm for Real-world Face Restoration**. It leverages rich and diverse priors encapsulated in a pre-trained face GAN (*e.g.*, StyleGAN2) for blind face restoration.

Often bundled with Stable Diffusion installs.

- **CODEFORMER (2022)**

Based on BasicSR (a GAN-based image restoration toolbox), along with three other facial data models.

- **GOPEN (2021 - UPDATED 2023)**

GAN Prior Embedded Network for Blind Face Restoration in the Wild.
Promising results, but not as well-known as GPGAN or CodeFormer.



Input
From real life

HiFaceGAN
ACMMM 20

DFDNet
ECCV 20

Wan *et al.*
CVPR 20

PULSE
CVPR 20

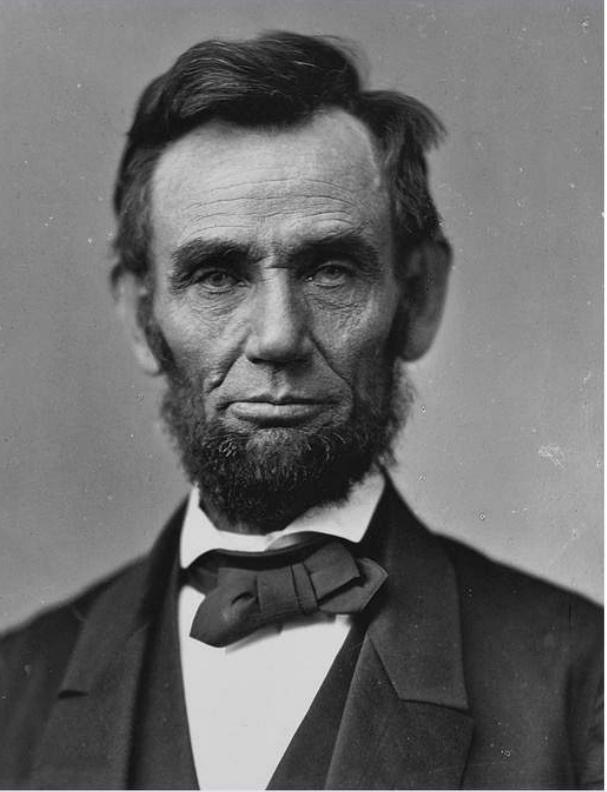
GFP-GAN
Ours

It can be used to restore your **old photos** or improve **AI-generated faces**.

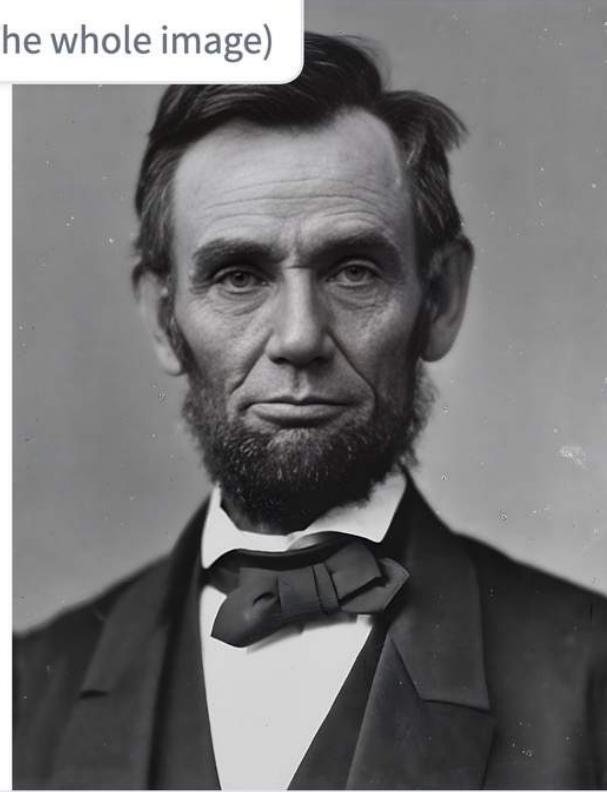
To use it, simply upload your image.

If GFPGAN is helpful, please help to  the [Github Repo](#) and recommend it to your friends 😊

Input



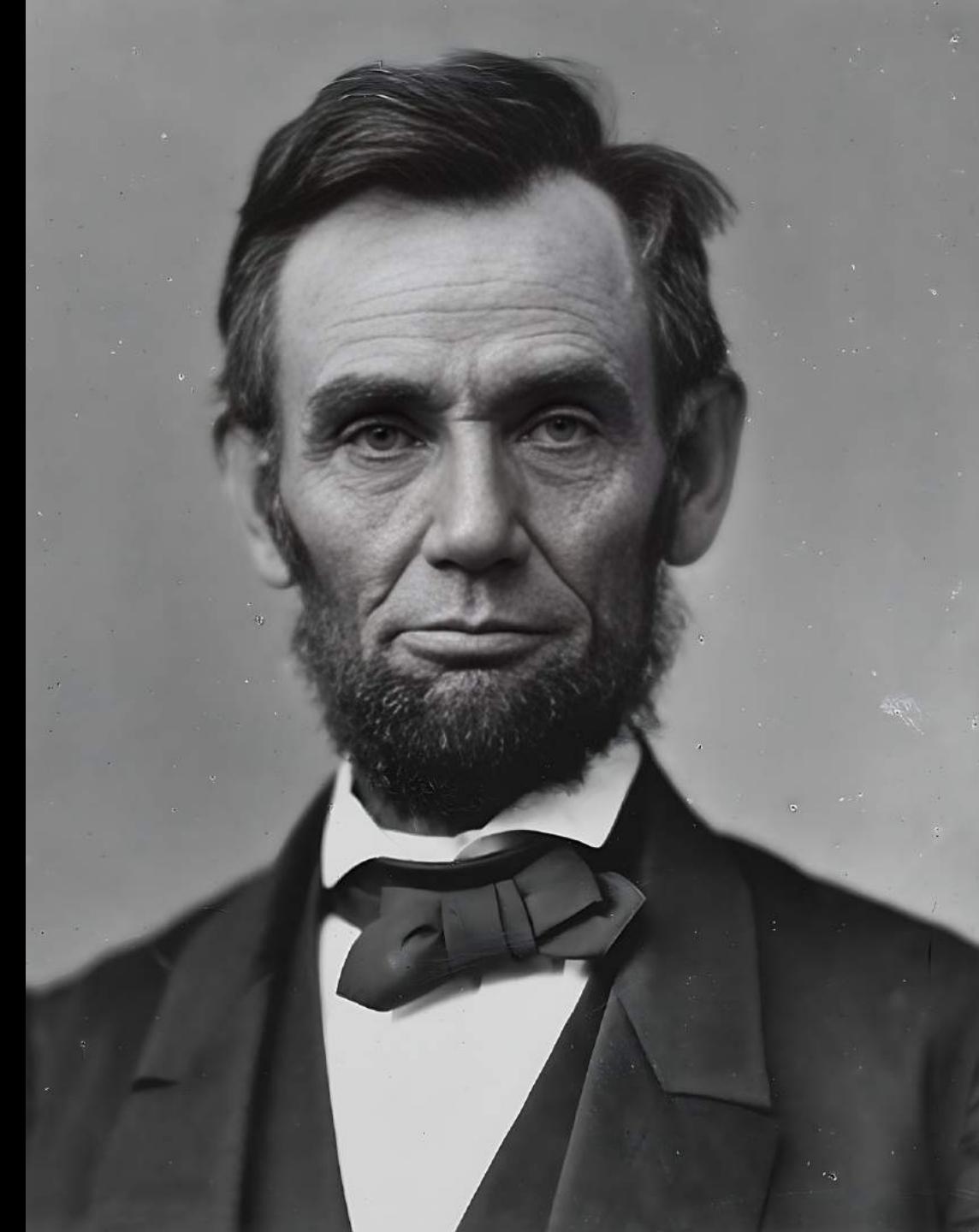
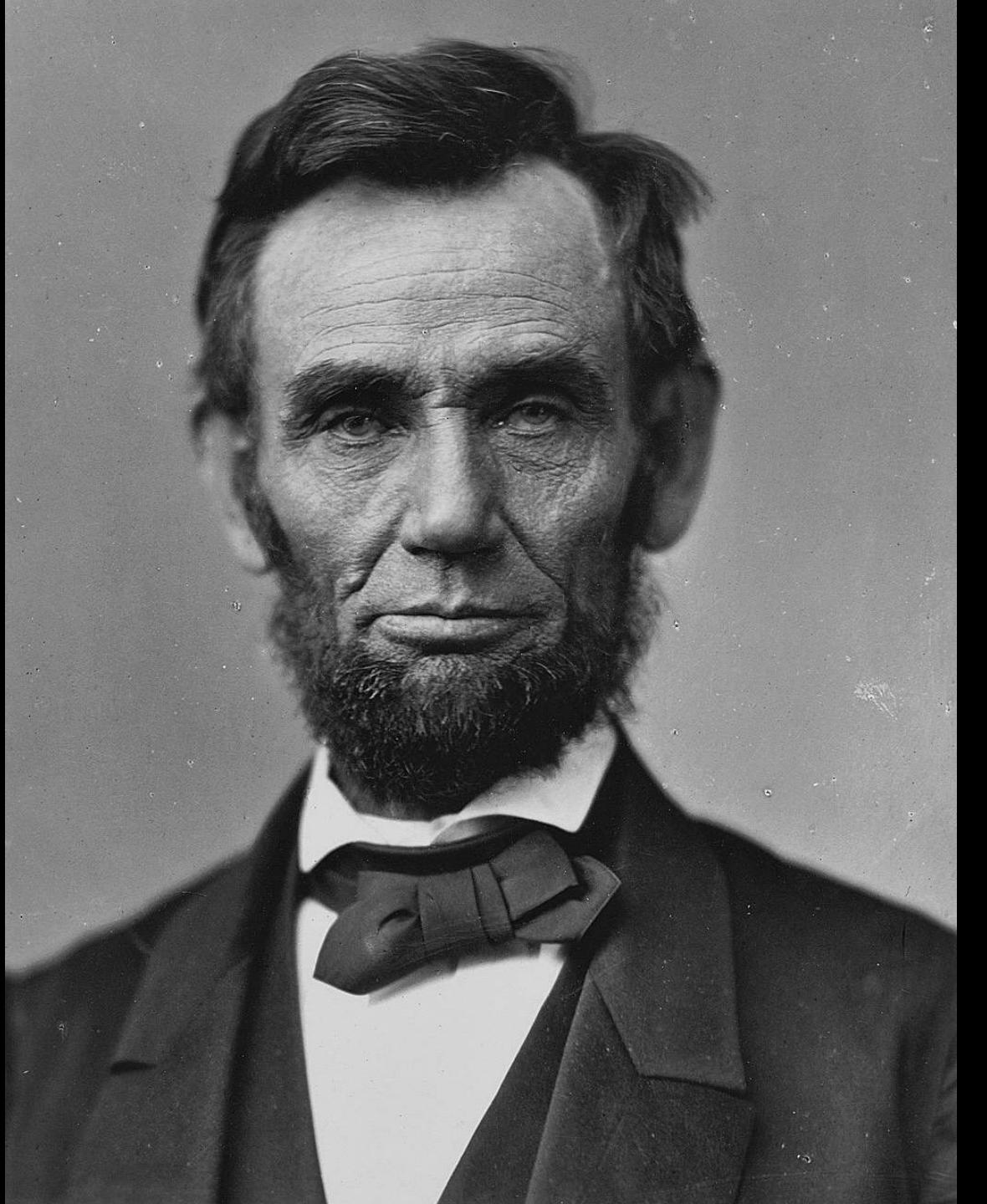
Output (The whole image)



Edit

X

Download

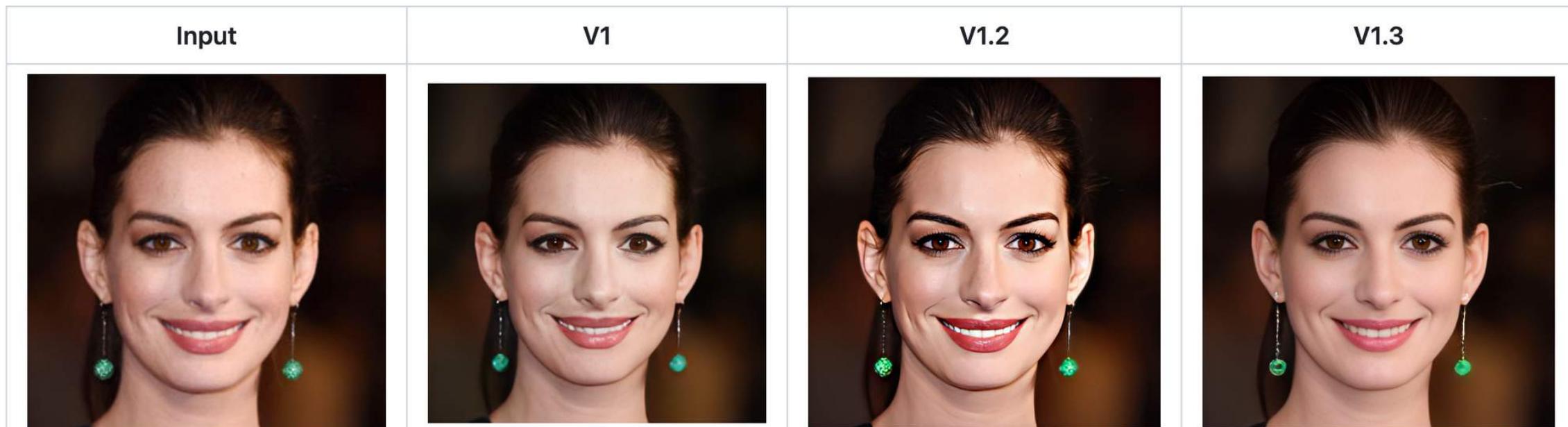


Comparisons among different model versions

Note that V1.3 is not always better than V1.2. You may need to try different models based on your purpose and inputs.

