Lecture 1: Introduction

Jun-Yan Zhu

16-726, Spring 2025

Teaching Staff

Instructors

Jun-Yan Zhu

junyanz at
cs.cmu.edu

Teaching Assistants

Zhixuan Liu

zhixuan2 at
andrew.cmu.edu

Jun-Yan Zhu

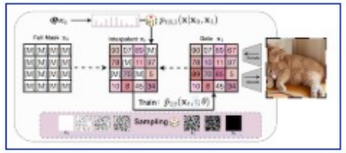
- Computer Vision, Computer Graphics, Machine Learning, Computational Photography
- Love pets (cat & dog)
- Swimming, tennis, traveling, gaming.

Cat Paper Collection

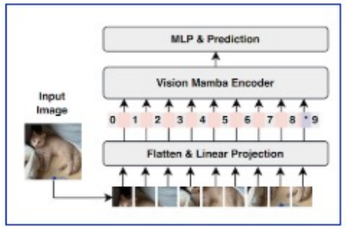
As reported by Cisco, 90% of net traffic will be visual, and indeed, most of the visual data are cat photos and videos. Thus, understanding, modeling, and synthesizing our feline friends becomes a more and more critical research problem these days, especially for our cat lovers.

Cat Paper Collection is an academic paper collection that includes computer graphics, computer vision, and machine learning papers that produce experimental results related to **cats**. If you would like to add/remove an article, please send an email to **Jun-Yan Zhu** (junyanz at cs dot cmu dot edu). We thank all the authors for their contribution and support.

See also **GitHub** | **CSV file**



MASK is All You Need
Vincent Tao Hu, Björn Ommer
In arXiv 2024
[Paper] [Project]

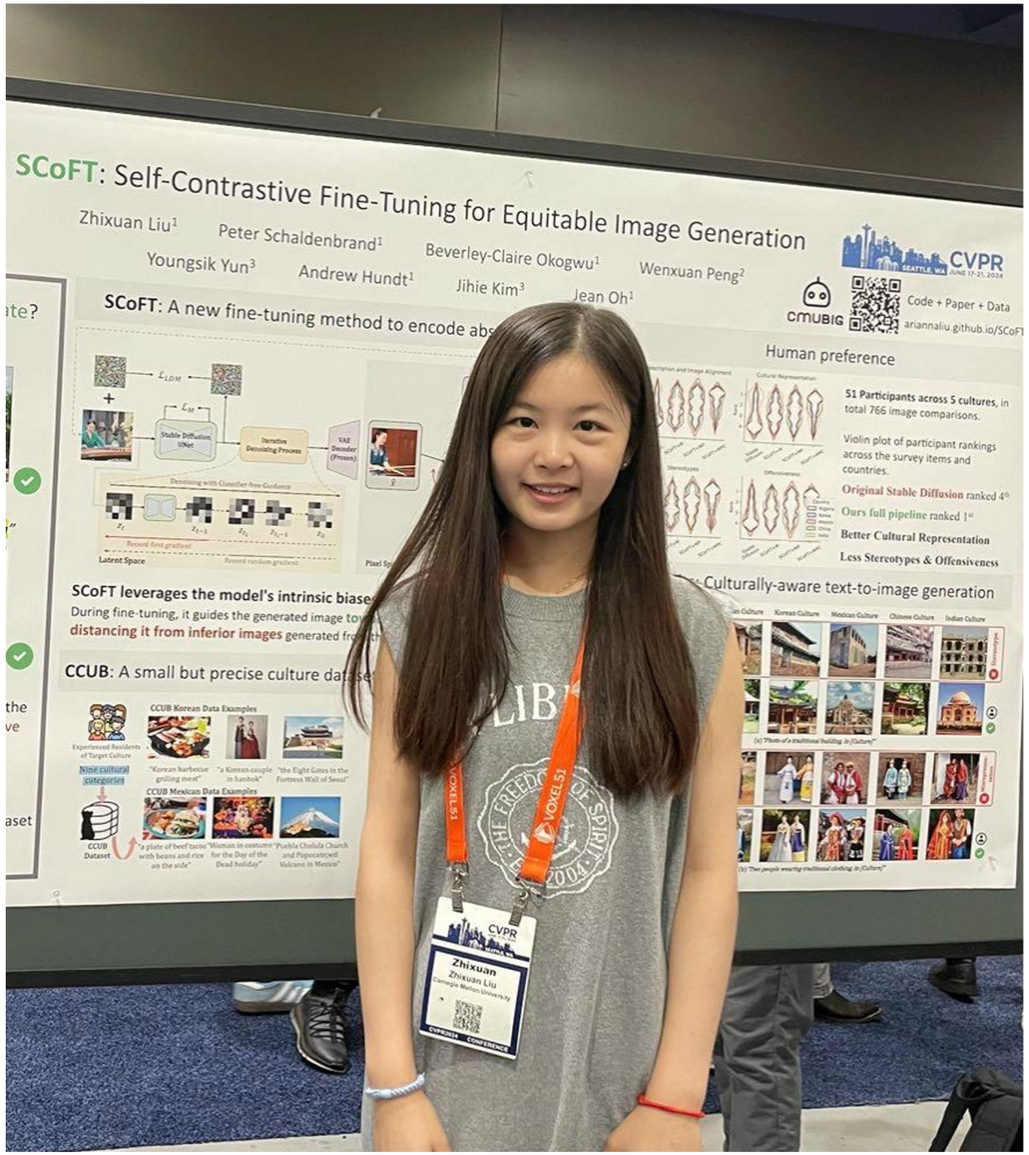


ZIGMA: A DiT-style Zigzag Mamba Diffusion Model
Vincent Tao Hu, Stefan Andreas Baumann, Ming Gui, Olga Grebenkova, Pingchuan Ma, Johannes Fischer, Björn Ommer
In ECCV 2024
[Paper] [Project]

Zero-shot Referring Expression Comprehension via Structural Similarity Between Images and Captions
Zeyu Han, Fangrui Zhu, Qianru Lao, Huaizu Jiang
In CVPR 2024
[Paper]

Zhixuan Liu

- PhD student at the Robotics Institute
- Advised by Prof. Jean Oh
- Interested in generative models with robotics



Visual Content Creation

Cave art

Time

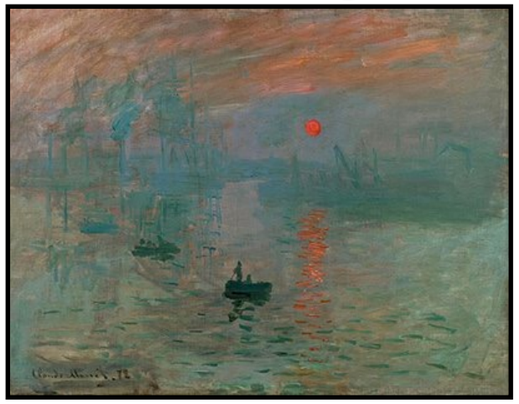
32,000 BC

Visual Content Creation

Cave art

Sculpture

Painting



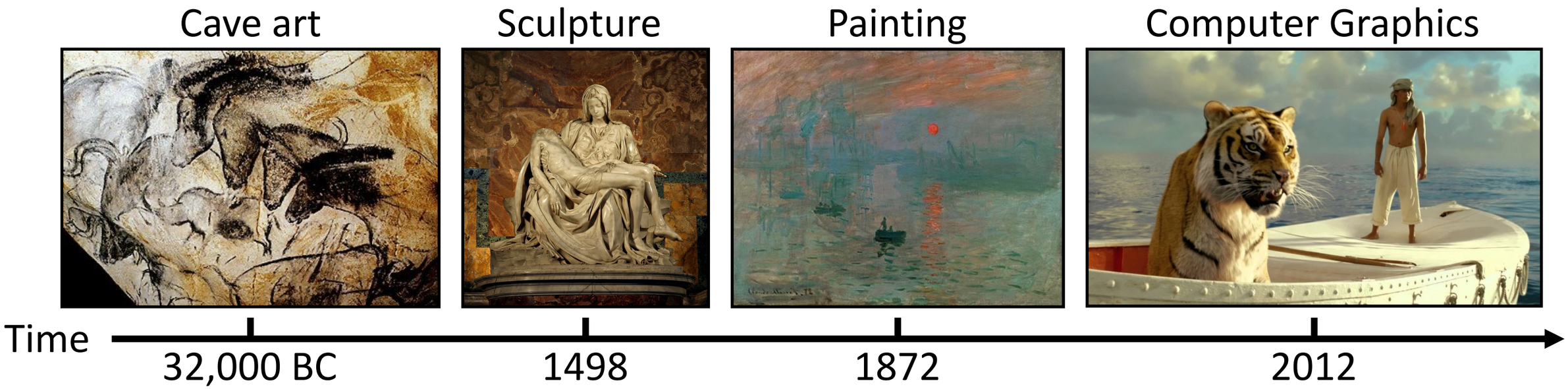
Time

32,000 BC

1498

1872

Visual Content Creation

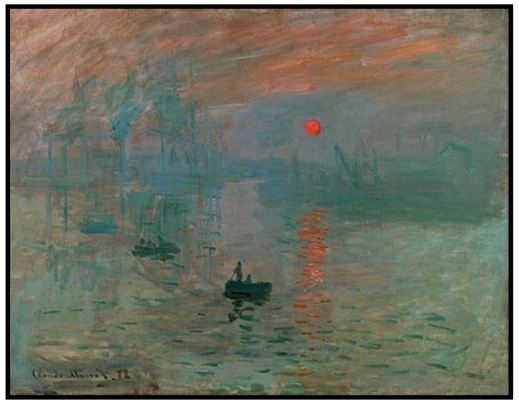


Who is creating visual content?

Cave art

Sculpture

Painting



Computer Graphics

Time

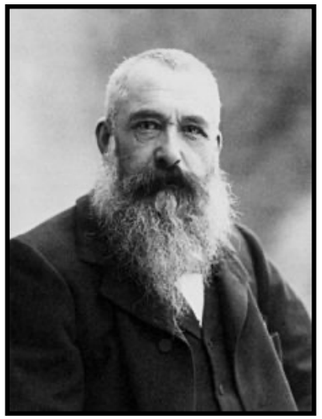
32,000 BC

1498

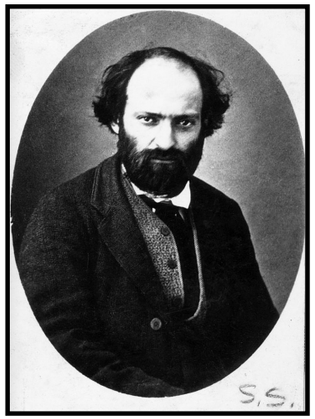
1872

2012

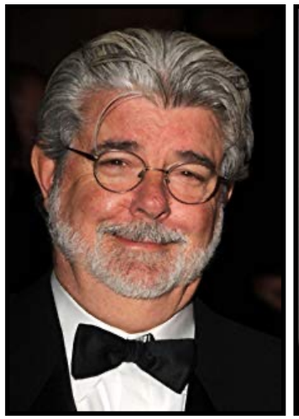
Michelangelo



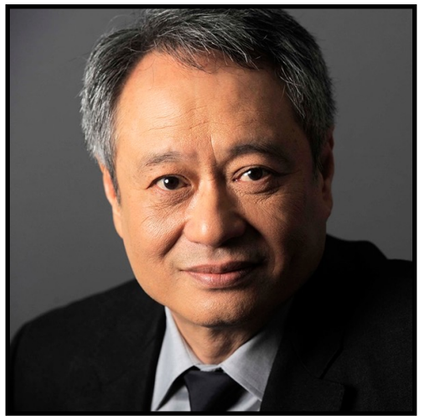
Claude Monet



Paul Cezanne

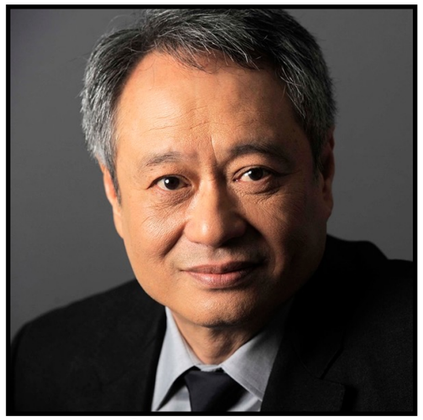


George Lucas

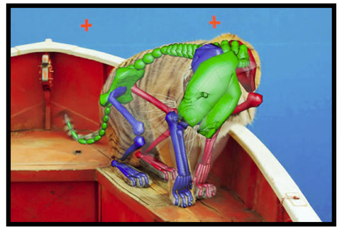


Ang Lee

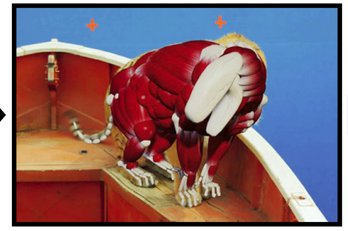
Who is creating visual content?



