GenAI for Creatives: A Few Projects from Adobe Research, Lessons Learned and Avenues for Future Work

Sylvain Paris

Adobe

I live in Boston. Thanks a lot for inviting me!

Dall•E 2 Got People's Attention 4 Years Ago

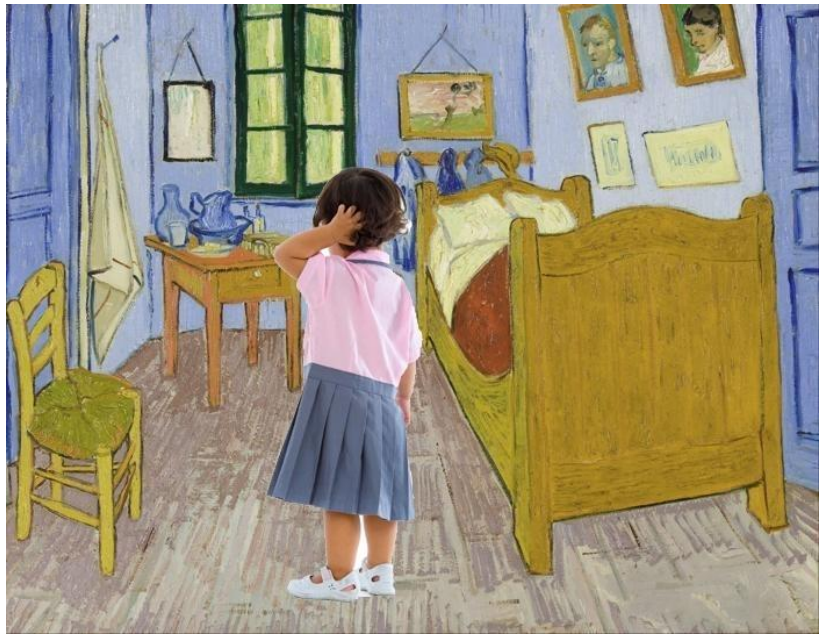
A photorealistic image of an astronaut riding a horse

Progress Happens Faster Than Ever

A photorealistic image of an astronaut riding a horse

[Nano Banana 2]

Generative AI Models Solve Hard Problems Without Even Trying



Naïve compositing

Deep Painterly Harmonization
[Luan et al., EGSR 2018]

6 ~ 12 months of work
by a talented student
and 3 senior researchers

Nano Banana 2:
Make the girl look like she belongs to the painting

Adobe

Naïve compositing

Deep Painterly Harmonization

Nano Banana 2

Then Sora Did it for Video 2 Years Ago

A large part of Adobe's business is about tools for creatives.

This is a momentous change.

This impacts Adobe Research.

A large part of Adobe's business
is about tools for creatives.

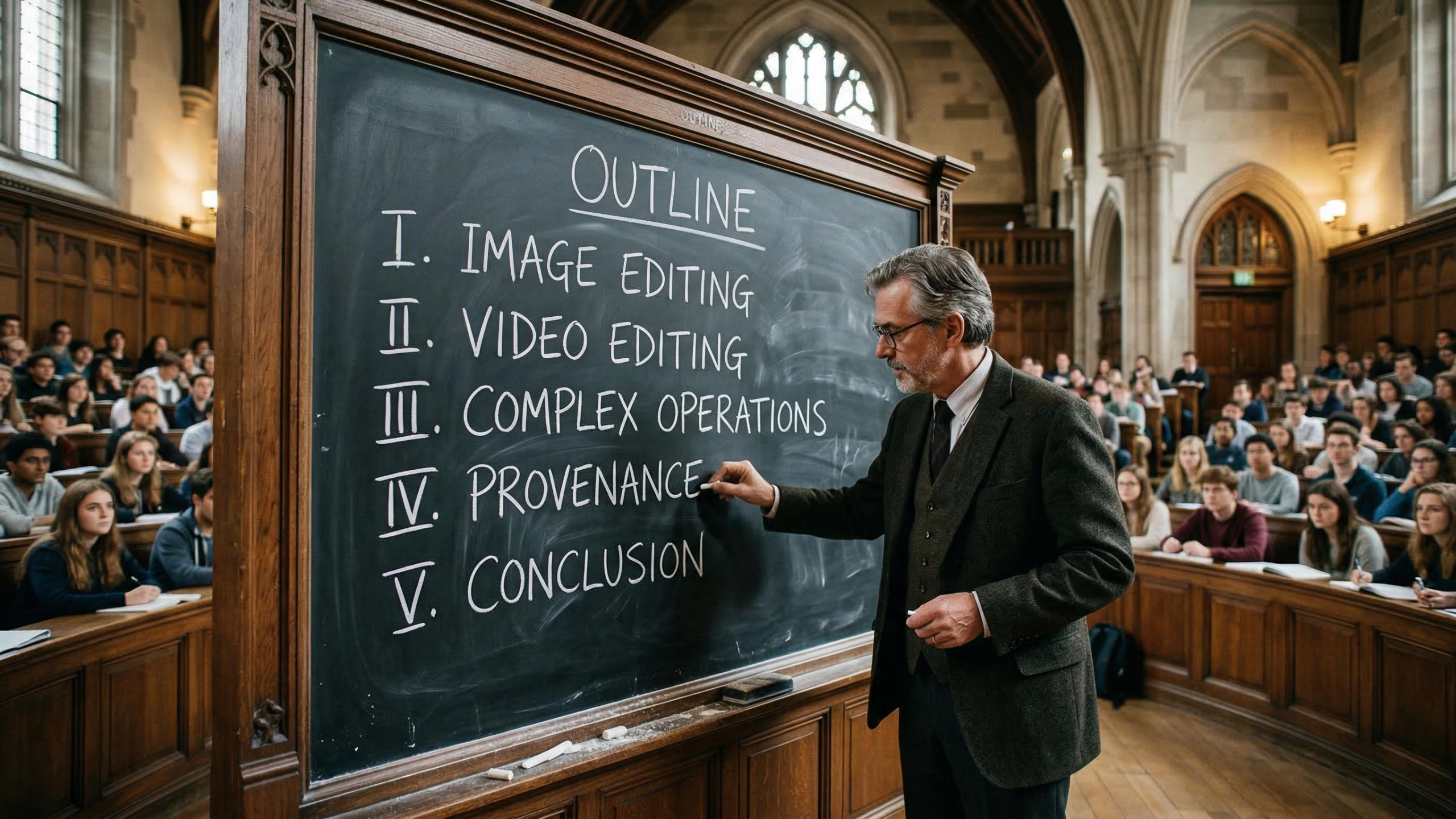
This is a momentous change.

This impacts Adobe Research.

Is there anything left to do?

What if this is not exactly what I have in mind?

Control matters.


OUTLINE
I. IMAGE EDITING
II. VIDEO EDITING
III. COMPLEX OPERATIONS
IV. PROVENANCE
V. CONCLUSION

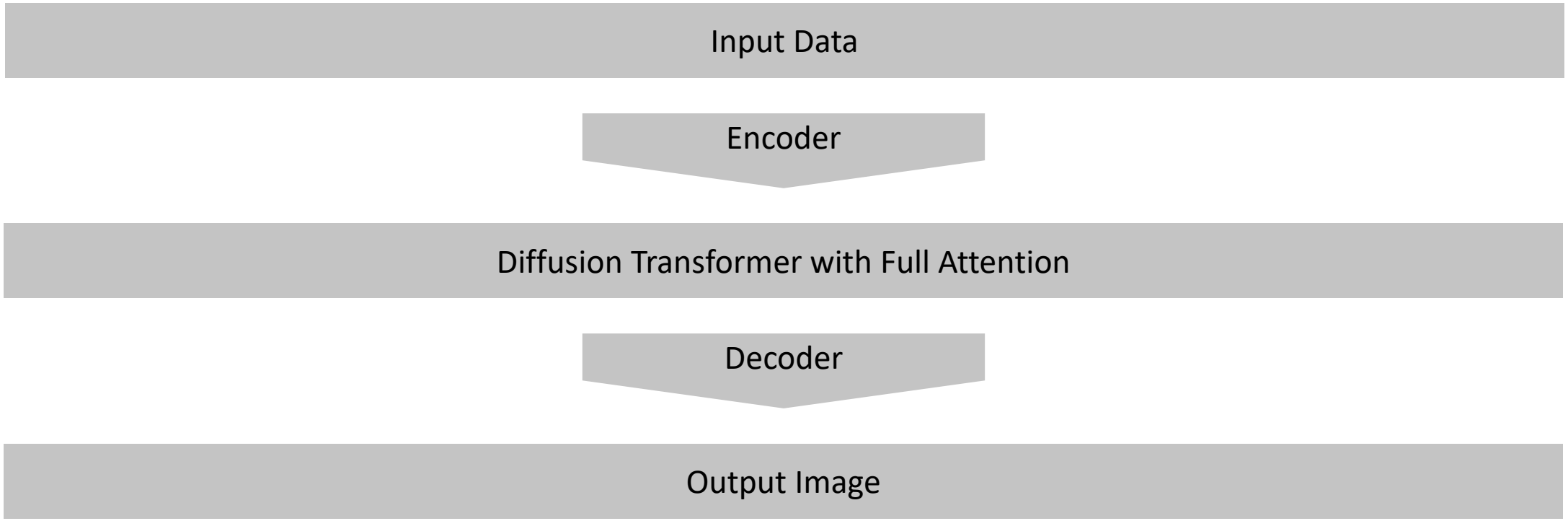
UniReal: Advanced Image Editing

- Goal: support many advanced image edits with a **single** model

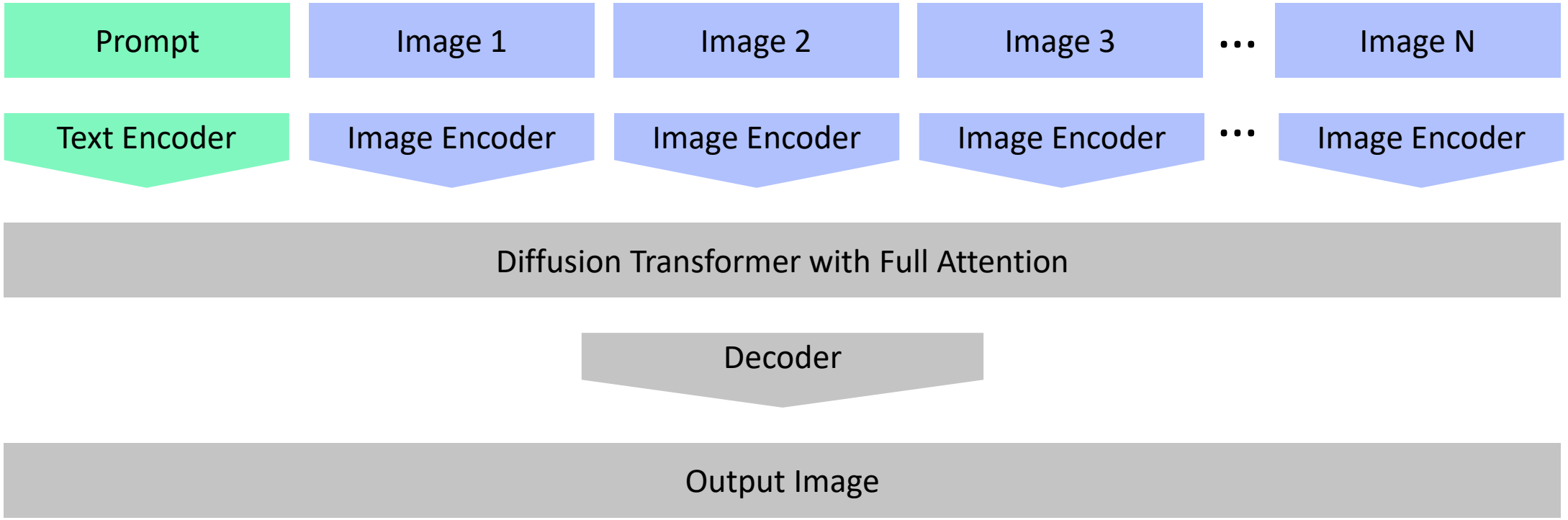
Xi Chen[1] Zhifei Zhang[2] He Zhang[2] Yuqian Zhou[2] Soo Ye Kim[2] Qing Liu[2] Yijun Li[2] Jianming Zhang[2] Nanxuan Zhao[2] Yilin Wang[2] Hui Ding[2] Zhe Lin[2] Hengshuang Zhao[1]
[1]The University of Hong Kong [2]Adobe

Adobe

Simplified View of the Model Architecture



Naïve Approach: Serialized All The Inputs



Naïve Approach: Serialized All The Inputs

It might work with enough data and compute resources.