Version	Strengths	Weaknesses
V1.3	<ul style="list-style-type: none">✓ natural outputs✓ better results on very low-quality inputs✓ work on relatively high-quality inputs✓ can have repeated (twice) restorations	<ul style="list-style-type: none">✗ not very sharp✗ have a slight change on identity
V1.2	<ul style="list-style-type: none">✓ sharper output✓ with beauty makeup	<ul style="list-style-type: none">✗ some outputs are unnatural

For the following images, you may need to **zoom in** for comparing details, or **click the image** to see in the full size.



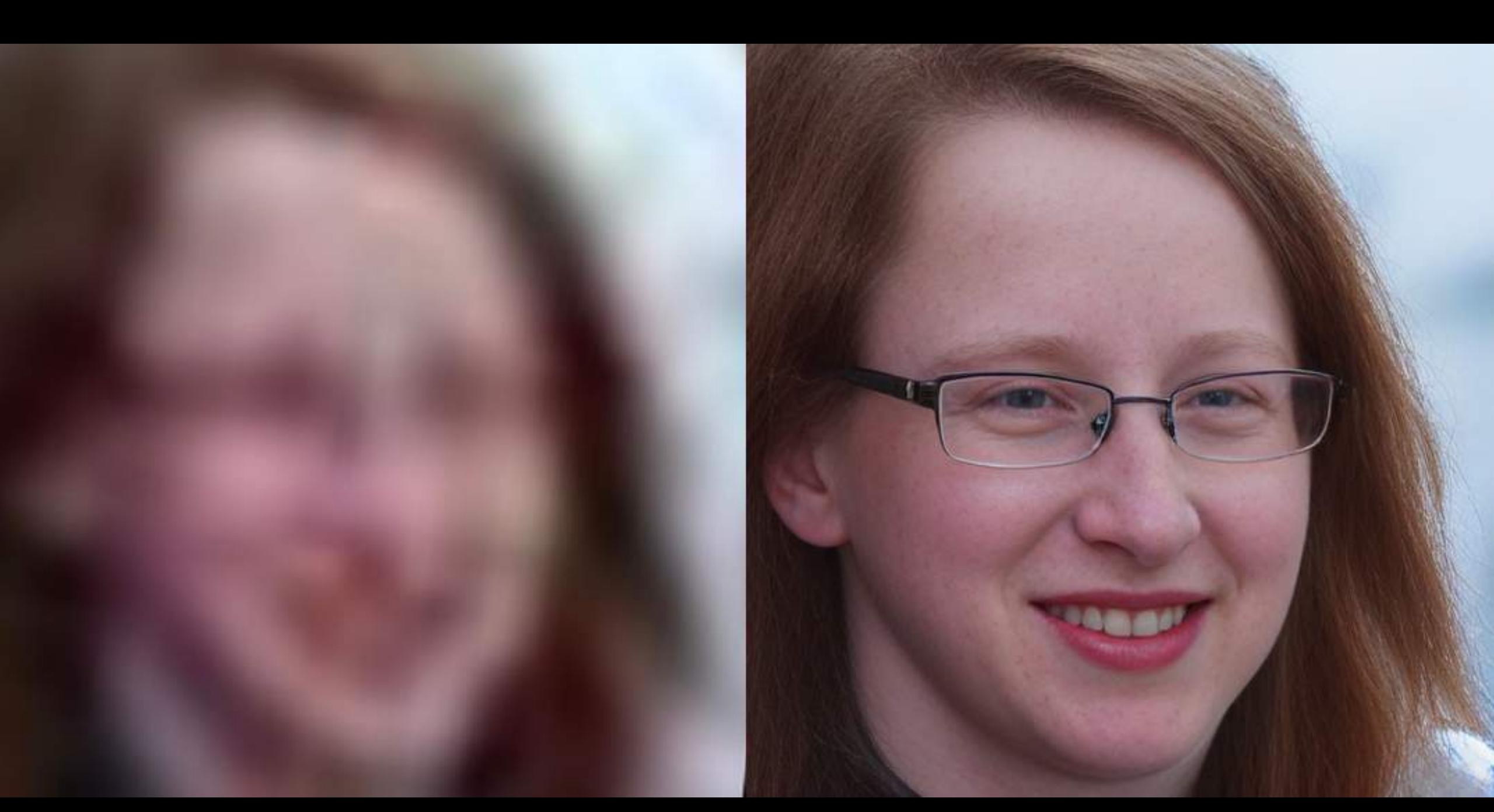
Old Photo

CodeFormer



Real Face Restoration







Wish Me



GOPEN

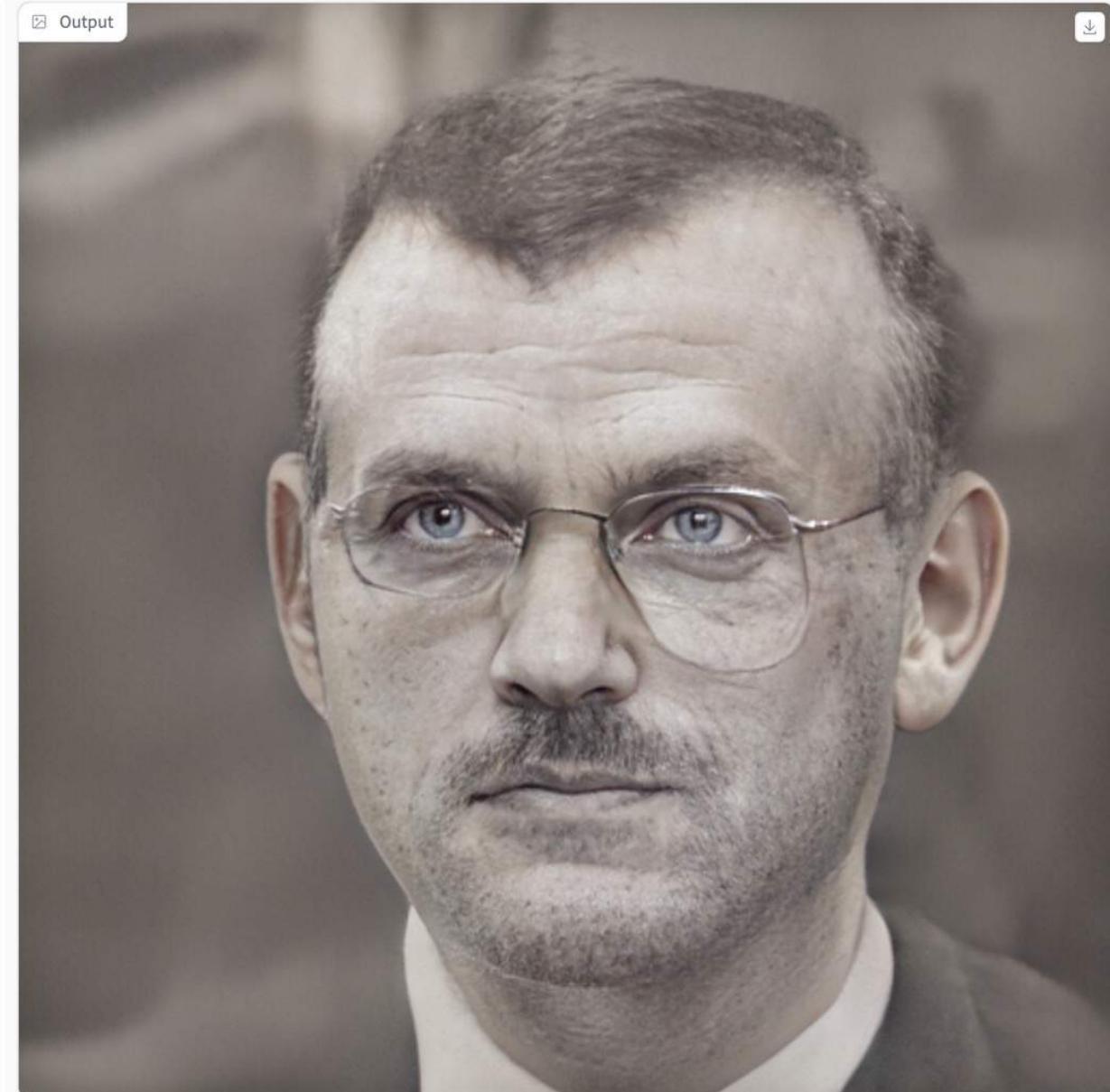
Gradio demo for GAN Prior Embedded Network for Blind Face Restoration in the Wild. To use it, simply upload your image, or click one of the examples to load them. Read more at the links below.



Input

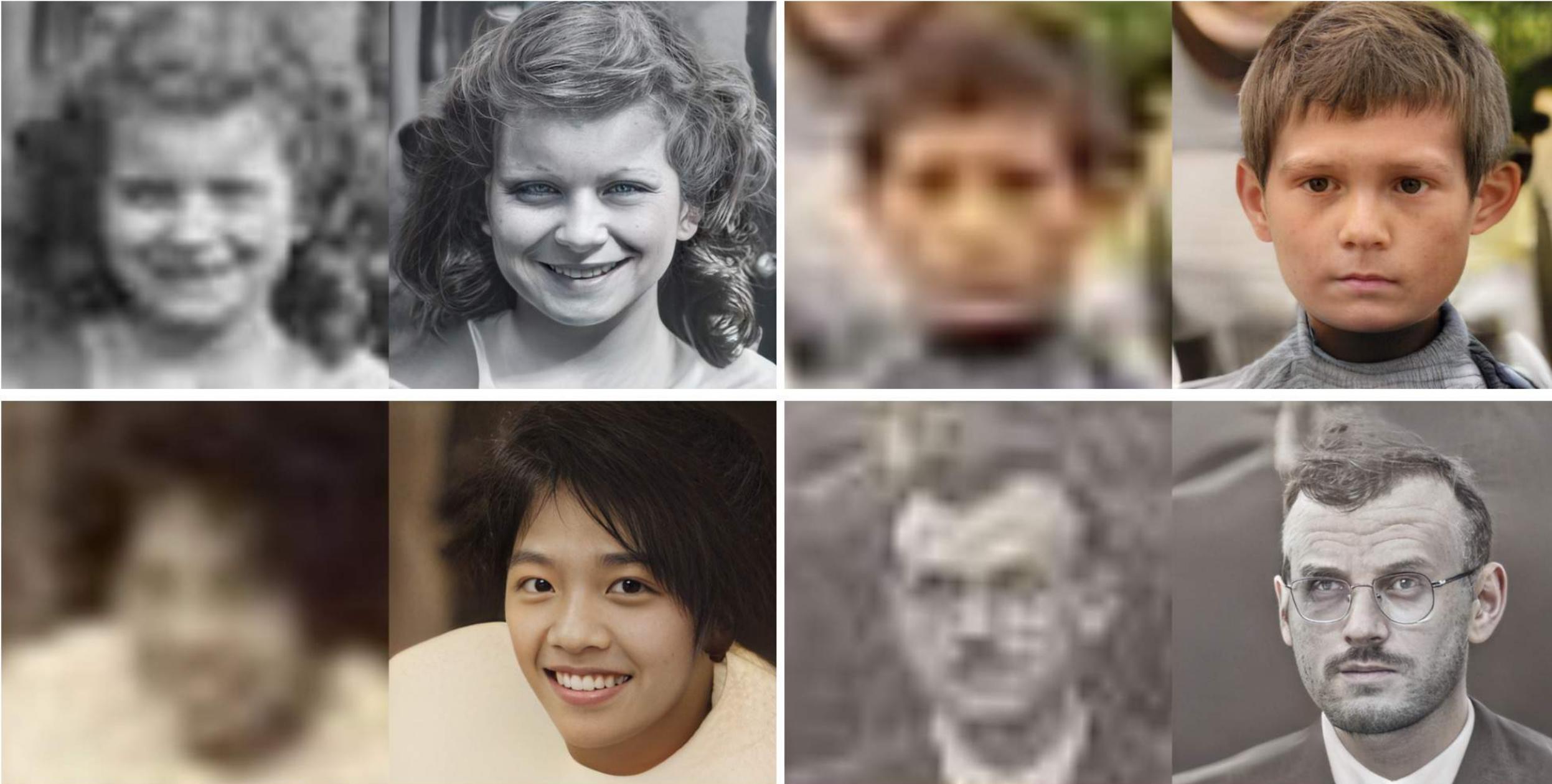
Clear

Submit



Output

Face Restoration



Face Colorization



Face Inpainting



Conditional Image Synthesis (Seg2Face)



Colorization

- **GPGAN**
- **CODEFORMER**
- **DEOLDIFY (2022)**
Uses a new type of GAN model (NoGAN).
- **PALETTE.FM (2022)**
Uses a proprietary machine learning algorithm.