Ang Lee



Skeleton



Geometry

Texture

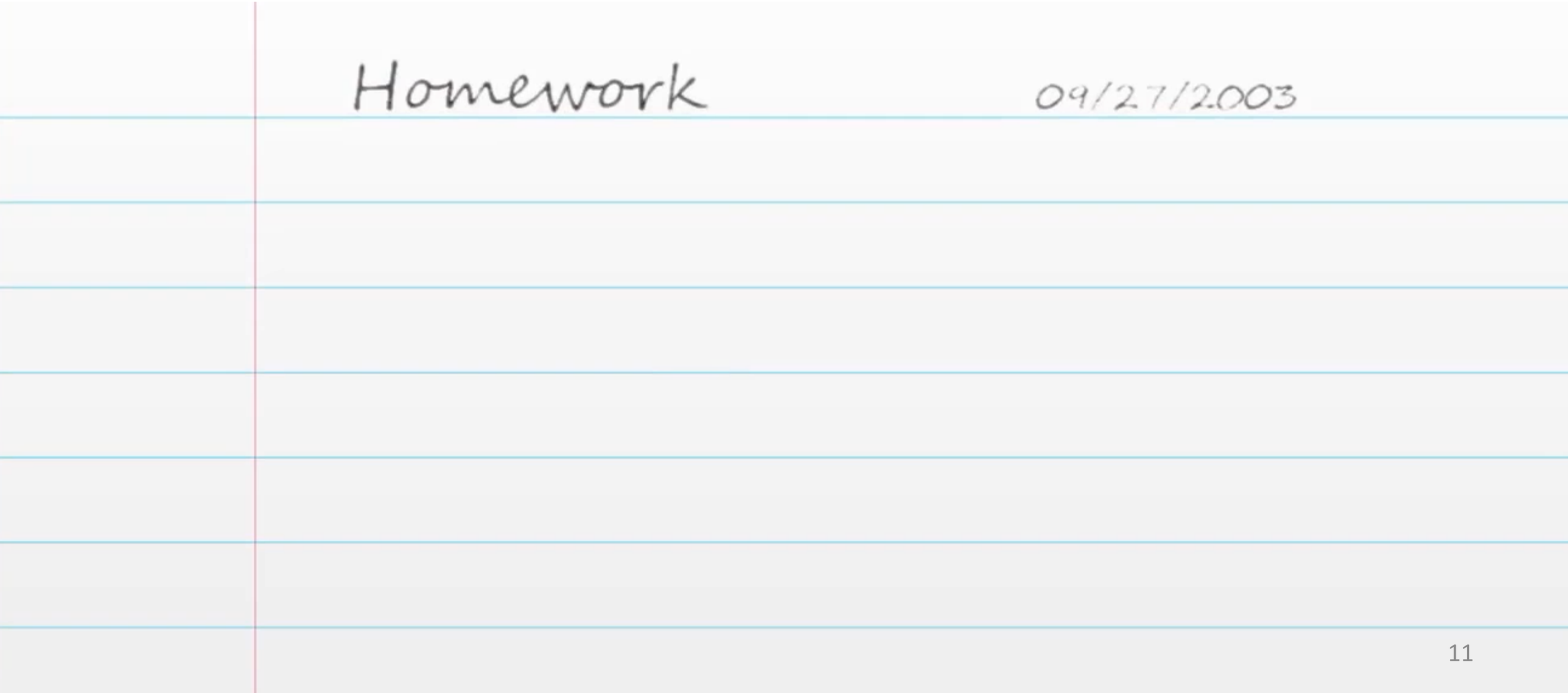
Details

Image

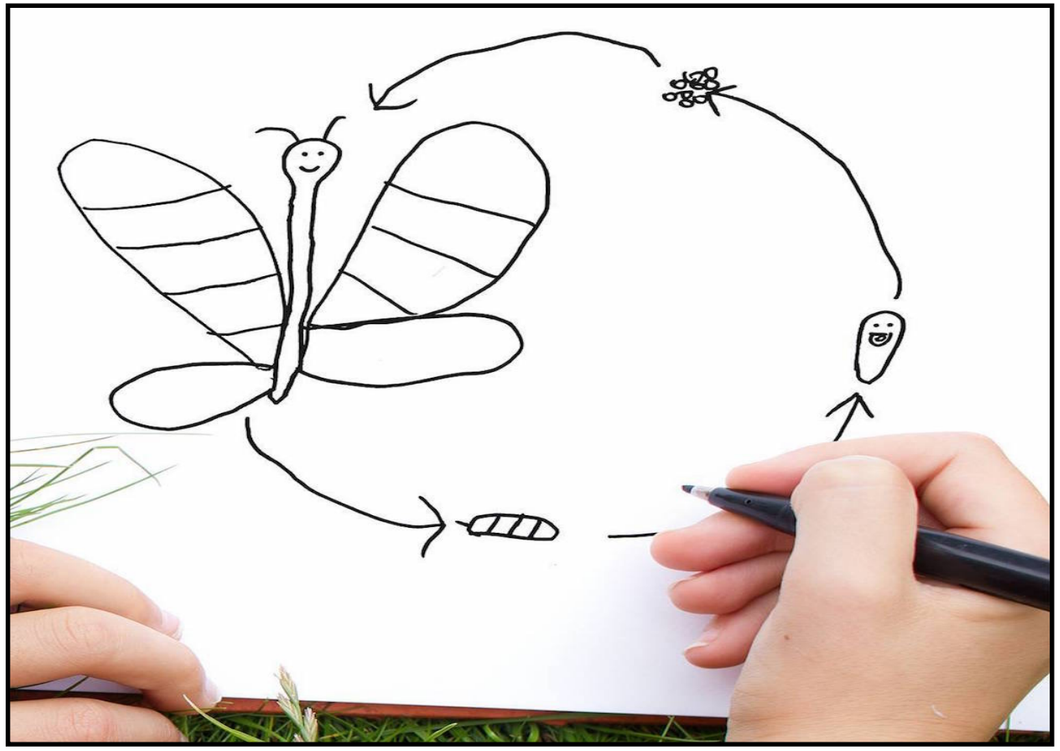
Idea

Visual Content

Who is creating visual content?



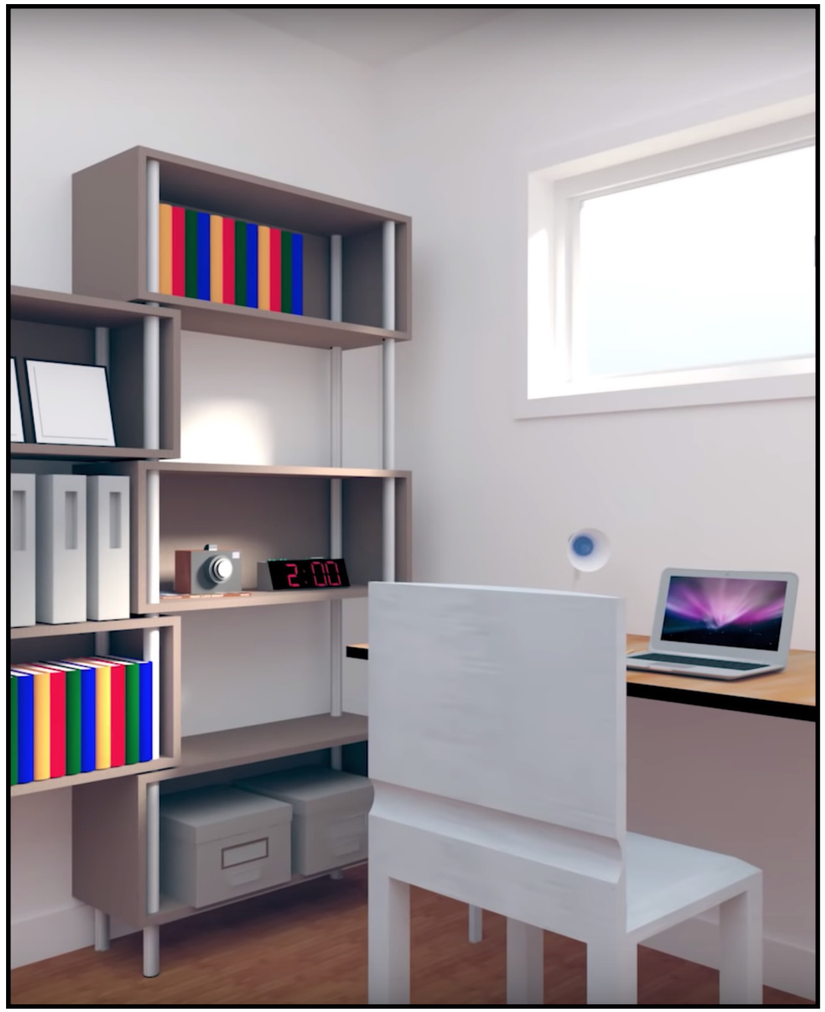
Who is creating visual content?



Kid’s drawing

Photoshop result by his father

Creating Visual Realism Manually



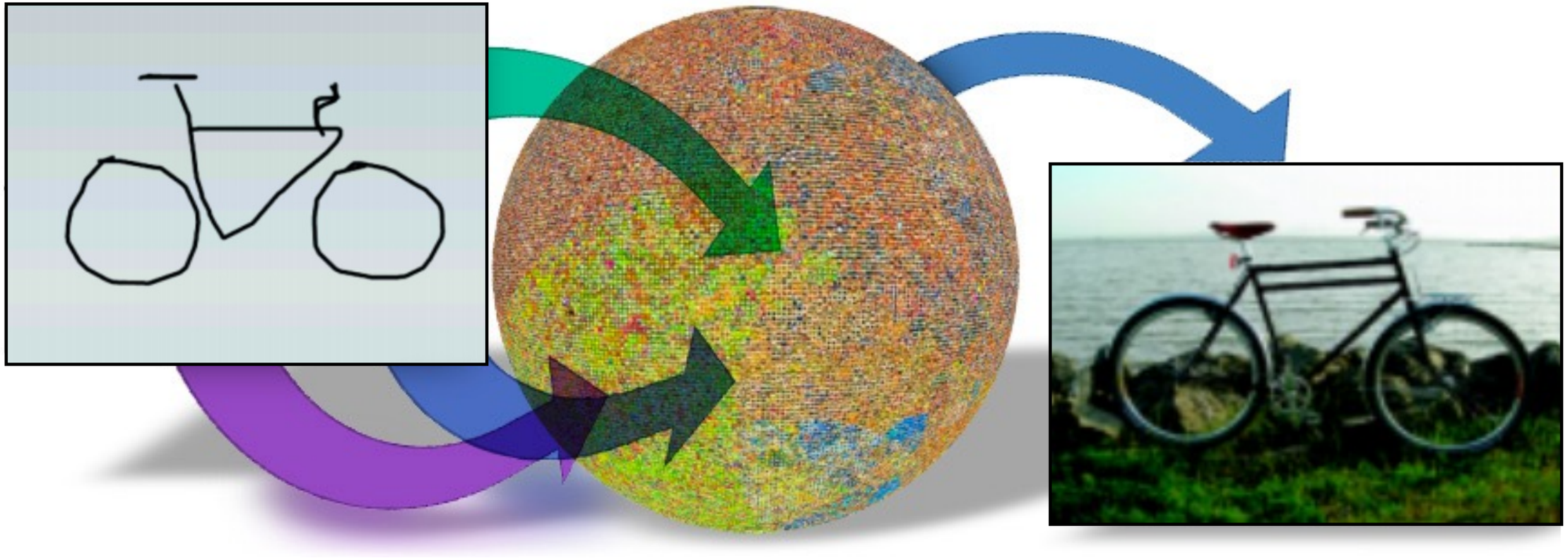
CG office

CG office (more details)