But it is a hard task: learning general image editing in all its diversity and ambiguity.

For instance:

- An image may contain an element to be composited on another tone.
- Or it may be the background.
- Or it may be the mask indicated where that element is in the picture.

Most Tasks Have Some Structure

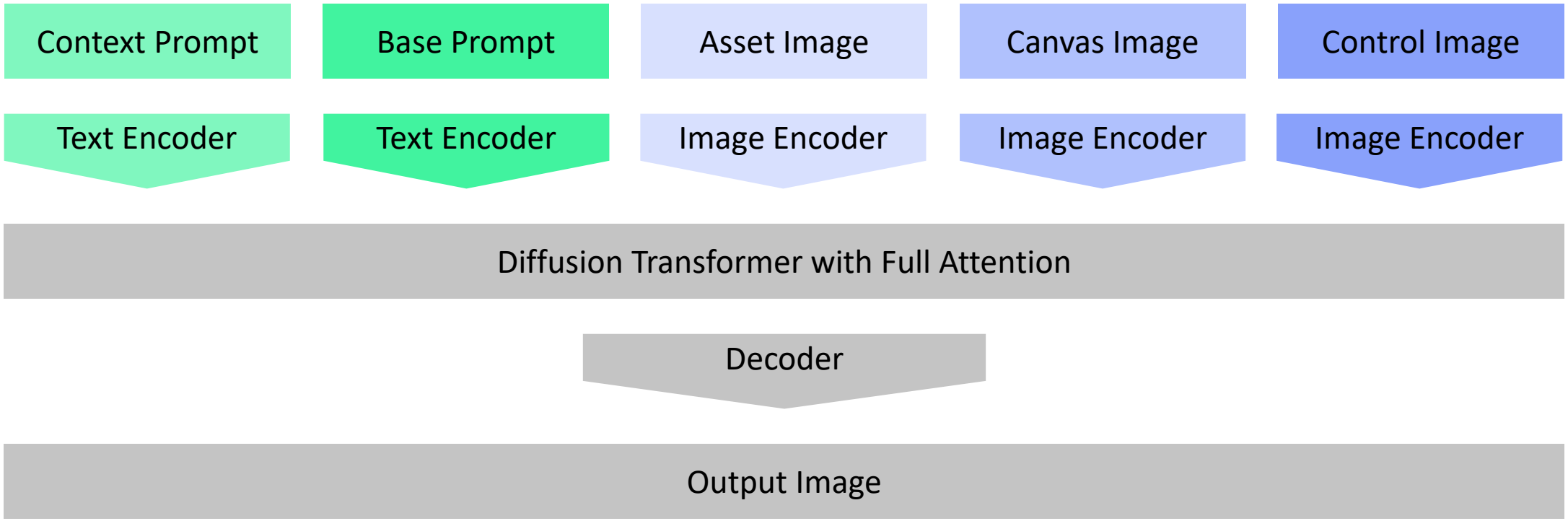
Prompt combines:

- Context, e.g., “realistic” and “uses a reference image”
- Instruction, e.g., description of the content or edits

Images can be:

- An “asset”: an element with an important role, e.g., the object to be added
- The “canvas”: the background or scene
- A “control”: complementary information to execute the prompt, e.g., a mask or a location

Specialize for the Most Popular Tasks and Exploit the Structure



Training Data from Video

- Video frames provide pseudo before-after pairs
- MLLM for generating labels

Make the person lift the weights closer together and shift the angle to focus more on the hands holding the dumbbells.

Change the background to a beach.

Change the background to a forest.

Make it snow.

Make it a sketch.

Make it watercolor.

Make it Picasso‘s style.

Add the sunglasses from IMG1 to the woman of IMG2.

Add the garment from IMG1 to the man of IMG2.

Take-Away Messages

Less in More

Videos as Training Data for Image Editing

The Path Towards Real-Time Video Editing

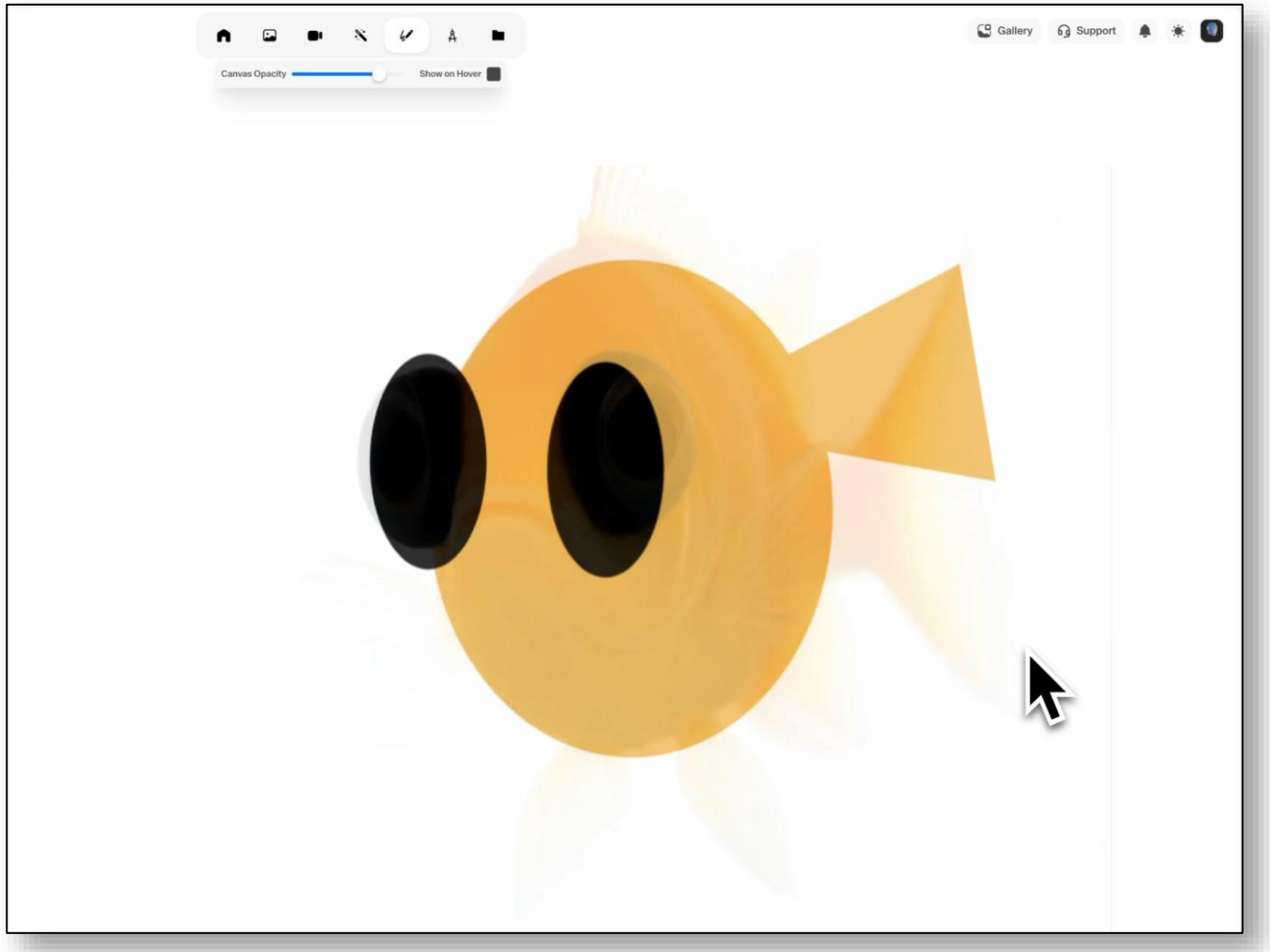
Adobe

Example of Existing Solutions: Google’s Genie 3

Fast, good visual quality but “frozen” world, limited controllability

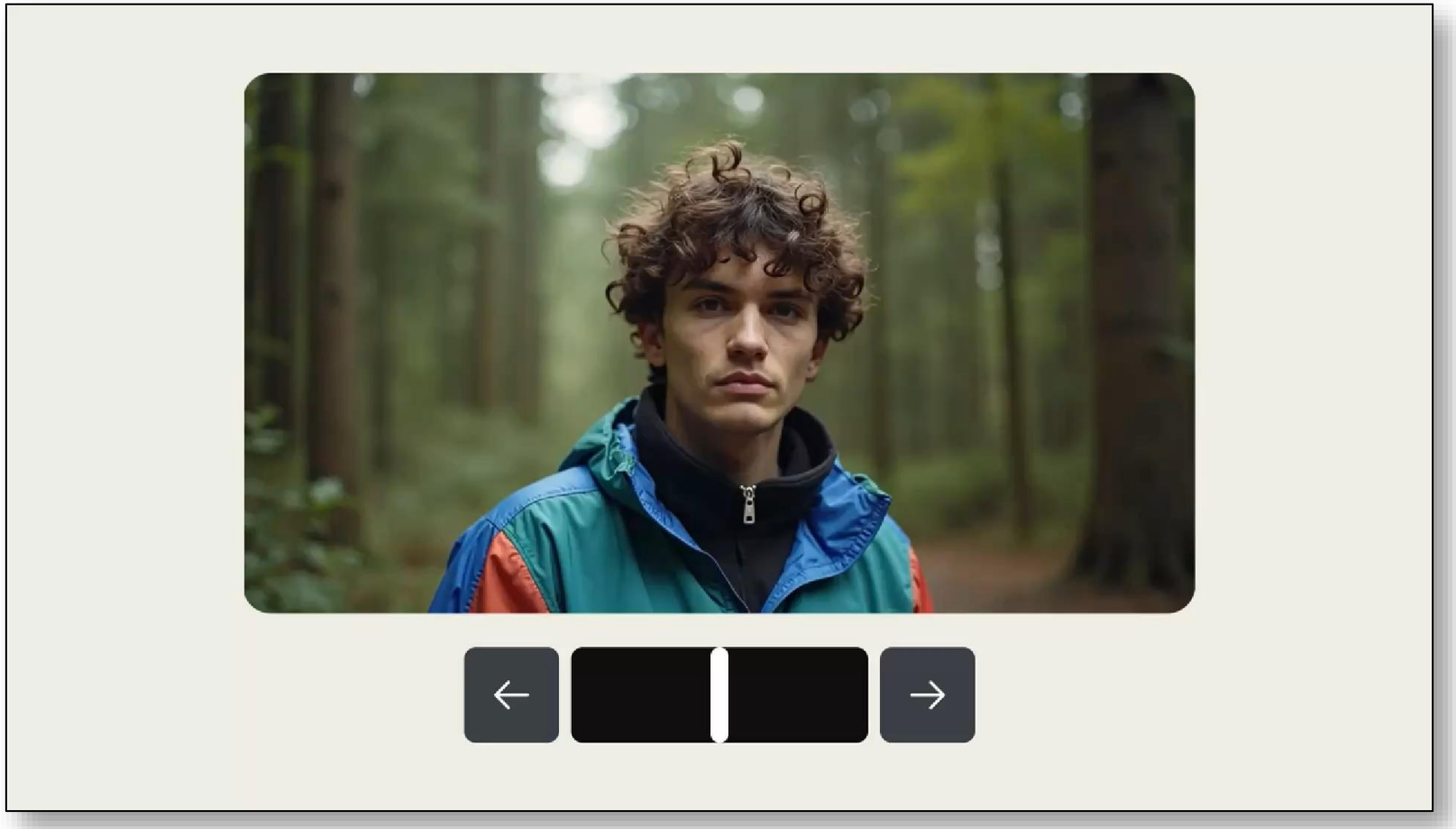
Example of Existing Solutions: Krea’s Realtime Video