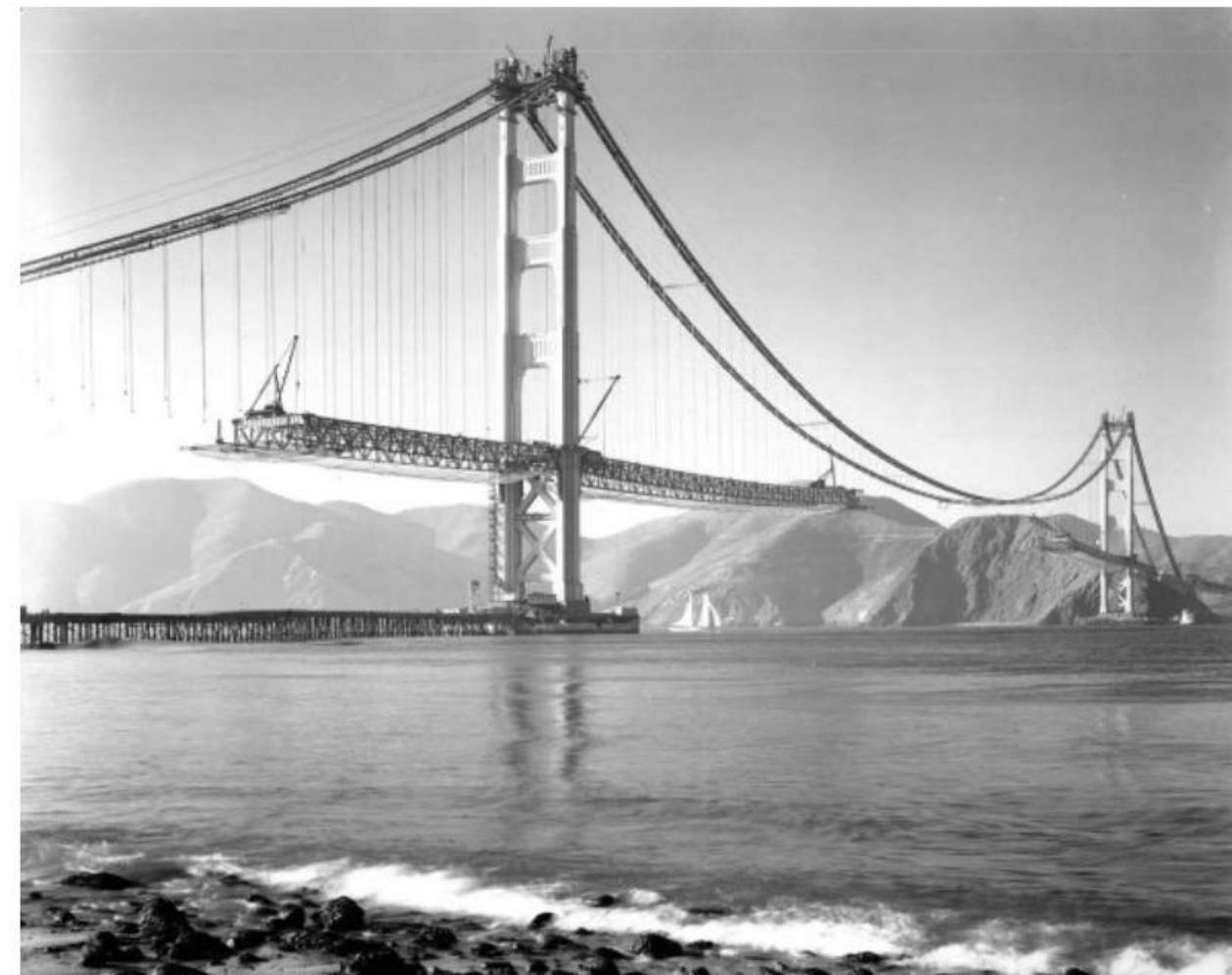
"Migrant Mother" by Dorothea Lange (1936)

DeOldify



"Building the Golden Gate Bridge" (est 1937)

DeOldify



Below is a quick comparison between the first generation of AI technology compared to Palette's second generation colorizer.



Original



First Gen



Palette



Upscaling & Image Restoration

- **REAL-ESRGAN (2022)**

Often bundled with Stable Diffusion installs.

Used by both GFPGAN & CodeFormer to upscale image backgrounds.

- **REAL-ESRGAN (CUSTOM MODEL)**

Super-Resolution Model from ruDALL-E

- **SWINIR (2021)**

Uses a Shifted Window Transformer model.

- **LATENT DIFFUSION SR (2022)**

Upscales images with the latent diffusion super resolution model.

- **STABLE DIFFUSION UPSCALERS**

- **ULTIMATE SD UPSCALE FOR AUTOMATIC1111**

- **TOPAZ GIGAPIXEL AI**



ClownVamp

@ClownVamp

...

Been on an upscaling/resolution obsession.

My new SuperRare 1/1 is 10752x7168

Latest pipeline is based around combining Gigapixel, Ultimate SD Upscale via Automatic1111 (I use the ultrasharp model), and hand-added texture overlays.

More on the piece in the thread!

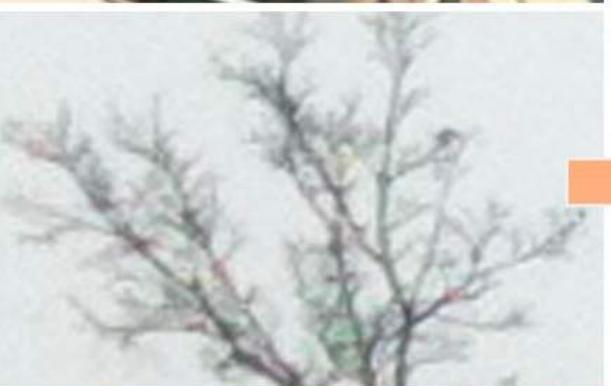


1:59 PM · Mar 29, 2023 · 5,360 Views

Input (Bicubic)



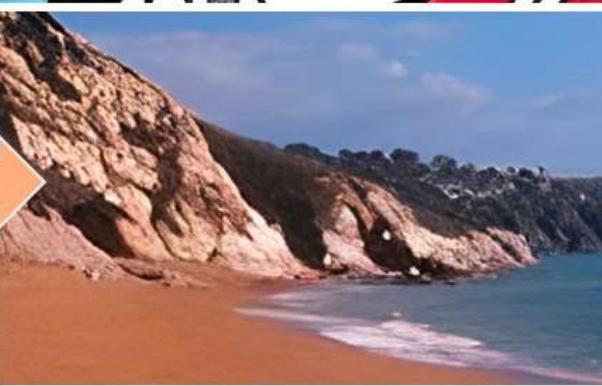
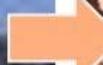
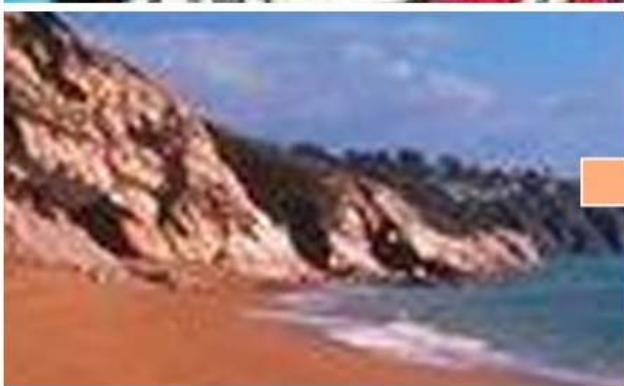
Real-ESRGAN Output



Input (Bicubic)



Real-ESRGAN Output



Kandinsky 2.1 model drew:a lion looking out over the savannah



a lion looking out over
the savannah

Share



0



0

You can also generate
images with our
models in the our
Android app [Salute](#) or
on a smart device
running by SalutTV



This material is licensed under the [Public License Attribution-](#)

[ShareAlike](#), SberBank©, 2022. All rights reserved

Try again

[Demo](#)[API](#)[Examples](#)[Versions \(32fdb223\)](#)

Input

image



scale

8

[Tweak It](#)

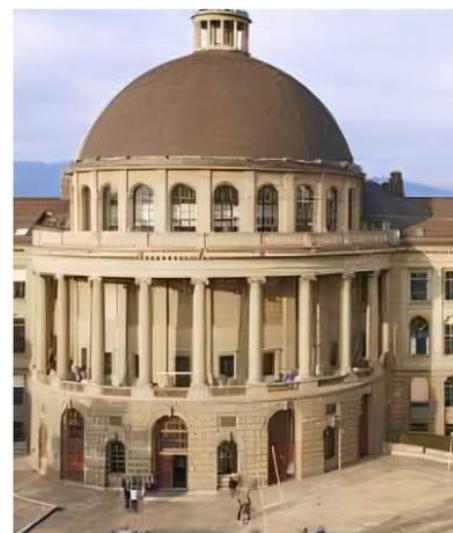
Output



**Real-World
Image (x4)**



**BSRGAN,
ICCV2021**



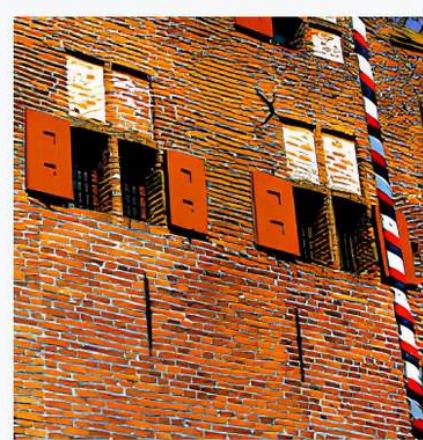
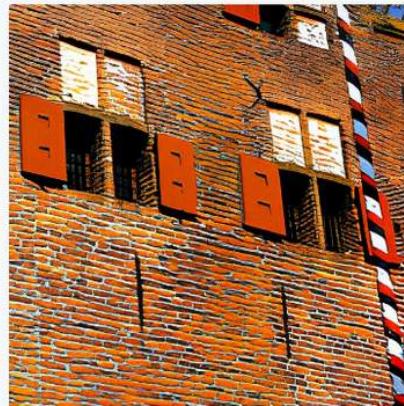
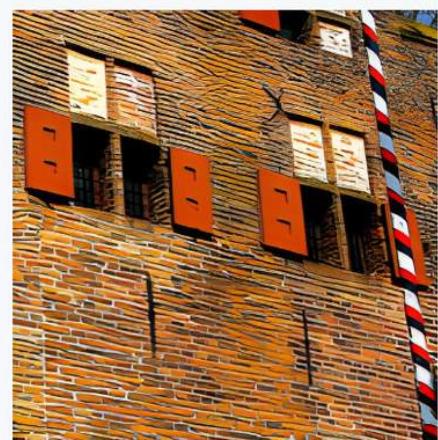
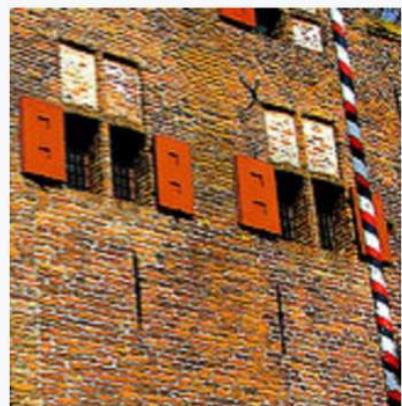
Real-ESRGAN



SwinIR (ours)



**SwinIR-Large
(ours)**



jpeg

40

image



noise

15

task_type

Real-World Image Super-Resolution-Large

This example was created by [@cjwbw](#)

