UCF

Institute for Artificial Intelligence

Better Learning Signals for Visual Generation

Dr. Chen Chen
Associate Professor
UCF Institute of Artificial Intelligence (IAI)
Department of Computer Science
University of Central Florida
chen.chen@ucf.edu
https://www.crcv.ucf.edu/chenchen/

What Visual Generation Does



Samples from a Data Distribution

Train



Neural Network



Sample



Art & Design



Content Generation



Representation Learning



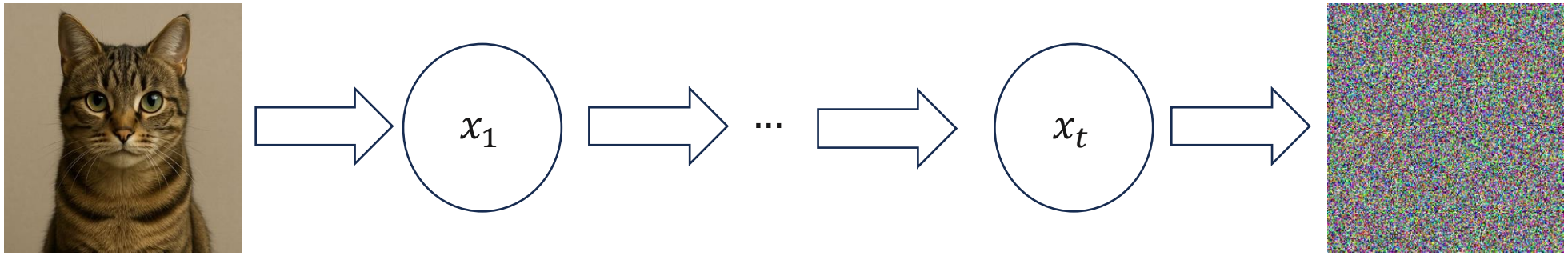
Entertainment



Visual generative models learn to synthesize images and video from data.

Background: Diffusion Models

- Forward diffusion process that gradually adds noise to input



x_0

Diffusion Process: Gradually add noise to the image until random noise

x_T

$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon$$

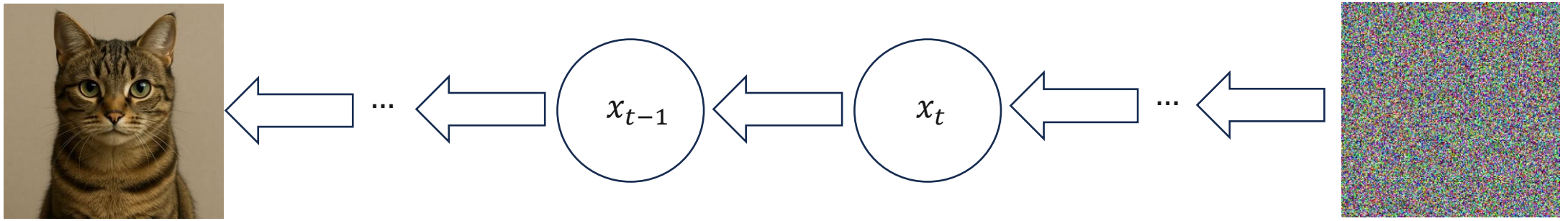
$$\epsilon \sim N(0, I)$$

Timestep (t) takes values in 1, …, T (e.g., T = 1000)

Ho, Jonathan, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models." *Advances in neural information processing systems* 33 (2020): 6840-6851.

Background: Diffusion Models

- Reverse denoising process that learns to generate data by denoising



x_0

x_T

Sampling Process: Gradually denoise a random noise to the image

$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon$$

$$\epsilon \sim N(0, I)$$

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\right) + \sigma_t \mathbf{z}$$

Neural Network predicts ϵ

The loss is **mean squared error**: $\mathcal{L} = \|\epsilon - \hat{\epsilon}\|^2$

Ho, Jonathan, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models." *Advances in neural information processing systems* 33 (2020): 6840-6851.

Text-to-Image Diffusion Models

- Text prompts guide image content.

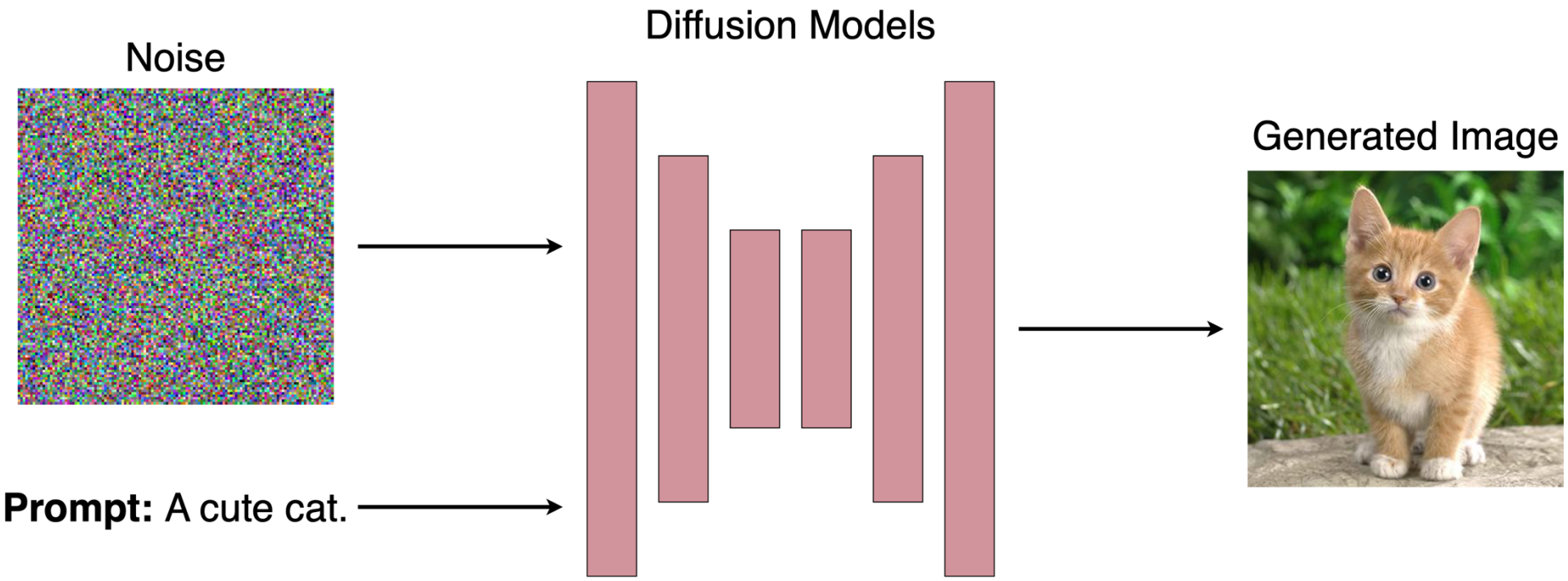
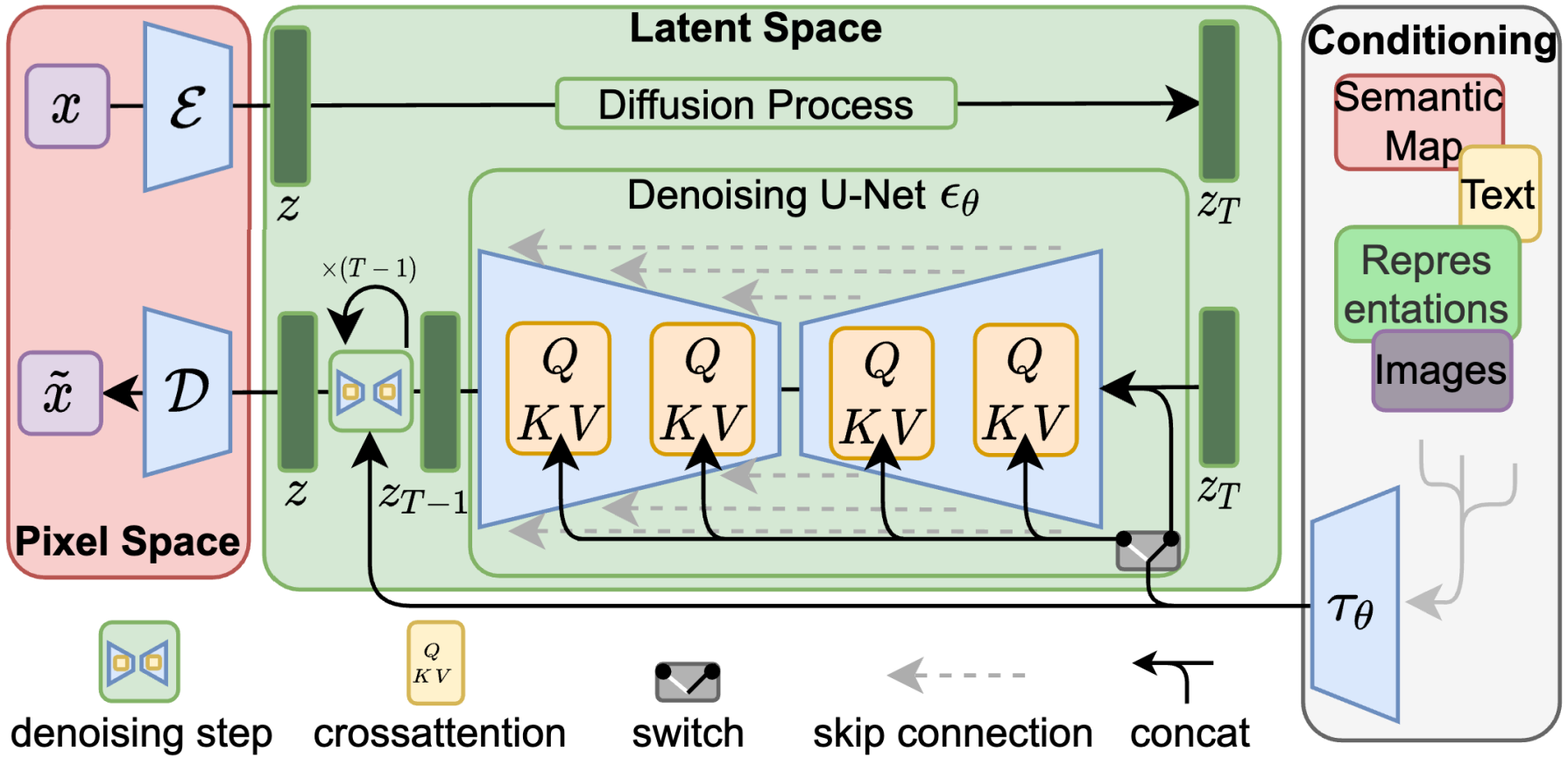


Image Source: https://hanlab.mit.edu/projects/can

Latent Diffusion Models



Edges, depth, and semantic maps specify visual structure beyond text.

High-resolution image synthesis with latent diffusion models, CVPR 2022

Technical Overview

PART I

Direct feedback for visual control

How can training directly enforce condition fidelity?

CHALLENGE

Condition fidelity

WORK

ControlNet++

ECCV 2024

PART II

Reliable and scalable preference learning

How can we construct, trust, and scale preference signals?

CHALLENGE

Cleaner comparisons

WORK

CPO

NeurIPS 2025

CHALLENGE

Conflicting preferences

WORK

Semi-DPO

ICLR 2026

CHALLENGE

Conflicting preferences
Learning at scale

WORK

ViPO

ICLR 2026

ControlNet++

Explicit Consistency Feedback for Visual Control

ControlNet++: Improving Conditional Controls with Efficient Consistency Feedback | ECCV 2024

Text Prompts Leave Visual Details Underspecified



Overall content

A deer stands on a grassy slope against misty mountains.

Object attributes

Deer: antlers and a brown coat.
Terrain: grass and small shrubs.
Background: misty mountains.

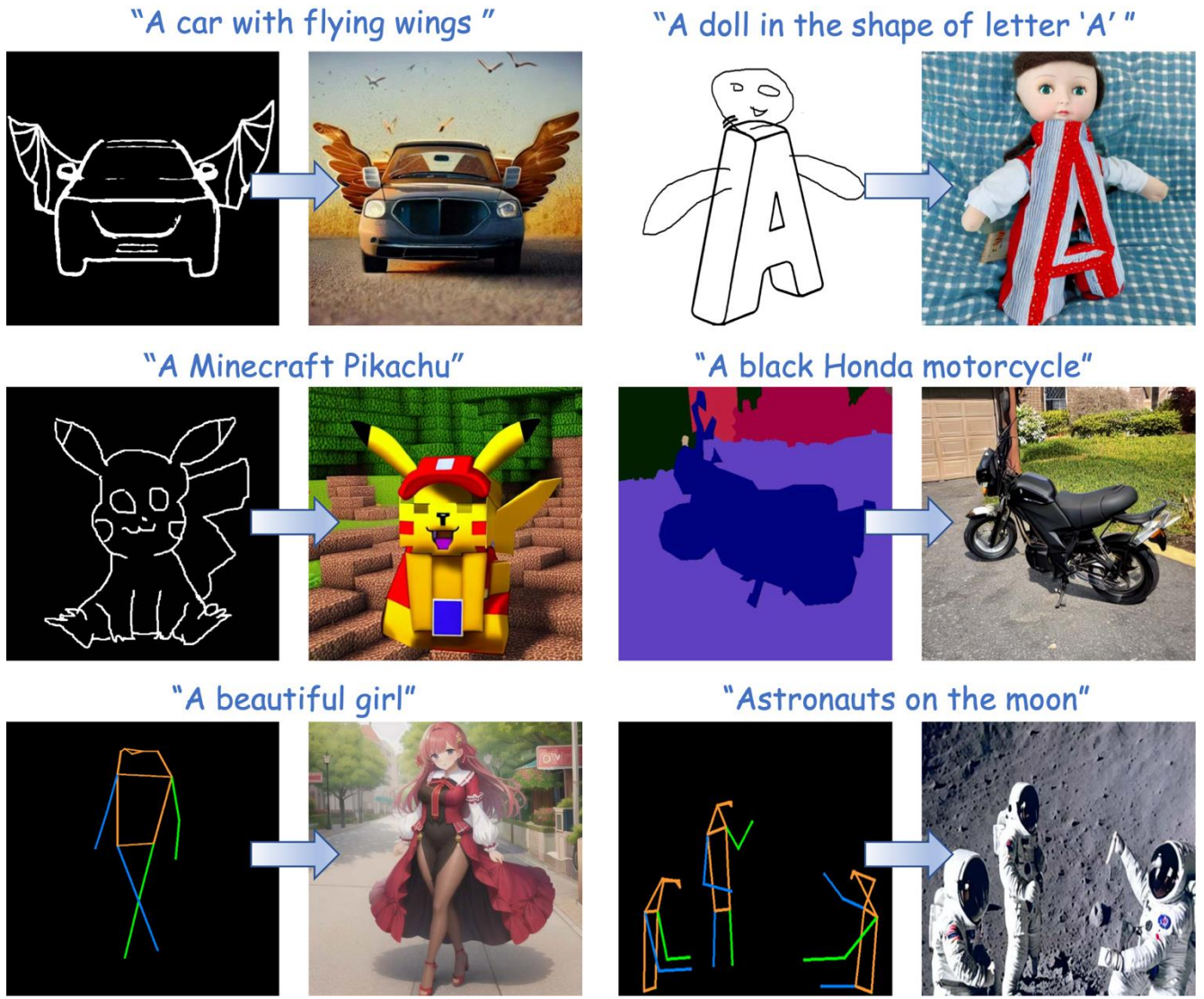
Hard to specify precisely

Exact pose and location
Fine contours and textures
Precise size and spatial relationships

An image is worth a thousand words

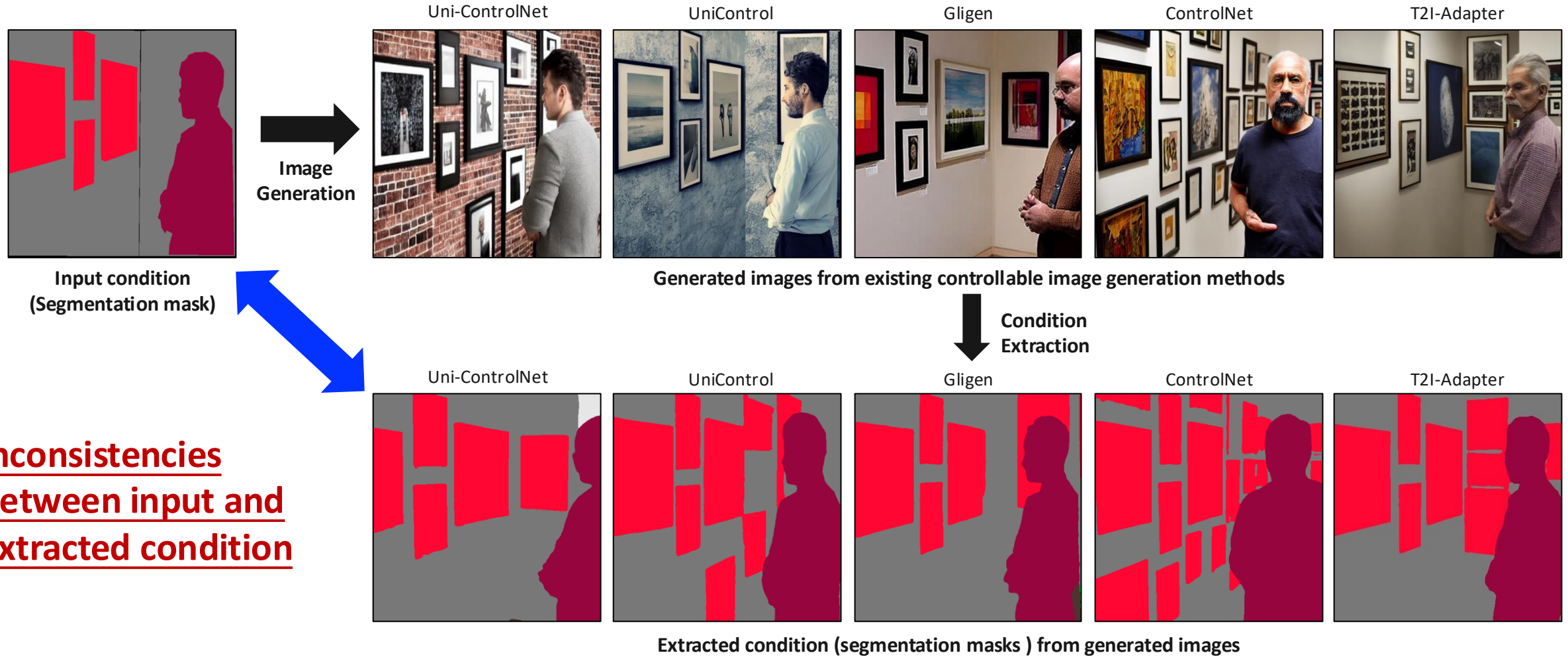
Image Control Signals Improve Controllability