Fast and promising but limited, not at Adobe’s level yet



Example of Existing Solutions: Runway's Realtime Controls

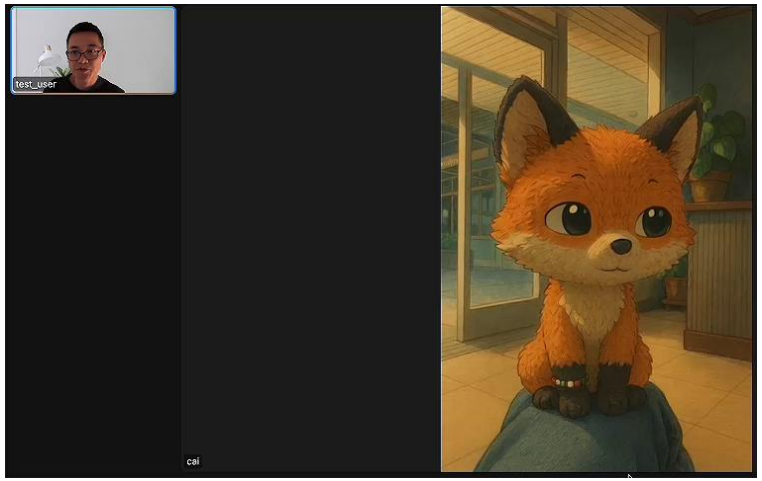
Great control but paused



And there are more…
All with similar pros & cons

Mirage, from Decart

Mirage2

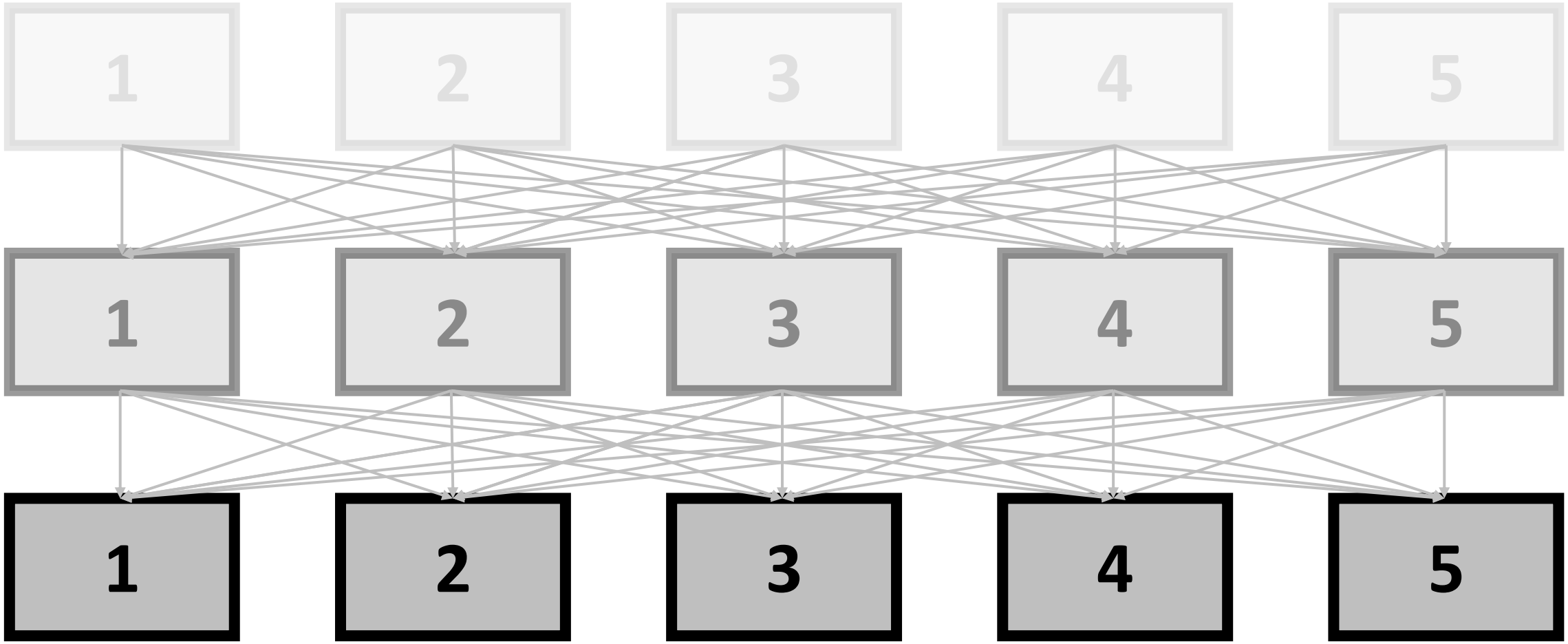


Character.AI

Yan, from Tencent

SeedWeed APT2, from ByteDance

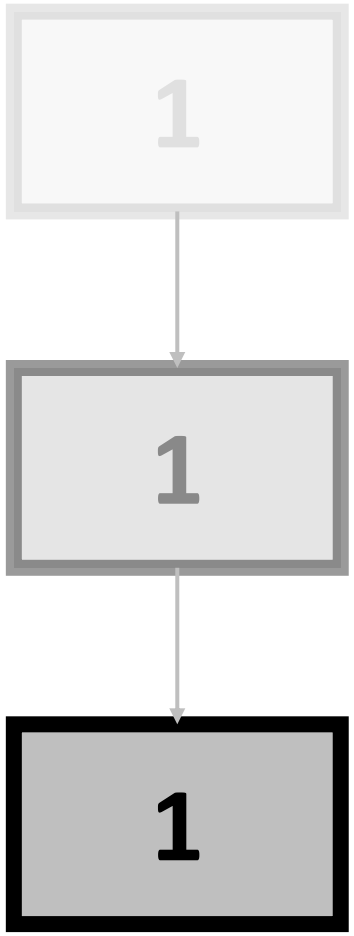
“Traditional” Video Generation with Bidirectional Attention



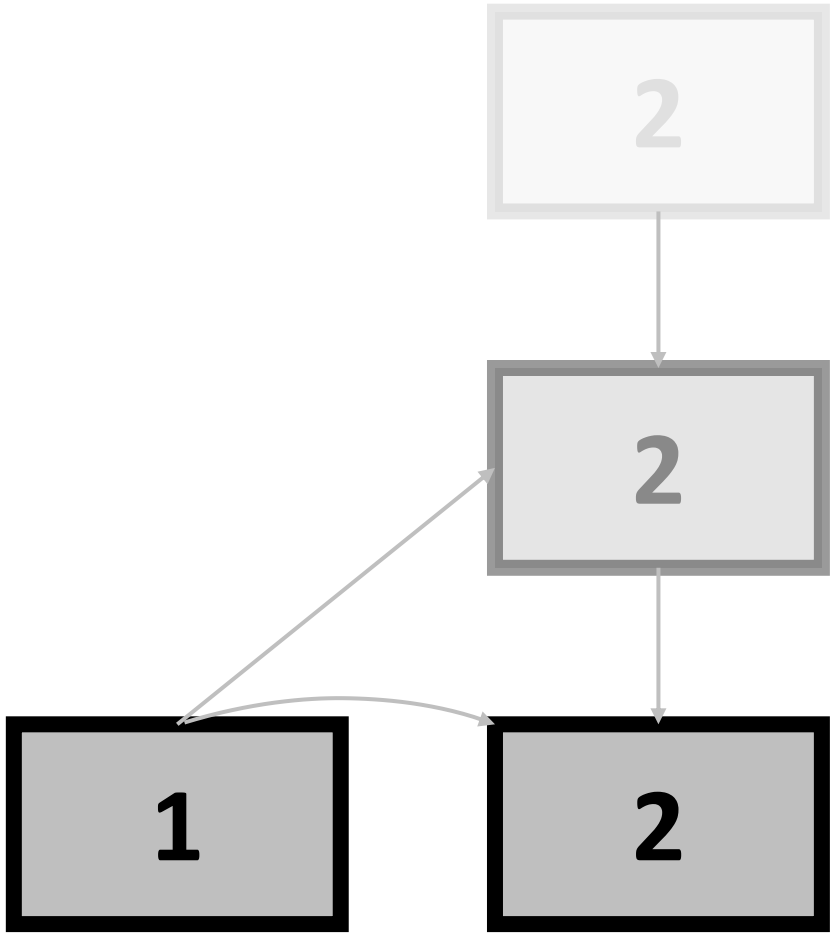
Good: maximal use of the available information