» Tweak It



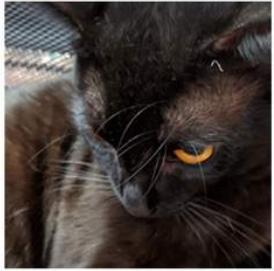
[Demo](#)[API](#)[Examples](#)[Versions \(9117a98d\)](#)

Input

up_f

4

image



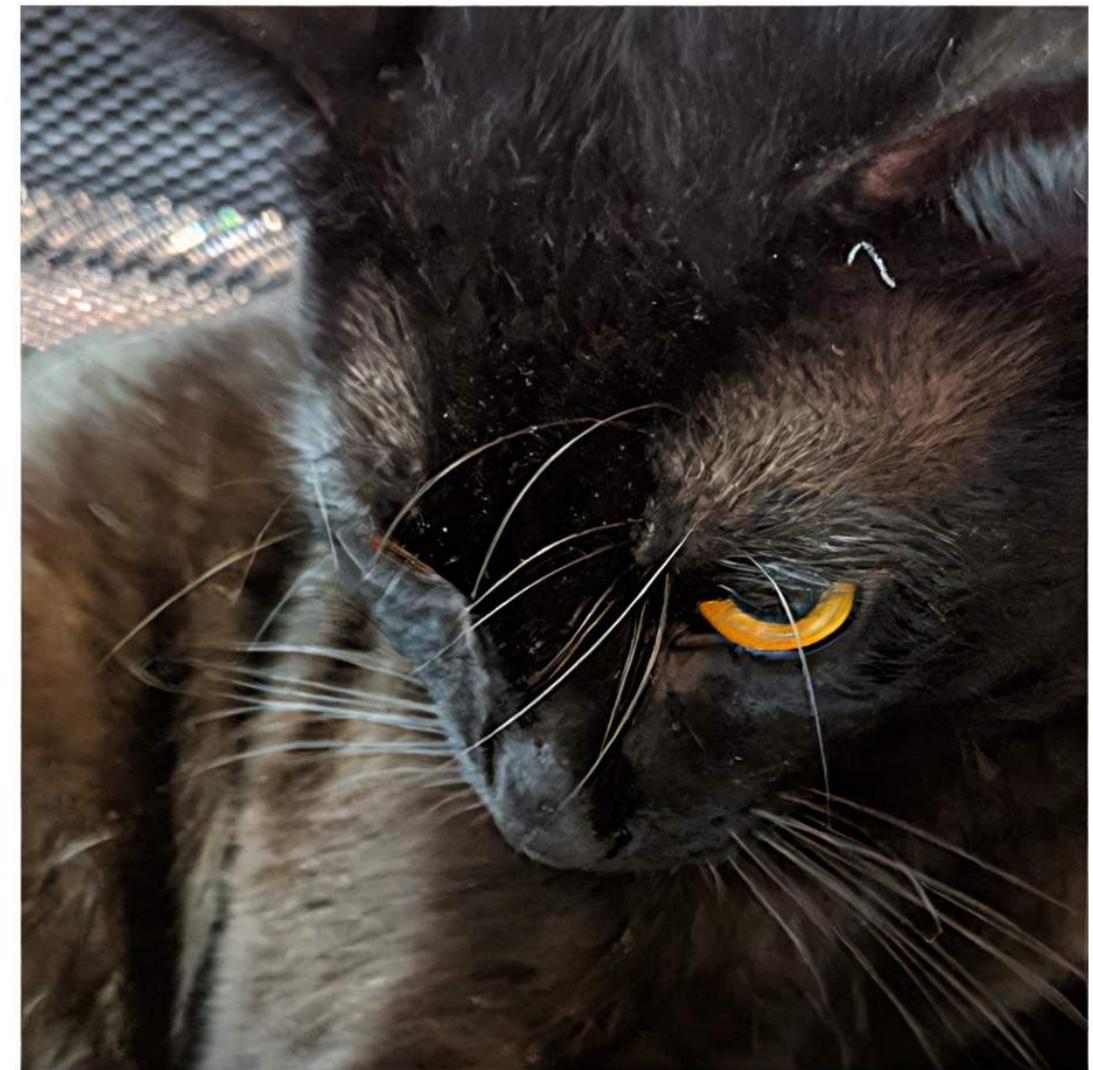
steps

100

This example was created by @palp

[Tweak It](#)

Output



Results from testing so far show it to be very good with textures and fine details on generated images, but poor on real-world jobs.

SD Upscaler

2x Latent Upscaler using SD 2.0 And 4x Upscaler using SD 2.1. This version runs on CPU or GPU and is currently running on a T4 GPU. For 4x Upscaling use images lower than 512x512, ideally 128x128 or smaller for 512x512 output. For 2x Upscaling use up to 512x512 images for 1024x1024 output.

Notice: Largest Accepted Resolution is 512x512



Models

Upscaler 2x Upscaler 4x

Optional: Enter a Prompt to Guide the AI's Enhancement, this can have an Img2Img Effect

Experimental: Influence What you do not want the AI to Enhance. Such as Blur, Smudges, or Pixels

Guidance Scale: How much the AI influences the Upscaling.
 1

Number of Iterations
 5

Seed

Clear **Submit**



FAQ

Anton Antonov edited this page on Mar 18 · 6 revisions

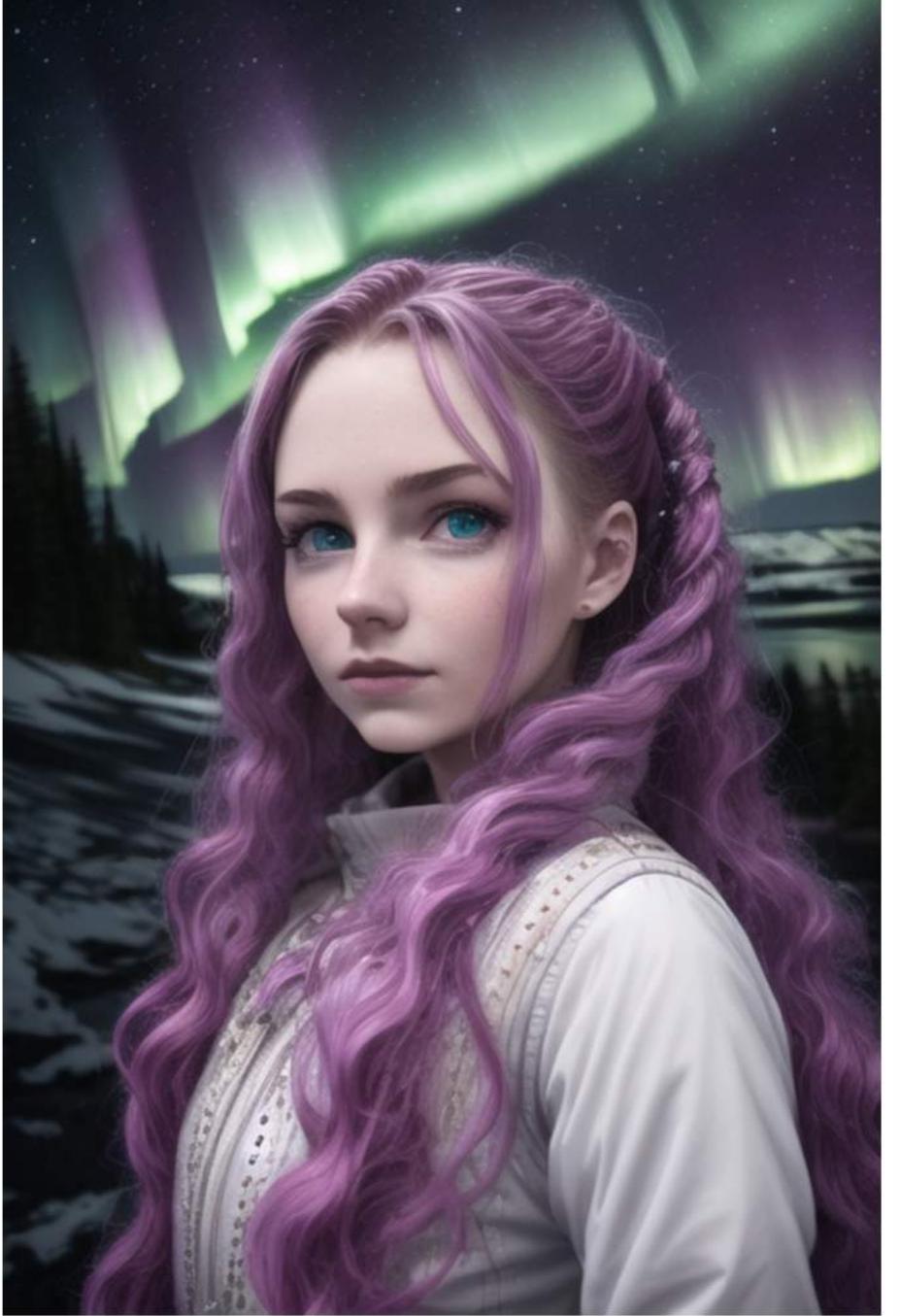
How it works

Enlarges the image with the selected upscaler, then divides the image into tiles and inpaints it. If the "Seams fix" option is set, then it additionally goes along the seams and redraws them.

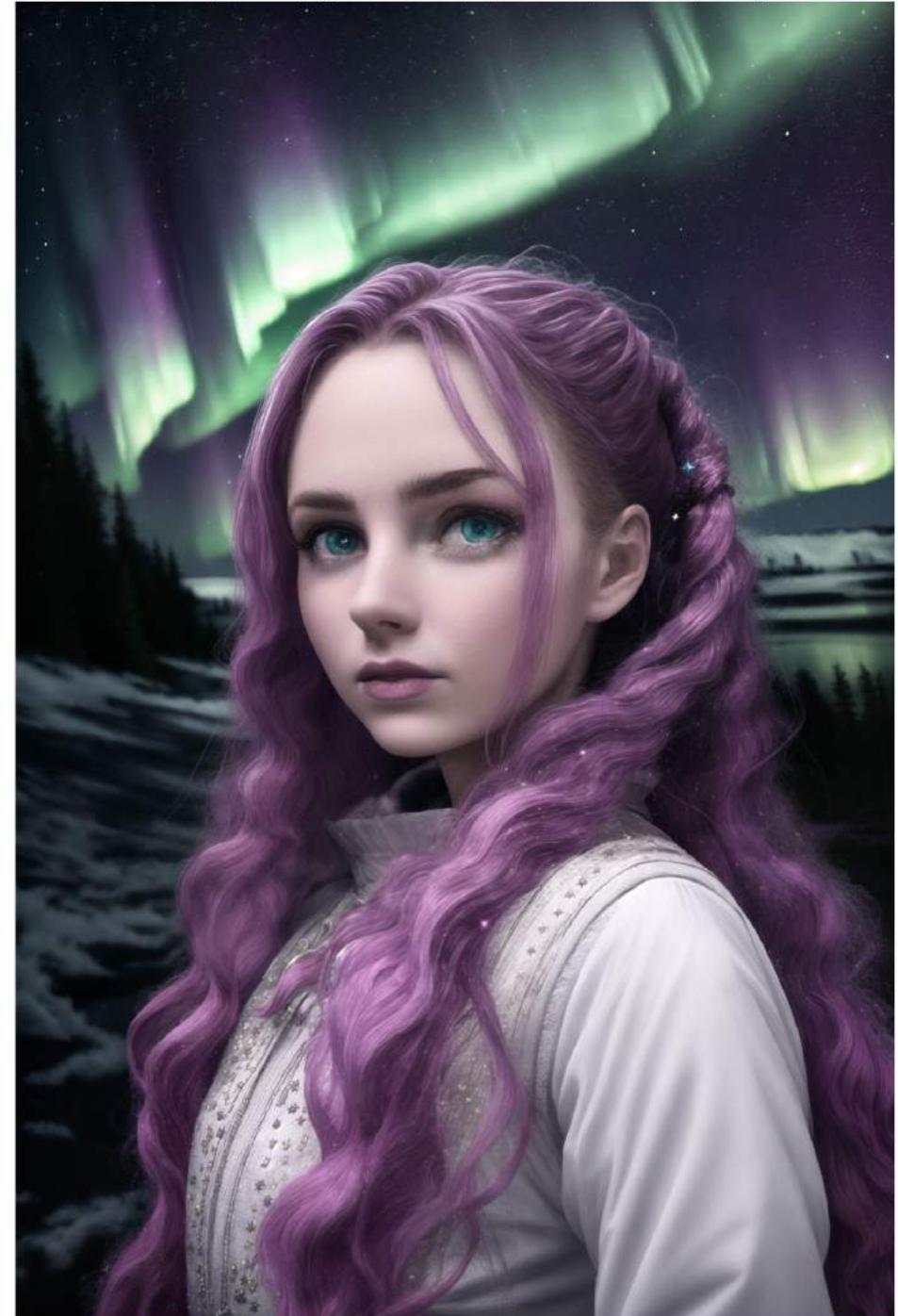
Parameters descriptions

- **Denoise** - uses from default img2img field. We recommend 0.35 value for image enhancements, but if you don't want changes use 0.15-0.20
- **Target size type** - Where to get the size of the final image
 - **From img2img settings** - Default img2img width and height sliders
 - **Custom size** - Built in width and height sliders. Max values - 8192
 - **Scale from image size** - Init image size multiplied by scale factor
- **Redraw**
 - **Upscaler** - upscale image before redrawing. Use what you like. Our recommendation - ESRGAN to photorealistic images, R-ESRGAN 4x+ for others (less requirements)
 - **Type**
 - **Linear** - All tiles processed one by one. column by column, row by row
 - **Chess** - All tiles are processed in a checkerboard pattern. Reduces the chance of seam artifacts
 - **None** - Disabled redraw. Use it when run generation without seam fix, see visible overlays or artifacts on seams and want to run just seam pass. Don't forget to put upscaled image as the source before
 - **Tile width** - The width of the tile to be processed. The larger the tile, the less artifacts will be in the final image. For 2k 512px is usually enough
 - **Tile height** - The height of the tile to be processed. The default is 0, in which case it is equal to the width. The larger the tile, the less artifacts will be in the final image. For 2k 512px is usually enough
 - **Padding** - How many pixels of neighboring tiles will be taken into account when processing a tile