ControlNet
T2I-Adapter



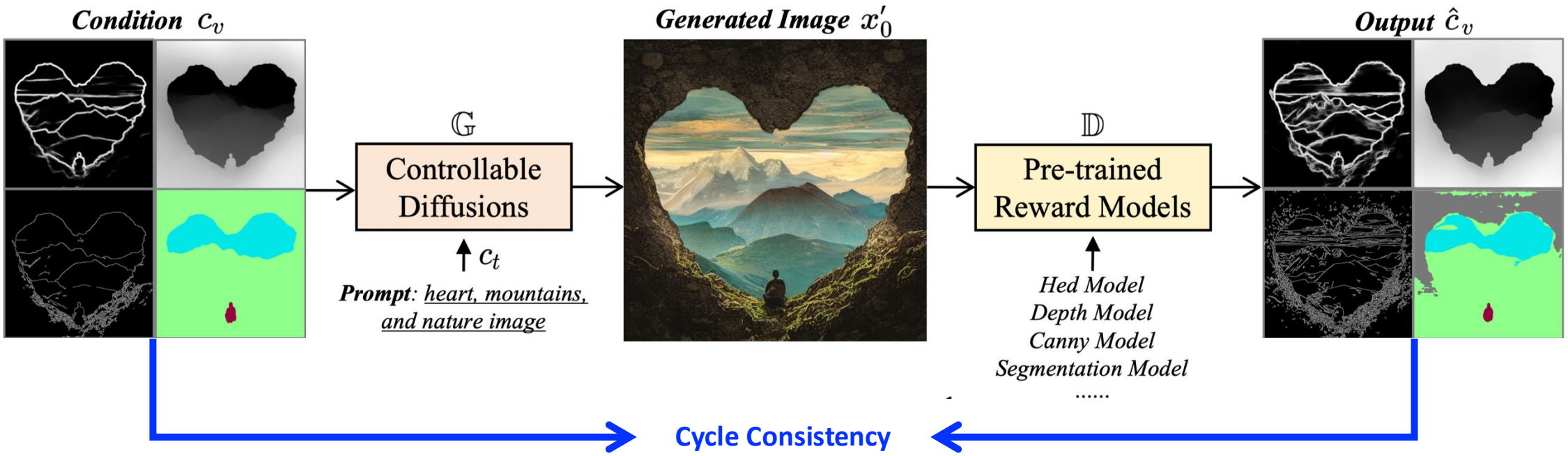
Adding Conditional Control to Text-to-Image Diffusion Models, ICCV 2023
T2I-Adapter: Learning Adapters to Dig out More Controllable Ability for Text-to-Image Diffusion Models, AAAI 2024

Visual Conditions Do Not Guarantee Precise Control



Improving Controllability by Cycle Consistency

Key idea: the condition extracted from a generated image should match the input condition.

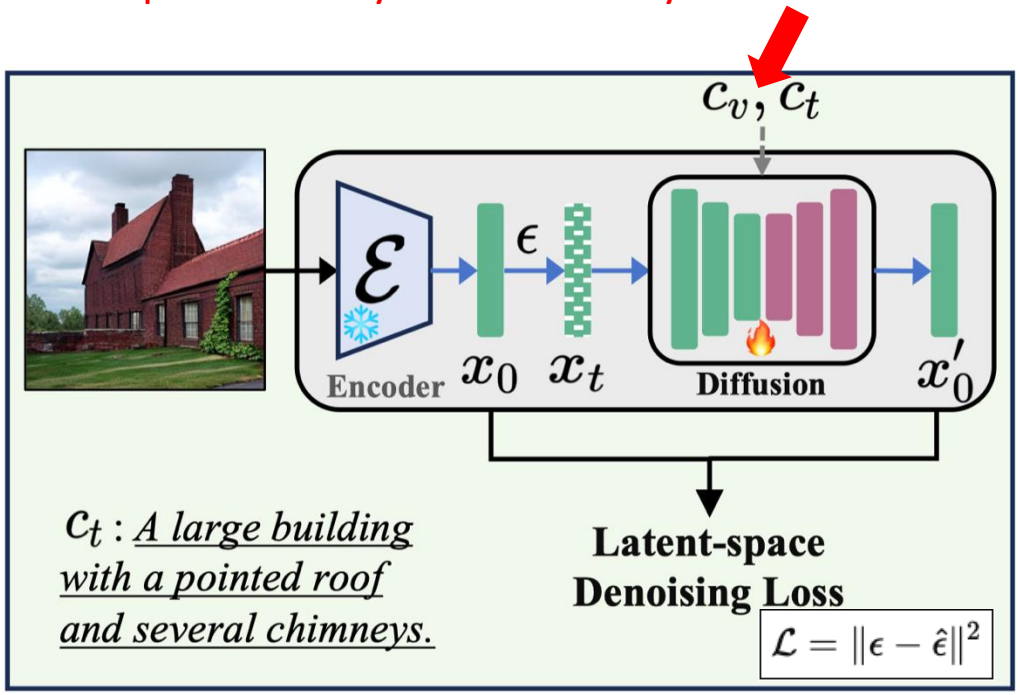


Measure pixel-wise differences

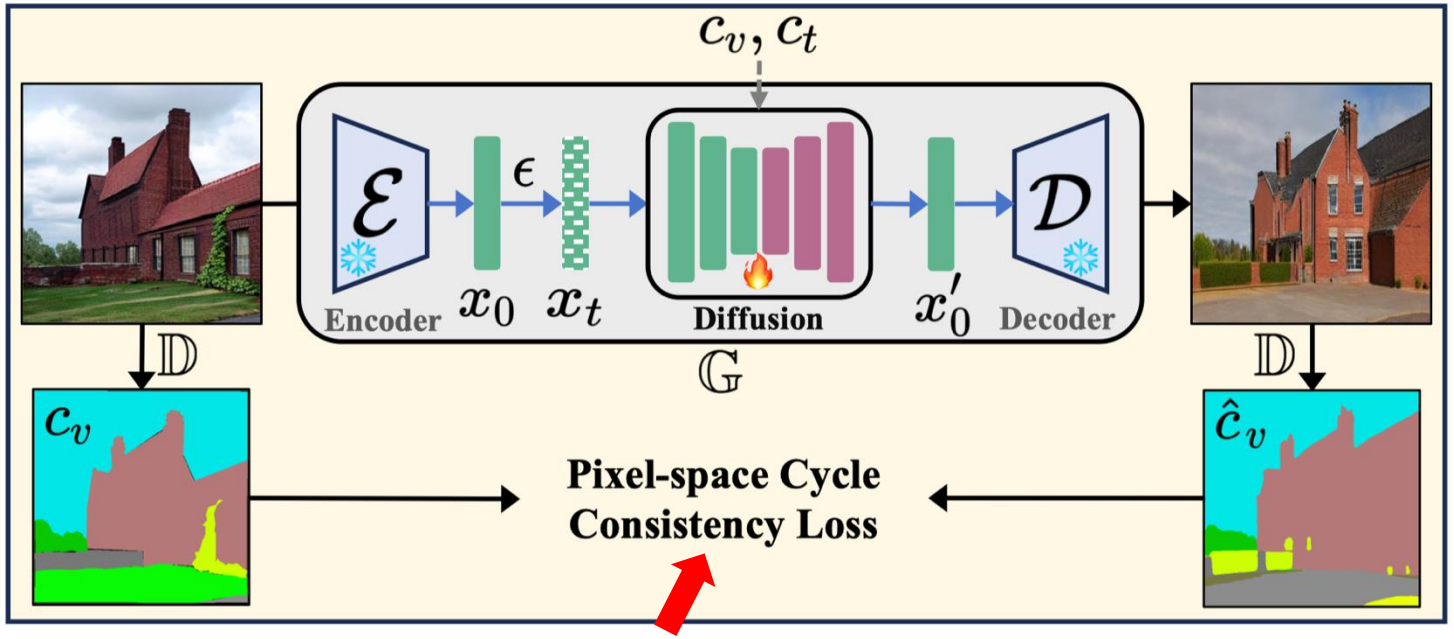
What Makes Our Method (ControlNet++) More Controllable?

- Existing methods achieve **implicit** controllability by introducing image-based conditional control c_v into the denoising process of diffusion models, with the guidance of latent-space denoising loss.

- We utilize discriminative reward models $\boldsymbol{D}$ to **explicitly** optimize the controllability of the diffusion model via pixel-level cycle consistency loss.

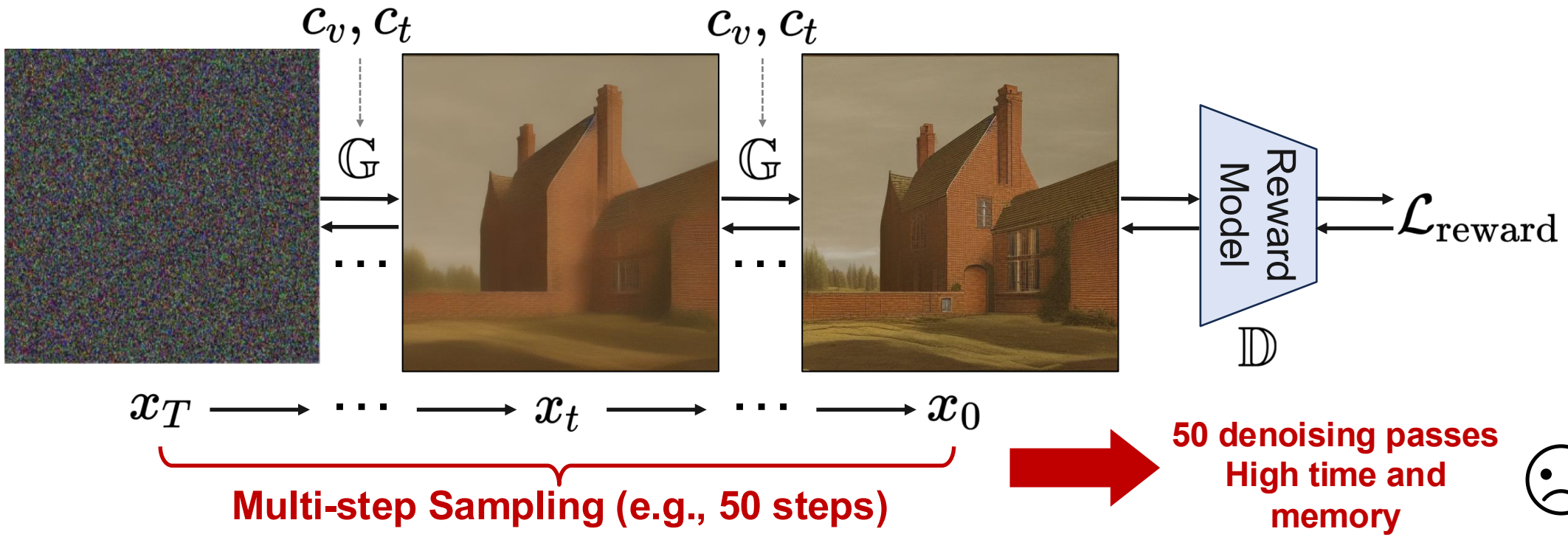


(a) Existing Methods



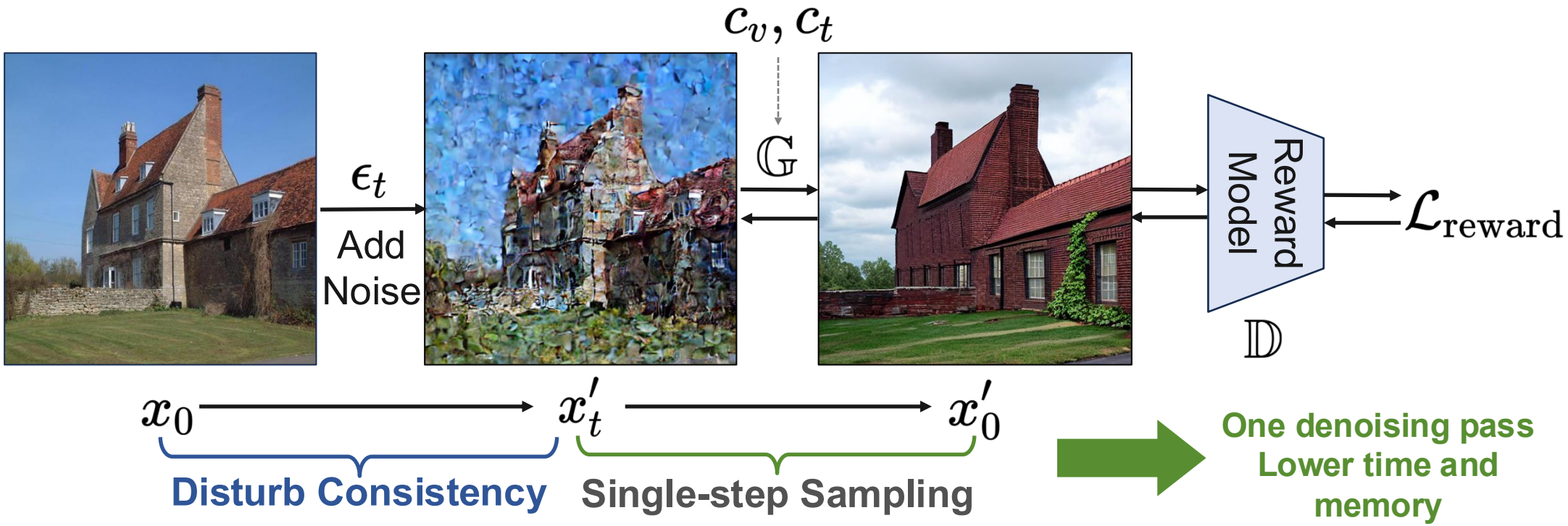
(b) Our Solution

Reward Feedback Through Full Sampling



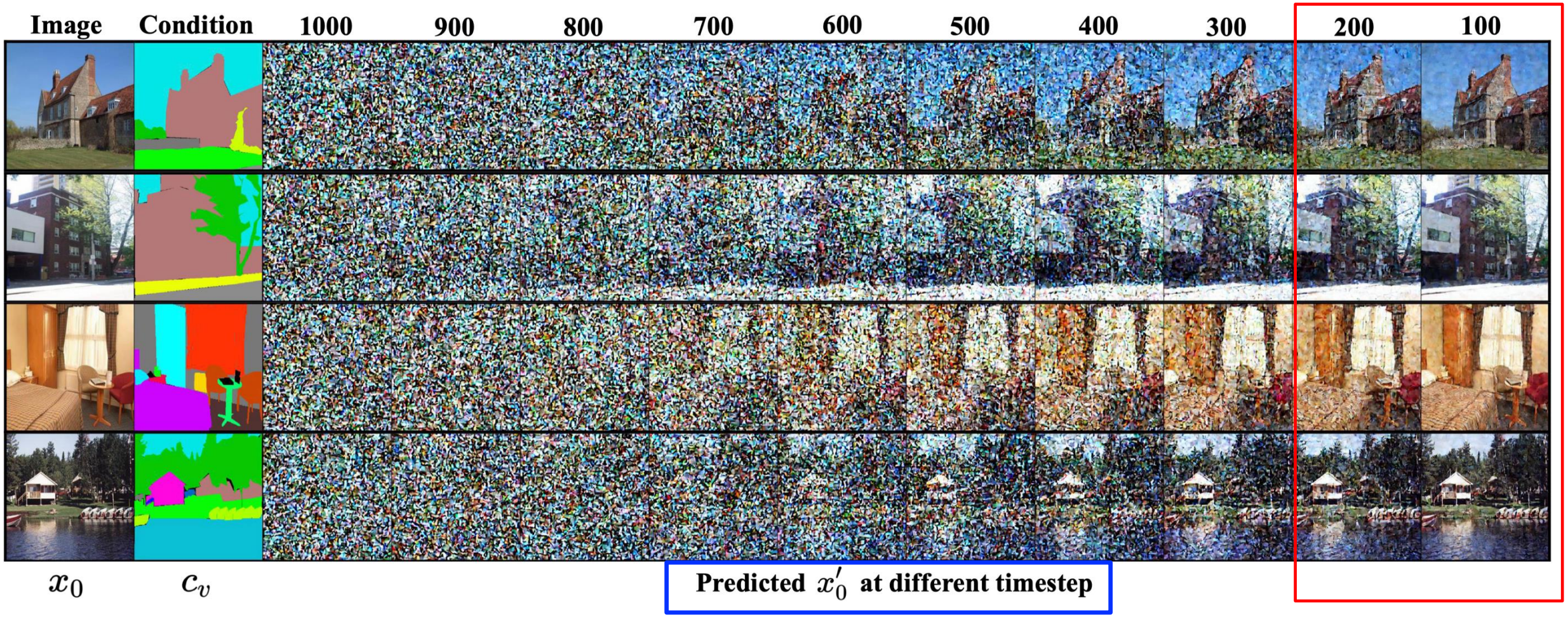
Backpropagation retains activations across the sampling trajectory.