Bad: nothing is available until the very end → high latency

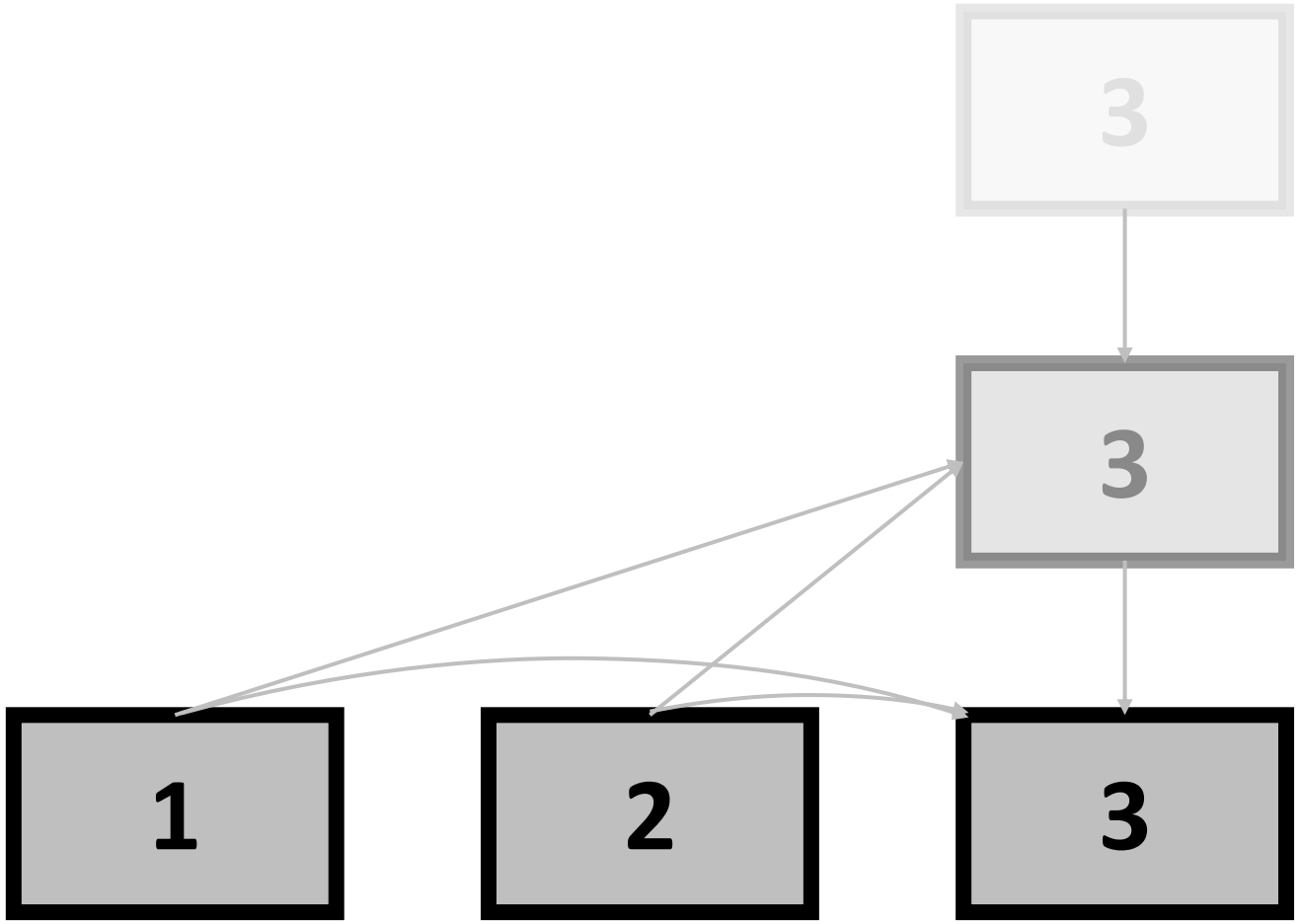
Autoregressive Video Generation with Causal Attention



Autoregressive Video Generation with Causal Attention



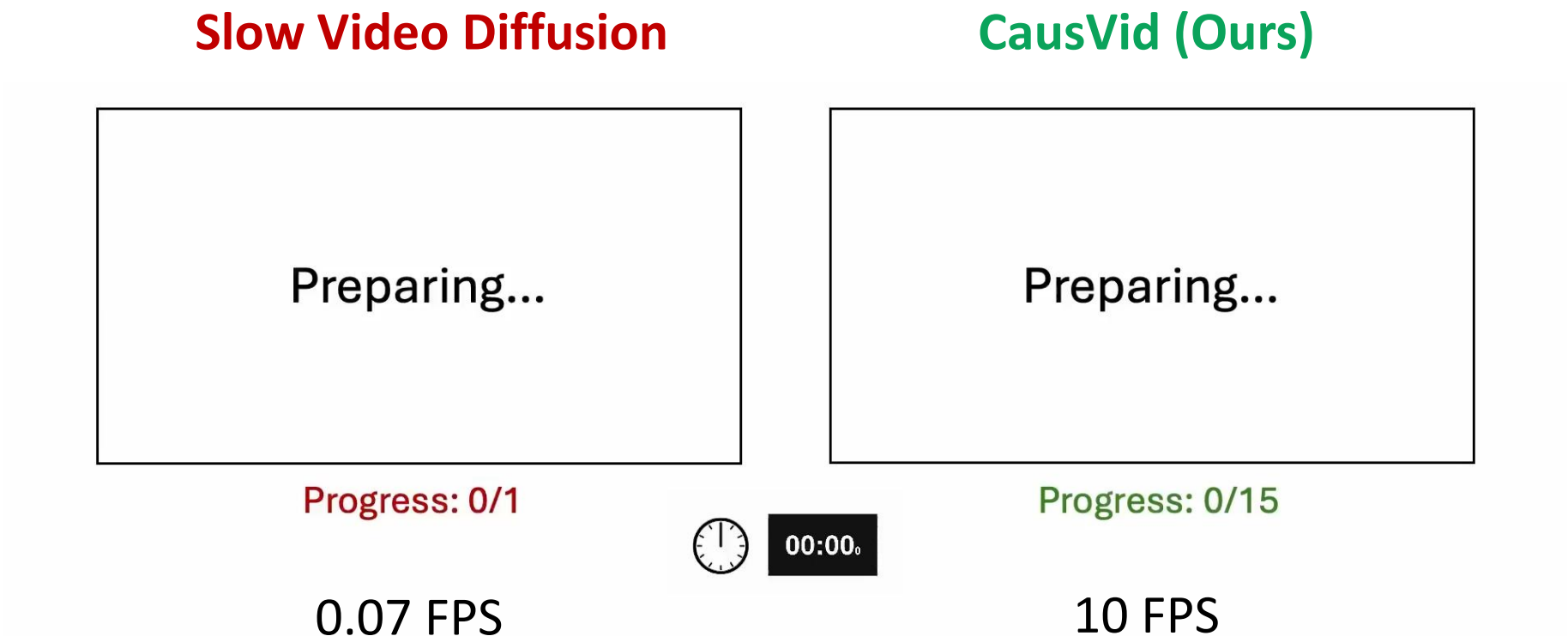
Autoregressive Video Generation with Causal Attention



Frames are available right away one by one → real time is possible.

CausVid: A First Step Towards the New Paradigm

- **CausVid model is distilled from a “traditional” slow teacher video diffusion model**
- **CausVid is two order magnitude faster than video diffusion model at 360P resolution on a single H100**



Yin et al., From Slow Bidirectional to Fast Autoregressive Video Diffusion Models, CVPR 2025

Real-Time Streaming Video Generation

The camera flies across a beach.

Upload Image | Connecting... | Send | Disconnect

Video Output

Statistics

Target FPS: 8
Frames Processed: 0
Status: Connecting...
Queue Size: 0
Recording: Yes

Prompt History

6:20:35 PM
🔵 Initial: The camera flies across a beach.

Logs

Self-Forcing: Bridging the Train-Test Gap in AR Video Diffusion

- Autoregressive roll-out process **during training** to mimic test-time inference

Generation from Self-Forcing

Generation from Self-Forcing

Adobe

Huang et al. Self Forcing: Bridging the Train-Test Gap in Autoregressive Video Diffusion, arxiv 2025

Comparison Self-Forcing with CausVid

- Oversaturation addressed
- Much less temporal flickering
- More natural motion

CausVid

Self-Forcing

Adobe

Huang et al. Self Forcing: Bridging Training and Inference in Autoregressive Video Diffusion, arxiv 2025

Comparing Self-Forcing with CausVid

- Self-Forcing runs **17 FPS** at **480P** on a single H100 GPU

CausVid

Self-Forcing

Adobe

Huang et al. Self Forcing: Bridging Training and Inference in Autoregressive Video Diffusion, arxiv 2025

MotionStream

[1000, 927, 771, 0]

Enter as list format, e.g., [1000, 0] or [1000, 927, 771, 0]

START ↵ PAUSE STOP ↵ CLEAR STOPPED Start on click Save on stop Adaptive FPS

Interactive Canvas

Generating video...

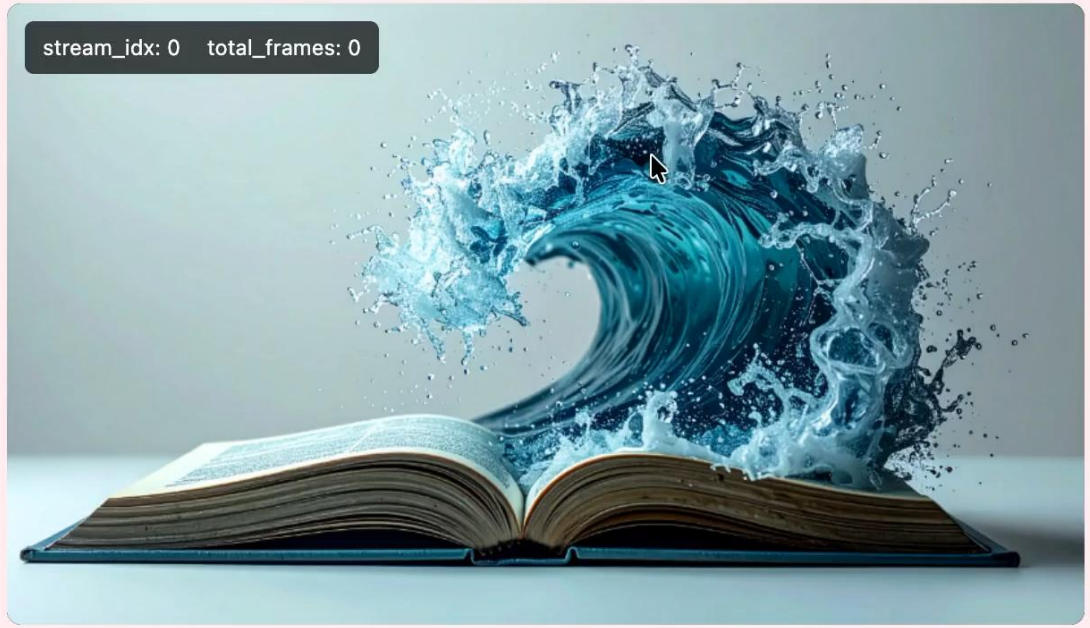
Generated Video

Enable "Save on stop" - videos will be saved locally!

MotionStream

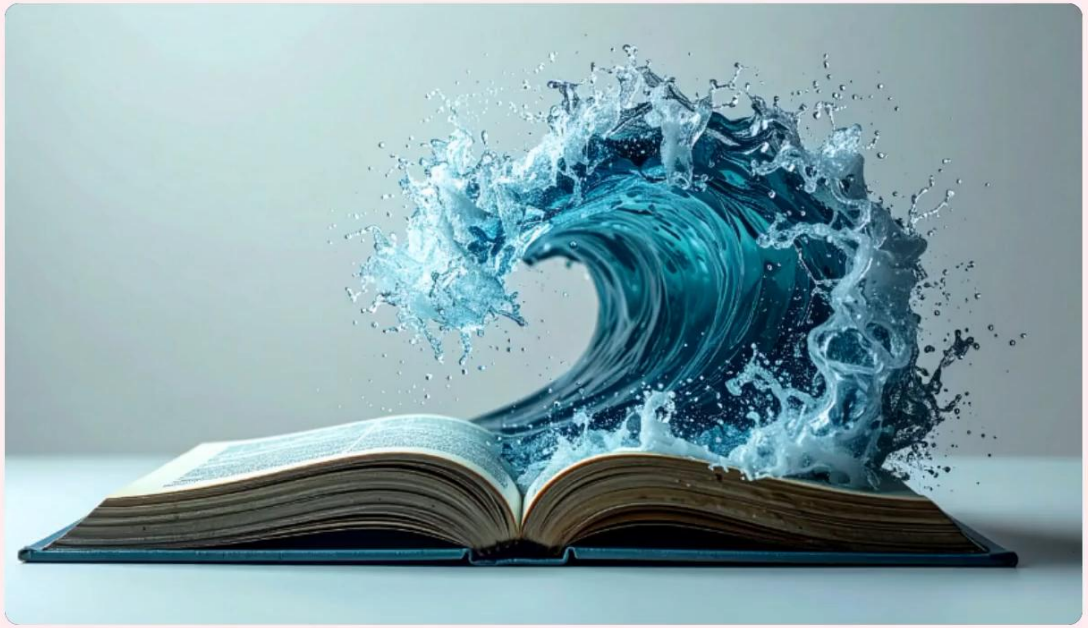
START ↵ PAUSE STOP ↵ CLEAR STARTING GENERATION... Start on click Save on stop Adaptive FPS

Interactive Canvas



Generating video...