Original image (512x768)



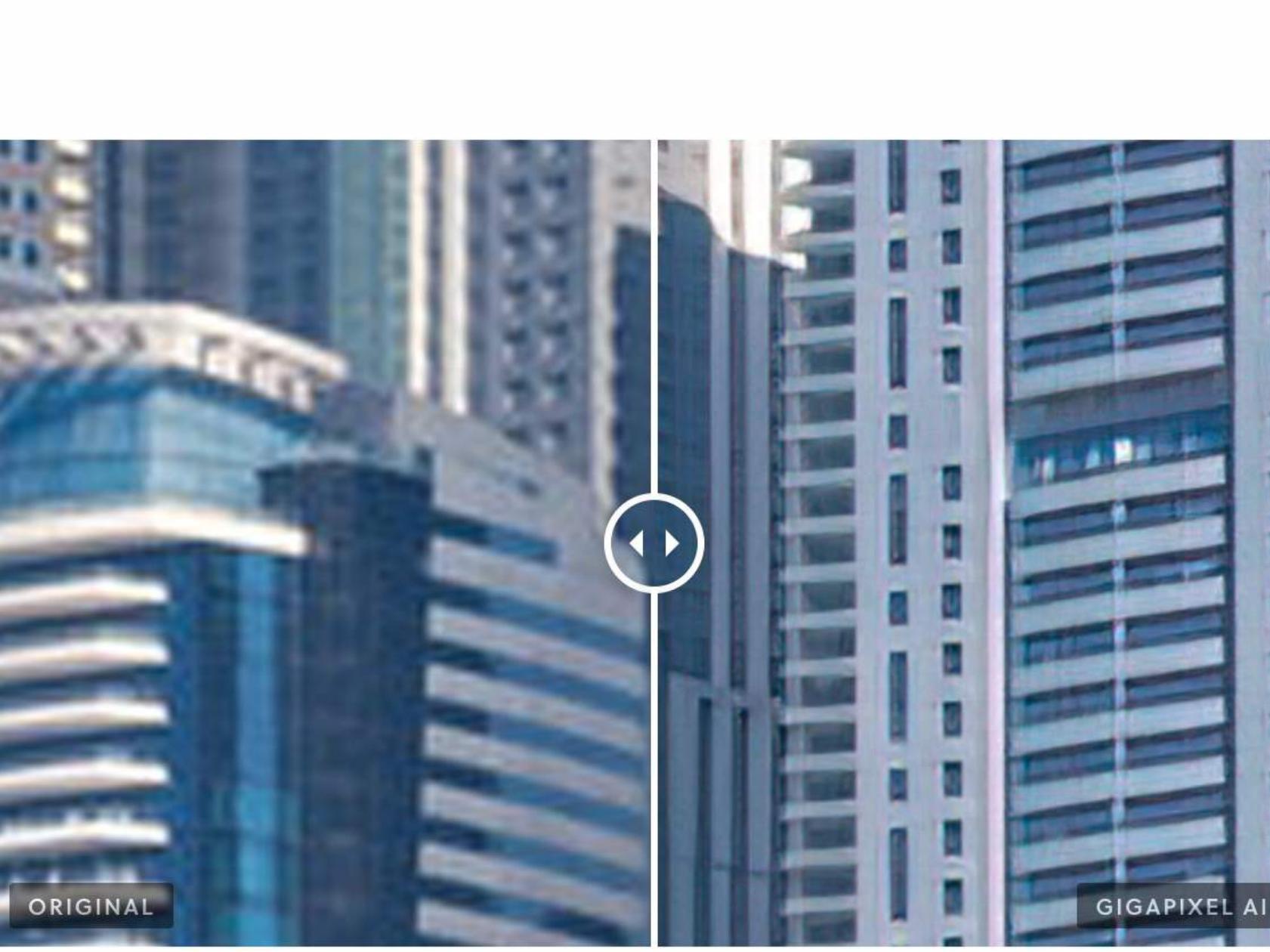
x4 upscaled with seams fix (2048x3072). Type: Half tile offset pass, Denoise:0.3, Mask blur: 16, Padding: 32





ORIGINAL

GIGAPIXEL AI



USED BY

NETFLIX



Disney



LUCASFILM
Ltd



amazon



WB WARNER BROS.

Style Transfer

- **SPECIFICALLY TRAINED TO MIMIC A CERTAIN AESTHETIC - LIMITED TO SPECIFIC USE CASES**
- **EXAMPLES:**
 - JoJoGAN (2021)
 - AnimeGANv2 (2021)
 - ArcaneGAN (2021)

JoJoGAN

Gradio Demo for JoJoGAN: One Shot Face Stylization. To use it, simply upload your image, or click one of the examples to load them. Read more at the links below.

img

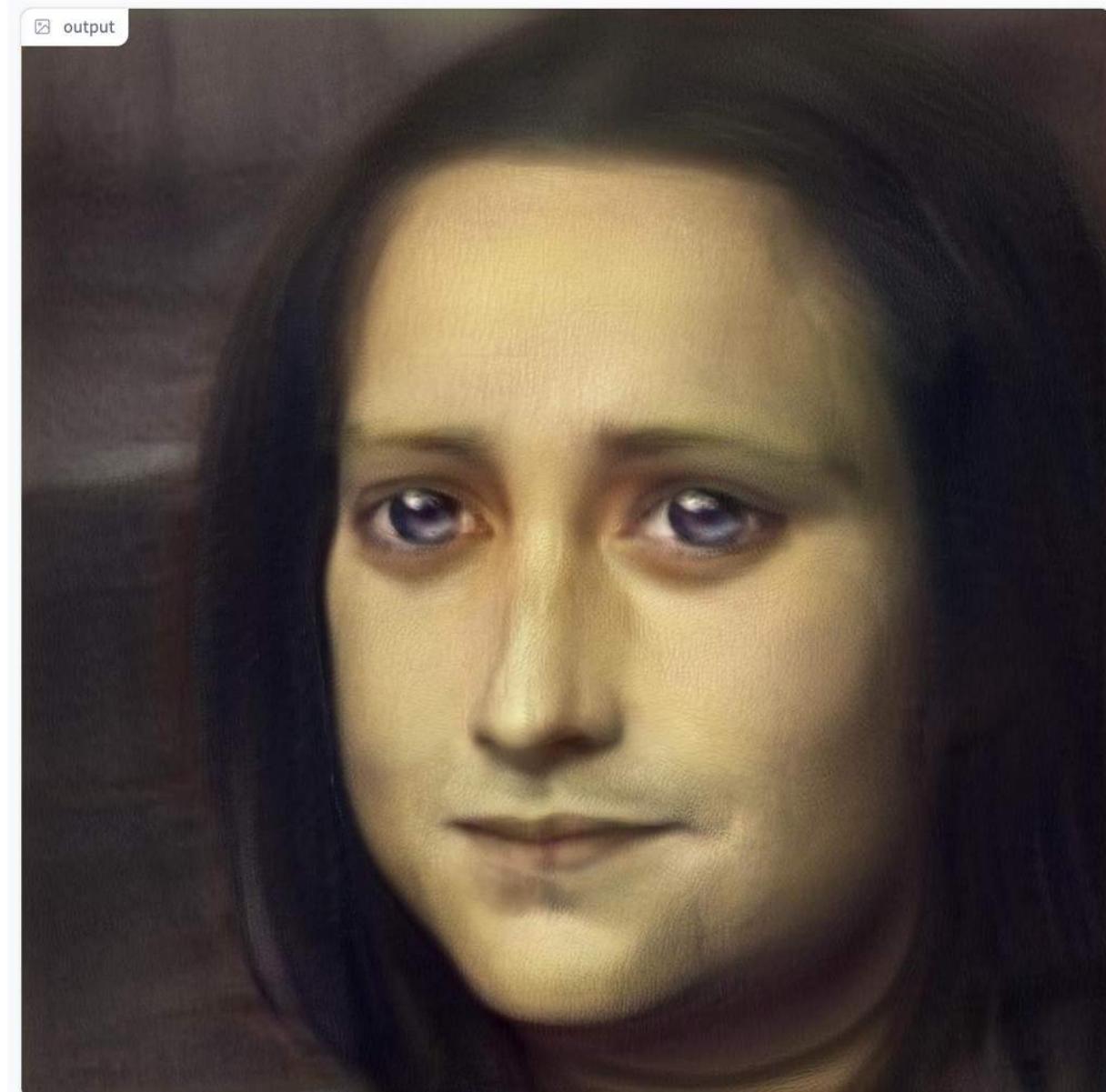


Model

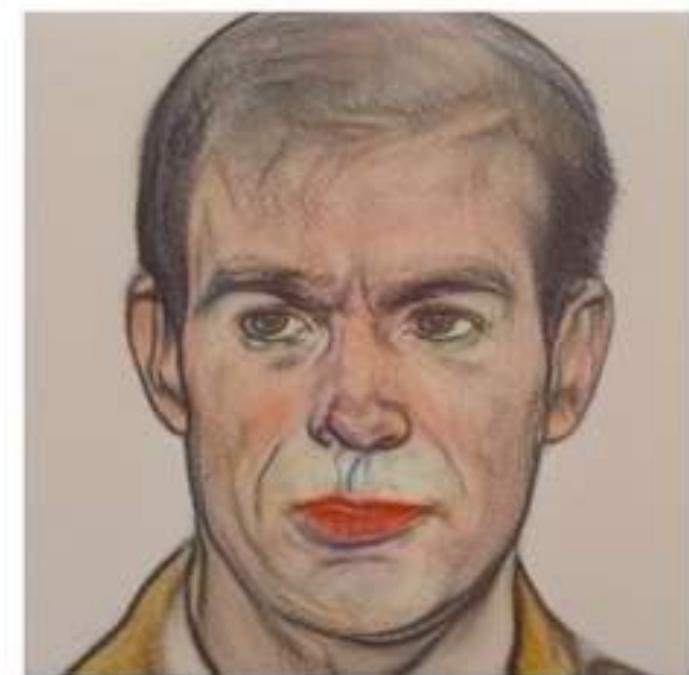
Jinx

Clear

Submit



COLLAGE OF ART ILLUSTRATIONS



AnimeGANv2

Gradio Demo for AnimeGanv2 Face Portrait. To use it, simply upload your image, or click one of the examples to load them. Read more at the links below. Please use a cropped portrait picture for best results similar to the examples below.