Our Efficient Reward Strategy



$$
\begin{aligned}
\mathcal{L}_{\text{reward}} &= \mathcal{L}(c_v, \hat{c}_v) \\
&= \mathcal{L}\big(c_v, \mathbb{D}(x'_0)\big) \\
&= \mathcal{L}\big(c_v, \mathbb{D}\big[\mathbb{G}(c_t, c_v, x'_t, t)\big]\big),
\end{aligned}
$$

Single-Step Reconstruction Is Reliable at Low Noise



Diffusion Model Training

1. Sample a clean image

$x_0 \sim \text{data}$

2. Add random noise at a random timestep

$x_t = \alpha_t x_0 + \sigma_t \epsilon,$

$\epsilon \sim \mathcal{N}(0, I)$

3. Train the network to predict the noise

Input: x_t, timestep t, (optional condition c)

Output: predicted noise $\hat{\epsilon}$

4. Optimize with MSE loss

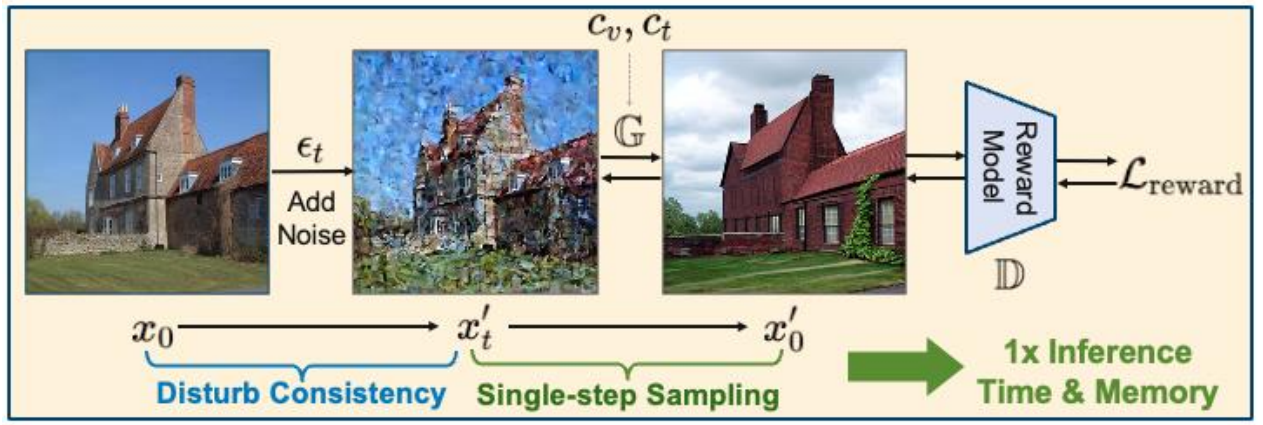
$\|\epsilon - \hat{\epsilon}\|^2$

ControlNet++ Training Loss

$$\mathcal{L}_{\text{total}} = \begin{cases} \mathcal{L}_{\text{train}} + \lambda \cdot \mathcal{L}_{\text{reward}}, & \text{if } t \leq t_{\text{thre}}, \\ \mathcal{L}_{\text{train}}, & \text{otherwise,} \end{cases}$$

Diffusion Loss: $\mathcal{L}_{\text{train}} = \mathbb{E}_{x_0,t,c_t,c_v,\epsilon\sim\mathcal{N}(0,1)} \left[\|\epsilon_\theta(x_t, t, c_t, c_v) - \epsilon\|_2^2 \right]$

The timestep threshold limits consistency feedback to low-noise reconstructions.

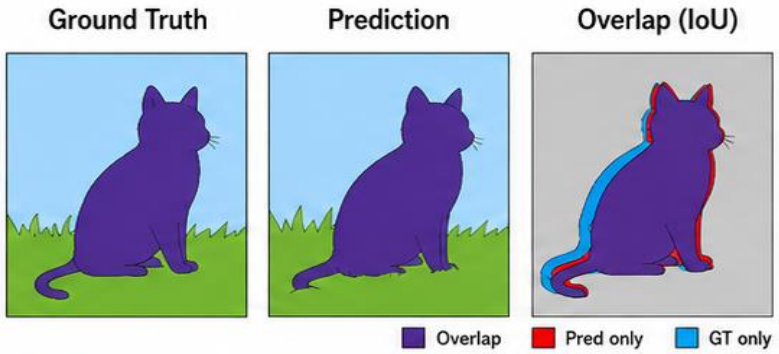


Evaluation

- Controllability
  - The consistency between the input condition and the condition extracted from the generated image.
  - The specific metric depends on each image condition
- Image Quality
  - FID compares feature distributions of generated and real images.
- Text-Image Alignment
  - CLIP-Score, measuring the image-text alignment between the input text and the generated image.

Evaluation Metrics for Controllability

1 Segmentation: mIoU

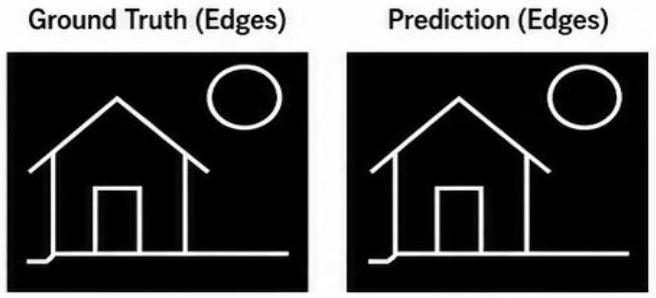


$$\mathrm{IoU}_c = \frac{|\mathrm{Pred}_c \cap \mathrm{GT}_c|}{|\mathrm{Pred}_c \cup \mathrm{GT}_c|}$$

$$\mathrm{mIoU} = \frac{1}{C}\sum_{c=1}^{C} \mathrm{IoU}_c$$

- Compute IoU for each class
- Average across classes
- Higher mIoU = better segmentation overlap

2 Edge Maps: F1-Score and SSIM



F1-Score

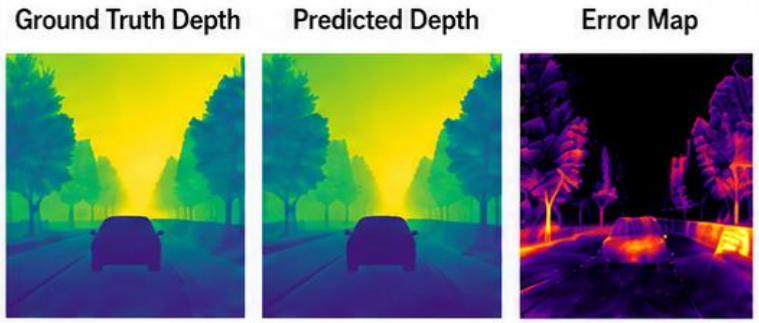
$$\mathrm{F1} = \frac{2TP}{2TP + FP + FN}$$

SSIM: structural similarity

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

- F1 evaluates edge detection accuracy
- SSIM compares structural similarity
- Higher is better for both

3 Depth Maps: RMSE



$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (D_{pred}(i) - D_{gt}(i))^2}$$

- Compute pixel-wise depth error
- Square, average, then take square root
- Lower RMSE = better depth estimation

★ mIoU measures segmentation overlap, F1/SSIM evaluate edge quality, and RMSE measures depth error.

Cycle Consistency Improves Visual Control

Metric	ControlNet	ControlNet++	Absolute gain
ADE20K mIoU (↑)	32.55	43.64	+11.09
COCO-Stuff mIoU (↑)	27.46	34.56	+7.10
Canny F1 (↑)	34.65	37.04	+2.39
HED SSIM (↑)	0.7621	0.8097	+0.0476
LineArt SSIM (↑)	0.7054	0.8399	+0.1345
Depth RMSE (↓)	35.90	28.32	-7.58

ControlNet++ improves all six control metrics over ControlNet.

Evaluation Metrics for Image Quality and Image-Text Alignment

1. FID

Fréchet Inception Distance



$$\text{FID} = \|\mu_r - \mu_g\|^2 + Tr(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{1/2})$$

- Compares generated images with real images
- Measures realism / distribution match
- Lower is better

2. CLIP-Score

Text-image alignment



$$\text{CLIP-Score} = \cos(E_{text}(\text{prompt}), E_{img}(\text{image}))$$

- Measures how well the image matches the prompt
- Focuses on semantic alignment
- Higher is better

FID measures realism. CLIP-Score measures prompt alignment.

Control Gains with Competitive FID and CLIP Scores

Tab. FID (↓) comparison

Method	T2I Model	Seg. Mask		Canny Edge	Hed Edge	LineArt Edge	Depth Map
		ADE20K	COCO	MultiGen-20M	MultiGen-20M	MultiGen-20M	MultiGen-20M
Gligen	SD1.4	33.02	-	18.89	-	-	18.36
T2I-Adapter	SD1.5	39.15	-	15.96	-	-	22.52
UniControlNet	SD1.5	39.70	-	17.14	17.08	-	20.27
UniControl	SD1.5	46.34	-	19.94	15.99	-	18.66
ControlNet	SD1.5	33.28	21.33	**14.73**	15.41	17.44	17.76
Ours	SD1.5	**29.49**	**19.29**	18.23	**15.01**	**13.88**	**16.66**

Tab. CLIP-score (↑) comparison

Method	T2I Model	Seg. Mask		Canny Edge	Hed Edge	LineArt Edge	Depth Map
		ADE20K	COCO	MultiGen-20M	MultiGen-20M	MultiGen-20M	MultiGen-20M
Gligen	SD1.4	31.12	-	31.77	-	-	31.75
T2I-Adapter	SD1.5	30.65	-	31.71	-	-	31.46
UniControlNet	SD1.5	30.59	-	31.84	31.94	-	31.66
UniControl	SD1.5	30.92	-	31.97	32.02	-	**32.45**
ControlNet	SD1.5	31.53	**13.31**	**32.15**	**32.33**	**32.46**	**32.45**
Ours	SD1.5	**31.96**	13.13	31.87	32.05	31.95	32.09

Better control with competitive FID and CLIP scores.

Better Control Produces Better Segmentation Data

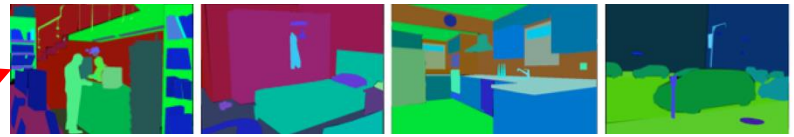


ADE20K Segmentation Dataset



...

Take the ground truth masks in ADE20K as condition



Generate synthetic images

Models: Baseline – ControlNet

Our ControlNet++

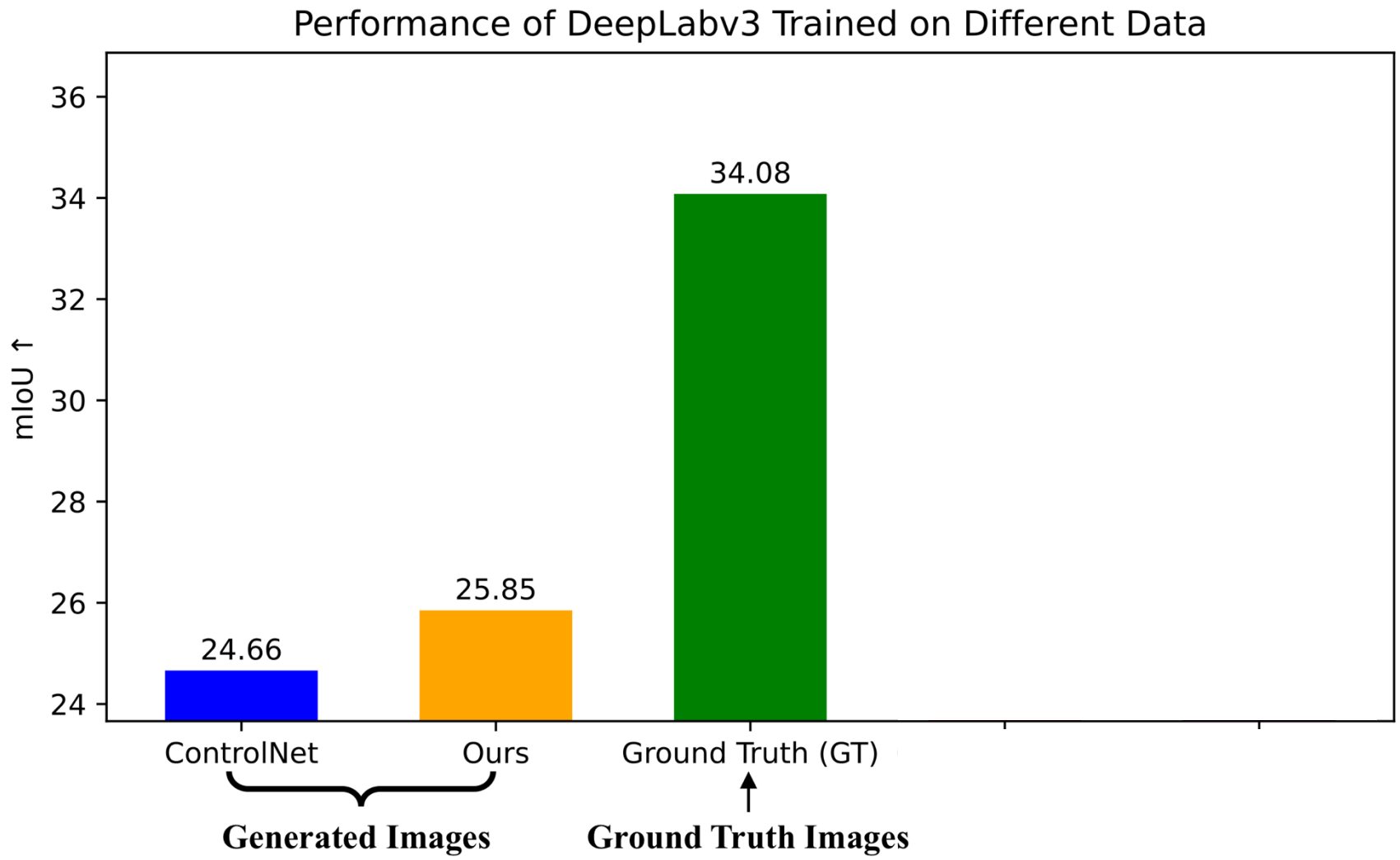
Synthetic image–mask pairs



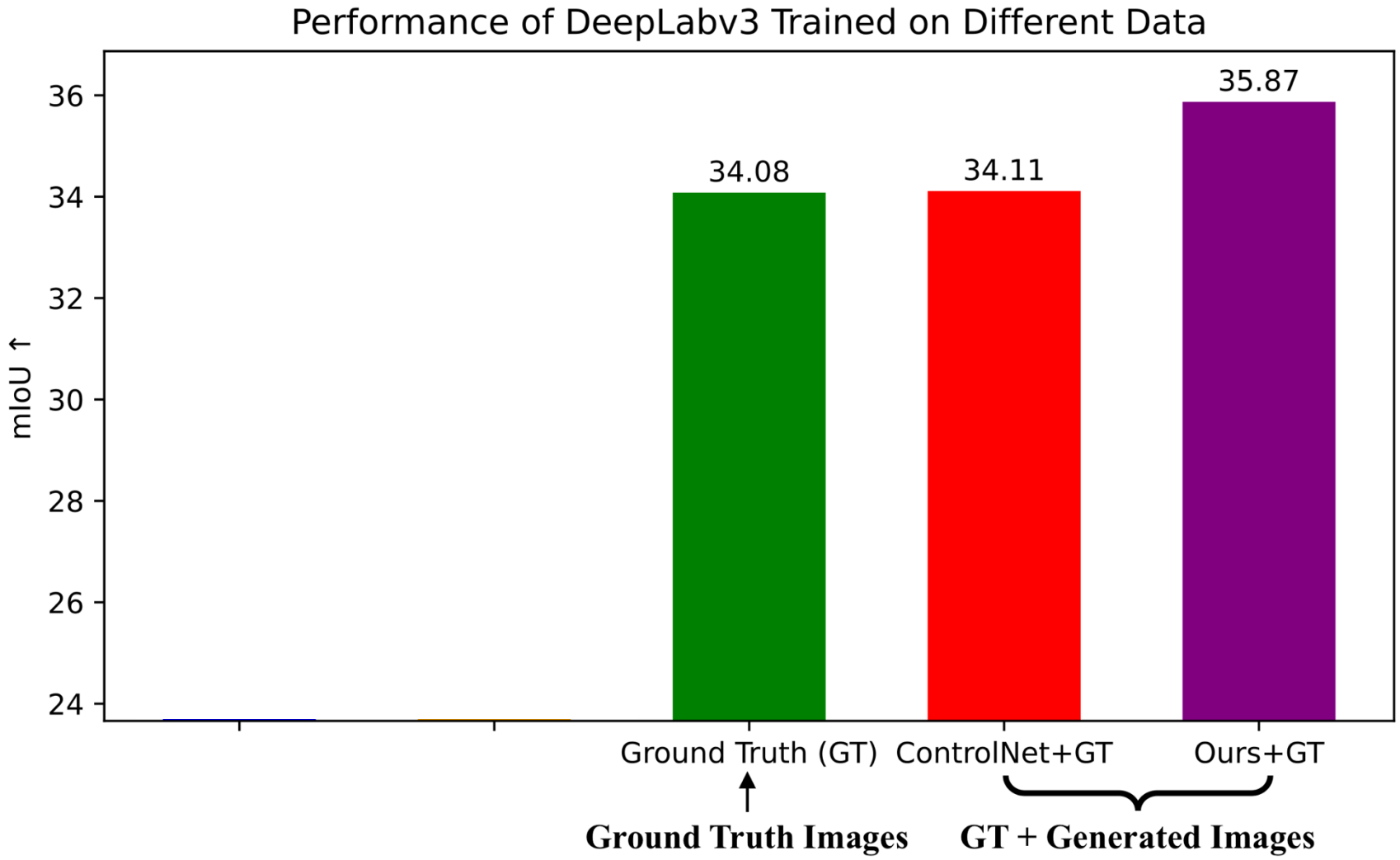
...