My advisor's office

Data-Driven Graphics (2000s)

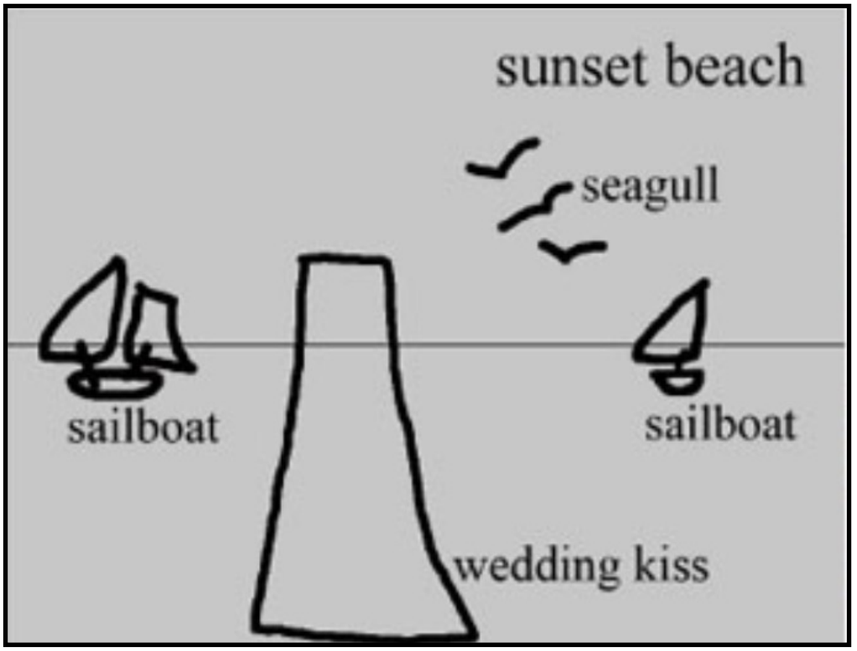
Graphics → Image Retrieval



Picture from James Hays

Data-Driven Graphics (2000s)

Compositing multiple parts



User Input

Database images

Output

Sketch2Photo [Tao et al. SIGGRAPH Asia 2009]

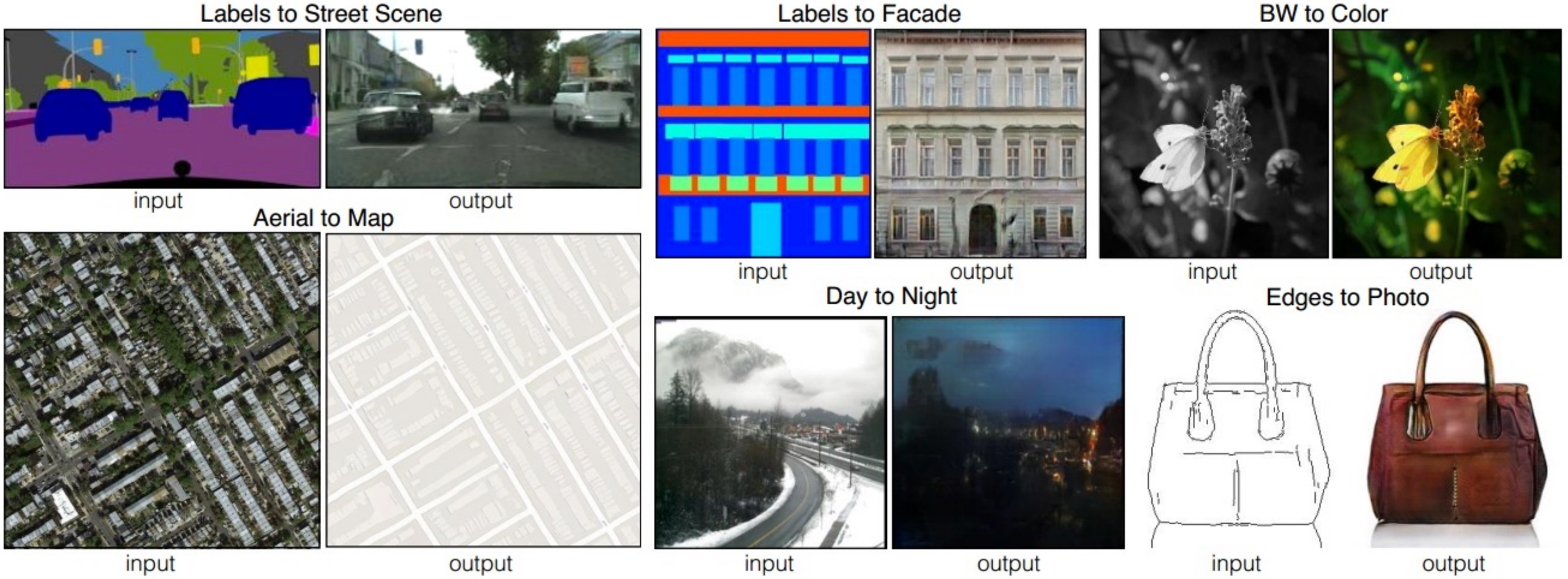
Data-Driven Graphics (2000s)

- Hard to combine pieces
- No understanding of visual realism

Help everyone
easily create visual content

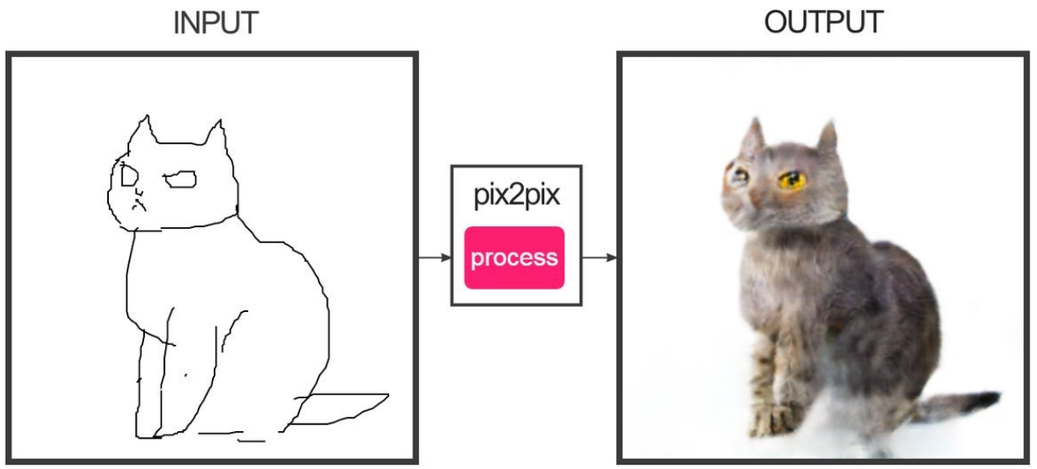
Teach machines
how to create realistic content

Image-to-Image Translation with **pix2pix**

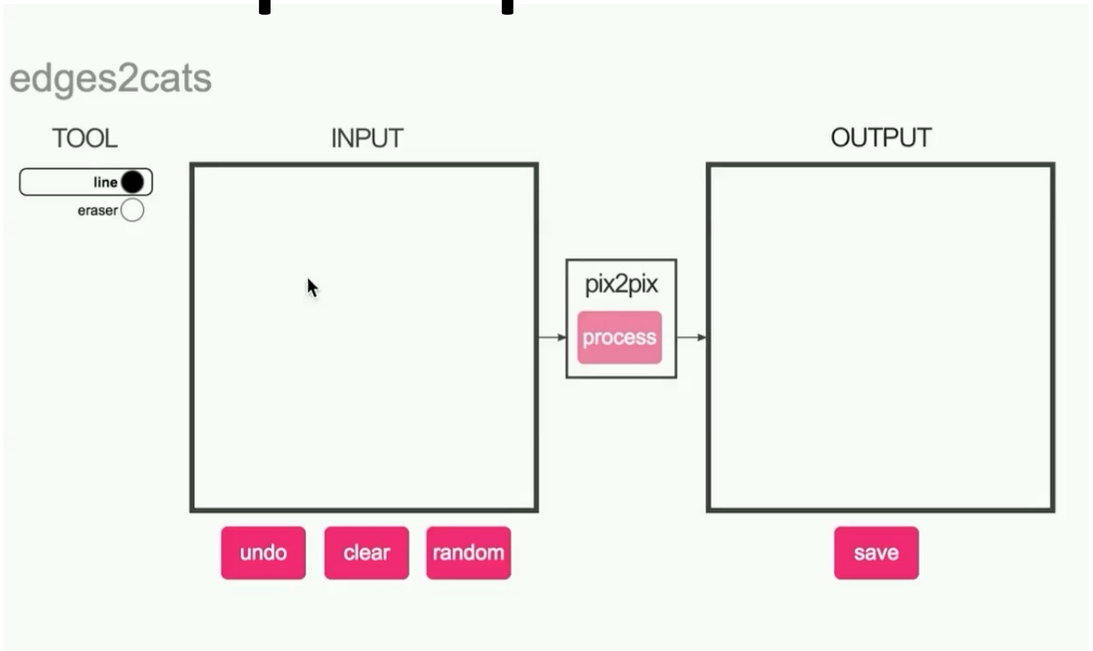


pix2pix [Isola, Zhu, Zhou, Efros. CVPR 2017]