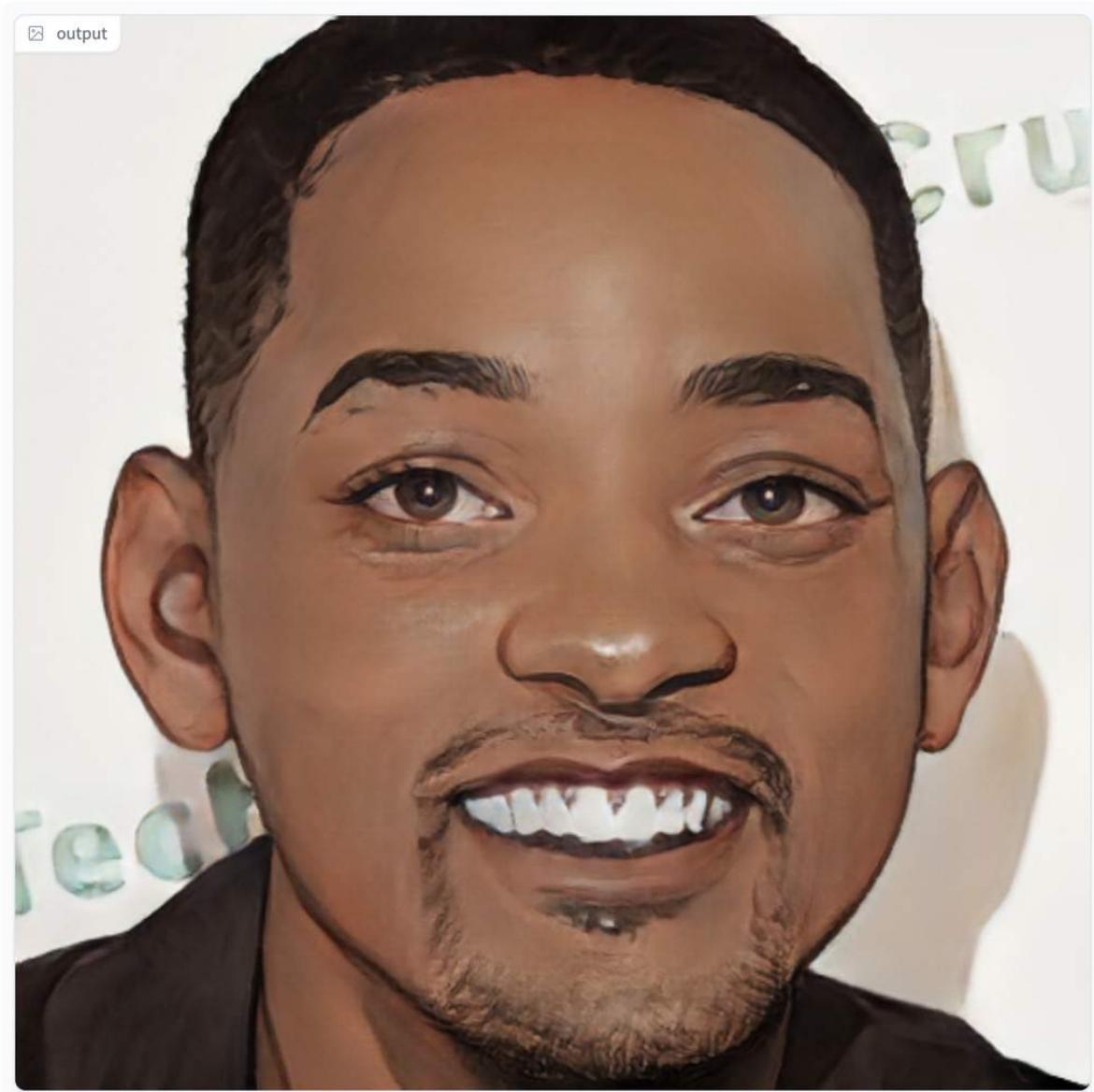
img



version

version 1 (▲ stylization, ▼ robustness) version 2 (▲ robustness, ▼ stylization)

Clear Submit



AnimeGANv2

Gradio Demo for AnimeGanv2 Face Portrait. To use it, simply upload your image, or click one of the examples to load them. Read more at the links below. Please use a cropped portrait picture for best results similar to the examples below.

img

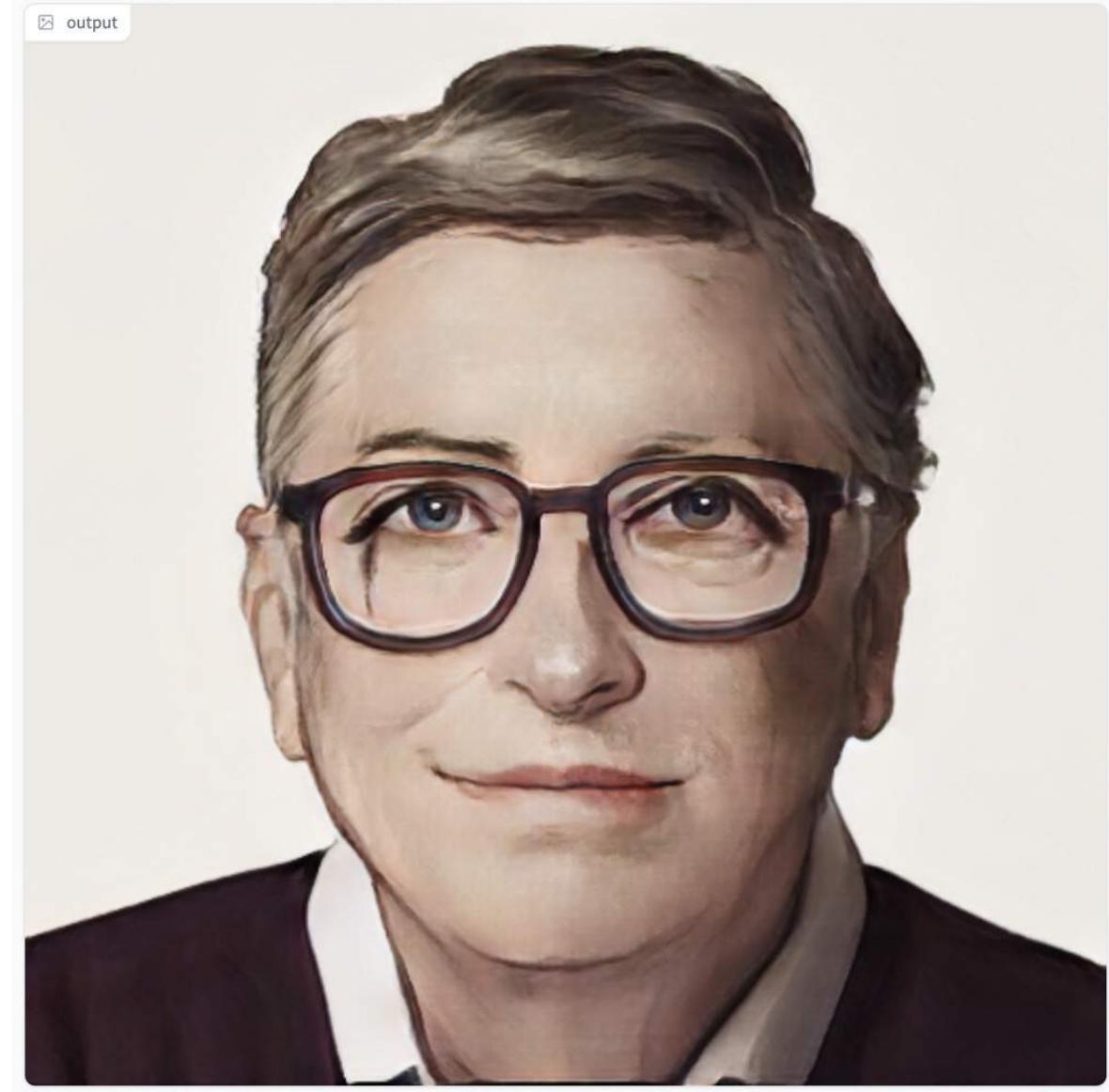


version

version 1 (▲ stylization, ▼ robustness) version 2 (▲ robustness, ▼ stylization)

Clear

Submit



ArcaneGAN

Gradio demo for ArcaneGAN, portrait to Arcane style. To use it, simply upload your image, or click one of the examples to load them. Read more at the links below.

Input



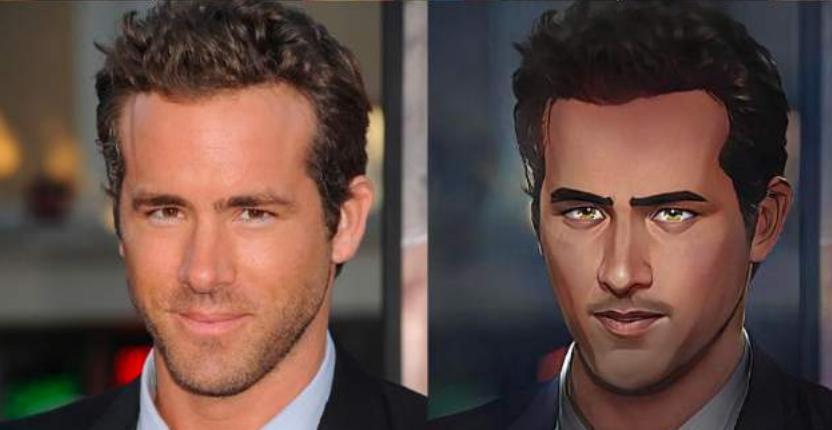
version

version 0.2 version 0.3 version 0.4

Clear

Submit





What's Next for GANs?

**Computer Science > Computer Vision and Pattern Recognition***[Submitted on 9 Mar 2023]*

Scaling up GANs for Text-to-Image Synthesis

Minguk Kang, Jun-Yan Zhu, Richard Zhang, Jaesik Park, Eli Shechtman, Sylvain Paris, Taesung Park

The recent success of text-to-image synthesis has taken the world by storm and captured the general public's imagination. GANs used to be the de facto choice, with techniques like StyleGAN. With DALL-E 2, a rapid shift raises a fundamental question: can we scale up GANs to benefit from large datasets like LAION? We find GigaGAN, a new GAN architecture that far exceeds this limit, demonstrating GANs as a viable option for text-to-image synthesis. First, it can generate images at high resolution in a short time, taking only 0.13 seconds to synthesize a 512px image. Second, it can synthesize high-resolution images, for applications such as latent interpolation, style mixing, and vector arithmetic operations.

Comments: CVPR 2023. Project webpage at [this https URL](#)

Subjects: **Computer Vision and Pattern Recognition (cs.CV)**; Graphics (cs.GR); Machine Learning (cs.LG)

Cite as: [arXiv:2303.05511 \[cs.CV\]](#)

(or [arXiv:2303.05511v1 \[cs.CV\]](#) for this version)

<https://doi.org/10.48550/arXiv.2303.05511>

GigaGAN: Large-scale GAN for Text-to-Image Synthesis

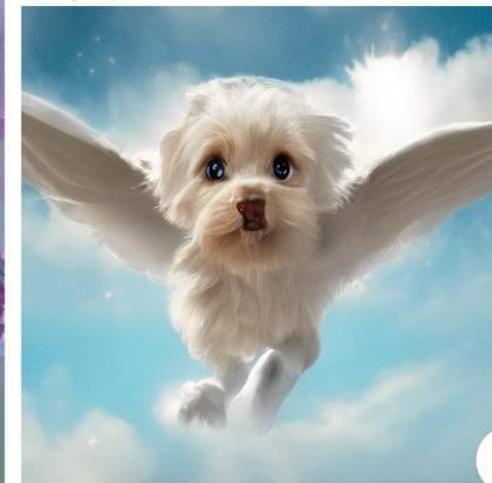
Can GANs also be trained on a large dataset for a general text-to-image synthesis task? We present our 1B-parameter GigaGAN, achieving lower FID than Stable Diffusion v1.5, DALL·E 2, and Parti-750M. It generates 512px outputs at 0.13s, orders of magnitude faster than diffusion and autoregressive models, and inherits the disentangled, continuous, and controllable latent space of GANs. We also train a fast upampler that can generate 4K images from the low-res outputs of text-to-image models.



A portrait of a human growing colorful flowers from her hair. Hyperrealistic oil painting. Intricate details.



A golden luxury motorcycle parked at the King's palace. 35mm f/4.5.



a cute magical flying maltipoo at light speed, fantasy concept art, bokeh, wide sky

Upscaling to 16-megapixel photos with GigaGAN

Our GigaGAN framework can also be used to train an efficient, higher-quality upsampler. This can be applied to real images, or to the outputs of other text-to-image models like diffusion. GigaGAN can synthesize ultra high-res images at 4k resolution in 3.66 seconds.



one more thing...