Train a segmentation model (e.g. DeepLabv3) from scratch and evaluate its performance

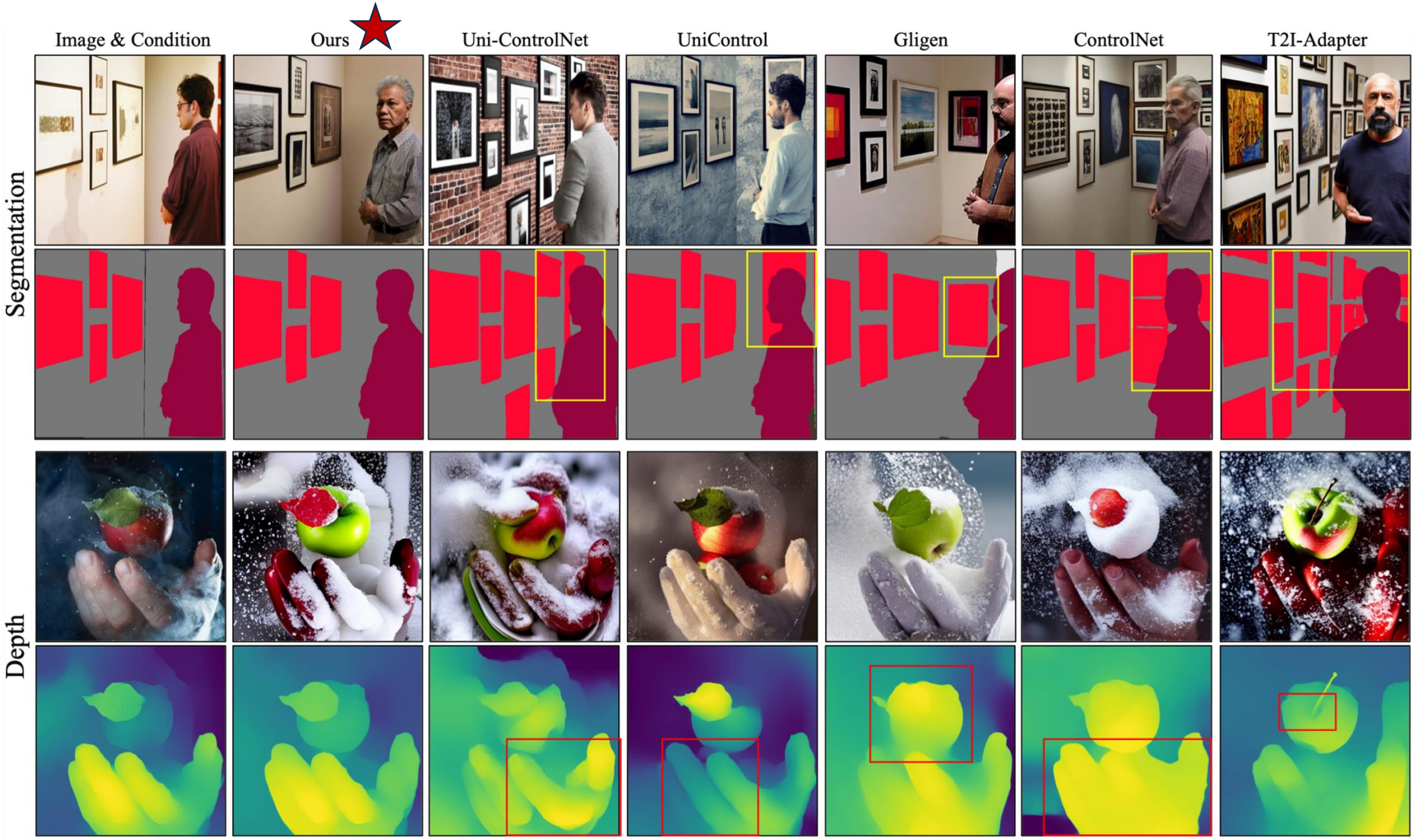
Segmentation Training with Synthetic Images



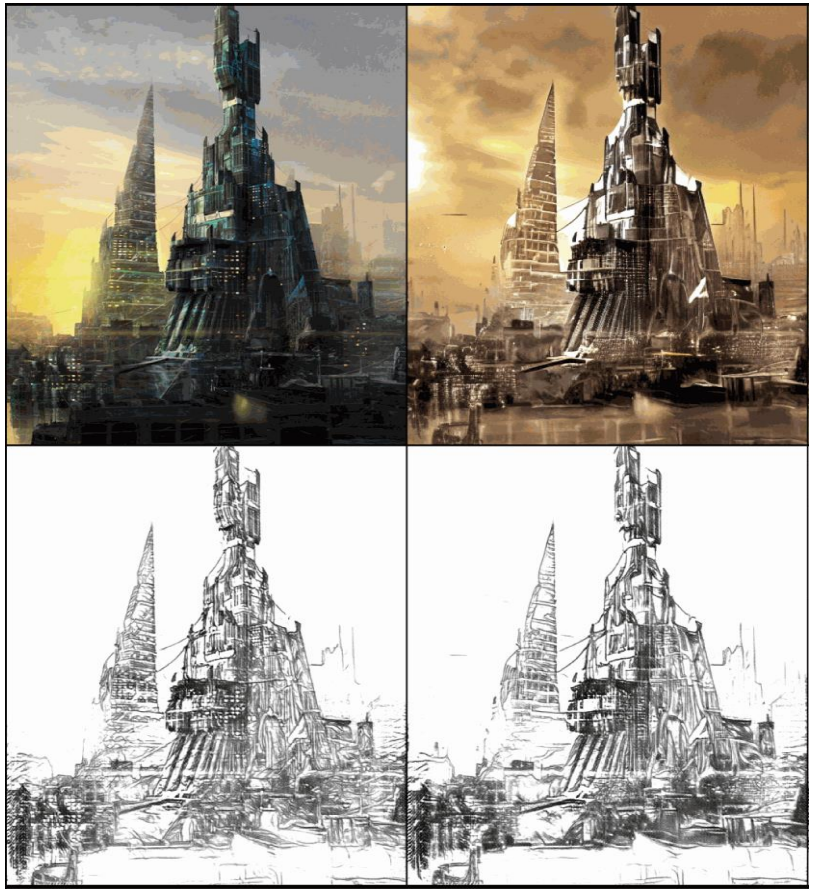
Synthetic Data Improves Real-Data Training



Qualitative Results



More Visualization Results







From Direct Feedback to Preference Optimization

- ControlNet++ directly rewards condition fidelity, but its single-step reconstruction is reliable only at low-noise timesteps.
- Preference optimization offers another path: learn from relative comparisons at arbitrary diffusion timesteps.

The next challenge is constructing reliable preference signals.

Technical Overview

PART I

Direct feedback for visual control

How can training directly enforce condition fidelity?

CHALLENGE

Condition fidelity

WORK

ControlNet++

ECCV 2024

PART II

Reliable and scalable preference learning

How can we construct, trust, and scale preference signals?

CHALLENGE

Cleaner comparisons

WORK

CPO

NeurIPS 2025

CHALLENGE

Conflicting preferences

WORK

Semi-DPO

ICLR 2026

CHALLENGE

Conflicting preferences Learning at scale

WORK

ViPO

ICLR 2026

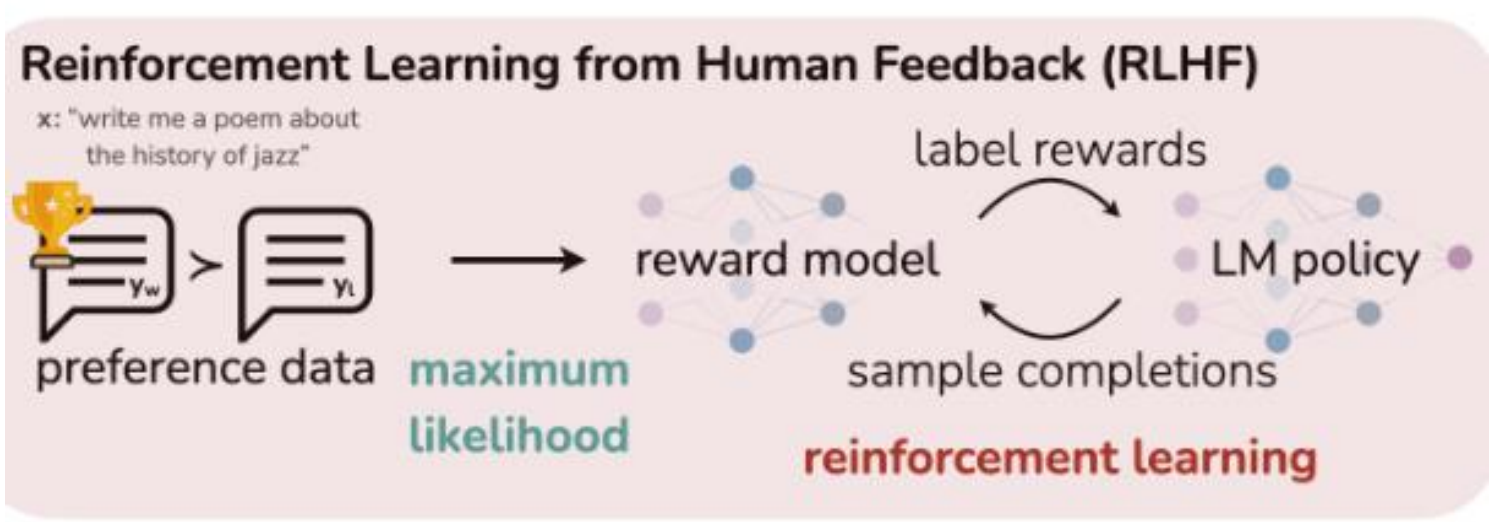
CPO

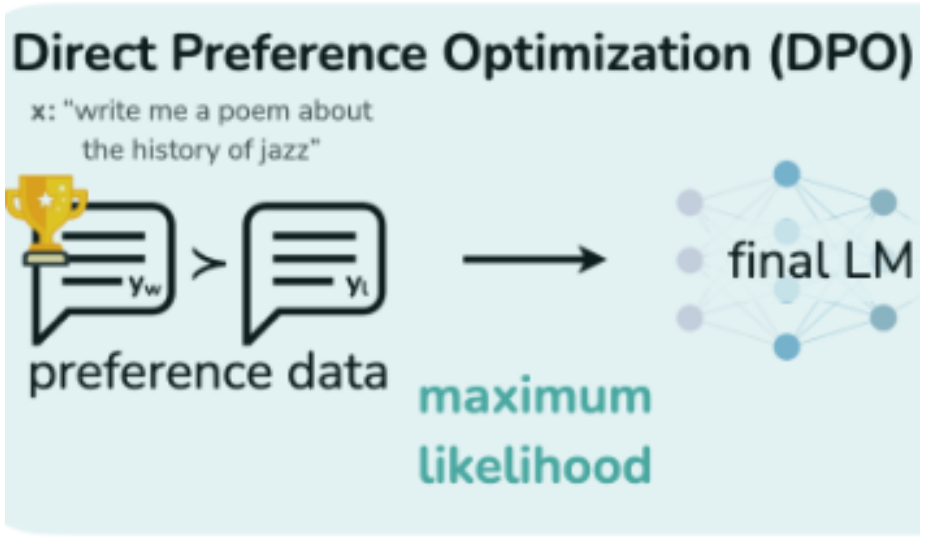
Cleaner Condition Comparisons Across Diffusion Timesteps

CPO: Condition Preference Optimization for Controllable Image Generation | NeurIPS 2025

From RLHF to Direct Preference Optimization

DPO learns directly from preference pairs without an RL training loop.





- Fit a reward model to a dataset of prompts and human preferences over pairs of responses
- Use RL to find a policy that maximizes the learned reward

- DPO directly optimizes for the policy best satisfying the preferences with a simple classification objective

Rafailov, Rafael, et al. "Direct preference optimization: Your language model is secretly a reward model." Advances in neural information processing systems 36 (2023): 53728-53741.

Diffusion-DPO

- DPO for text-to-image diffusion models
- Compare a preferred and rejected image under the same prompt
- Train with reference-relative denoising-error differences



I^l: less preferred I^w: Preferred

Wallace, Bram, Meihua Dang, Rafael Rafailov, Linqi Zhou, Aaron Lou, Senthil Purushwalkam, Stefano Ermon, Caiming Xiong, Shafiq Joty, and Nikhil Naik. "Diffusion model alignment using direct preference optimization." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8228-8238. 2024.

Image Preferences Can Mix Control and Quality

A more controllable image may still contain visual artifacts.



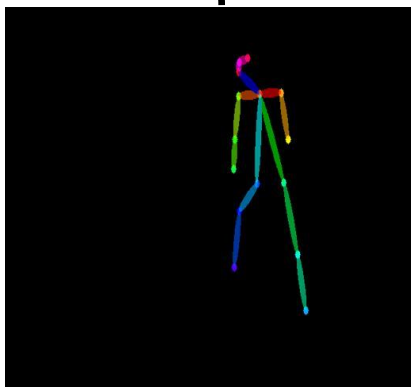
Better pose alignment



I^w



I^l



c

Hard to guarantee quality

How about using the real image as the winner?

A tempting image-pair construction

Same pose condition c

WINNER

Real image

(I_{real}, c)

Pose matches c

vs.

LOSER

Generated image

$(I_{generated}, c)$

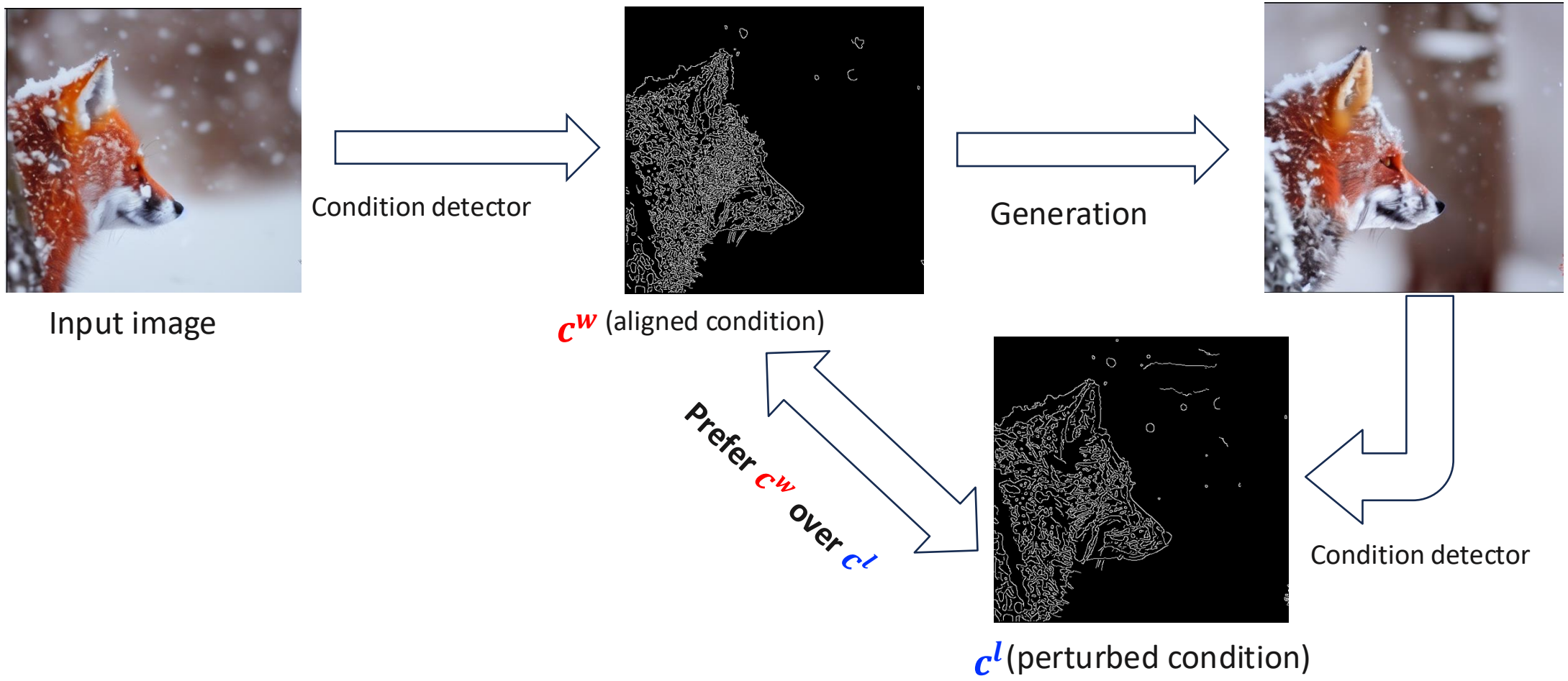
Pose may also match c

The generated image may follow the pose but differ in texture, color, or realism.

The preference label does not isolate controllability.

CPO Data: One Image, Two Conditions

- Condition Preference Optimization
- Compare correctly aligned and perturbed conditions
- Keep the training image fixed

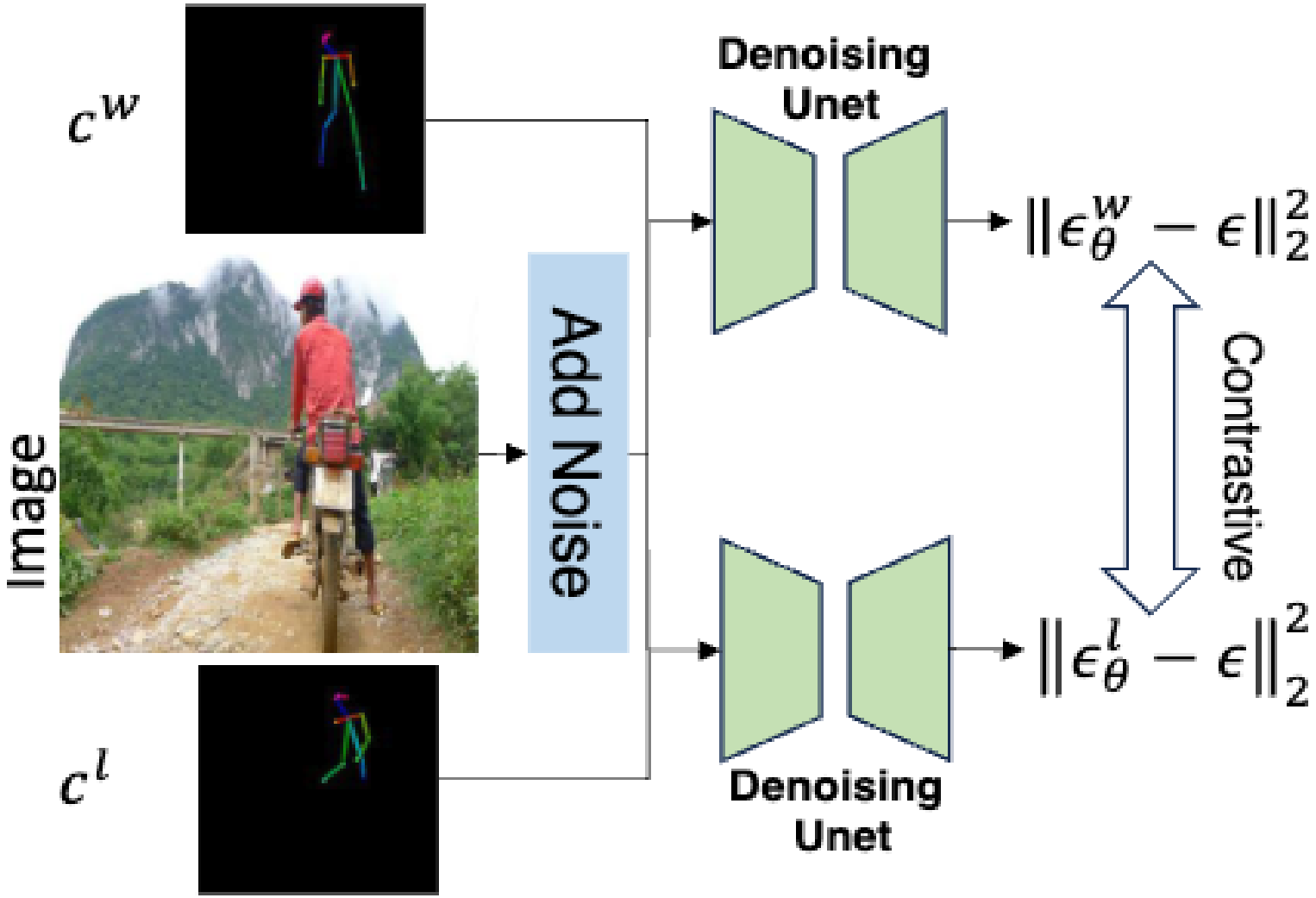


Why condition preference is cleaner

	Image-based DPO	CPO
Comparison	Two images	**Two conditions**
Variation	Control + visual factors vary	**Image remains fixed**
Curation	Sample and filter image pairs	**One generated sample**
Signal	Confounded, higher variance	**Focused, lower variance**

Keep the image fixed to isolate condition fidelity.