Generated Video

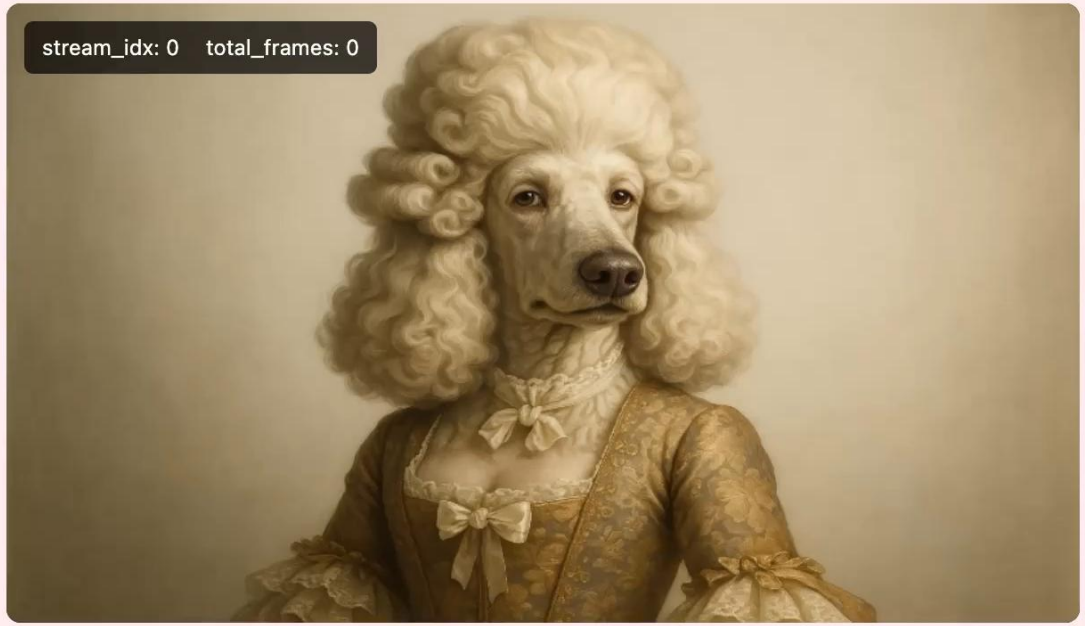


Enable "Save on stop" - videos will be saved locally!

MotionStream

START ↵ PAUSE STOP ↵ CLEAR STOPPED Start on click Save on stop Adaptive FPS

Interactive Canvas



Generating video...

Generated Video

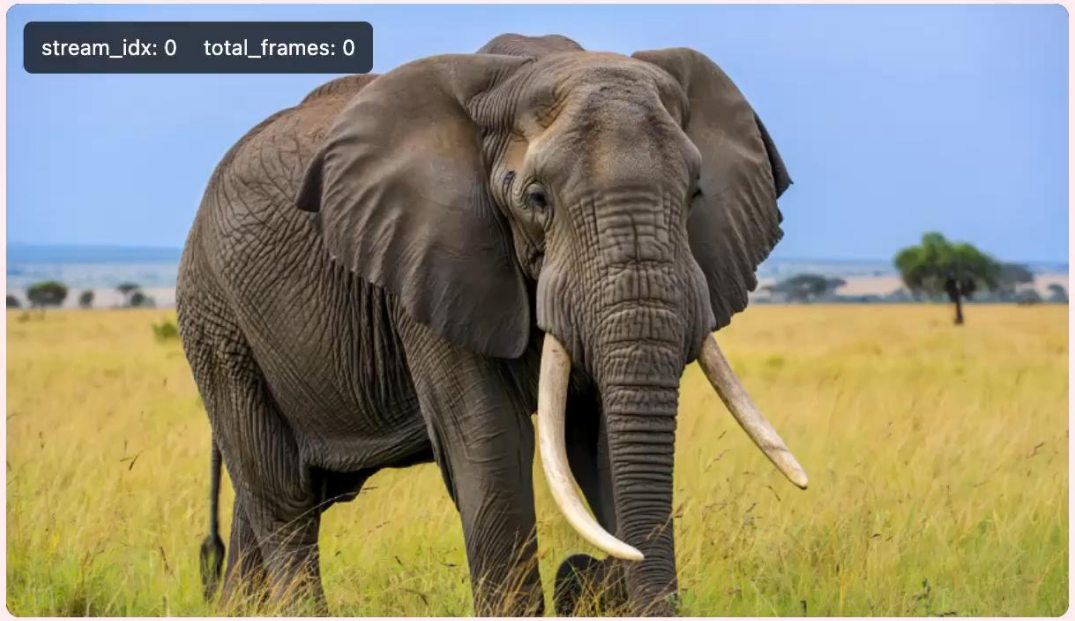
Enable "Save on stop" - videos will be saved locally!

Performance Metrics

MotionStream

START ↵ PAUSE STOP ↵ CLEAR STOPPED Start on click Save on stop Adaptive FPS

Interactive Canvas

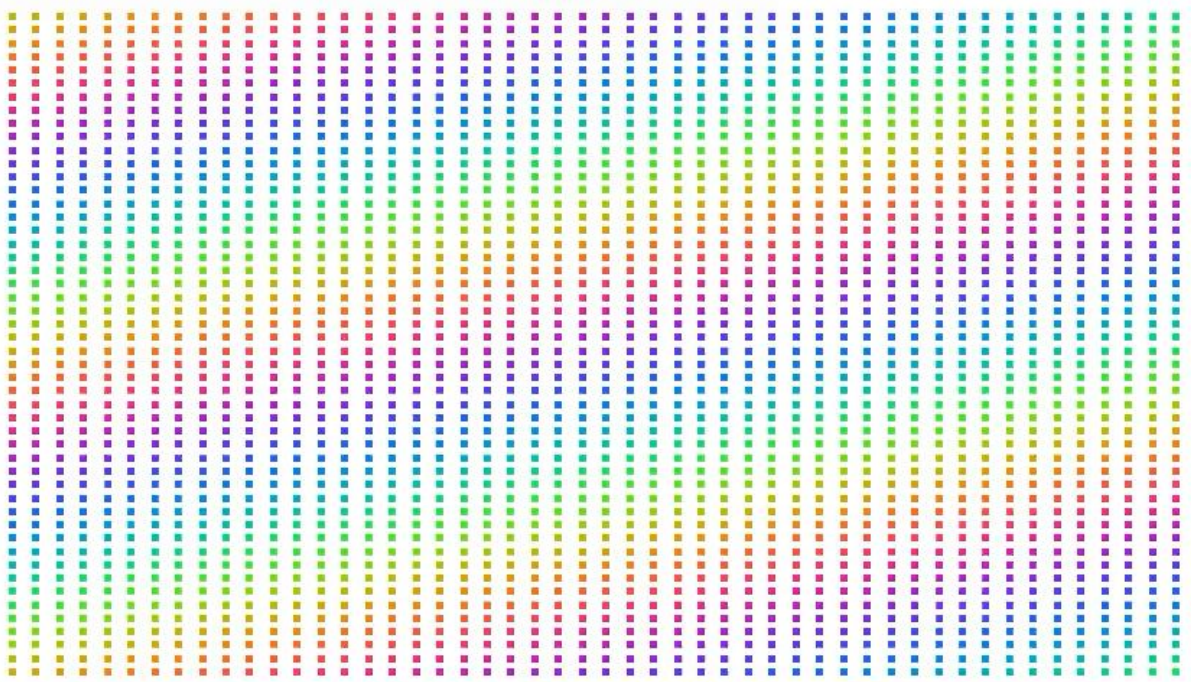


Click & drag to control

Generated Video

Enable "Save on stop" - videos will be saved locally!

Camera Control



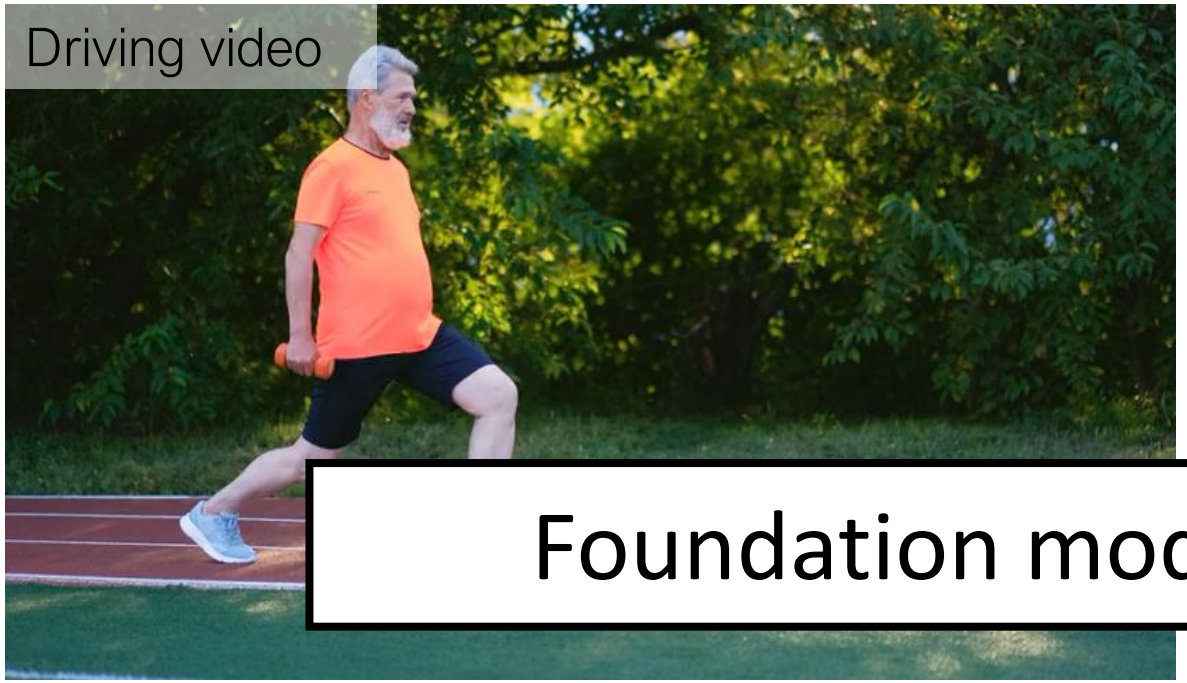
Points “lifted” to 3d

Input image

Synthesized camera control

Pose Control

Pose Control



Foundation model for interaction

Take-Away Messages

Performance as a Feature

Simple Low-Level Controls Are Expressive

Beyond Simple Edits: X-Planner

Adobe

Chun-Hsiao Yeh, Yilin Wang, Nanxuan Zhao, Richard Zhang, Yuheng Li, Yi Ma, Krishna Kumar Singh

Current Editing Models **Fall Short** on Complex Instruction-Based Image Editing

Input Image:

"Could you make this image Christmas?"