New Models (2023)

- GLIGEN
- INSTRUCTPIX2PIX
- DEEP-FLOYD IF

GLIGEN:

Open-Set Grounded Text-to-Image Generation

Yuheng Li, Haotian Liu, Qingyang Wu, Fangzhou Mu, Jianwei Yang, Jianfeng Gao,
Chunyuan Li*, Yong Jae Lee*,

► University of Wisconsin-Madison; ► Columbia University; ► Microsoft *Equal Advising

arXiv

Code

Demo

Video



Caption: "a photo of a hybrid between a bee and a rabbit"
Grounded text: **hybrid between a bee and a rabbit, flower**



Caption: "Pikachu is under a tree, digital art"
Grounded text: **Pikachu, tree**; Grounded style image: **blue inset**



Caption: "A dog / bird / helmet / backpack is on the grass"
Grounded image: **red inset**



Caption: "superman / monkey / Homer Simpson / is scratching its head"
Grounded keypoints: **plotted dots on the left image**

Teddy bear

parrot



Caption: A teddy bear sitting next to a parrot

Teddy bear

parrot



Caption: A teddy bear sitting next to a parrot

Teddy bear

parrot



Caption: A teddy bear sitting next to a parrot

Text Prompt

a painting of a fox sitting in a field at sunrise in the style of Claude Monet

Grounding Prompt



GLIGEN



DALLE 1



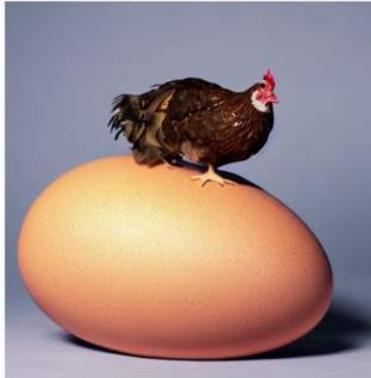
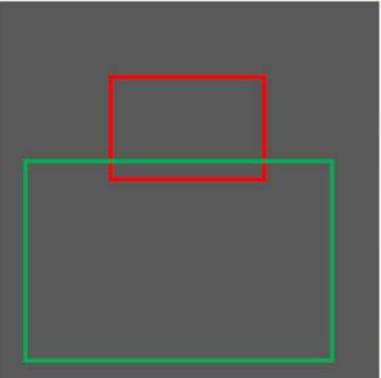
DALLE 2



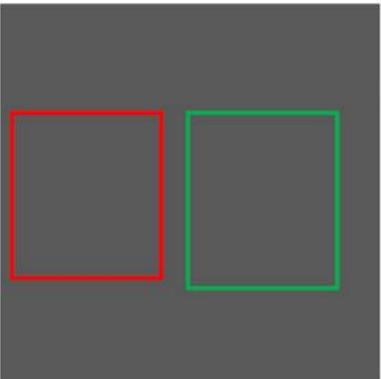
Compared with existing text2img models such as DALLE1 and DALLE2, GLIGEN enables the new capability to allow grounding instruction. The text prompt and DALLE generated images are from [OpenAI Blog](#).

Spatially counterfactual generation

GLIGEN



Stable-Diffusion

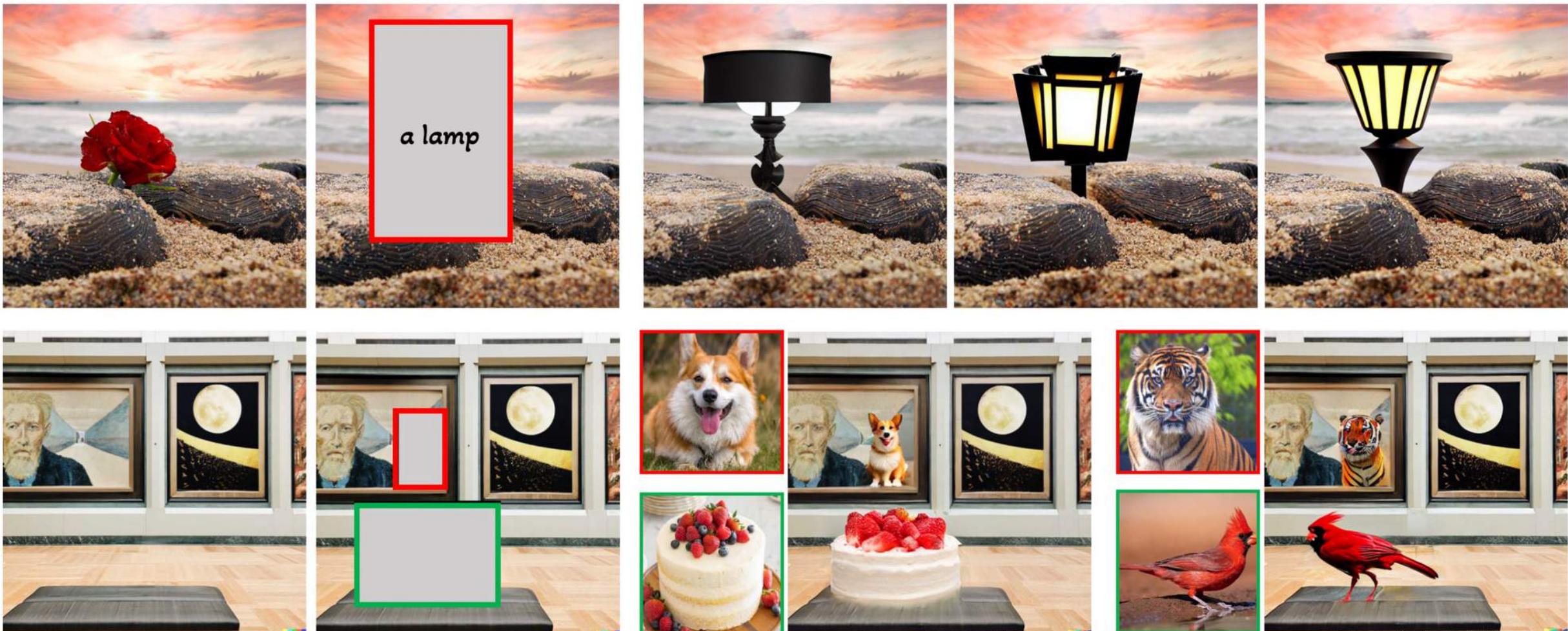


Caption: "an apple and a same size dog"

Grounded text: **apple**, **dog**

By explicitly specifying object size and location, GLIGEN can generate spatially counterfactual results which are difficult to release through text2img model (e.g., Stable Diffusion).

Grounded Inpainting



Like other diffusion models, GLIGEN can also perform grounded image inpaint, which can generate objects tightly following provided bounding boxes.

Canny Map Grounded T2I Generation



Caption: "A Humanoid Robot Designed for Companionship"

Grounded canny map: [the left image](#)



Caption: "a chair and a table"

Grounded canny map: [the left image](#)

GLIGEN results on canny maps.

GLIGen: Open-Set Grounded Text-to-Image Generation

[Project Page] [Paper] [GitHub]

To ground concepts of interest with desired spatial specification, please (1) enter the concept names in *Grounding Instruction*, and (2) draw their corresponding bounding boxes one by one using *Sketch Pad* -- the parsed boxes will be displayed automatically.

For faster inference without waiting in queue, you may duplicate the space and upgrade to GPU in settings.  **Duplicate Space**

Task

Grounded Generation Grounded Inpainting

Language instruction

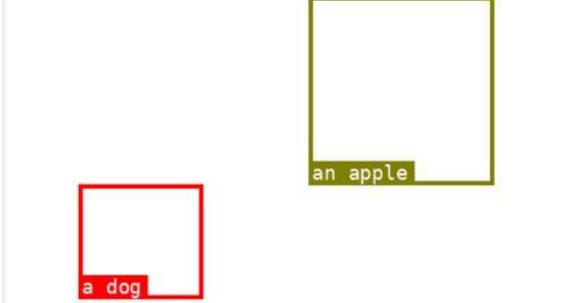
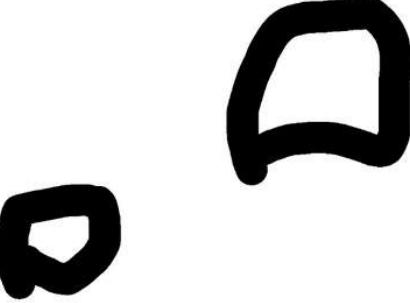
a dog and an apple

Grounding instruction (Separated by semicolon)

a dog;an apple

 Sketch Pad   

 Parsed Sketch Pad



Clear **Generate**

Generated Images



InstructPix2Pix

Learning to Follow Image Editing Instructions

Tim Brooks*, Aleksander Holynski*, Alexei A. Efros

University of California, Berkeley

*Denotes equal contribution

CVPR 2023 (Highlight)

arXiv

Code

Demo



Given an image and a written instruction, our method follows the instruction to edit the image.

Abstract

We propose a method for editing images from human instructions: given an input image and a written instruction that tells the model what to do, our model follows these instructions to edit the image. To obtain training data for this problem, we combine the knowledge of two large pretrained models---a language model (GPT-3) and a text-to-image model (Stable Diffusion)---to generate a large dataset of image editing examples. Our conditional diffusion model, InstructPix2Pix, is trained on our generated data, and generalizes to real images and user-written instructions at inference time. Since it performs edits in the forward pass and does not require per-example fine-tuning or inversion, our model edits images quickly, in a matter of seconds. We show compelling editing results for a diverse collection of input images and written instructions.



Input



"Add boats on the water"



"Replace the mountains with a city skyline"

Note that isolated changes also bring along accompanying contextual effects: the addition of boats also adds wind ripples in the water, and the added city skyline is reflected on the lake.



Input

"Apply face paint"

"What would she look like as a
bearded man?"

"Put on a pair of sunglasses"

"She should look 100 years old"



"What if she were in an anime?"

"Make her terrifying"

"Make her more sad"

"Make her James Bond"

"Turn her into Dwayne The Rock
Johnson"

Vermeer's *Girl with a Pearl Earring* with a variety of edits.

InstructPix2Pix: Learning to Follow Image Editing Instructions

For faster inference without waiting in queue, you may duplicate the space and upgrade to GPU in settings.