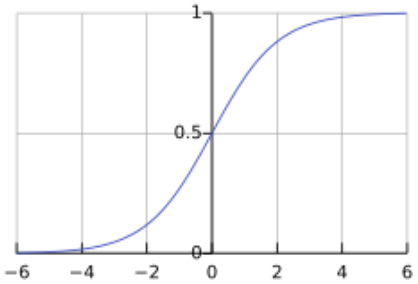
CPO Training: One Image, Two Conditions



Condition Preference Optimization

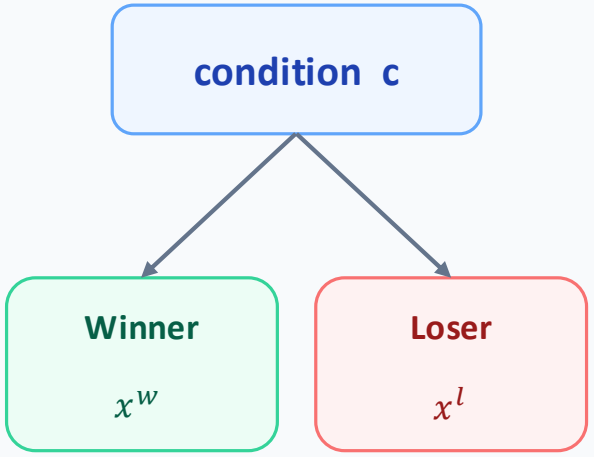
Direct Preference Optimization: From Pair to Loss

One condition, two outputs, one labeled preference



Sigmoid function

Preference pair



Label: the winner should rank above the loser

1 **Compare the current generator with a frozen reference**

$$r_\theta(x, c) = \beta \log \frac{p_\theta(x \mid c)}{p_{\text{ref}}(x \mid c)}$$

2 **Convert the reward difference into a model preference score**

$$p \equiv P_\theta\left(x^w \succ x^l \mid c\right) = \sigma\left(r_\theta(x^w, c) - r_\theta\left(x^l, c\right)\right)$$

3 **Minimize negative log-probability of the labeled preference**

$$\mathcal{L}_{\text{DPO}} = -\mathbb{E}_{(c,x^w,x^l)\sim\mathcal{D}} \log \sigma\left(\beta\left[\log \frac{p_\theta(x^w \mid c)}{p_{\text{ref}}(x^w \mid c)} - \log \frac{p_\theta(x^l \mid c)}{p_{\text{ref}}(x^l \mid c)}\right]\right)$$

DPO increases the current model's relative preference for the labeled winner, while remaining anchored to the reference policy

From Image Preference to Condition Preference

DPO compares two images under one condition; CPO compares two conditions for one fixed image.

$$\mathcal{L}_{\text{DPO}} = -\mathbb{E}_{\mathbf{c},\mathbf{x}_0^w,\mathbf{x}_0^l} \log \sigma \left[\mathbb{E}_{\substack{\mathbf{x}_{1:T}^w \sim p_\theta(\mathbf{x}_{1:T}^w|\mathbf{x}_0^w) \\ \mathbf{x}_{1:T}^l \sim p_\theta(\mathbf{x}_{1:T}^l|\mathbf{x}_0^l)}} \left(\log \frac{p_\theta(\mathbf{x}_0^w|\mathbf{c})}{p_{\text{ref}}(\mathbf{x}_0^w|\mathbf{c})} - \log \frac{p_\theta(\mathbf{x}_0^l|\mathbf{c})}{p_{\text{ref}}(\mathbf{x}_0^l|\mathbf{c})} \right) \right]$$

$$\mathcal{L}_{\text{CPO}} = -\mathbb{E}_{\mathbf{c}^w,\mathbf{c}^l,\mathbf{x}_0} \log \sigma \left[\mathbb{E}_{\mathbf{x}_{1:T} \sim p_\theta(\mathbf{x}_{1:T}|\mathbf{x}_0)} \left(\log \frac{p_\theta(\mathbf{x}_0|\mathbf{c}^w)}{p_{\text{ref}}(\mathbf{x}_0|\mathbf{c}^w)} - \log \frac{p_\theta(\mathbf{x}_0|\mathbf{c}^l)}{p_{\text{ref}}(\mathbf{x}_0|\mathbf{c}^l)} \right) \right]$$

Quantitative Results: Controllability

Method	T2I Model	Segmentation		Pose		Canny	HED	Lineart	Depth
		ADE20K	COCO-Stuff	COCO-Pose	HumanArt	MultiGen-20M			
		mIoU ↑		mAP ↑		F1 ↑	SSIM ↑	SSIM ↑	RMSE ↓
ControlAR (CFG = 4.0)	LlamaGen-XL	39.95†	**37.49**	-	-	36.78*	0.8184*	0.7922†	29.33*
Ours (CFG = 4.0)	SD1.5	**46.38**	36.10	-	-	**39.68**	**0.8299**	**0.8559**	**25.98**
T2I-Adapter	SD1.5	12.61	-	-	-	23.65	-	-	48.40
Gligen	SD1.4	23.78	-	87.84	39.88	26.94	0.5634	-	38.83
Uni-ControlNet	SD1.5	19.39	-	12.50	4.23	27.32	0.6910	-	40.65
UniControl	SD1.5	25.44	-	75.29	27.90	30.82	0.7969	-	39.18
ControlNet	SD1.5	32.55	27.46	72.53	34.52	34.65	0.7621	0.7054	35.90
ControlNet++	SD1.5	43.64	34.56	-	-	38.03*	0.8097	0.8399	28.32
Ours	SD1.5	**44.81**	**35.49**	**87.98**	**45.71**	**39.28**	**0.8201**	**0.8447**	**27.49**
Ours-DINOv2	SD1.5	43.72	39.25	88.45	49.98	44.81	0.8376	0.8301	26.92
Oracle	-	55.07	41.32	92.54	50.03	100.0	1.0	1.0	0.0

CPO improves its starting model across all tested control types.

CPO vs DPO

Quantitative comparison on COCO-Pose

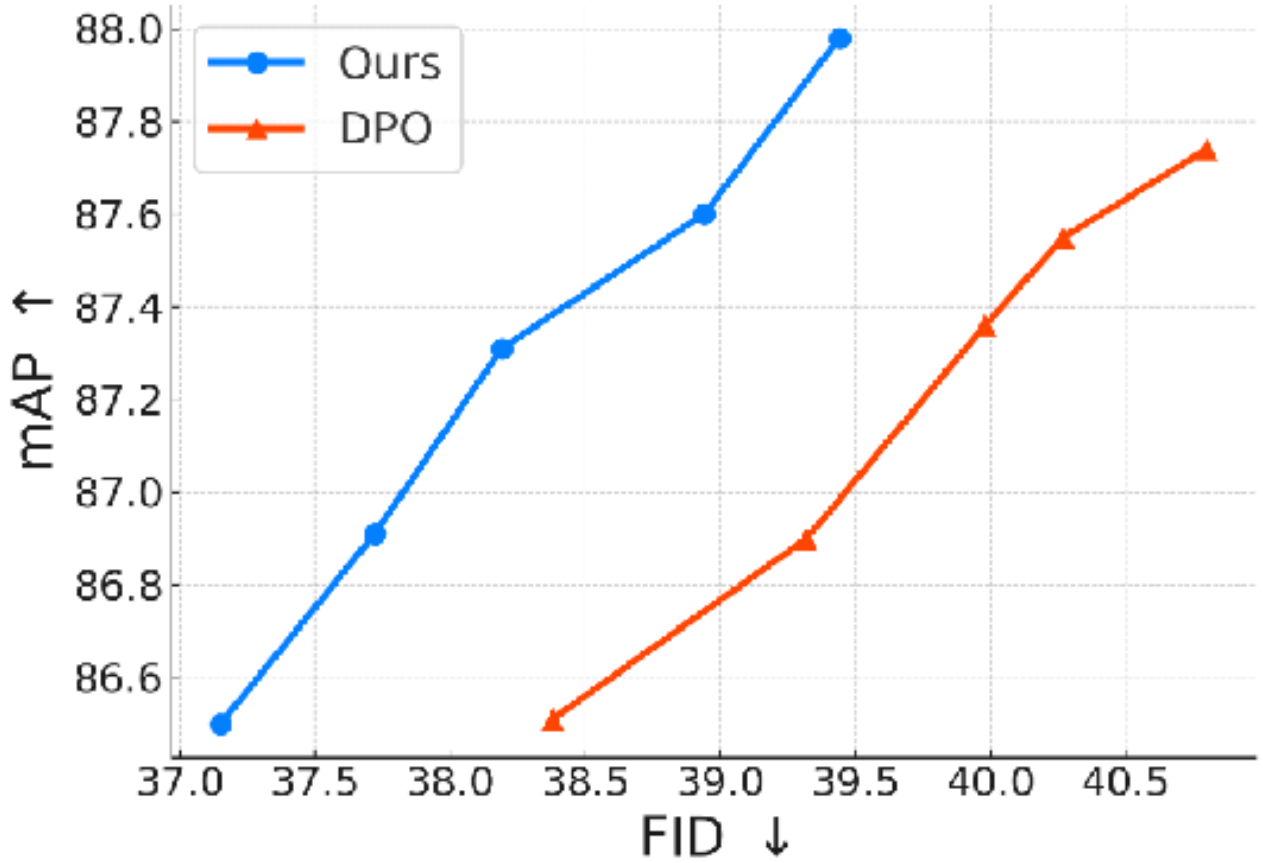
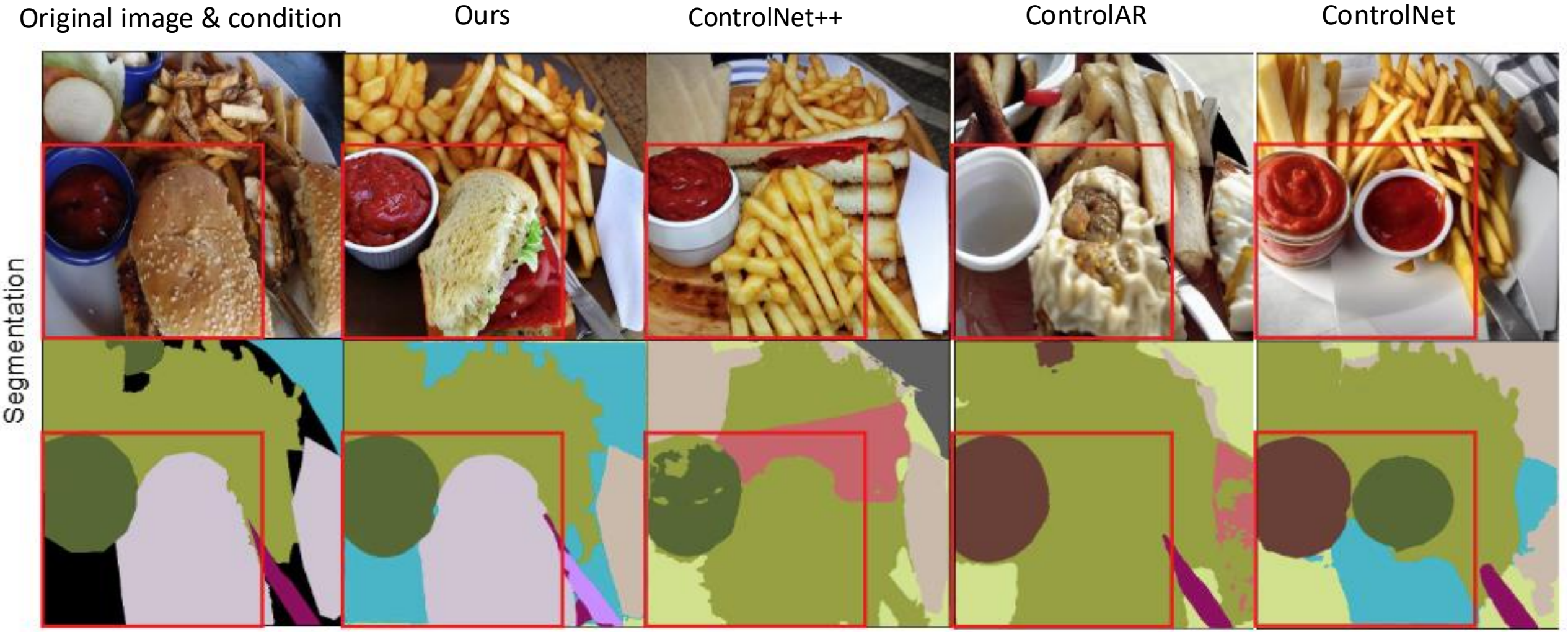


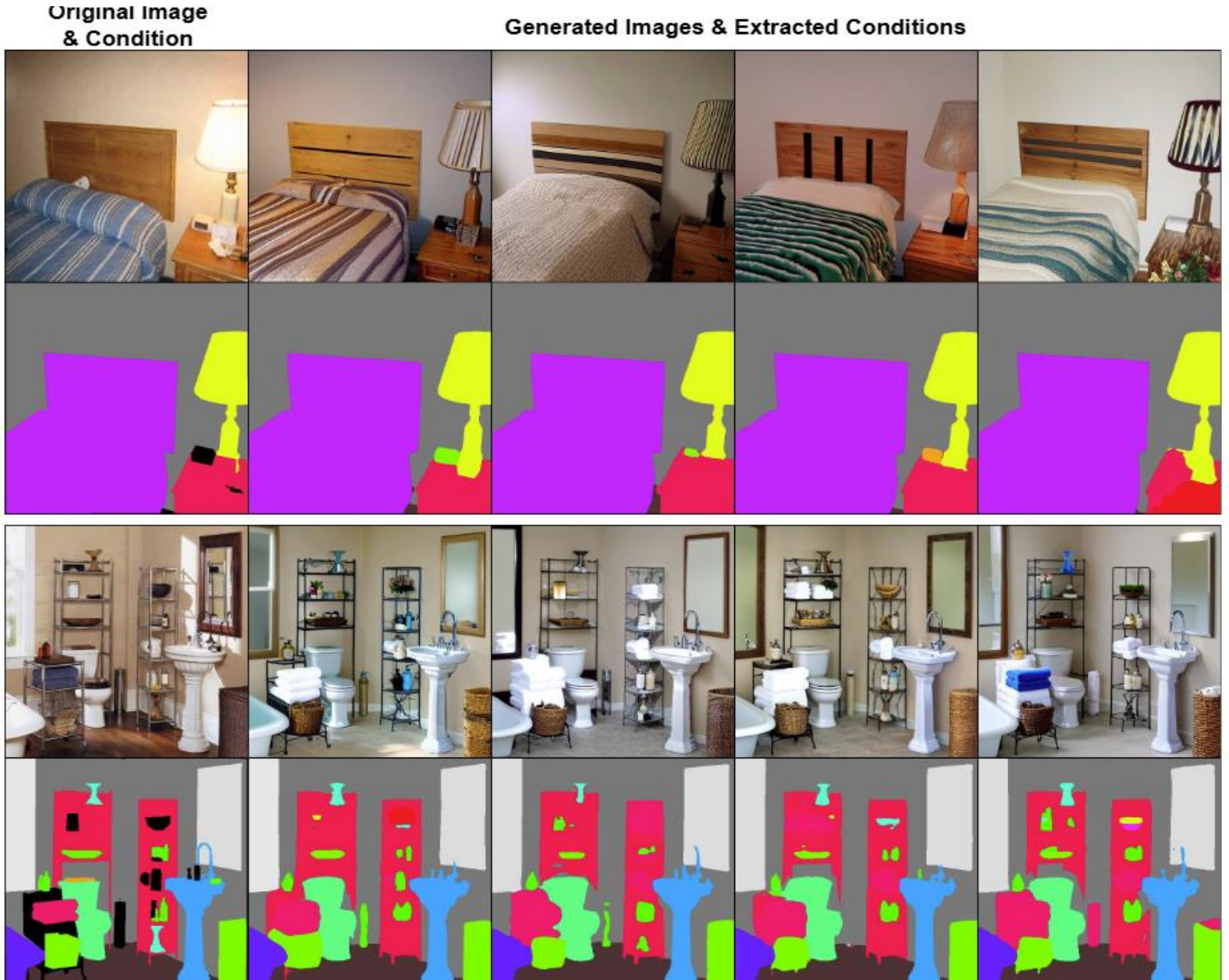
Figure 6: Comparison with DPO. We evaluated 5 checkpoints equally throughout training and observed that under the same mAP level, our method gets better FID.

At matched pose accuracy, CPO achieves lower FID than DPO

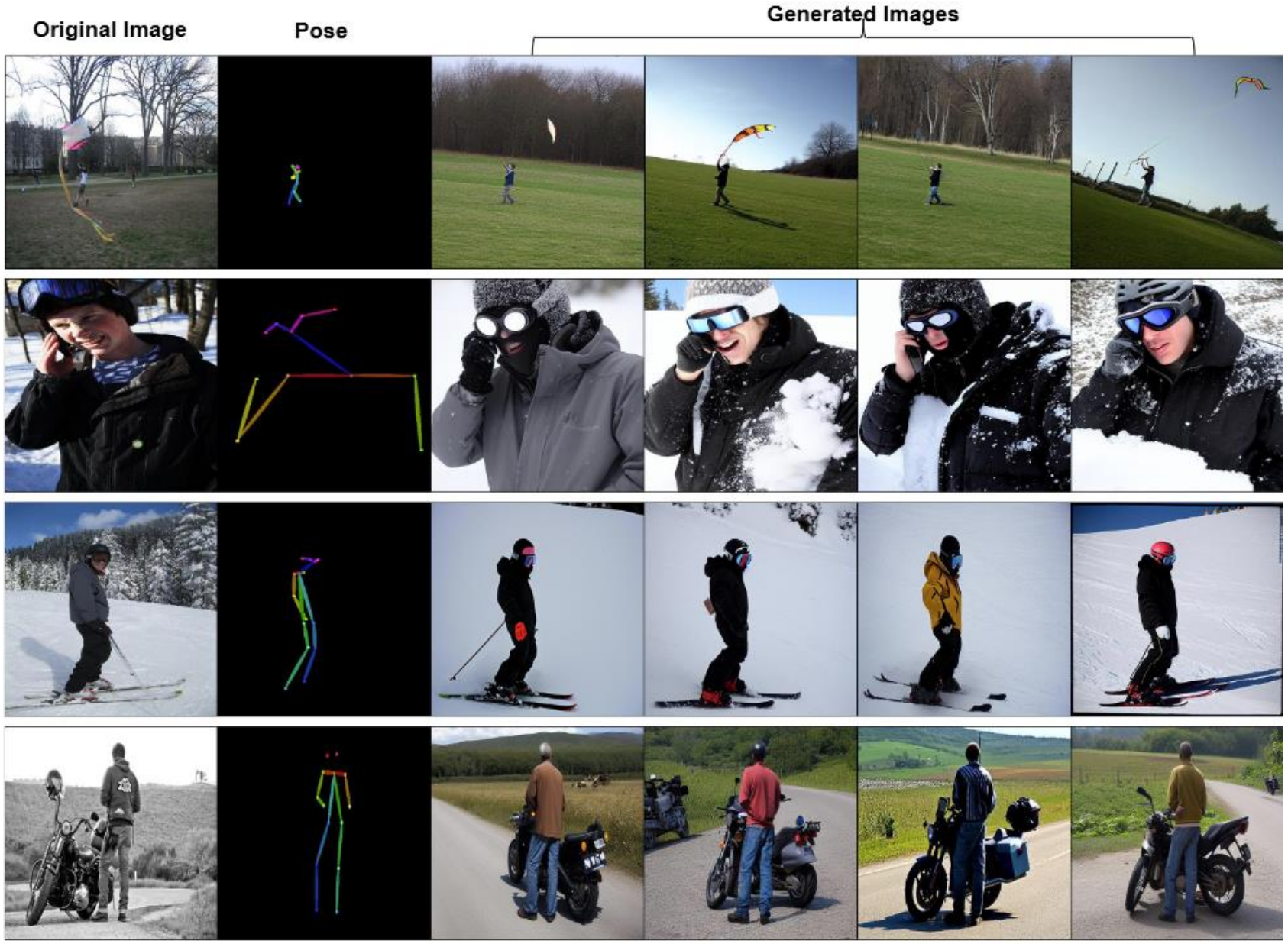
Qualitative Comparison: Segmentation Control



More Segmentation-Control Examples



Pose-Control Examples



Technical Overview

PART I

Direct feedback for visual control

How can training directly enforce condition fidelity?