MGIE (ICLR'24)

Edited Image:

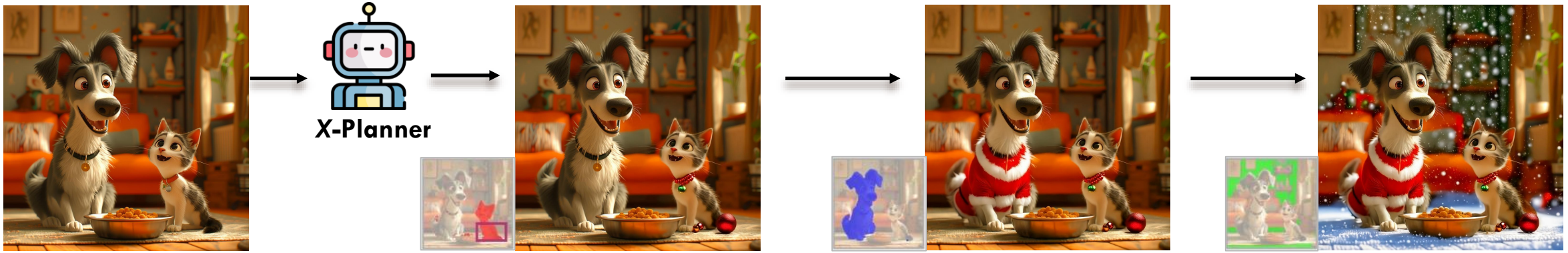
LEDITS++ (CVPR'24)

InstructPix2Pix

- **[Planning]** MLLMs (e.g., GPT4) is needed for analyzing the complex instruction
- **[Automatic Control Guidance]** Localizing the area of interest (mask, box) is necessary for editing

Adobe

Qualitative Results of X-Planner



Make all animals look like they are celebrating Christmas?

insertion: <box> Add Christmas ornaments around the cat

local texture: Change the dog to have a red and white Christmas suit

background: Make the background look like a cozy snowy Christmas setting

Qualitative Results of X-Planner

Make all animals look like they are celebrating Christmas?

insertion: <box> Add Christmas ornaments around the cat

Qualitative Results of X-Planner

insertion: <box> Add Christmas ornaments around the cat

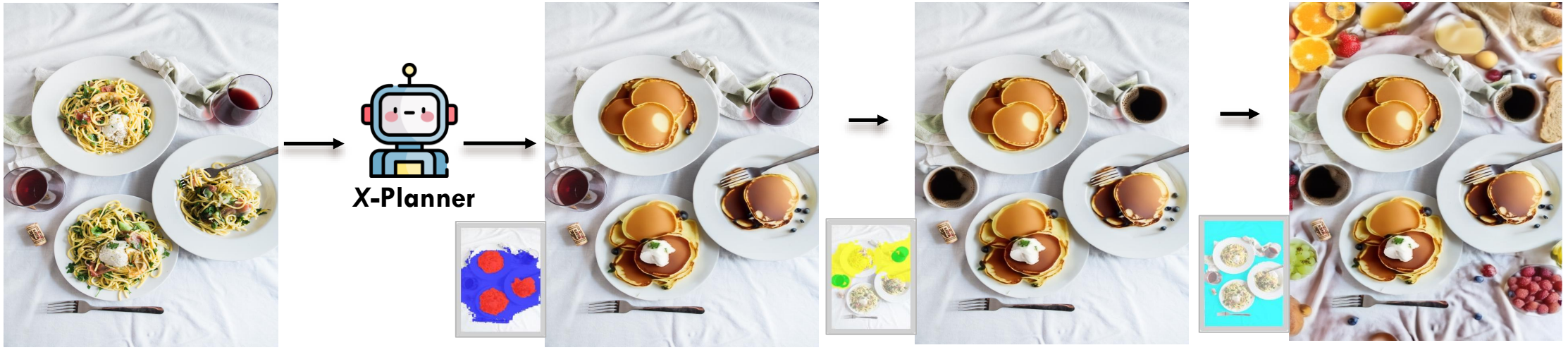
local texture: Change the dog to have a red and white Christmas suit

Qualitative Results of X-Planner

local texture: Change the dog to have a red and white Christmas suit

background: Make the background look like a cozy snowy Christmas setting

Qualitative Results of X-Planner



Could you change this meal setting to a breakfast theme?

replace: Replace the pasta with pancakes

replace: Change the wine glass to a coffee mug

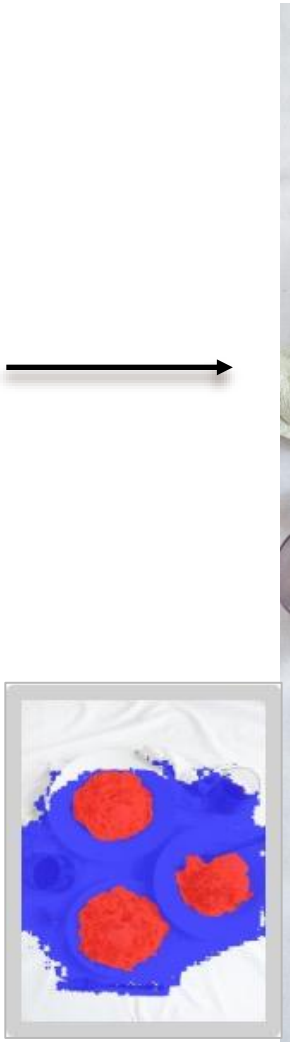
background: adjust the background to a cozy morning setting

Qualitative Results of X-Planner

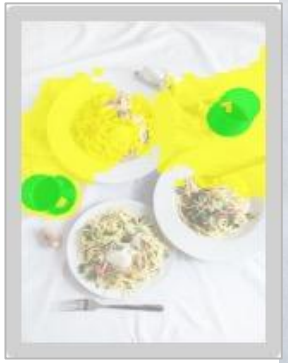
Could you change this meal setting to a breakfast theme?

replace: Replace the pasta with pancakes

Qualitative Results of X-Planner

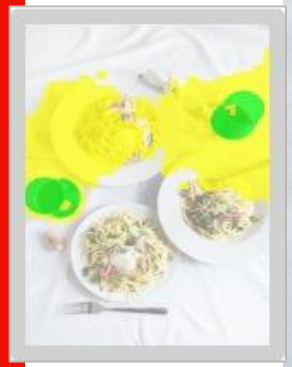


replace: Replace the pasta with pancakes



replace: Change the wine glass to a coffee mug

Qualitative Results of X-Planner



replace: Change the wine glass to a coffee mug

background: adjust the background to a cozy morning setting

Under the Hood

- MLLM trained with automatically created training data

Additional Benefits

- Transparent process: edit list helps user understand what was done
- More control: each step can be individually adjusted or removed

Take-Away Messages

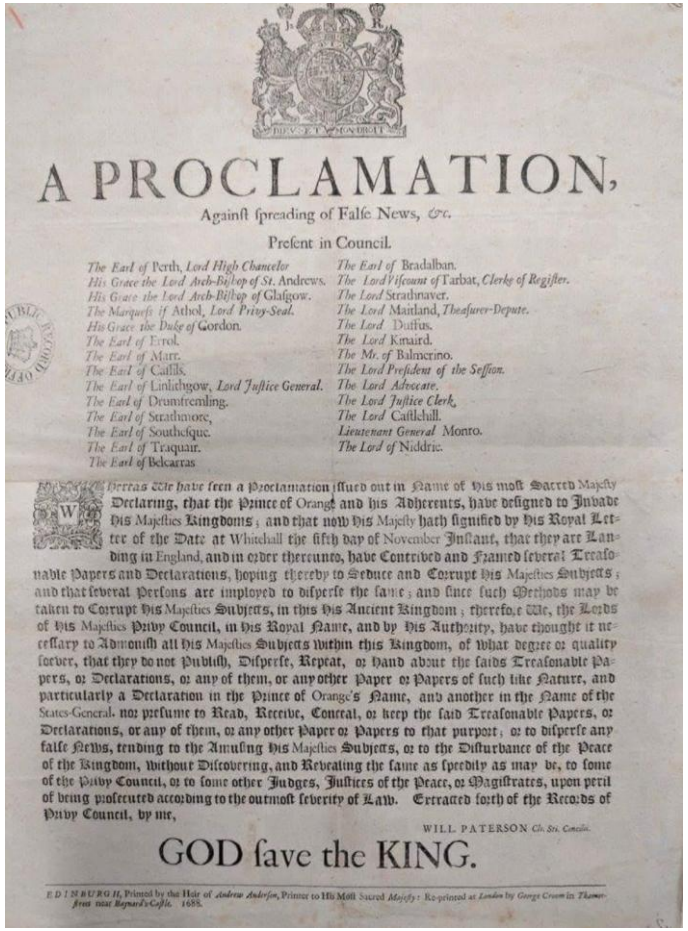
Breaking Down Complex Tasks
Into Simple Ones Helps

Provenance

Adobe

Tu Bui, Shruti Agarwal, John Collomosse

Fake News Has Always Been a Problem…



A PROCLAMATION,

Against ſpreading of Falſe News, &c.

Preſent in Council.

The Earl of Perth, *Lord High Chancellor*
His Grace the *Lord Arch-Biſhop* of St. Andrews.
His Grace the *Lord Arch-Biſhop* of Glaſgow.
The Marqueſs of Athol, *Lord Privy-Seal.*
His Grace the Duke of Gordon.
The Earl of Errol.
The Earl of Marr.
The Earl of Caſſils.
The Earl of Linlithgow, *Lord Juſtice General.*
The Earl of Drumfremling.
The Earl of Strathmore,
The Earl of Southeſque.
The Earl of Traquair.
The Earl of Belcarras
The Earl of Bradalban.
The *Lord Viſcount* of Tarbat, *Clerke of Regiſter.*
The *Lord* Strathnaver.
The *Lord* Maitland, *Theaſurer-Depute.*
The *Lord* Duffus.
The *Lord* Knaird.
The *Mr. of* Balmerino.
The *Lord Preſident of the Seſſion.*
The *Lord Advocate.*
The *Lord Juſtice Clerk,*
The *Lord* Caſtlehill.
Lieutenant General Monro.
The *Lord* of Niddrie.

Whereas We have ſeen a Proclamation iſſued out in Name of His moſt Sacred Majeſty Declaring, that the Prince of Orange and his Adherents, have deſigned to Invade his Majeſties Kingdoms; and that now His Majeſty hath ſignified by His Royal Letter of the Date at Whitehall the fifth day of November Inſtant, that they are Landing in England, and in order thereunto, have Contrived and Framed ſeveral Treaſonable Papers and Declarations, hoping thereby to Seduce and Corrupt His Majeſties Subjects; and that ſeveral Perſons are imployed to diſperſe the ſame; and ſince ſuch Methods may be taken to Corrupt His Majeſties Subjects, in this His Ancient Kingdom; therefore We, the Lords of His Majeſties Privy Council, in his Royal Name, and by His Authority, have thought it neceſſary to Admoniſh all His Majeſties Subjects within this Kingdom, of what degree or quality ſoever, that they do not Publiſh, Diſperſe, Repeat, or Hand about the ſaids Treaſonable Papers, or Declarations, or any of them, or any other Paper or Papers of ſuch like Nature, and particularly a Declaration in the Prince of Orange's Name, and another in the Name of the States-General. nor preſume to Read, Receive, Conceal, or keep the ſaid Treaſonable Papers, or Declarations, or any of them, or any other Paper or Papers to that purport; or to diſperſe any falſe News, tending to the Amuſing His Majeſties Subjects, or to the Diſturbance of the Peace of the Kingdom, without Diſcovering, and Revealing the ſame as ſpeedily as may be, to ſome of the Privy Council, or to ſome other Judges, Juſtices of the Peace, or Magiſtrates, upon peril of being proſecuted according to the outmoſt ſeverity of Law. Extracted forth of the Records of Privy Council, by me,

WILL. PATERSON *Cl. Sti. Concilii.*

GOD ſave the KING.