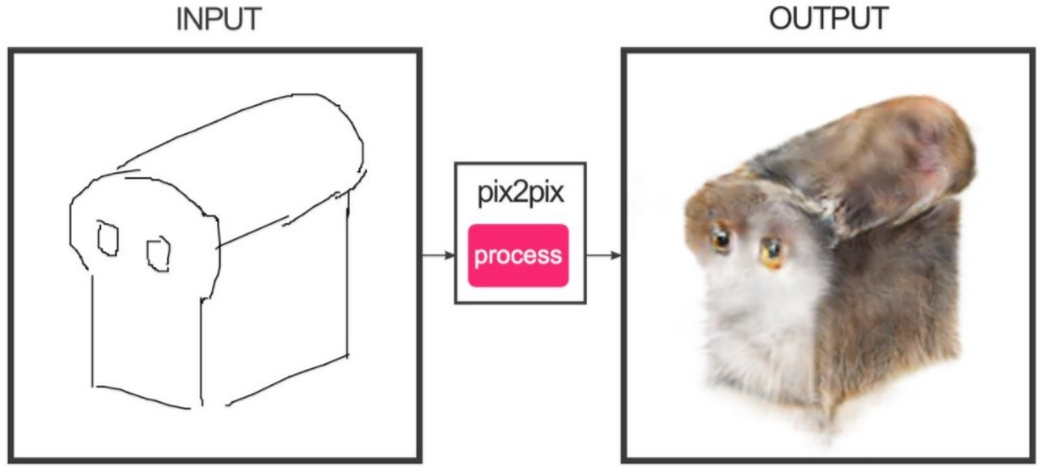
#edges2cats with pix2pix



@gods_tail



@matthematician

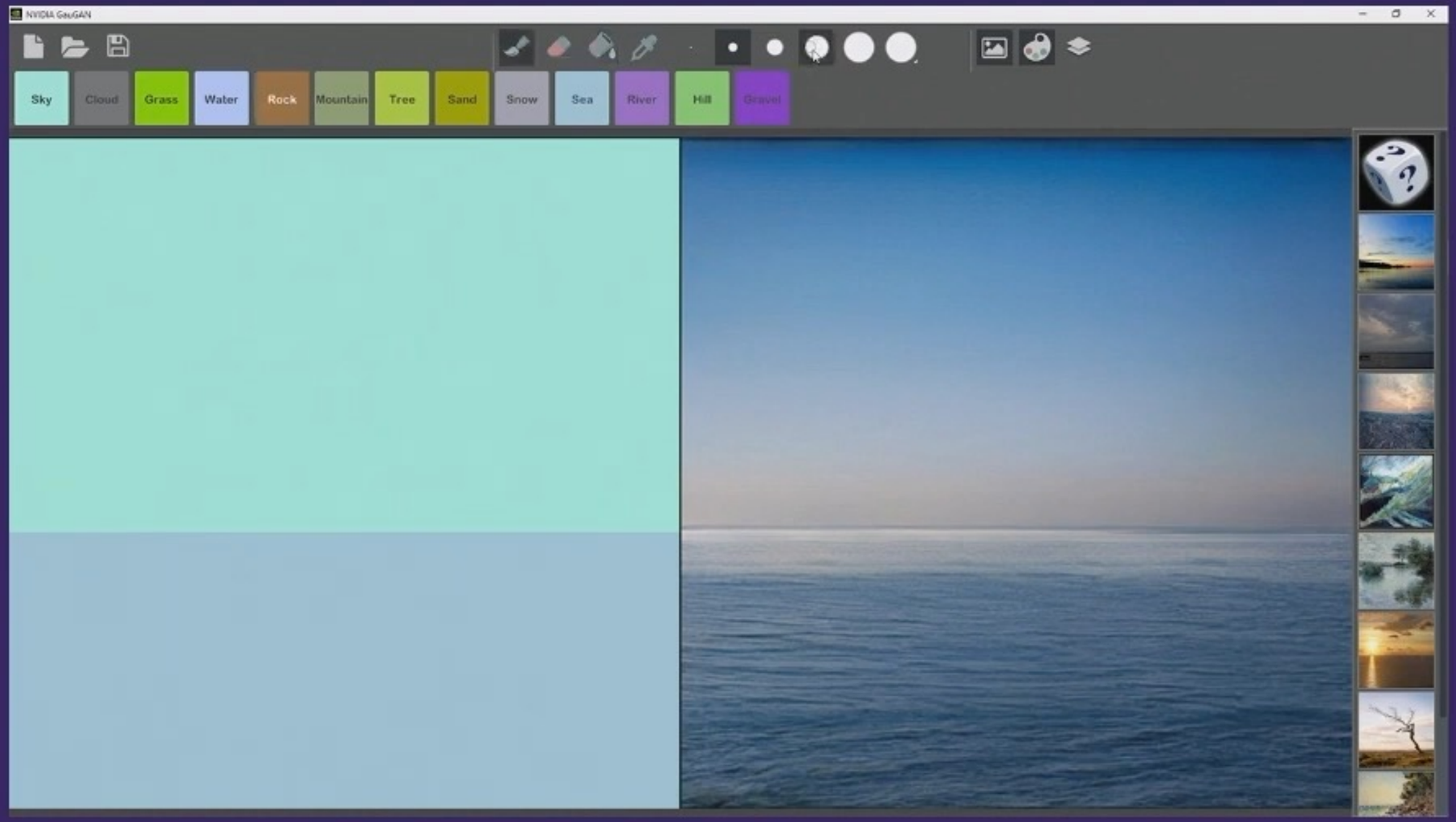


Ivy Tasi @ivymyt

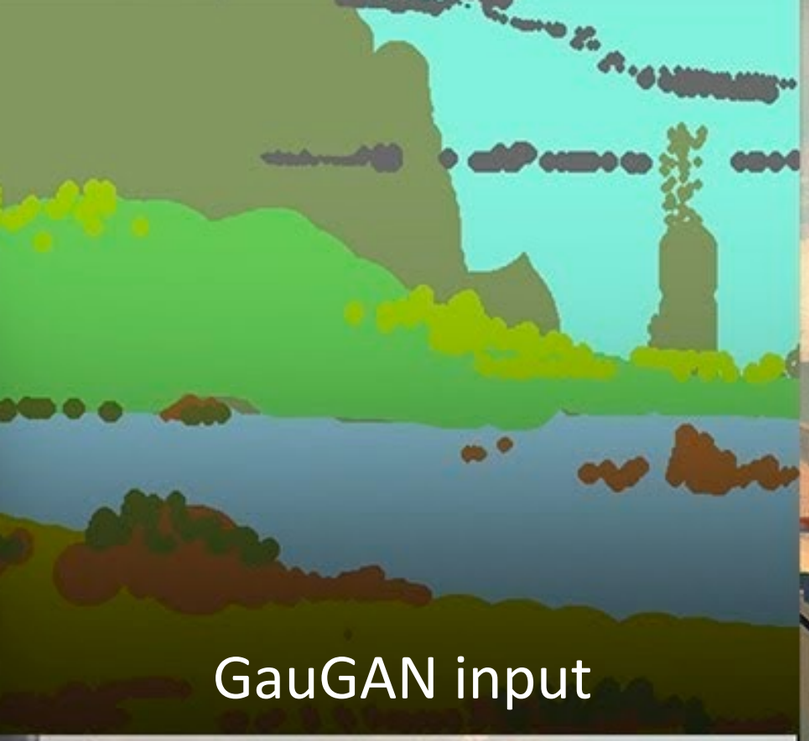
Vitaly Vidmirov @vvid

By Christopher Hesse https://affinelayer.com/pixsrv/

GauGAN [Park, Liu, Wang, Zhu. 2019]



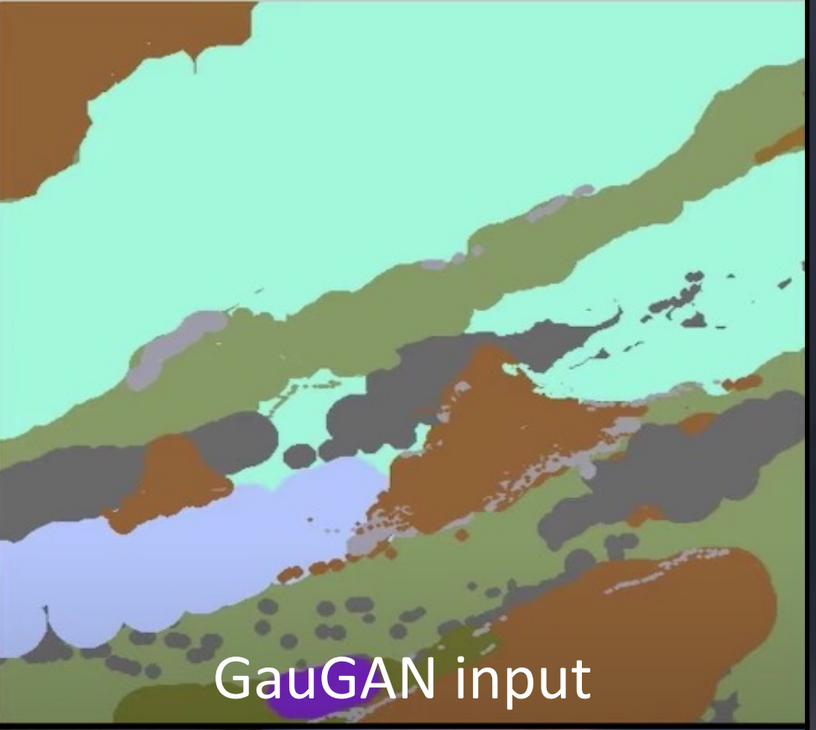
SIGGRAPH 2019 Real-time Live! “Best of Show Award" and "Audience Choice Award"



GauGAN input

GauGAN result

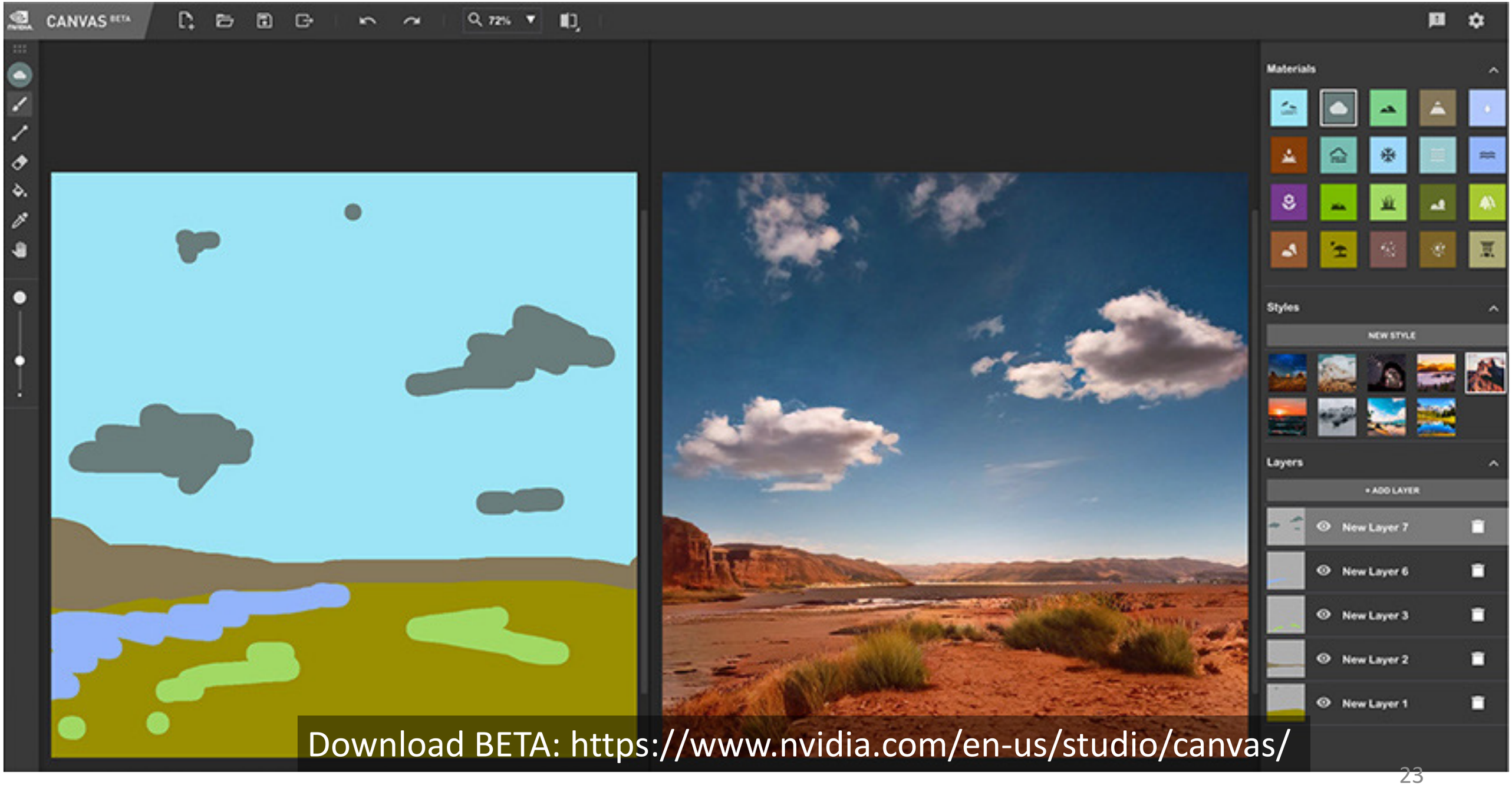
By Darek Zabrocki, Concept Designer and Illustrator



GauGAN input

GauGAN result

By Darek Zabrocki, Concept Designer and Illustrator



Download BETA: https://www.nvidia.com/en-us/studio/canvas/

Collection Style Transfer

Photograph ©Alexei Efros

Monet

Van Gogh

Cezanne

Ukiyo-e

CycleGAN [Zhu, Park, Isola, Efros. 2017]

Monet’s paintings → photographic style

CycleGAN [Zhu, Park, Isola, Efros. 2017]

Horse → Zebra

CycleGAN [Zhu, Park, Isola, Efros. 2017]