CHALLENGE

Condition fidelity

WORK

ControlNet++

ECCV 2024

PART II

Reliable and scalable preference learning

How can we construct, trust, and scale preference signals?

CHALLENGE

Cleaner comparisons

WORK

CPO

NeurIPS 2025

CHALLENGE

Conflicting preferences

WORK

Semi-DPO

ICLR 2026

CHALLENGE

Conflicting preferences Learning at scale

WORK

ViPO

ICLR 2026

CPO constructs cleaner comparisons.
Semi-DPO learns from conflicting preference labels.

Semi-DPO

Robust Preference Optimization via Semi-Supervised Learning

Learning from Noisy Preferences: A Semi-Supervised Learning Approach to Direct Preference Optimization | ICLR 2026

One Binary Preference Can Hide Conflicting Visual Attributes



Image A



Image B

Prompt : A vast green grassland with blue sky and two white clouds, a mother cow and her calf **both eating grass,** natural landscape, extremely detailed, 4k resolution, perfect lighting, **fine textures.**

Diffusion DPO

Conflict Signal
Composition: A
Semantic Alignment A
✗ Texture: A

Human Visual Preferences
Composition: A
Semantic Alignment A
Texture: B

Human Annotator

Label : A >B

Binary Label

Noisy Label
(Conflicting Supervision)



Diffusion Model

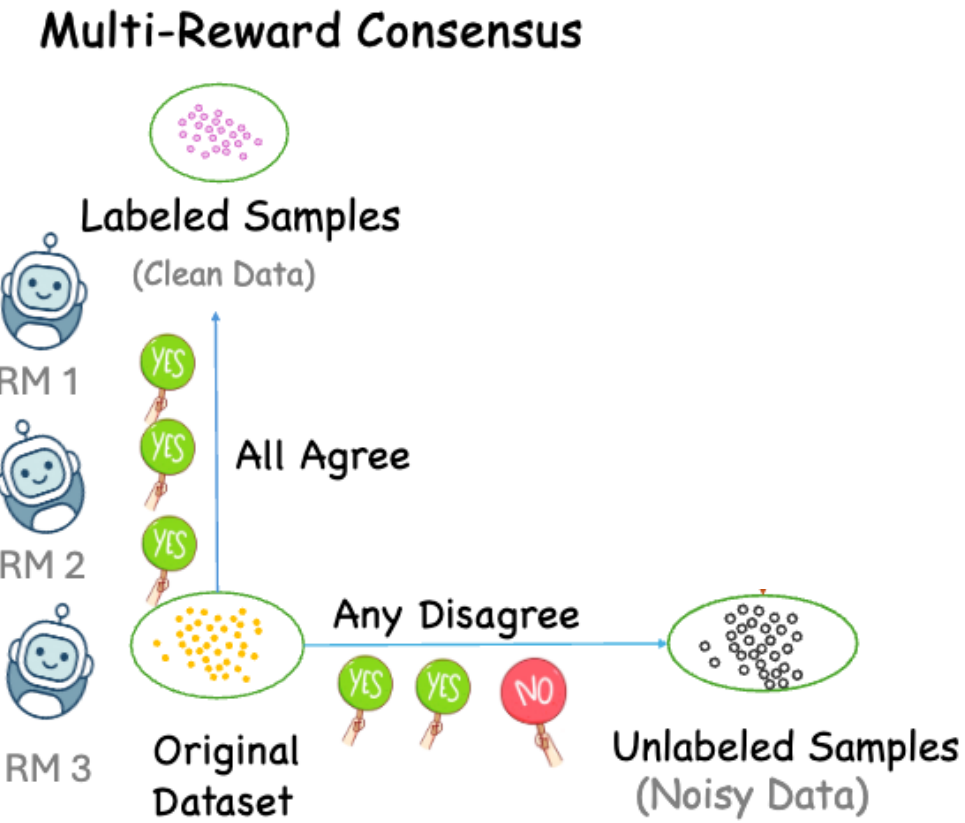
Collapsing multidimensional preferences into one label can create conflicting gradient signals.

Semi-DPO: Filter, Relabel, and Retrain

- Reframe learning with noisy labels as a semi-supervised learning (SSL) problem

- A two-stage framework
  - First stage: Multi-Reward Consensus
  - Second stage: Iterative self-training

Identify a Consensus-Labeled Subset

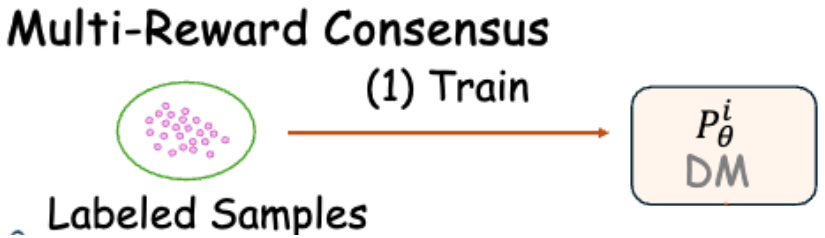


Start with labeled image pairs.
Keep pairs when all five reward models agree with the original label. Treat the rest as unlabeled.

Pick-a-Pic V2: about 21% of labels agree with all five proxy reward models.

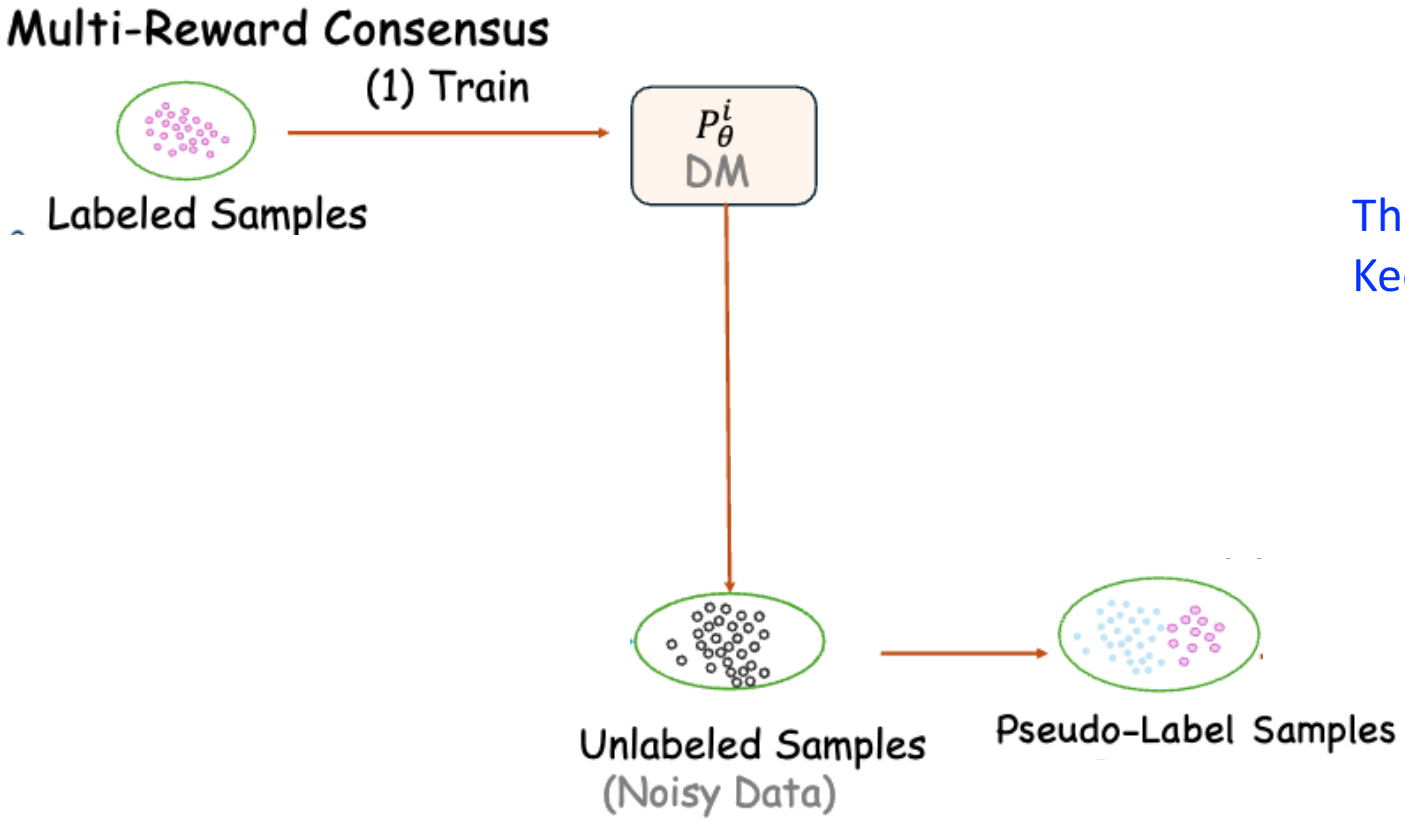
Separate consensus-labeled pairs from the remaining data

Train the Initial Preference Model



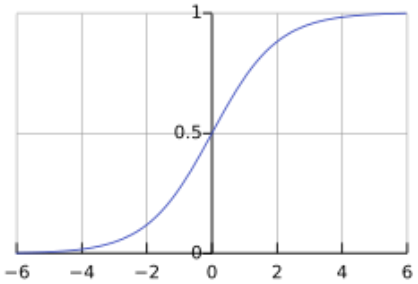
Train the initial diffusion model using clean labeled preference data

Generate Timestep-Conditional Pseudo-Labels



The model predicts pseudo-labels.
Keep only confident predictions.

Why Can a Diffusion Model Predict Pairwise Preferences?



Sigmoid function

Diffusion-DPO already learns a binary decision over image pairs.

c: shared prompt t: noise level σ: sigmoid

During DPO training

Known label $y = 1$ if $A \succ B$, $y = 0$ otherwise

$$p_t = P_\theta(A \succ B \mid c, t) = \sigma(z_t)$$

$$\mathcal{L}_{\text{BCE}} = -\left[y \log p_t + (1 - y) \log(1 - p_t)\right]$$

p_t model's sigmoid ranking score

z_t logit from current/reference denoising errors

BCE learns which image should rank higher.

During pseudo-labeling

The preference label is unknown

$z_t > 0$ **A wins at this timestep**

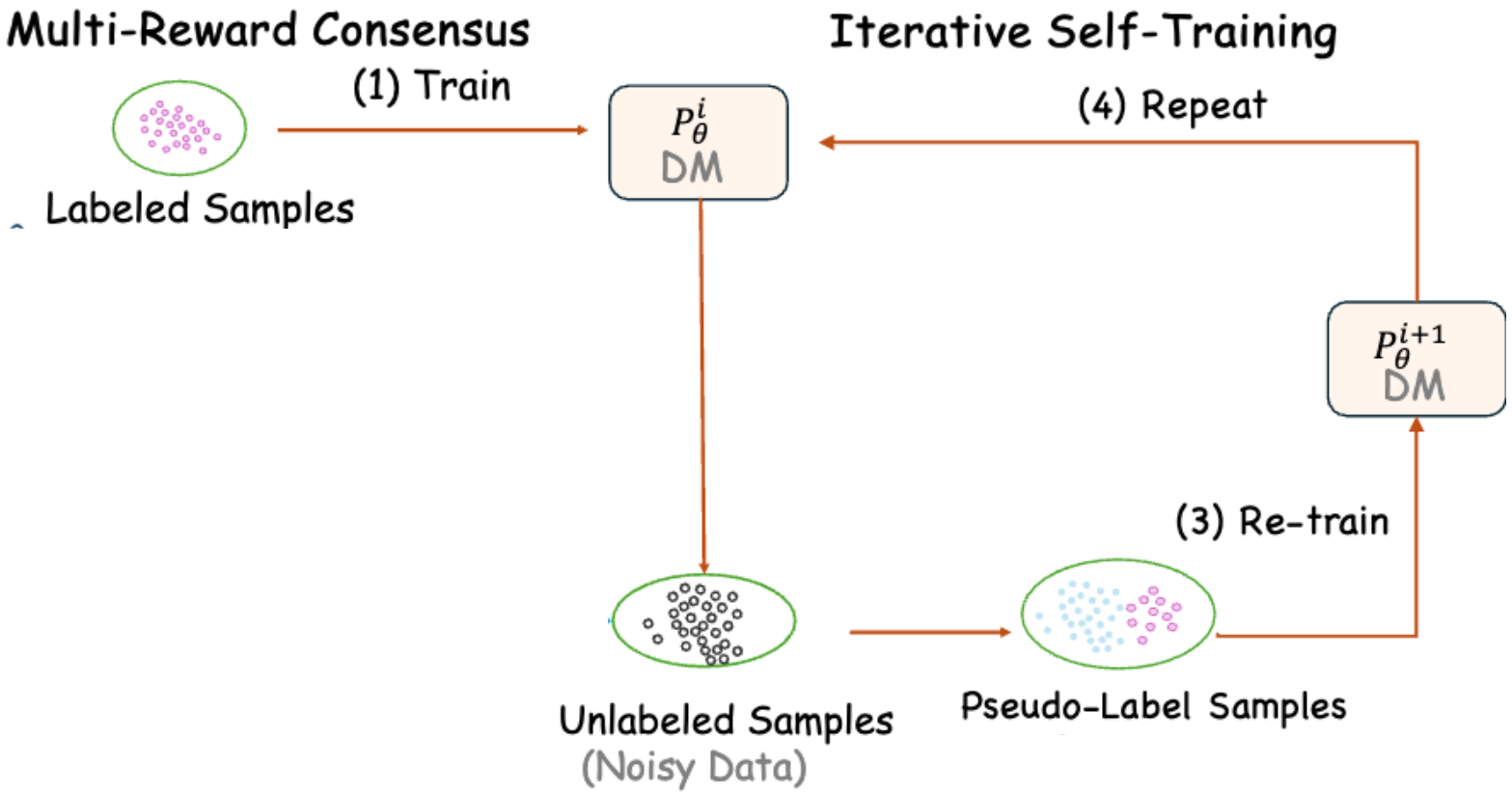
$z_t < 0$ **B wins at this timestep**

$|z_t|$ **supports confidence filtering**

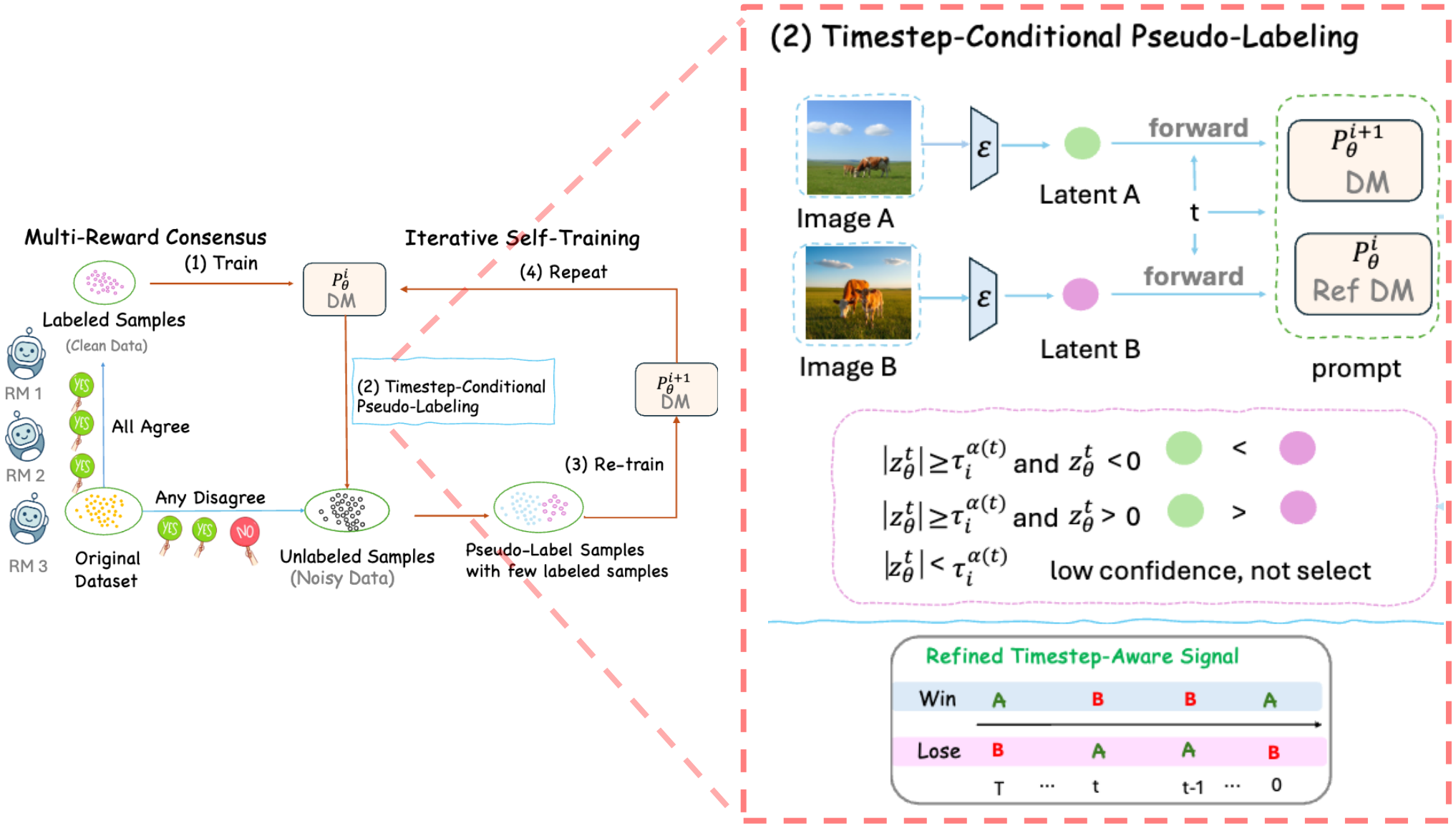
Keep only pairs above the timestep-specific confidence threshold.