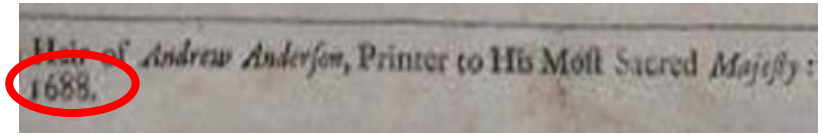
EDINBURGH, Printed by the Heir of *Andrew Anderſon*, Printer to His Moſt Sacred *Majeſty*: Re-printed at *London* by *George Croom* in *Thames-ſtreet* near *Bayard's-Caſtle.* 1688.

OCLAMA I

Againſt ſpreading o Falſe News, &c.

Preſent in Council.

Lord High Chancellor
Arch-Biſhop of St. Andrews.
Arch-Biſhop of Glaſgow.
The Earl of Bradalban.
The *Lord Viſcount of* Tarba
The *Lord* Strathnaver.



Andrew Anderſon, Printer to His Moſt Sacred *Majeſty*:
1688.

Fake news in the time of James II

Cottingley Fairies

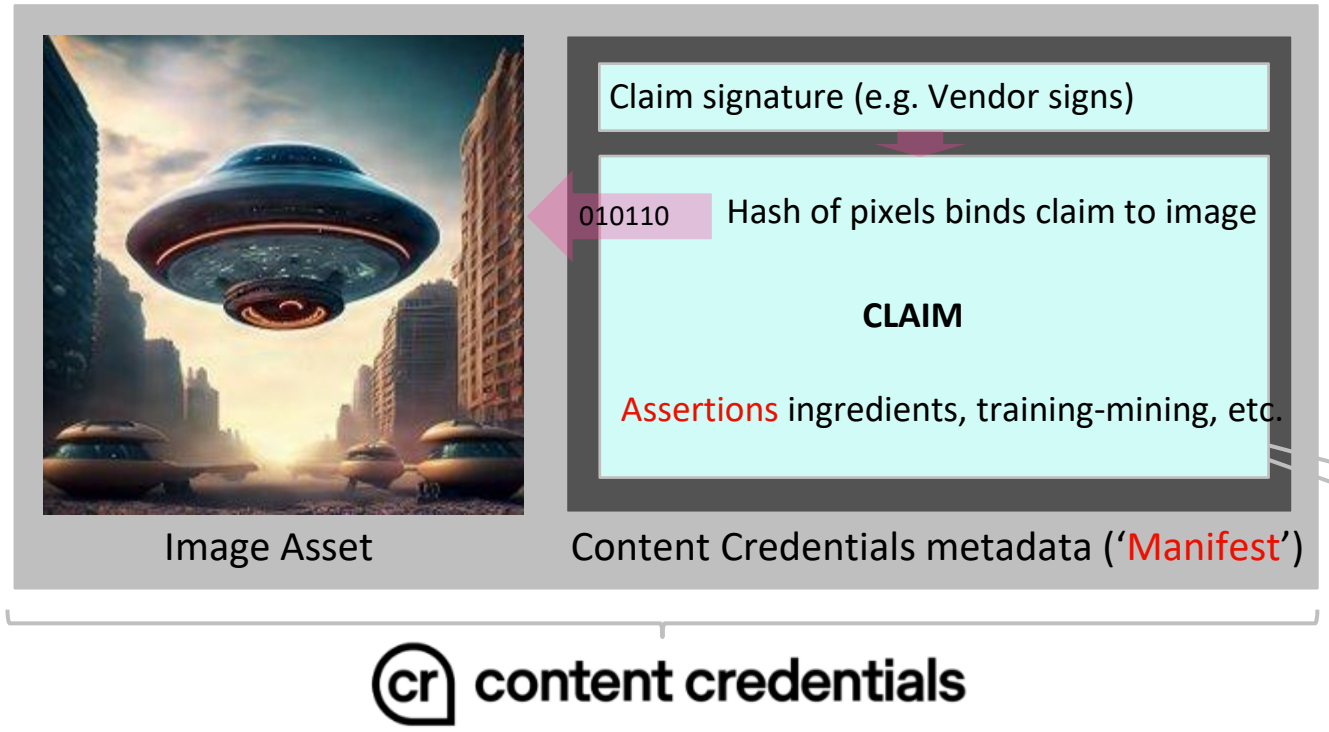
The UK National Archives

Photographs c. 1917 (Elsie Wright, UK)

How Can Provenance Help?

Content Credentials establish content provenance and authenticity at scale to give publishers, creators, and consumers the ability to trace the origin of media.

C2PA is an open cross-industry standard for specifying provenance of media



Manifest contains signed claim

Claim contain facts called assertions

Assertions describe how an asset was made, which sources, actions, etc.

Can link to manifests in ingredient assets to form a provenance history

Figure from Collomosse and Parsons, IEEE CG&A 2024

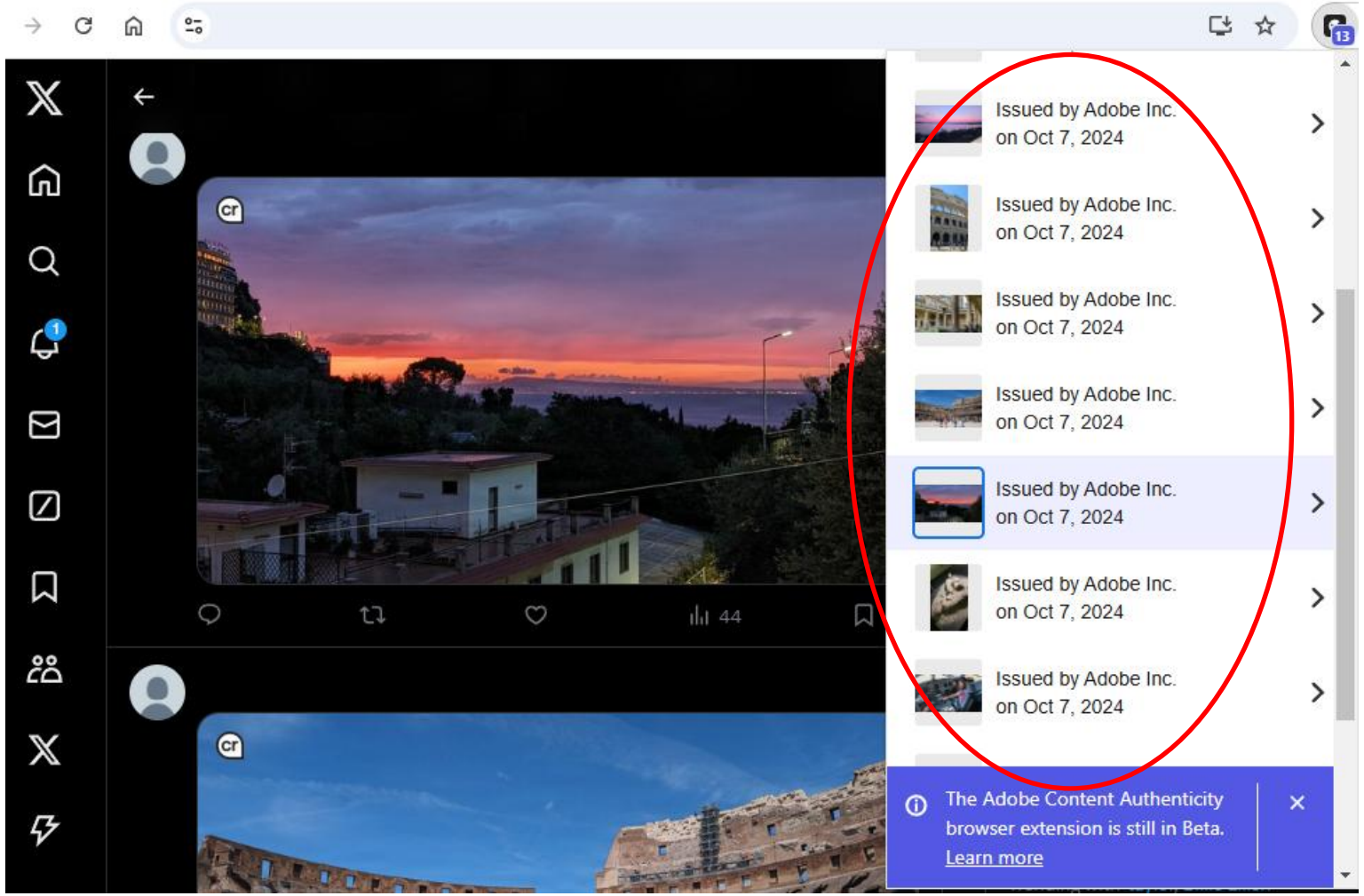
Why Do We Need Watermarking?

Content Credentials are often **stripped** by non-compliant content platforms or when screenshotted.

Currently all social platforms (even those that display presence of credentials) strip the metadata

Watermarking Makes Content Credentials Stick, Anywhere on the Web

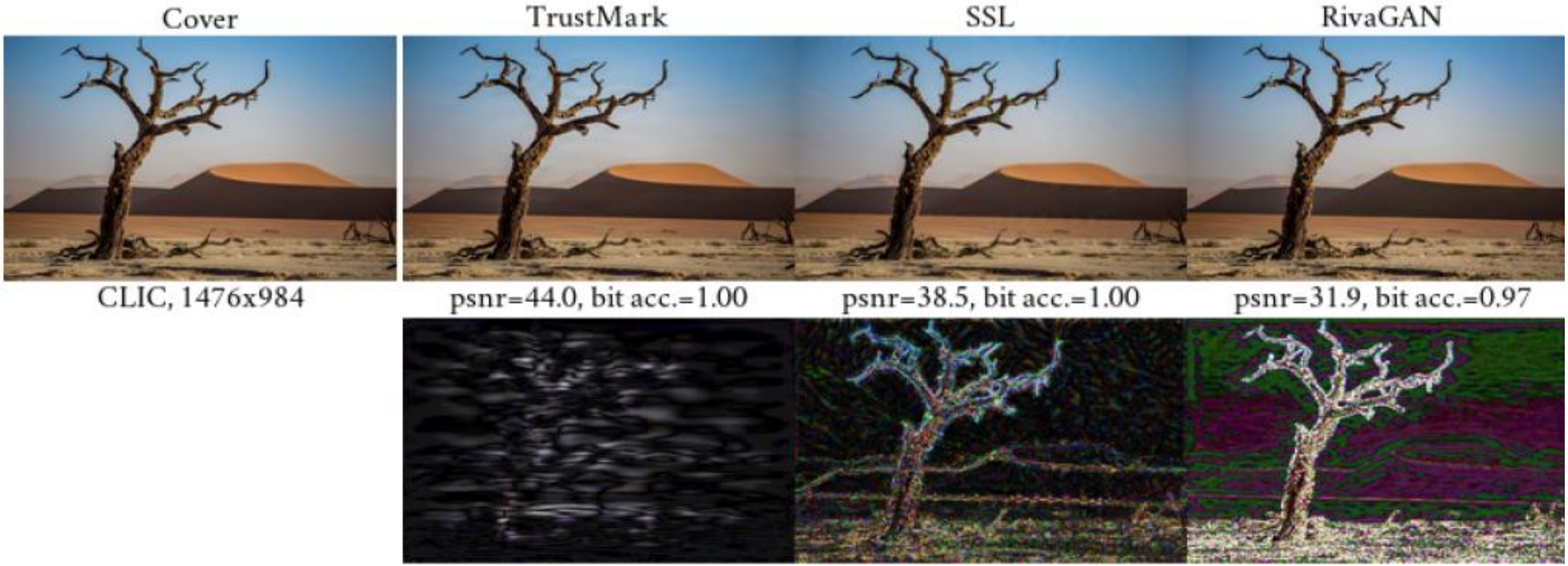
Browser detects invisible watermarks and **recovers stripped content credentials**