Duplicate Space

Generate

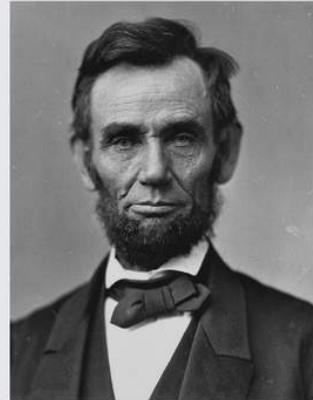
Load Example

Reset

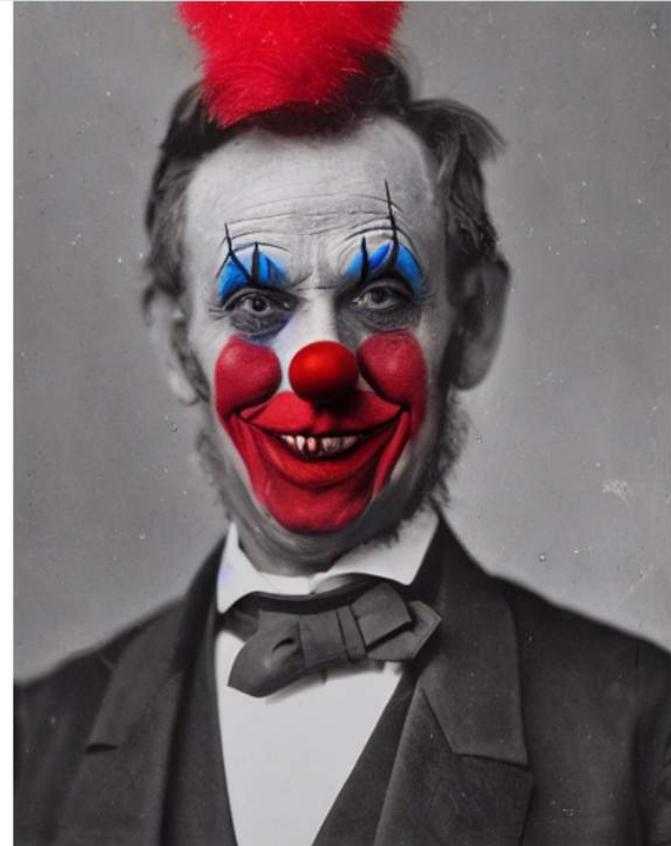
Edit Instruction

wearing a clown hat

Input Image



Edited Image



Steps

50

Fix Seed

Randomize Seed

Seed

9792

Fix CFG

Randomize CFG

Text CFG

7.5

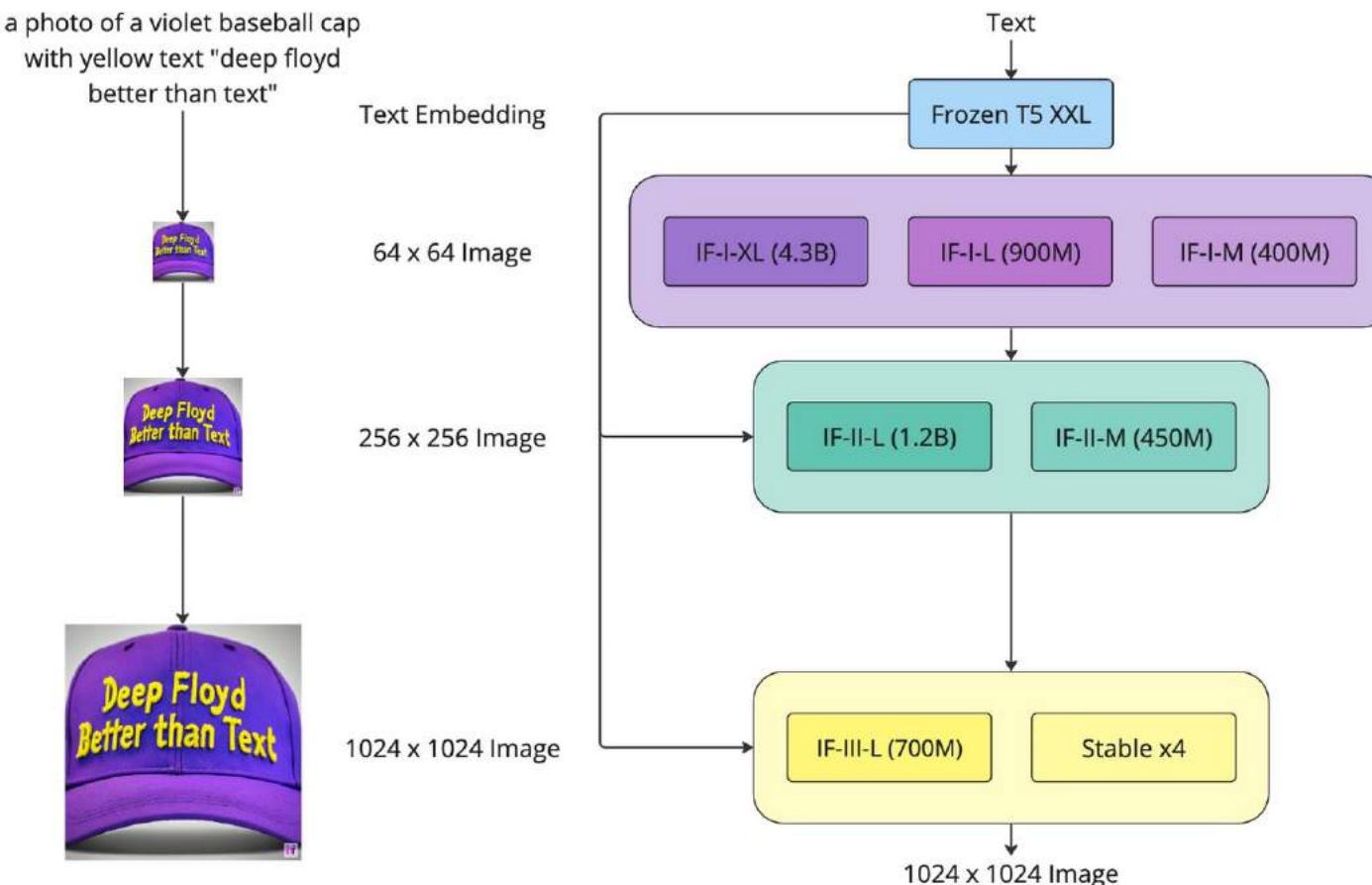
Image CFG

1.5

DeepFloyd IF by DeepFloyd, StabilityAI



We introduce DeepFloyd IF, a novel state-of-the-art open-source text-to-image model with a high degree of photorealism and language understanding. DeepFloyd IF is a modular composed of a frozen text encoder and three cascaded pixel diffusion modules: a base model that generates 64x64 px image based on text prompt and two super-resolution models, each designed to generate images of increasing resolution: 256x256 px and 1024x1024 px. All stages of the model utilize a frozen text encoder based on the T5 transformer to extract text embeddings, which are then fed into a UNet architecture enhanced with cross-attention and attention pooling. The result is a highly efficient model that outperforms current state-of-the-art models, achieving a zero-shot FID score of 6.66 on the COCO dataset. Our work underscores the potential of larger UNet architectures in the first stage of cascaded diffusion models and depicts a promising future for text-to-image synthesis.



I. Dream

Dream is the text-to-image mode of the IF model

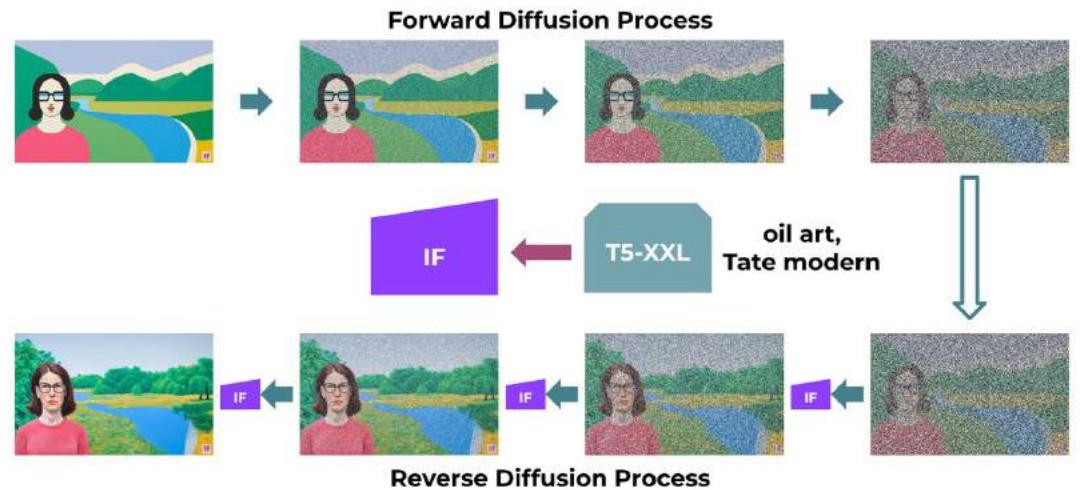
```
from deepfloyd_if.pipelines import dream

prompt = 'ultra close-up color photo portrait of rainbow owl with deer horns in the woods'
count = 4

result = dream(
    t5=t5, if_I=if_I, if_II=if_II, if_III=if_III,
    prompt=[prompt]*count,
    seed=42,
    if_I_kwarg={ "guidance_scale": 7.0, "sample_timestep_respacing": "smart100", },
    if_II_kwarg={ "guidance_scale": 4.0, "sample_timestep_respacing": "smart50", },
    if_III_kwarg={ "guidance_scale": 9.0, "noise_level": 20, "sample_timestep_respacing": "75", },
)
if_III.show(result['III'], size=14)
```



II. Zero-shot Image-to-Image Translation



In Style Transfer mode, the output of your prompt comes out at the style of the `support_pil_img`

```
from deepfloyd_if.pipelines import style_transfer

result = style_transfer(
    t5=t5, if_I=if_I, if_II=if_II,
    support_pil_img=raw_pil_image,
    style_prompt=[
        'in style of professional origami',
        'in style of oil art, Tate modern',
        'in style of plastic building bricks',
        'in style of classic anime from 1990',
    ],
    seed=42,
    if_I_kwargs={
        "guidance_scale": 10.0,
        "sample_timestep_respacing": "10,10,10,10,10,10,10,10,0,0",
        'support_noise_less_qsample_steps': 5,
    },
    if_II_kwargs={
        "guidance_scale": 4.0,
        "sample_timestep_respacing": 'smart50',
        "support_noise_less_qsample_steps": 5,
    },
)
if_I.show(result['II'], 1, 20)
```





resize



IF-II-L



IF-III-L

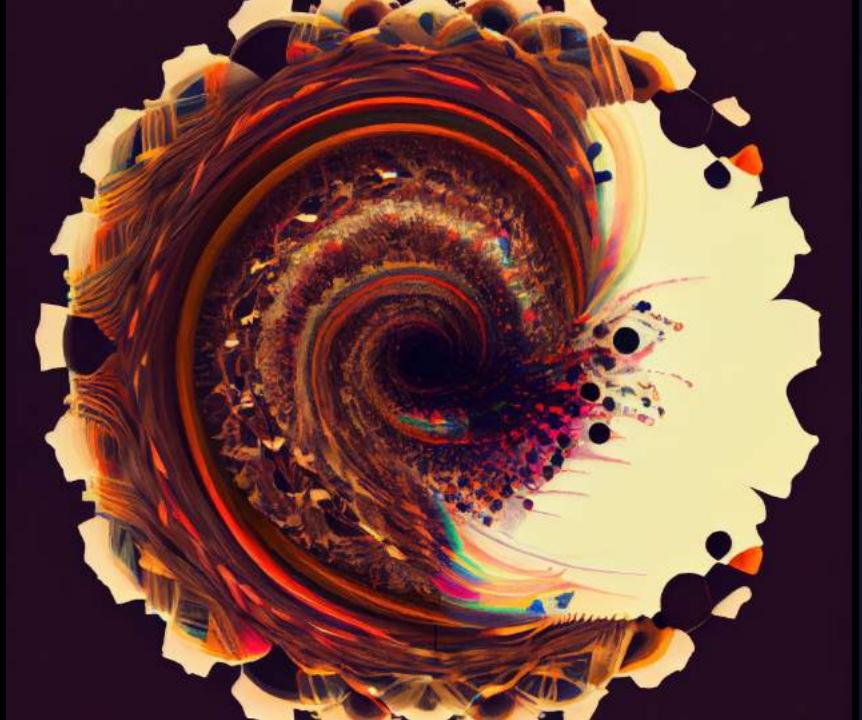


IV. Zero-shot Inpainting

```
from deepfloyd_if.pipelines import inpainting

result = inpainting(
    t5=t5, if_I=if_I,
    if_II=if_II,
    if_III=if_III,
    support_pil_img=raw_pil_image,
    inpainting_mask=inpainting_mask,
    prompt=[
        'oil art, a man in a hat',
    ],
    seed=42,
    if_I_kwargs={
        "guidance_scale": 7.0,
        "sample_timestep_respacing": "10,10,10,10,10,0,0,0,0,0",
        'support_noise_less_qsample_steps': 0,
    },
    if_II_kwargs={
        "guidance_scale": 4.0,
        'aug_level': 0.0,
        "sample_timestep_respacing": '100',
    },
    if_III_kwargs={
        "guidance_scale": 9.0,
        "noise_level": 20,
        "sample_timestep_respacing": "75",
    },
)
if_I.show(result['I'], 2, 3)
if_I.show(result['II'], 2, 6)
if_I.show(result['III'], 2, 14)
```





NEXT LESSON

- TO BE CONTINUED...
- REFERENCES: [BMEADOWS.XYZ/INPEAK](https://bmeadows.xyz/inpeak)

Lesson 5 - Getting Familiar with GAN

Summary - In this lesson, we will take a look at how GANs (Generative Adversarial Networks) are used to create and enhance art. We will go over GFPGAN, StyleGAN, ESRGAN, VQGAN+CLIP, and SFTGAN.

References

- [StackGAN GitHub code repo](#)
- [StyleGAN GitHub code repo](#)
- [VQGAN+CLIP GitHub code repo](#)
- [GFPGAN GitHub code repo](#)
- [GFPGAN hands-on web app](#)
- [CodeFormer GitHub code repo](#)
- [CodeFormer hands-on web app](#)
- [GFPGAN & CodeFormer hands-on web app](#)
- [GPEN GitHub code repo](#)
- [GPEN hands-on web app](#)
- [DeOldify GitHub code repo](#)
- [Palette.fm](#)
- [Real-ESRGAN GitHub code repo](#)
- [Real-ESRGAN hands-on web app](#)
- [Real-ESRGAN hands-on web app](#)
- [Real-ESRGAN hands-on web app](#)
- [ruDALL-E](#)
- [reDALL-E SR \(Real-ESRGAN\) hands-on web app](#)
- [SwinIR GitHub code repo](#)
- [SwinIR hands-on web app](#)
- [Latent SR hands-on web app](#)
- [SD Upscalers hands-on web app](#)
- [Ultimate Upscale for Automatic1111 GitHub code repo](#)
- [Topaz Gigapixel AI](#)
- [JoJoGAN hands-on web app](#)
- [AnimaGANv2 hands-on web app](#)
- [ArcaneGAN hands-on web app](#)
- [GigaGAN project page info](#)
- [GLIGEN project page info](#)
- [GLIGEN hands-on web app](#)
- [InstructPix2Pix project page info](#)
- [InstructPix2Pix hands-on web app](#)
- [InstructPix2Pix hands-on web app](#)
- [DeepFloyd IF GitHub code repo](#)
- [Midjourney: Multiprompt Image Blending April 15, 2023 Post](#) [twitter](#)
- [Five GANs for Better Image Processing article](#)
- [List of Projects that use CLIP reddit](#)
- [How Does CLIP Work? video](#)