No extra classification head: Semi-DPO reuses the DPO logit to predict and filter pseudo-labels.

Iteratively Refine Preference Learning



Confidence Filtering Across Diffusion Timesteps



Semi-DPO Improves Multi-Reward Alignment

Reward Score comparisons on HPS v2, Parti-Prompt and Pick-a-Pic V2 datasets.

Model	Dataset	Method	ImageReward	HPSv2.1	PickScore	Aesthetic	CLIP	MPS
SD1.5	HPS v2	SD1.5	0.139	0.246	20.862	5.578	0.293	12.211
		Diff-DPO	0.339 (+0.200)	0.259 (+5.3%)	21.308 (+2.1%)	5.714 (+2.4%)	0.297 (+1.4%)	12.739 (+4.3%)
		Diff-KTO	0.690 (+0.551)	0.284 (+15.4%)	21.454 (+2.8%)	5.803 (+4.0%)	0.298 (+1.7%)	13.016 (+6.6%)
		Semi-DPO	**0.816** (+0.677)	**0.287** (+16.7%)	**21.945** (+5.2%)	**5.899** (+5.8%)	**0.299** (+2.0%)	**13.514** (+10.7%)
	Parti Prompt	SD1.5	0.194	0.254	21.284	5.358	0.270	9.754
		Diff-DPO	0.352 (+0.158)	0.262 (+3.1%)	21.520 (+1.1%)	5.443 (+1.6%)	0.272 (+0.7%)	10.135 (+3.9%)
		Diff-KTO	0.615 (+0.421)	0.279 (+9.8%)	21.594 (+1.5%)	5.552 (+3.6%)	**0.277** (+2.6%)	10.202 (+4.6%)
		Semi-DPO	**0.798** (+0.604)	**0.284** (+11.8%)	**21.964** (+3.2%)	**5.706** (+6.5%)	0.276 (+2.2%)	**10.771** (+10.4%)
	Pick-a-Pic V2	SD1.5	0.085	0.250	20.566	5.421	0.273	9.635
		Diff-DPO	0.297 (+0.212)	0.261 (+4.4%)	20.948 (+1.9%)	5.549 (+2.4%)	0.279 (+2.2%)	10.144 (+5.3%)
		Diff-KTO	0.629 (+0.544)	0.281 (+12.4%)	21.064 (+2.4%)	5.659 (+4.4%)	**0.281** (+2.9%)	10.226 (+6.1%)
		Semi-DPO	**0.801** (+0.716)	**0.288** (+15.2%)	**21.524** (+4.7%)	**5.801** (+7.0%)	**0.281** (+2.9%)	**11.030** (+14.5%)

Qualitative Comparison



Qualitative Comparison



Technical Overview

PART I

Direct feedback for visual control

How can training directly enforce condition fidelity?

CHALLENGE

Condition fidelity

WORK

ControlNet++

ECCV 2024

PART II

Reliable and scalable preference learning

How can we construct, trust, and scale preference signals?

CHALLENGE

Cleaner comparisons

WORK

CPO

NeurIPS 2025

CHALLENGE

Conflicting preferences

WORK

Semi-DPO

ICLR 2026

CHALLENGE

Conflicting preferences Learning at scale

WORK

ViPO

ICLR 2026

ViPO: focused image and video preferences at scale.

ViPO

Reward Data Scaling: robust pairwise-preference visual preference learning

ViPO: Visual Preference Optimization at Scale | ICLR 2026

A Preference Needs a Clear Criterion



Image A



Image B

Prompt:

A vast green grassland, blue sky and white clouds, two cows are *grazing with their heads down*.



Human Annotator

Prompt alignment: A
Visual quality: B



Diffusion Model

Which image wins?

DPO Strongly Corrects Disagreement with the Label

p = model agreement with the labeled winner

DPO loss: $L = -\log p$

Logit-update magnitude: $1 - p$

Model disagrees with label

$p = 0.1$

Loss ≈ 2.30

Update magnitude = 0.90

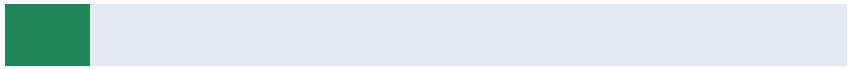


Model agrees with label

$p = 0.9$

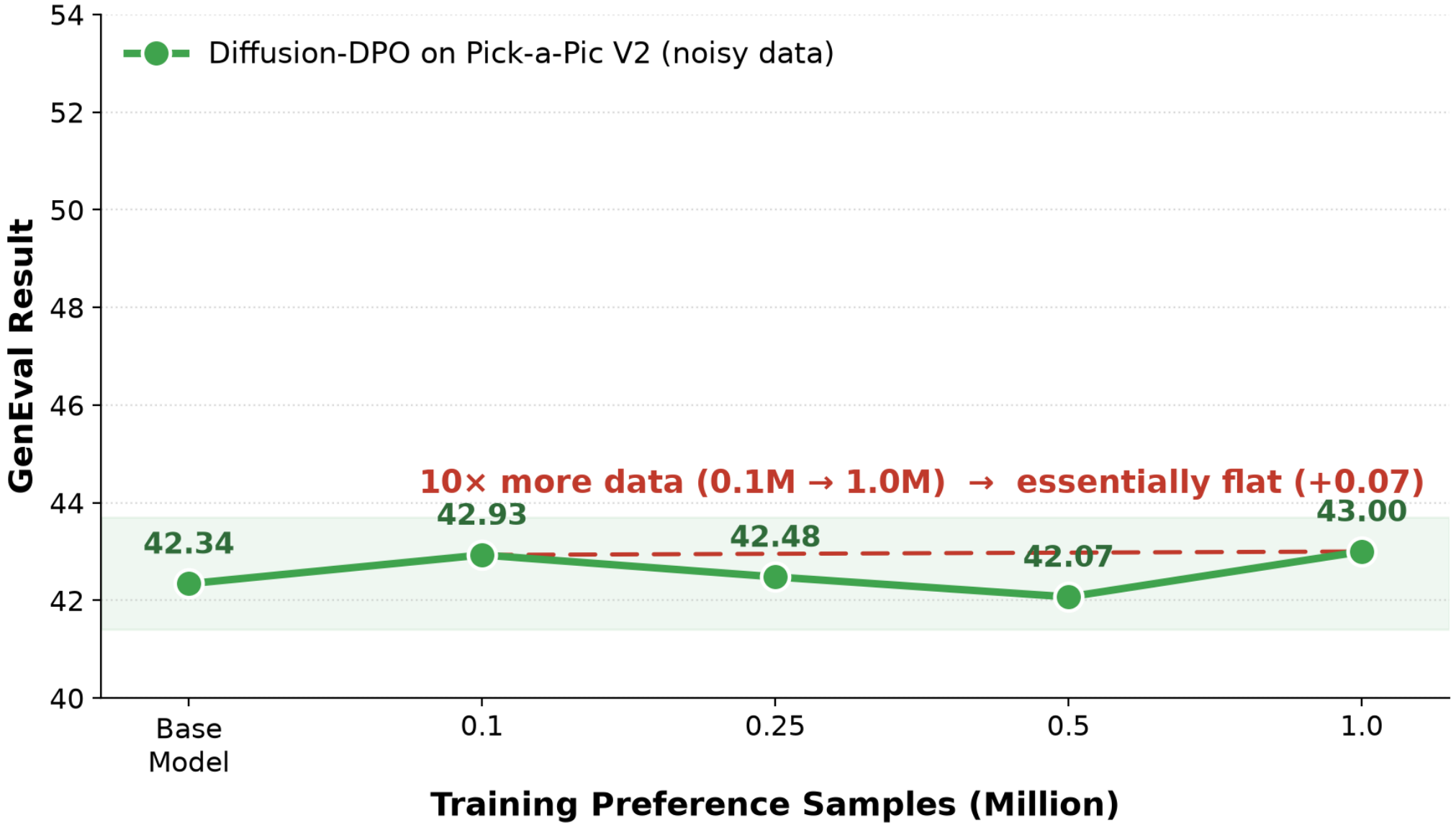
Loss ≈ 0.11

Update magnitude = 0.10



- **Low p can describe a valid, difficult pair.**
- **A conflicting label can also have low p.**

On Pick-a-Pic V2, More Data Barely Improves Standard DPO



Poly-DPO Adjusts How Preference Pairs Influence Learning

p = σ(z): agreement with labeled winner | α = tunable coefficient

Standard DPO $L = -\log p$

Taylor view $-\log p = (1-p) + \tfrac{1}{2}(1-p)^2 + \tfrac{1}{3}(1-p)^3 + \cdots$

x (1+ α)

Poly-DPO $L = -\log p + \alpha(1-p)$

α adjusts the first polynomial term; the full –log p remains.

Logit gradient: $\partial L/\partial z = -(1-p)(1+\alpha p)$

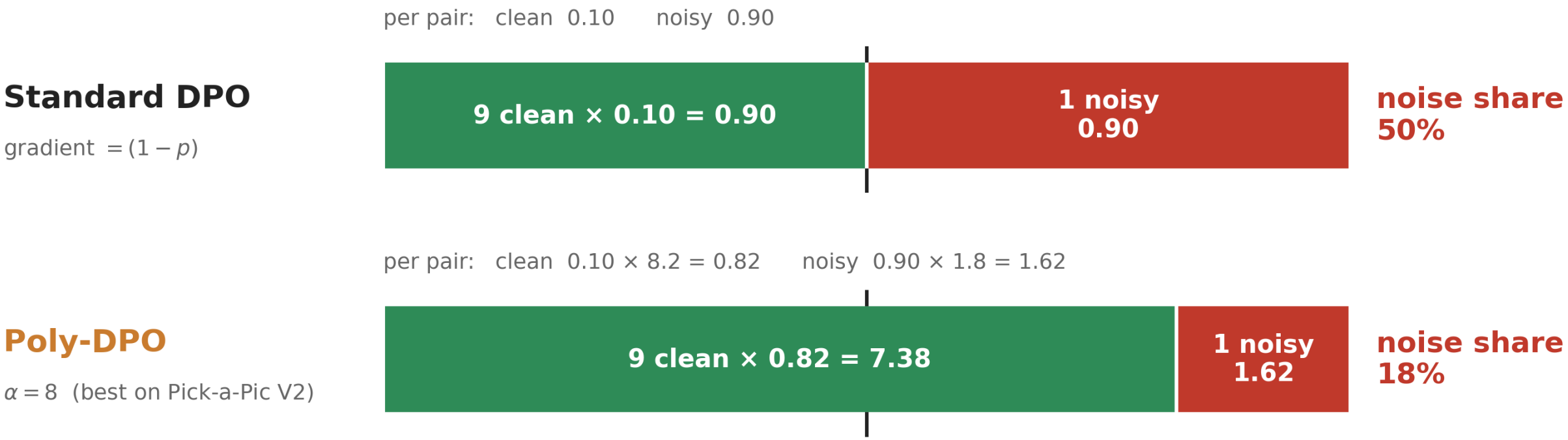
- **α > 0 gives a larger proportional boost to higher-p pairs.**
- **This changes pair influence.**

Poly-DPO Adjusts How Preference Pairs Influence Learning

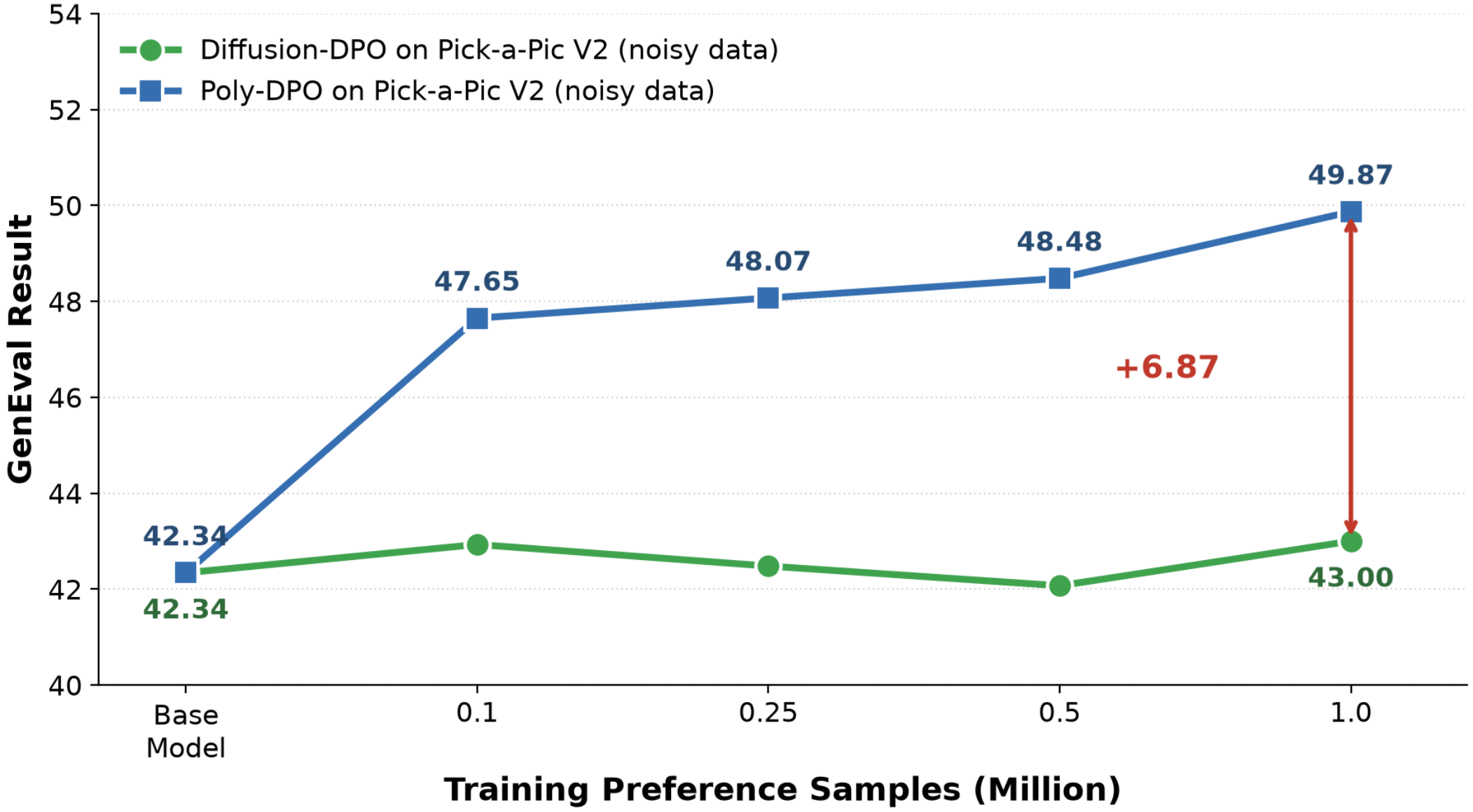
Poly-DPO $L = -\log p + \alpha(1-p) \Rightarrow \text{gradient} = (1-p) \times (1+\alpha p)$

Poly factor: grows with confidence p

A batch of 10 pairs: 9 clean ($p = 0.9$) + 1 noisy ($p = 0.1$)



Poly-DPO Scales on the Same Noisy Data



How ViPO Data Is Built: Best Generator + Best Judge for Each Dimension

Dimension	Pairs	How the pair is generated	Who labels the winner
ViPO-Image-1M (1024px, 5 × 200K pairs)			
Aesthetics	200K	Same prompt, two strong generators	3 VLMs vote
Composition	200K	Same prompt, two strong generators	3 VLMs vote
Text Rendering	200K	Only top text-rendering models	OCR checks spelling; VLM if tie
Human Quality	200K	11 open models vs. Seedream-3.0	VLM checks anatomy
Alignment	200K	Original prompt vs. slightly edited prompt	**No judge: original-prompt image wins**
ViPO-Video-300K (720p+, 3 × 100K pairs)			
Motion Quality	100K	Same first frame, different video models	VLM scores; drop too easy / ambiguous
Visual Quality	100K	Animate image pairs all 3 VLMs agreed on	Inherit the image label
Alignment	100K	Same image, two different motion prompts	**No judge: video matching its prompt wins**

Why high-quality: **pairs differ in ONE dimension** · **best judge per dimension** · **keep consensus, non-trivial pairs**

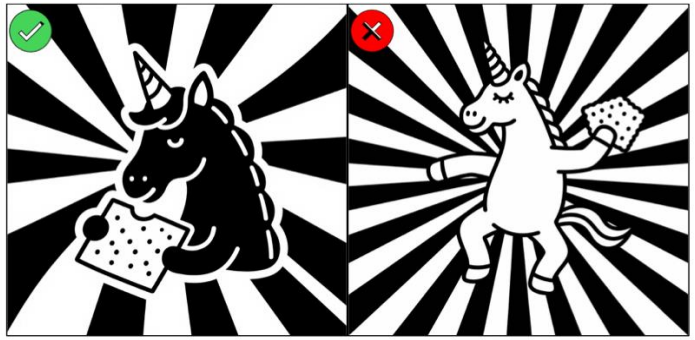
✓ chosen (winner) ✗ rejected (loser)



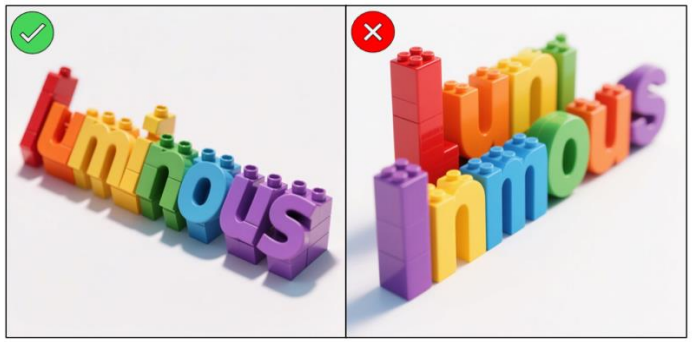
Aesthetics: *A wolf-headed figure donning heavy, armored attire wields a torch and sword amidst an ominously misty library environment.*



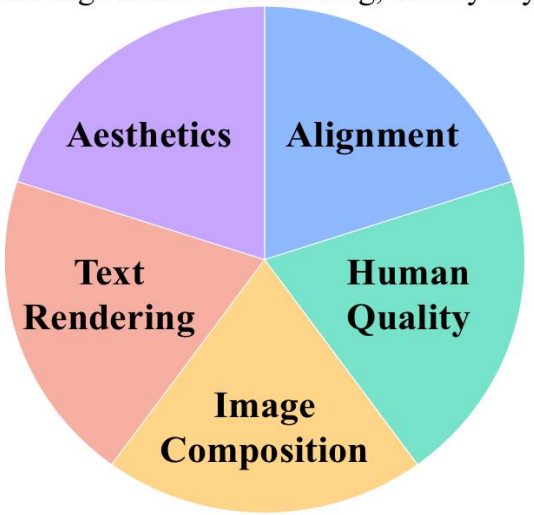
Aesthetics: Armored skeletal warriors and dogs under a threatening, cloudy sky.