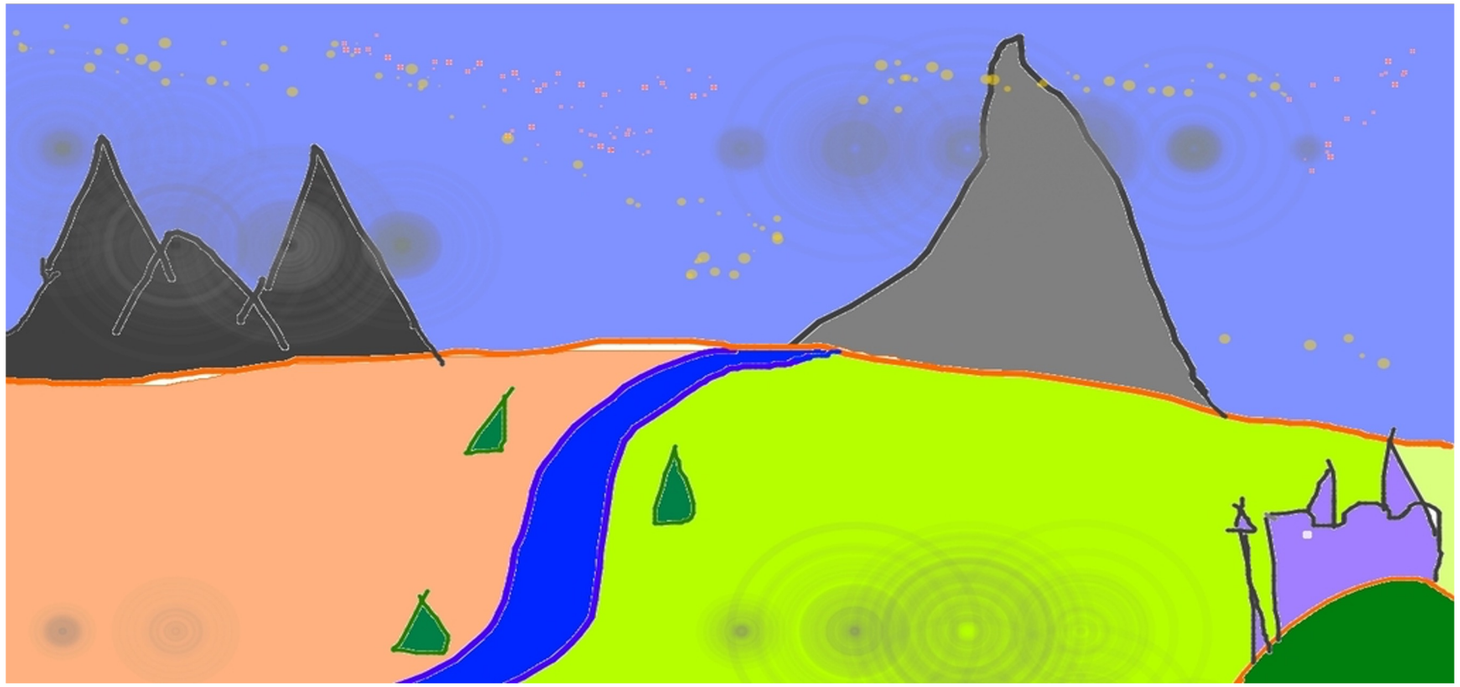
Failure case

Failure case

Swapping Autoencoder [Park et al. NeurIPS 2020]

SDEdit: Guided Image Synthesis with Diffusion

Input User Drawing



Used in Stable Diffusion Image-to-Image (“img2img’)

[Meng et al., ICLR 2022]

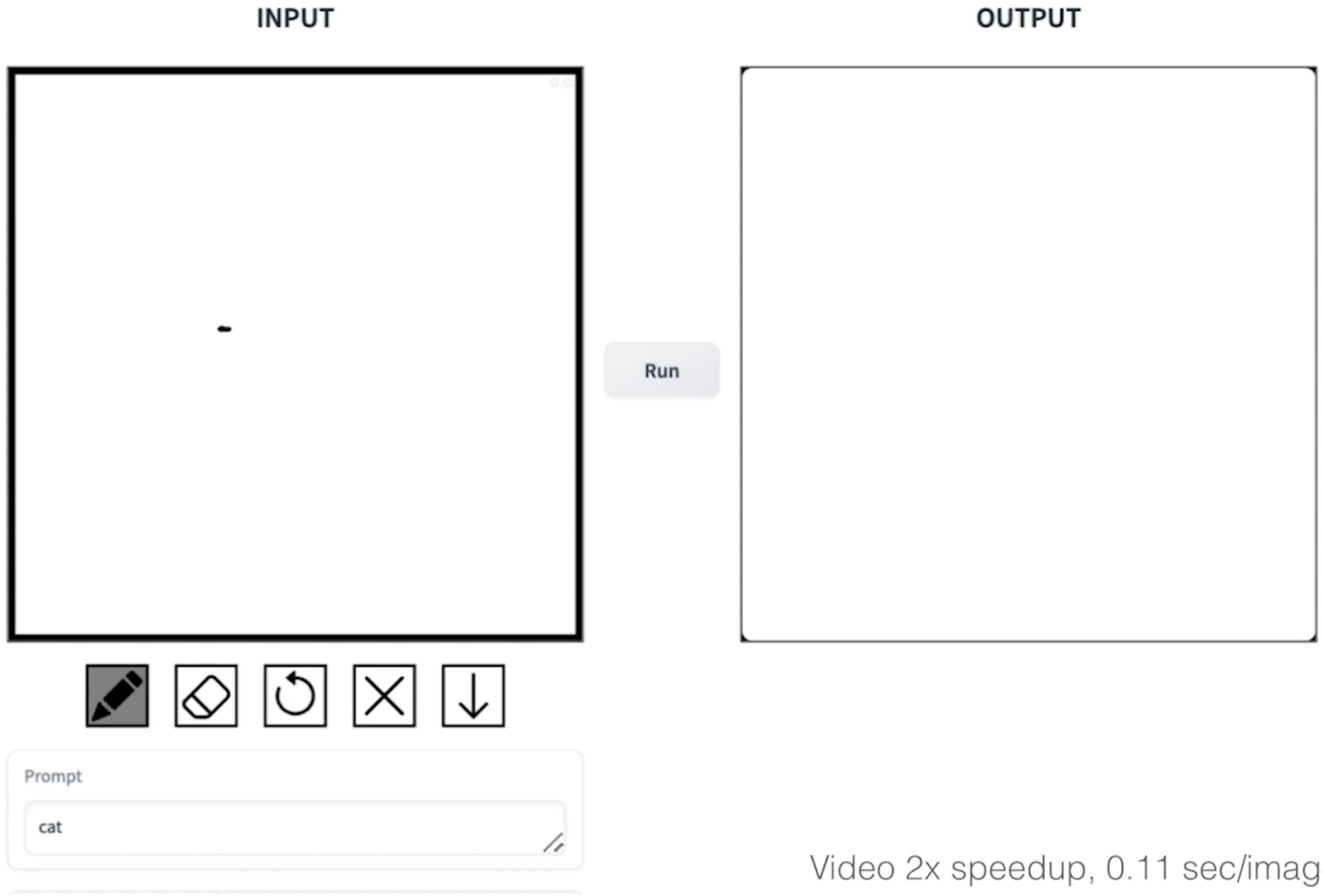
SDEdit: Guided Image Synthesis with Diffusion

Text prompt: “A fantasy landscape, trending on artstation”

Used in Stable Diffusion Image-to-Image (“img2img’)

[Meng et al., ICLR 2022]

pix2pix-turbo (2024)



Video 2x speedup, 0.11 sec/image on A100

[Parmer et al., 2024]

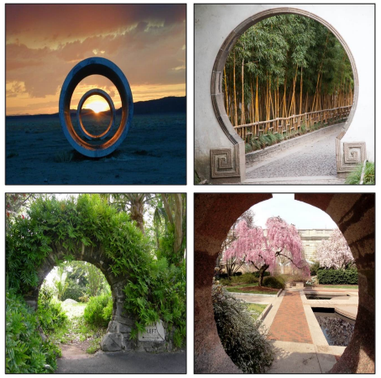
<not_ads>


Ps
Photoshop 22 Landscape Mixer

Photoshop 2021 Neural Filters

</not_ads>

Custom Stable Diffusion



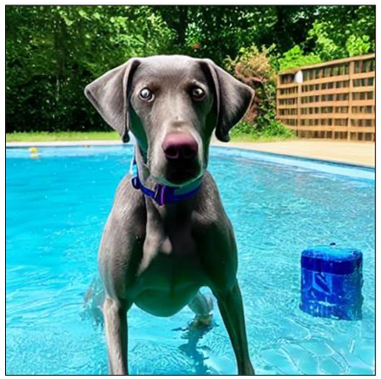
A photo of a **moongate**

A **moongate** in the snowy ice

A squirrel in front of **moongate**

A photo of a **V* dog**

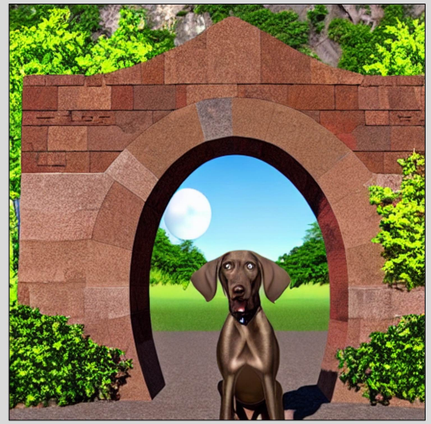
User input images



A **V* dog** in a swimming pool

A **V* dog** wearing sunglasses

Single-concept generation



A digital illustration of a **V* dog** in front of a **moongate**

V* dog wearing sunglasses in front of a **moongate**

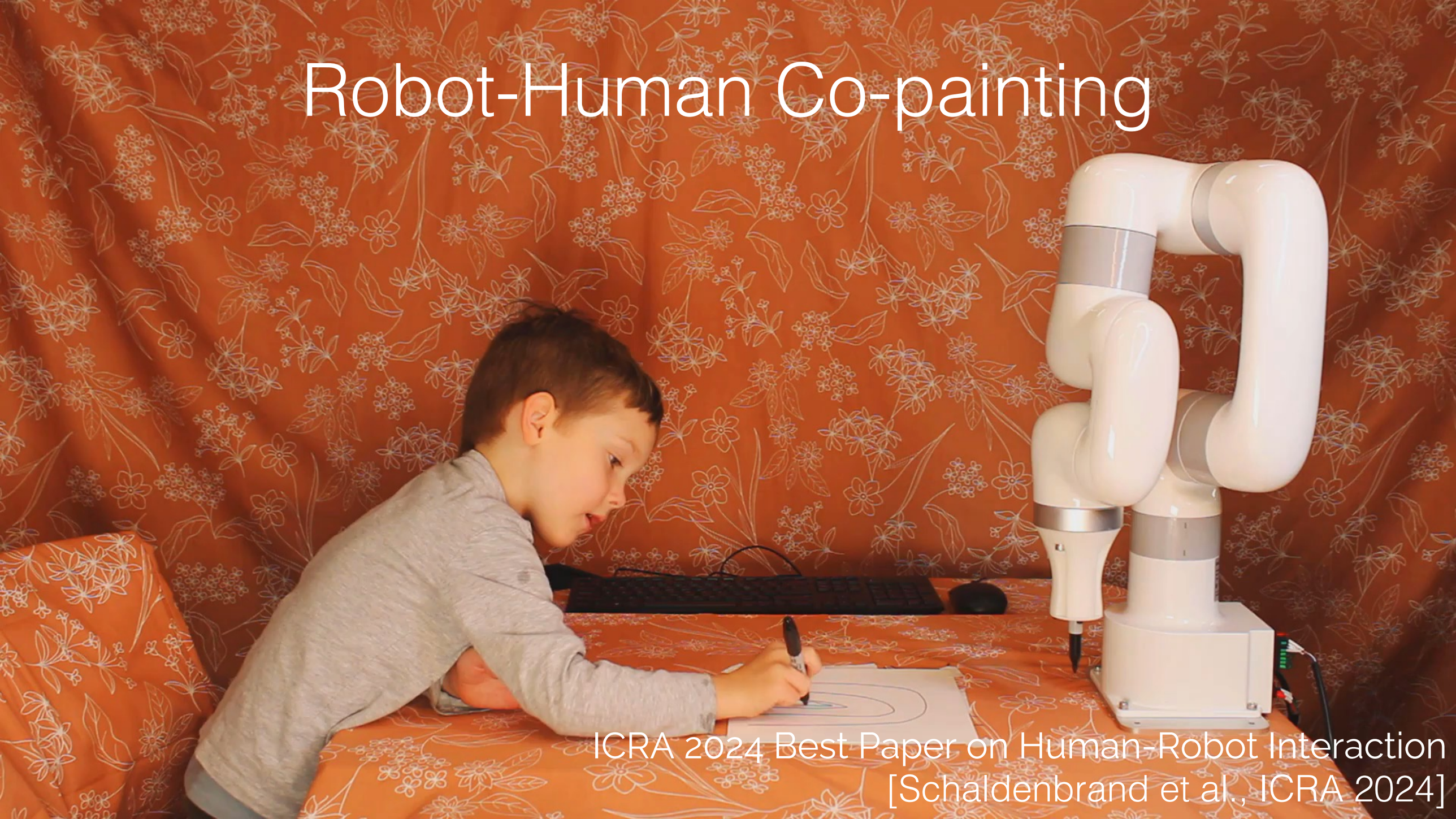
Multi-concept composition

custom-diffusion [Nupur Kumari et al., 2022]

FlashTex: Relightable Mesh Texturing

[Deng et al., ECCV 2024]

Robot-Human Co-painting



ICRA 2024 Best Paper on Human-Robot Interaction
[Schaldenbrand et al., ICRA 2024]

Diffusion2GAN

"Traditional gondolas lined up along the water, ready to transport visitors."