Content Credentials recovered

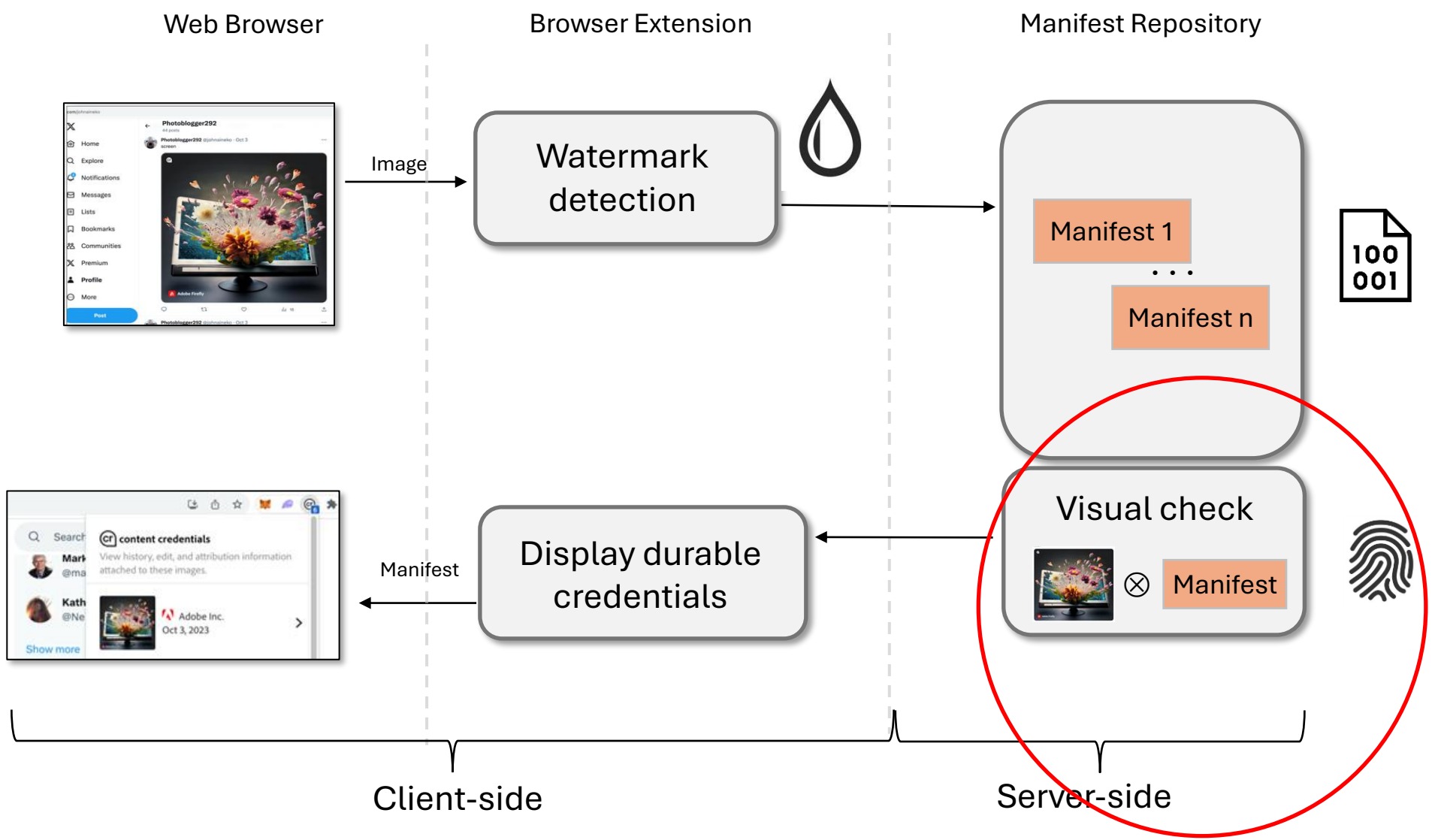
https://contentauthenticity.org/blog/durable-content-credentials

TrustMark: Open Source Watermarking for Provenance



TrustMark: Robust Watermarking and Watermark Removal for Arbitrary Resolution Images. Bui, Agarawal, Collomosse. Proc. ICCV 2025

Watermark Resolution with Visual Fingerprint Check



Security check against spoofing (transfer) of watermark

Take-Away Messages

Helping Good Actors Is Possible

What's Missing and What's Next?

Adobe

We Used to Have Rich, Fine-Grain Controls.

These Controls Came From Analysis.

- There is a large set of studies dedicated to understanding the latent spaces of GANs.

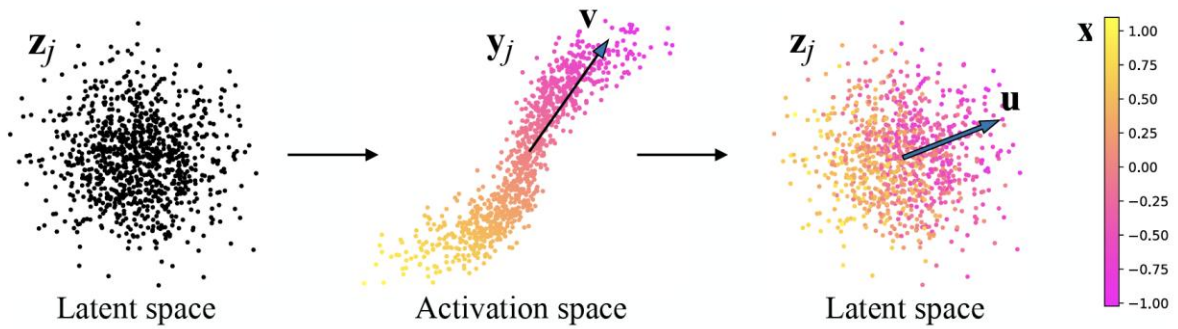


Figure 2: 2D Illustration of identifying a principal activation direction for BigGAN. Random latent vectors $\mathbf{z}_j$ are sampled, and converted to activations $\mathbf{y}_j$. The PCA direction $\mathbf{v}$ is computed from the samples, and PCA coordinates x_j computed, shown here by color-coding. Finally, back in the latent space, the direction $\mathbf{u}$ is computed by regression from $\mathbf{z}_j$ to x_j.

- Challenge: activation space of AI models is huge
- Mitigation: use CLIP space

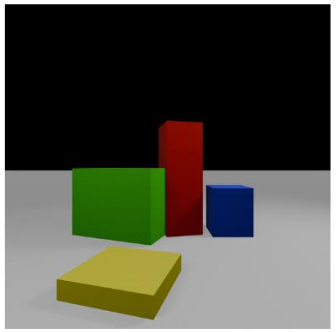
Similar Approach Is Possible with GenAI, but Identity Can Vary

More Broadly:
How to Control GenAI Model?

- Prompts are limited,
  “prompt engineering” is particularly awful.
- Dragging points à la MotionStream,
  elements & characters (see fig. on the right),
  3D proxies (see bottom fig.)…
- **What else?**

[NanoBanana 2]

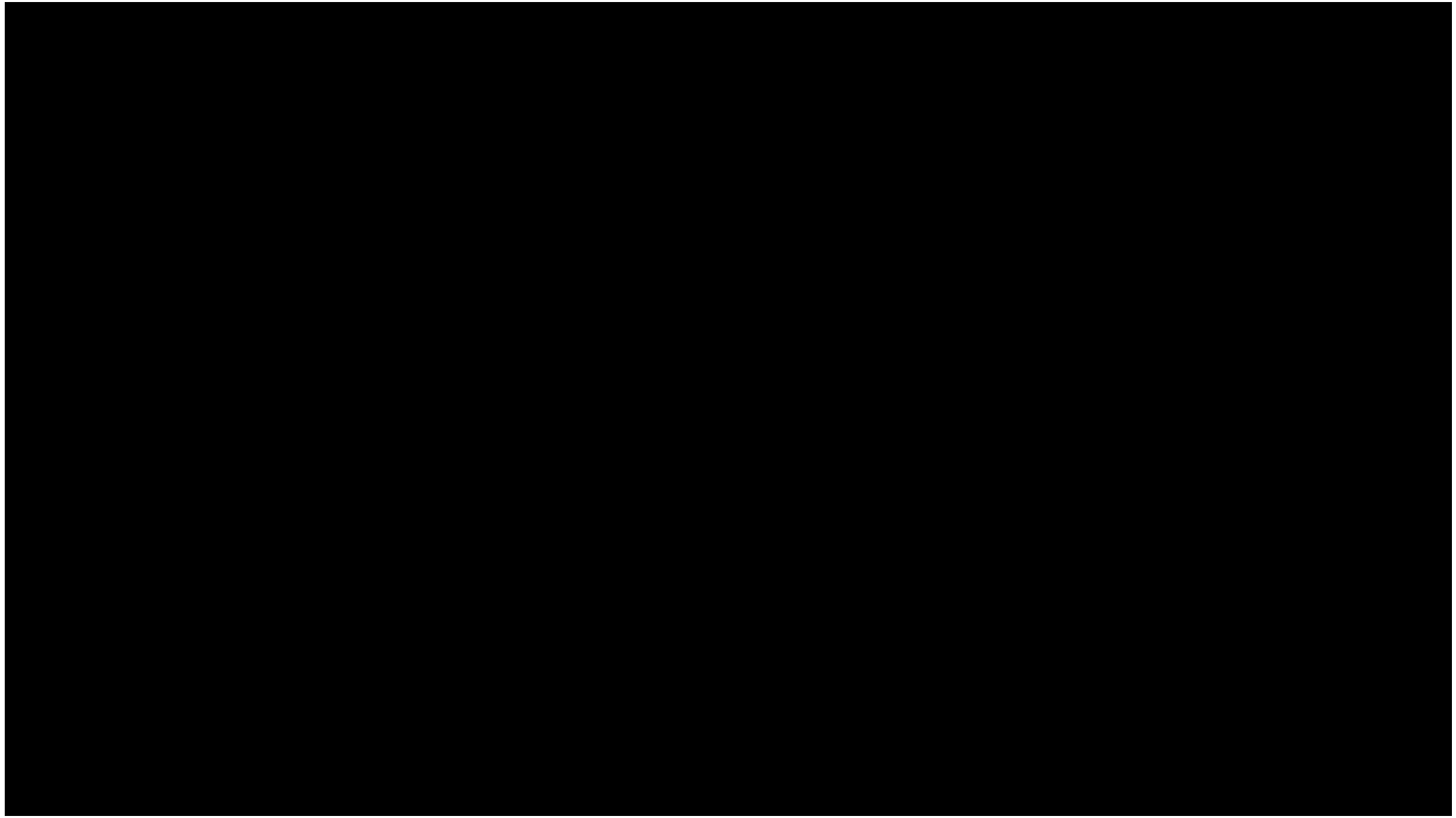
Identities



3D boxes

[SIGMA-GEN, Saha et al. 2025]

Remaining Challenge: Combining Everything Without Constrain



The structure of the teams reflects on the workflow they build.

Thanks!

- The authors of the projects: they did all the work.
- Aaron Hertzmann and Richard Zhang for helping prepare the presentation.
- Daniel, Stefan, the INSTICC team, and all the organizers.
- You!

Wrapping Up

AI opens a new frontier
for research on creativity.

Control matters.

A lot to explore.