"Skiers enjoying the pristine slopes of the Swiss Alps on a sunny day."

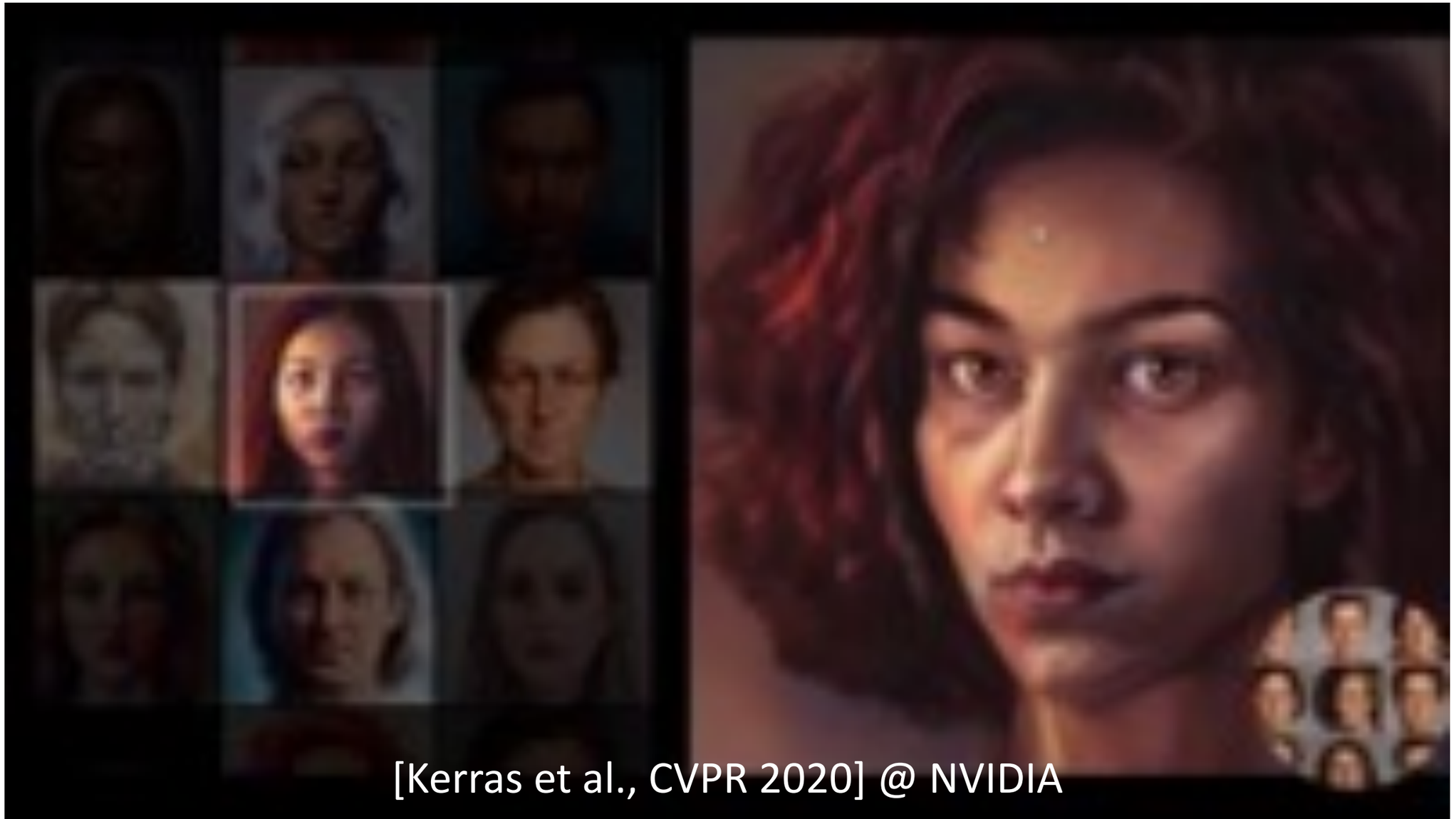
“Russian Blue cat exploring a garden, surrounded by vibrant flowers.”

Distilling Diffusion Models into Conditional GANs [Kang et al., ECCV 2024]

Research Highlights

from other universities & industry labs

Synthesizing High-res Portraits



[Kerras et al., CVPR 2020] @ NVIDIA

Everybody Dances Now

Caroline Chan , Shiry Ginosar, Tinghui Zhou, Alexei A. Efros, ICCV 2019. @ UC Berkeley

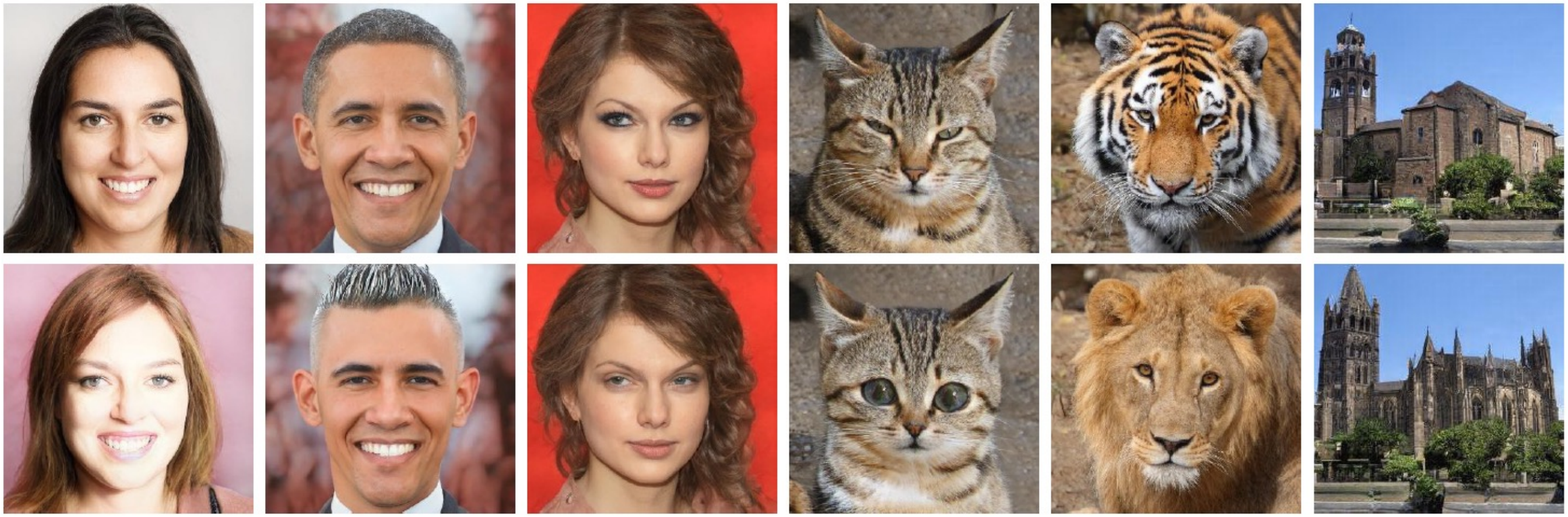
Neural Talking-Head Synthesis

face-vid2vid: One-Shot Free-View Neural Talking-Head Synthesis for Video Conferencing
Ting-Chun Wang, Arun Mallya, Ming-Yu Liu. CVPR 2021 @ NVIDIA

NeRF in the Wild

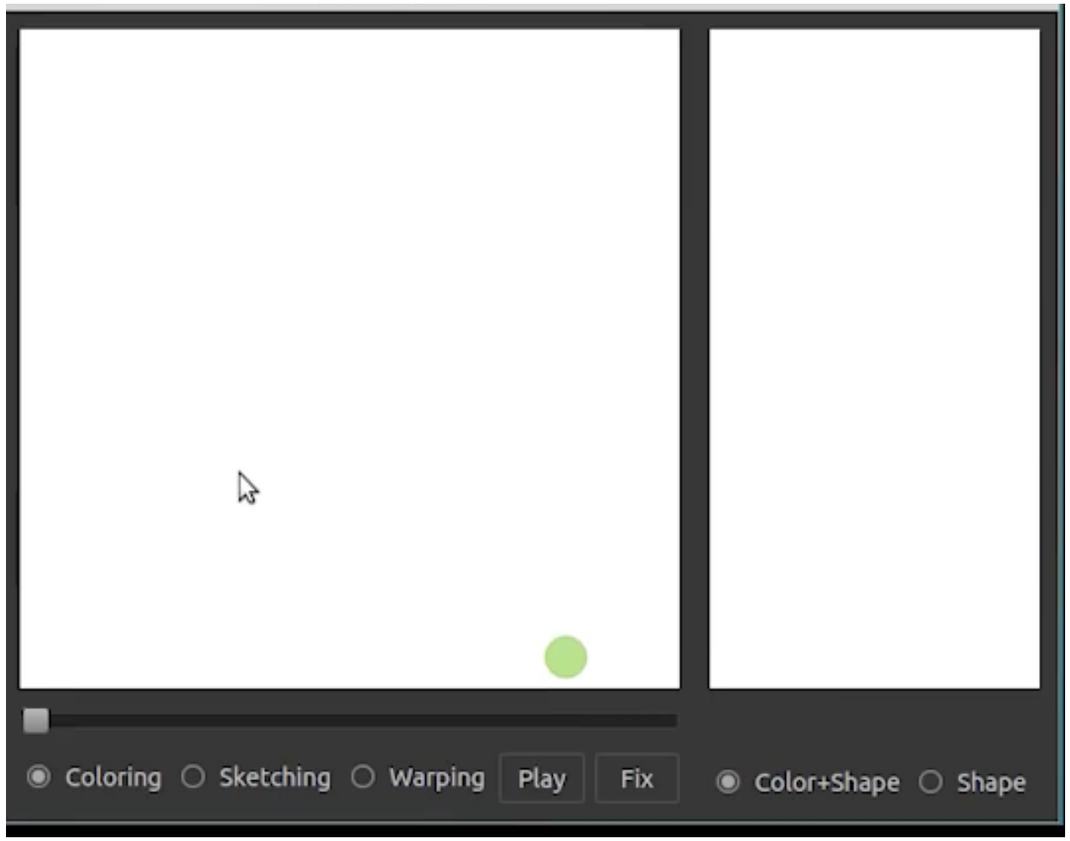
[Martin-Brualla et al., CVPR 2021] @ Google Research

Text-based Image Editing



StyleCLIP [Or Patashnik*, Zongze Wu*, et al., ICCV 2021]

Generative Models (2016)



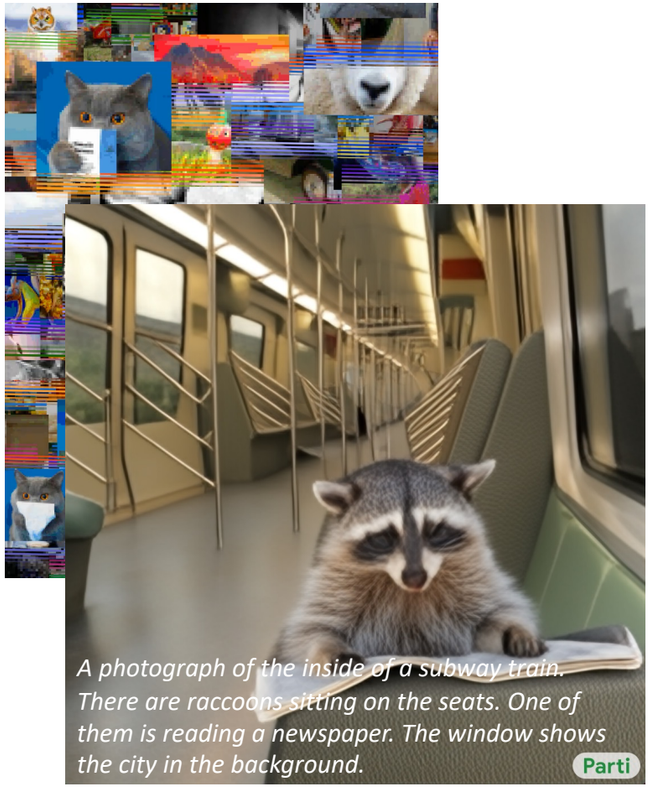
iGAN [Zhu et al., ECCV 2016]

A toilet seat sits open in the grass field.

Text2image [Mansimov et al., ICLR 2016]

Generative Models (2025)

Diffusion models
(DALL-E 2, Imagen, SD)



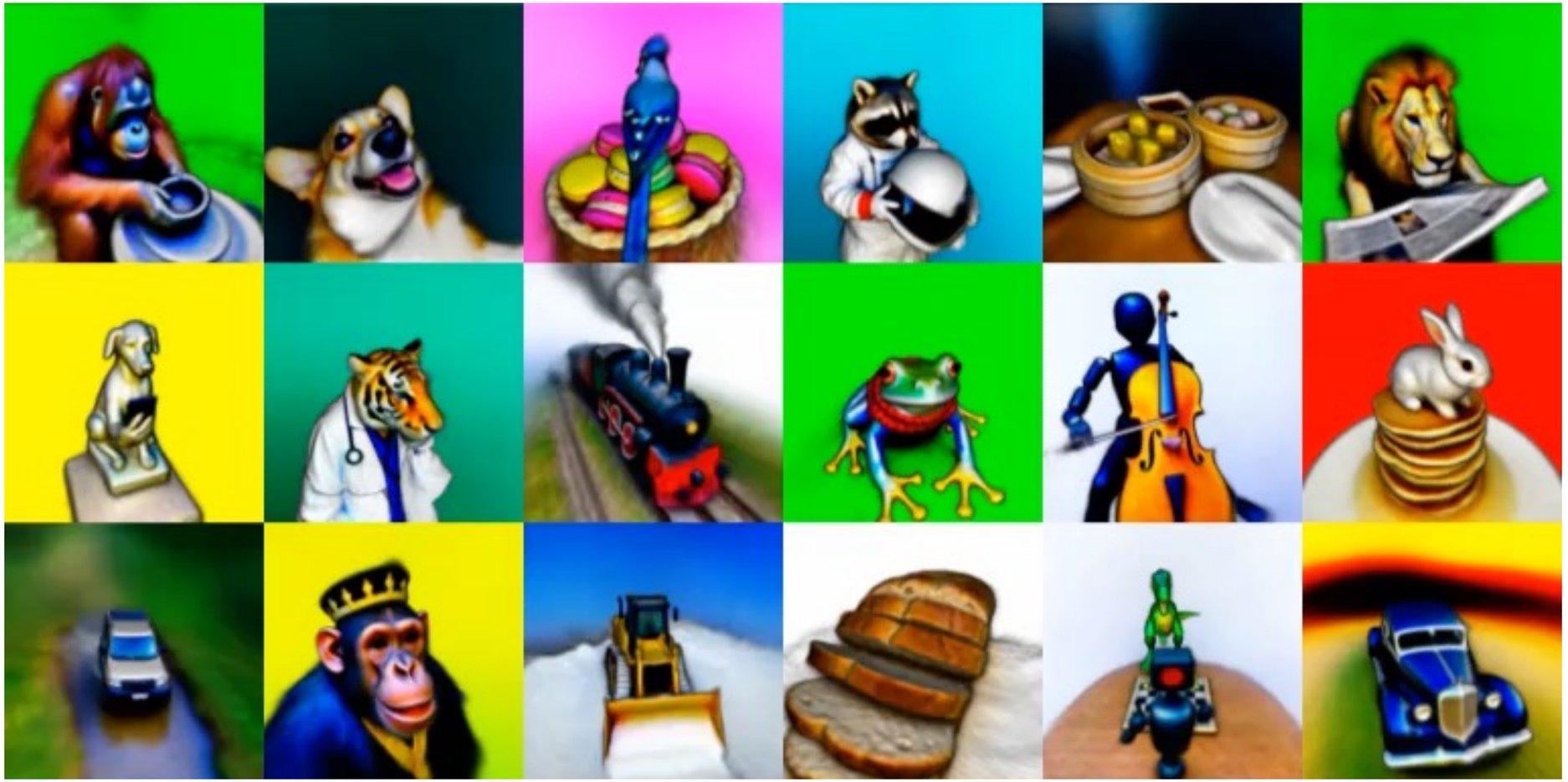
Autoregressive models
(Image GPT, Parti)

GANs, Masked GIT
(GigaGAN, MUSE)

Generative Models (2025)

By DALL·E 3

Generative Models (2025)



DreamFusion: Text-to-3D using 2D Diffusion [Poole et al., ICLR 2023] from UC Berkeley/Google

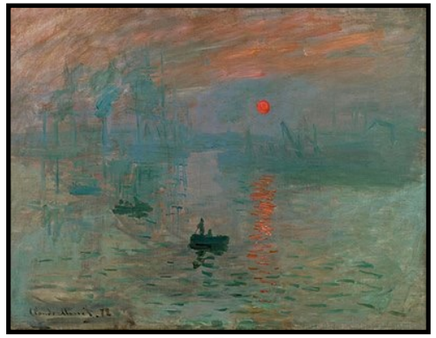
Cave art

32,000 BC

Sculpture

1498

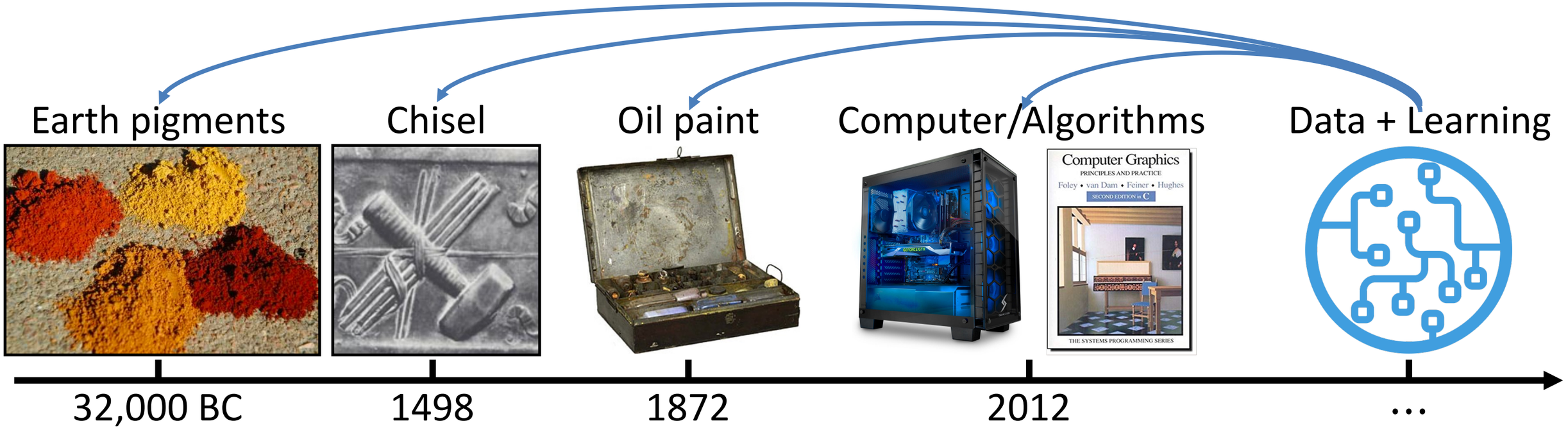
Painting



1872

Computer Graphics

2012



Course preview

- A modern machine learning perspective
- Widely-used learning algorithms
- Interactive content creation tools

Logistics

Course objectives

1. You will get a foundation in image editing and synthesis.
   - Texture synthesis and style transfer.
   - Face modeling and synthesis.
   - Image colorization and inpainting.
   - Video generation and editing.
   - Image-to-image translation.
   - Image and video editing. (warping, morphing, compositing)
   - Societal Implications: forensics, copyrights, biases.

Course objectives

2. You will get a foundation of machine learning concepts
   - (fast) Nearest neighbor search.
   - Principal component analysis, Gaussian Mixture model. Markov Random Field (MRF)
   - CNNs and transformers.
   - Deep generative models: Auto-encoder, Generative Adversarial Networks, Flow-based models, Variational Auto-encoder, Autoregressive Models, Diffusion Models.
   - Conditional generative models.
   - Neural Radiance Fields (NeRF).

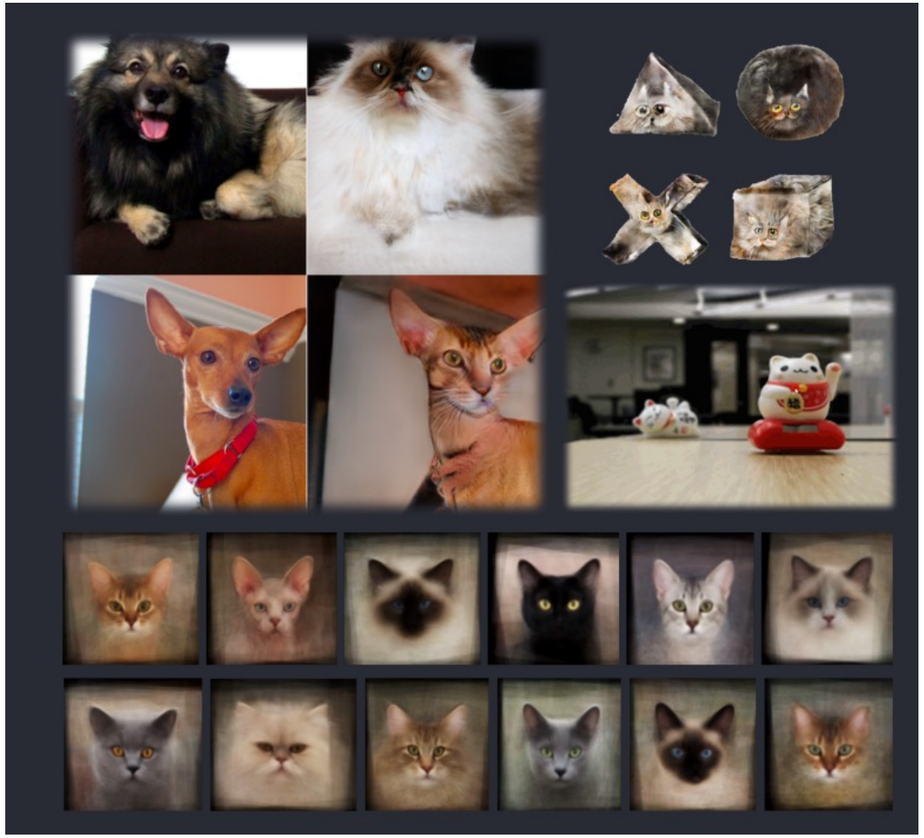
Course Schedule (Tentative)

Date	Topics	Assignments Due
M 01/13	Introduction	
W 01/15	Pointwise Processing and Image Filtering	HW0 out
M 01/20	No class (MLK Jr Day)	
W 01/22	Image Warping and Retargeting	
M 01/27	Data-driven Graphics	HW1 out
W 01/29	Convolutional Network for Image Synthesis	
M 02/03	Generative Adversarial Networks (part I)	
W 02/05	Generative Adversarial Networks (part II)	
M 02/10	Generative Models Zoo (part I)	
W 02/12	Generative Models Zoo (part II)	HW2 out, HW1 due
M 02/17	Generative Models Zoo (part III)	
W 02/19	Generative Models (student presentation)	
M 02/24	Image-to-Image Translation and Conditional Generative Models (part I)	
W 02/26	Image-to-Image Translation and Conditional Generative Models (part II)	HW3 out, HW2 due
M 03/03	No class (spring break)	
W 03/05	No class (spring break)	
M 03/10	Style and Content, Texture Synthesis	
W 03/12	Text-to-Image Synthesis	
M 03/17	Conditional Image Synthesis (student presentation)	HW4 out, HW3 due

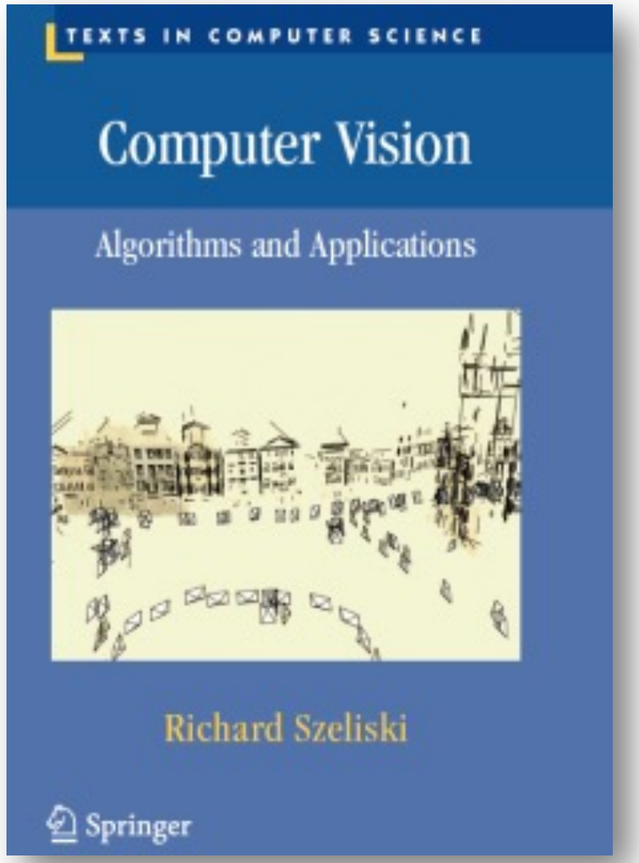
W 03/19	Image Editing with Optimization (part I)	
M 03/24	Image Editing with Optimization (part II)	Project proposal due
W 03/26	Image Editing (Student presentation)	
M 03/31	3D-aware Synthesis (part I)	HW4 due HW5 out
W 04/02	3D-aware Synthesis (part II)	
M 04/07	3D Content Creation (student presentation)	
W 04/09	Video Synthesis and Editing (part I)	
M 04/14	Video Synthesis and Editing (part II)	HW5 due
W 04/16	Societal Implications (part I) Guest Lecture	
M 04/21	Societal Implications (part II) Guest Lecture	
W 04/23	Final project presentation	Project Due (05/01)

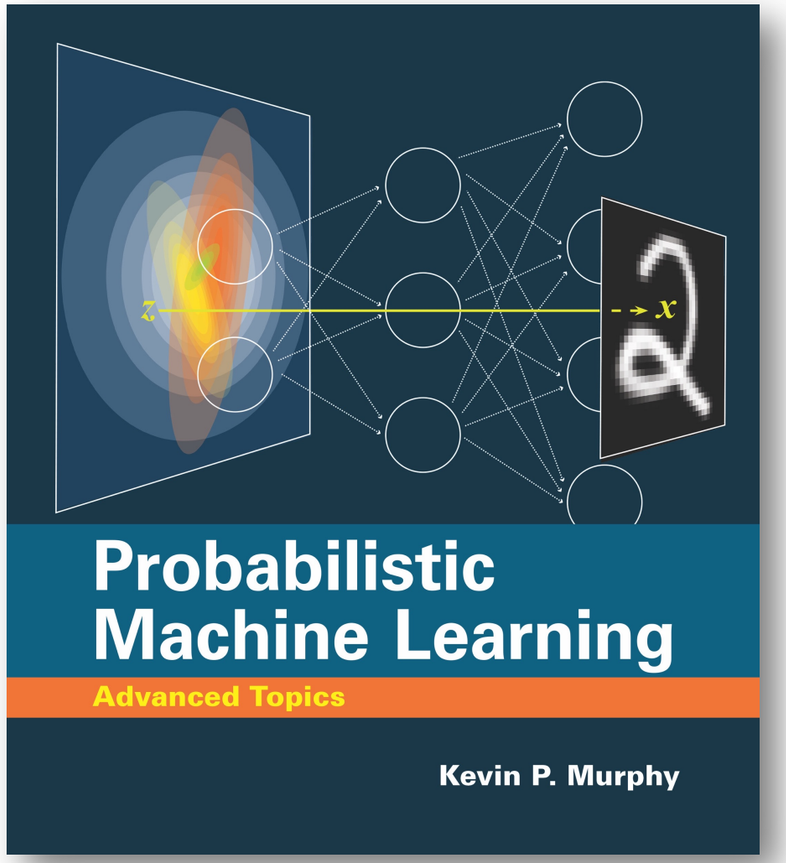
Course objectives

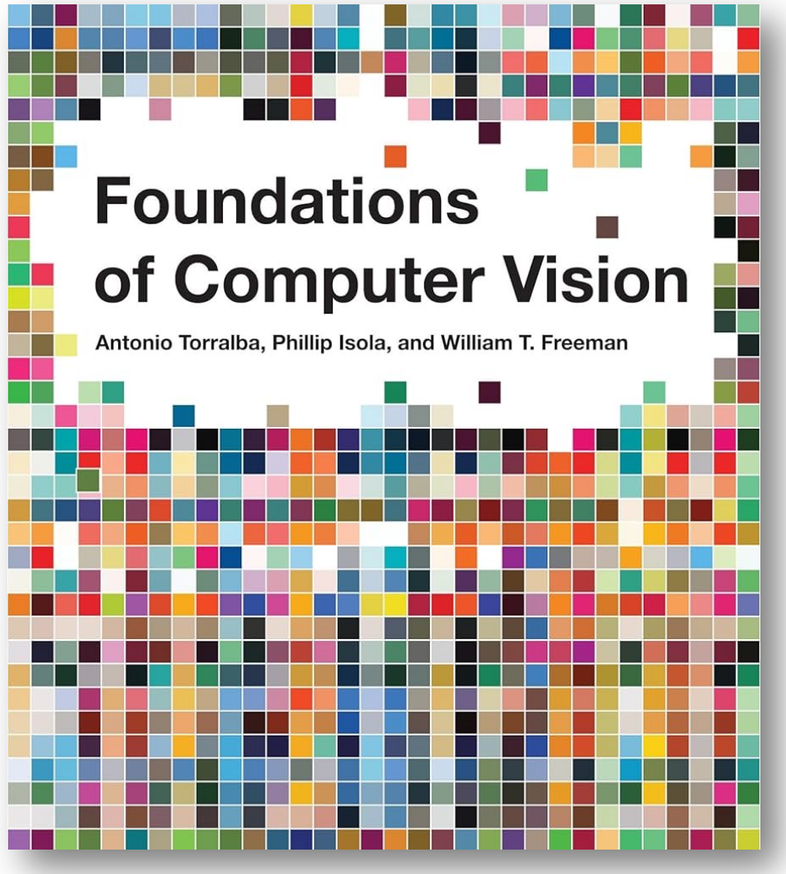
3. You will have some cool results with your own photos



Textbook (optional)







https://szeliski.org/Book/

https://probml.github.io/pml-book/book2.html

https://mitpress.mit.edu/9780262048972/foundations-of-computer-vision/

Grading

- Emphasis on programming projects (**60%**).
  - Two Classic assignments: 1. image alignment. 2. image blending
  - Deep learning: 3. Generative model training (GANs & Diffusion). 4. neural style transfer. 5. imageediting with GANs & Diffusion.
- Late Policy for programming assignments.
  - Five (5) emergency late days for semester, to be spent wisely
  - 10% of penalty per 24 hours afterwards
- Class attendance (**5%**)
- One paper presentation (**10%**):
  - 10-20 min, 2-3 people in a group.
  - Need to answer questions about this paper from now on.
- Final Project (**25%**)
  - A webpage-based report + a presentation.
  - No late day.
  - 2-3 people per group.

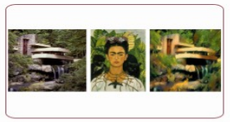
Assignments and Project

Assignment #0 - How to submit assignments?

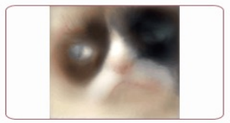
Assignment #1 - Colorizing the Prokudin-Gorskii Photo Collection
Winner: [Chiyuan Fu]
Honorable Mentions: [Seah Shao Xuan] [Junkai Huang] [Zoë LaLena]

Assignment #2 - Gradient Domain Fusion
Winner: [Max Grebinskiy] [Zoë LaLena]
Honorable Mentions: [Chiyuan Fu] [Yu-Chen Lin]

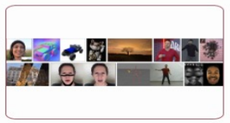
Assignment #3 - Cats Generator Playground
Winner: [Seah Shao Xuan]
Honorable Mentions: [Aviral Agrawal]



Assignment #4 - Neural Style Transfer
Winner: [Jing Gao] [Ming Chong Lim]
Honorable Mentions: [Michael Mu]



Assignment #5 - Cats Photo Editing
Winner: [Jing Gao]
Honorable Mentions: [Michael Mu] [Benran Hu] [Ming Chong Lim]



Final Project

For each assignment

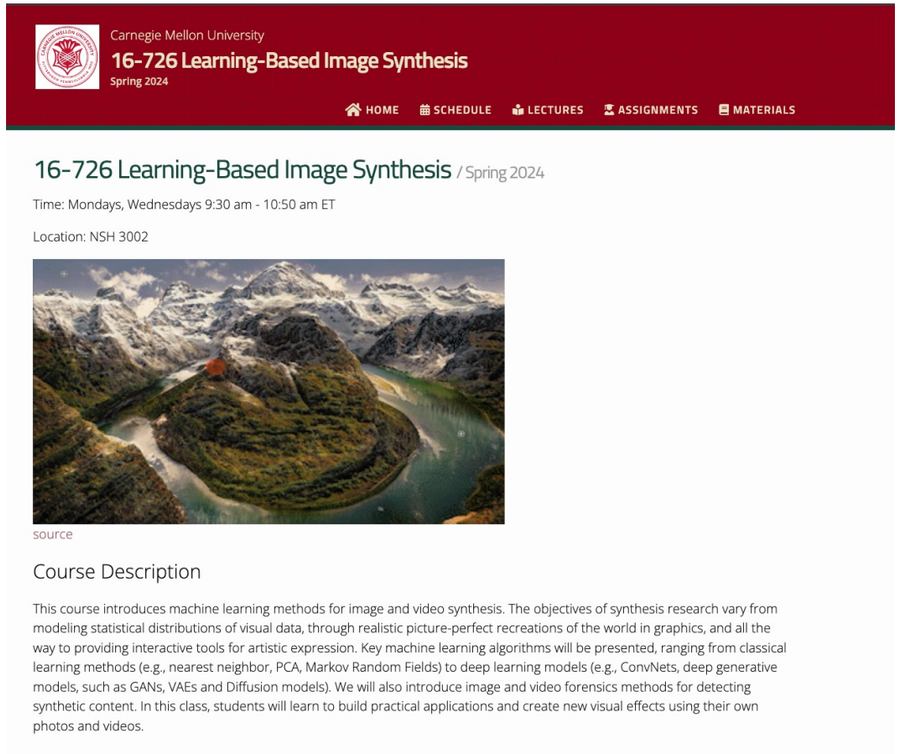
- Derive the math, implement stuff from scratch (+ starter code), and apply it to your own photos
- Every person does their own project (except final)
- Reporting via web page (+ submit code to Canvas)
- Afterwards, vote for class **favorite**(s)! **Gift**!
- Programming Language:
  - Python and PyTorch
  - you can use other languages, but you are on your own

Academic Integrity

- Can discuss projects, but don’t share code
- Don’t look up code or copy from a friend
- If you’re not sure if it’s allowed, ask
- Acknowledge any inspirations
- If you get stuck, come talk to us

Getting help outside of class

- Course Web Page
  - https://16726-image-synthesis.github.io/
- Discussion board:
  - Piazza.com
- Assignment submission
  - Canvas
- Office hours
  - See piazza posts



Why you should NOT take this class

- Project-based class
  - No canned problem sets.
  - Not theory-heavy.
  - will read many research papers.
  - Open-ended by design.
- Need time to think, not just hack
  - **Creativity** is a class requirement.
- Not worth it if you don’t enjoy it.

Now… reasons to take this class

- Not too many similar courses at other places.
- You get to create pictures and unleash your creative potential.
- Interested in grad school and research? ☺
- Interested in industry jobs? ☺

Become a friend with every pixel!

- Jitendra Malik, UC Berkeley

Thank You!

16-726, Spring 2025

https://learning-image-synthesis.github.io/