Alignment: *A unicorn with black & white stripes holding a piece of mozzarella with its two front hooves, placing it in front of its mouth.*



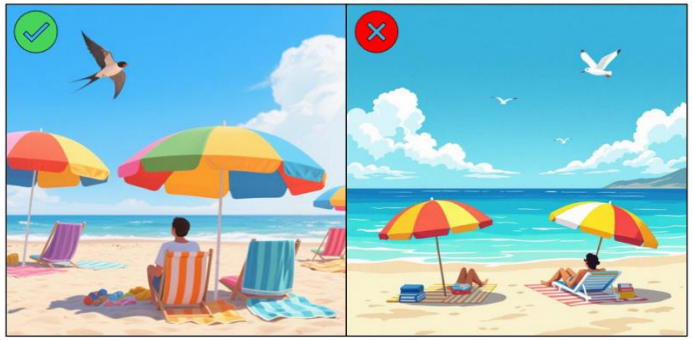
Text Rendering: *A vibrant display of the word 'luminous' spelled out in colorful Lego bricks.*



ViPO-Image-1M



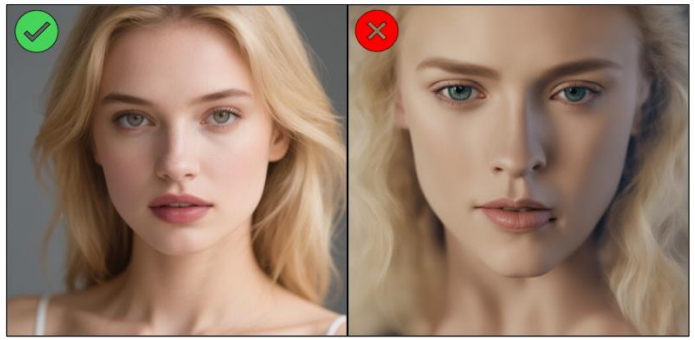
Human Quality: *A girl in purple animal-print pajamas sits cross-legged, touching her hair.*



Composition: *Peaceful beach scene with vibrant umbrellas, towels, a person under an umbrella.*



Composition: *A restroom with a green chair, mirror, sink and toilet.*



Human Quality: *A close-up portrait of a blonde woman, focusing on her expression and attire.*

✓ chosen (winner) ✗ rejected (loser)



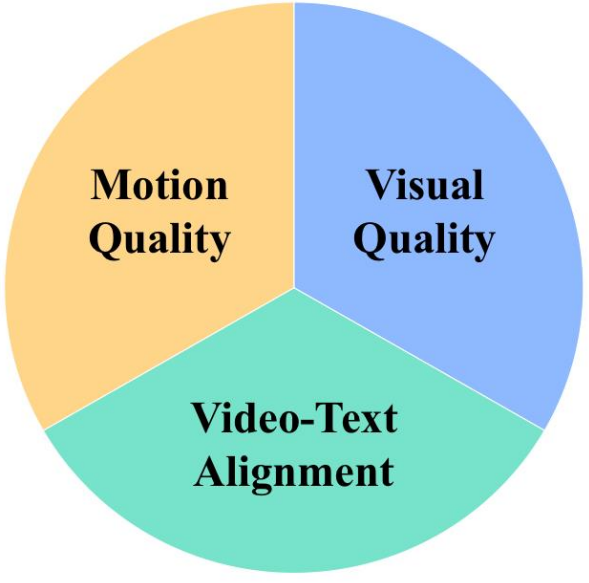


Motion Quality: *Shrek is showing his muscle..*



Spatial Alignment: *A couple walks hand in hand through a meadow. The man is wearing a dark blue suit and the woman is wearing a white wedding dress.*





ViPO-Video-300K



Visual Quality: *A heavily armed warrior holding a futuristic gun cautiously enters the cyberpunk arena from a smoky industrial area with a skull in the background.*

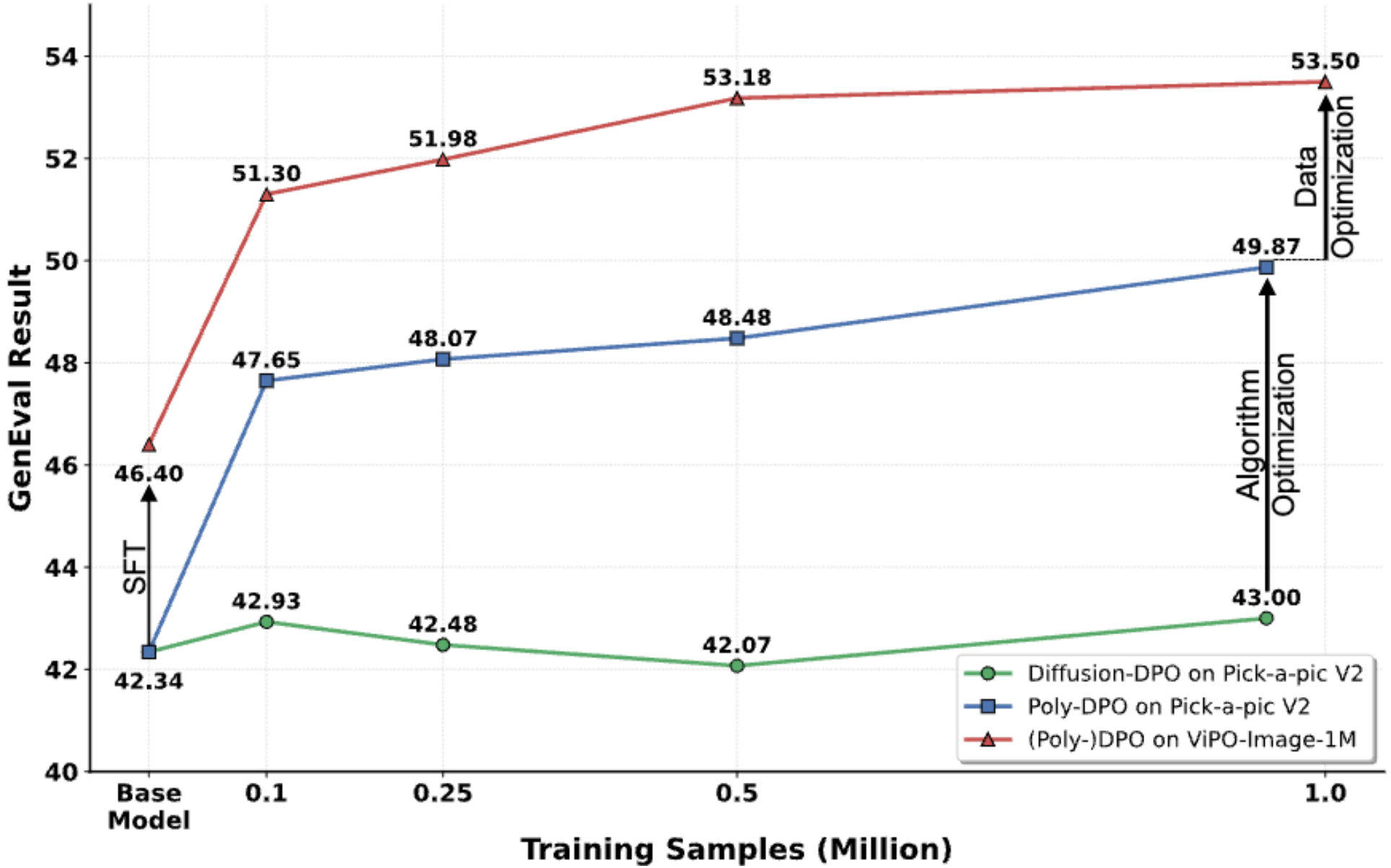




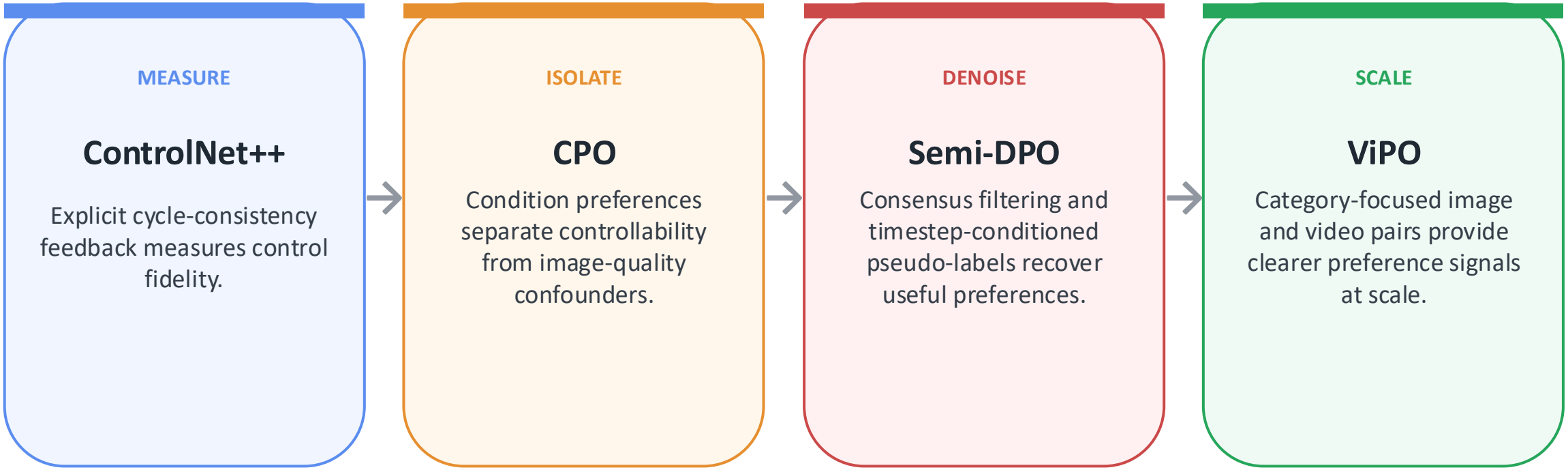


Motion Alignment: *People desperately run away from the giant spider chasing them.*

Poly-DPO Scales on Our ViPO Data



Summary: Better Signals, Better Visual Generation



Signal quality, structure, and scale shape visual generation.

Future Directions: Benefits and Risks

+ OPPORTUNITIES

Robotics

Generate rare, dangerous, and long-tail training scenarios

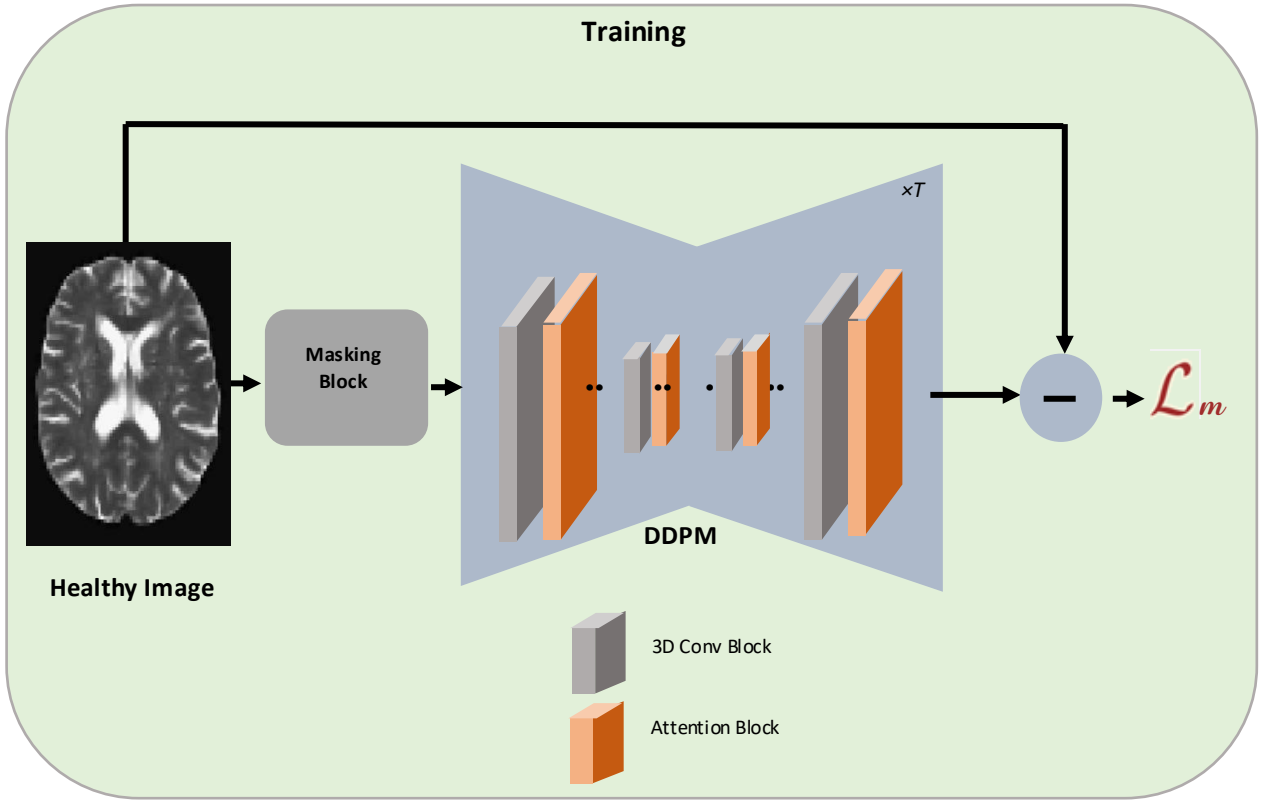
Healthcare

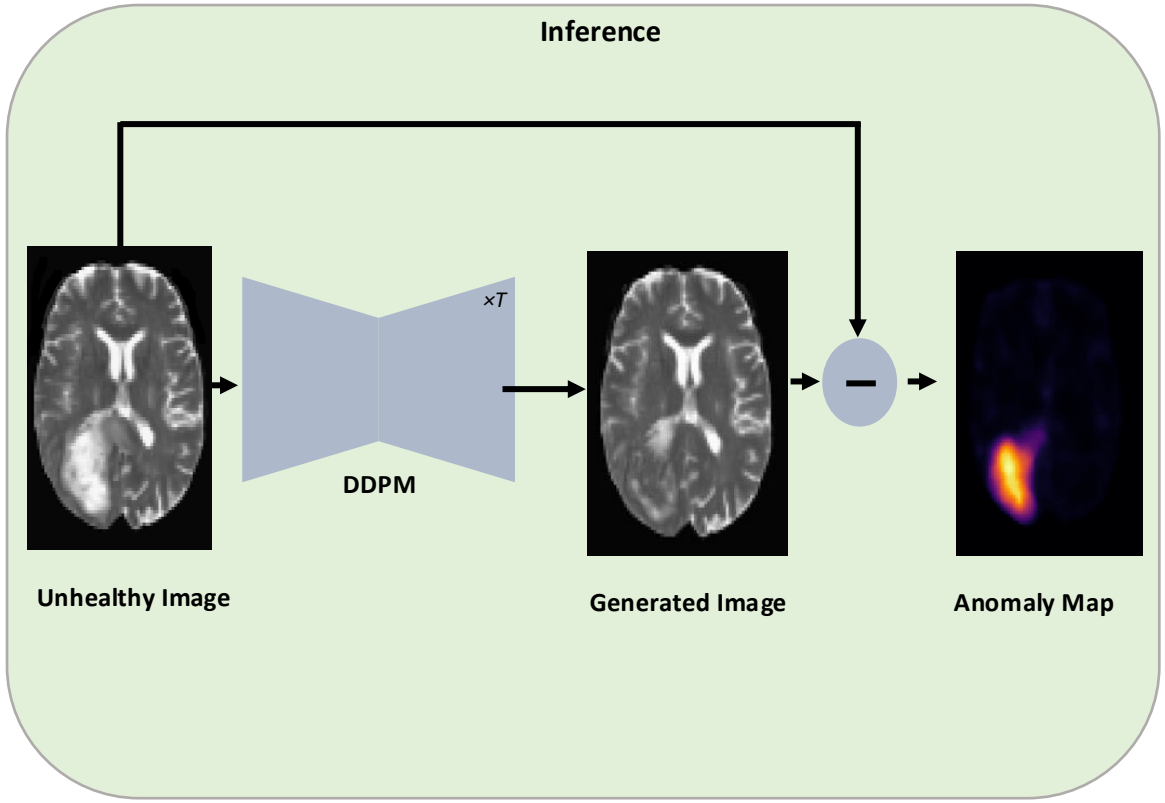
Augment scarce imaging data, with clinical validation

Science

Create controlled data with known factors

Diffusion Model for Unsupervised Medical Image Anomaly Detection





Iqbal, Hasan, Umar Khalid, Chen Chen, and Jing Hua. "Unsupervised anomaly detection in medical images using masked diffusion model." Machine Learning in Medical Imaging, pp. 372-381. Cham: Springer Nature Switzerland, 2023.

SciVideoBench



Prepare a 0.1 mg/ml stock solution of the standard, bovine serum albumin. Any other standard may be chosen, but note that the same standard must be used in all experiments



Dilute the unknown samples in deionized water. Aim to 5-50 µg/ml. Yet, higher or lower protein concentrations are acceptable, since there is no apparent limit for the linear range of the assay. However, the measurement must be within the linear range of the absorbance reader.

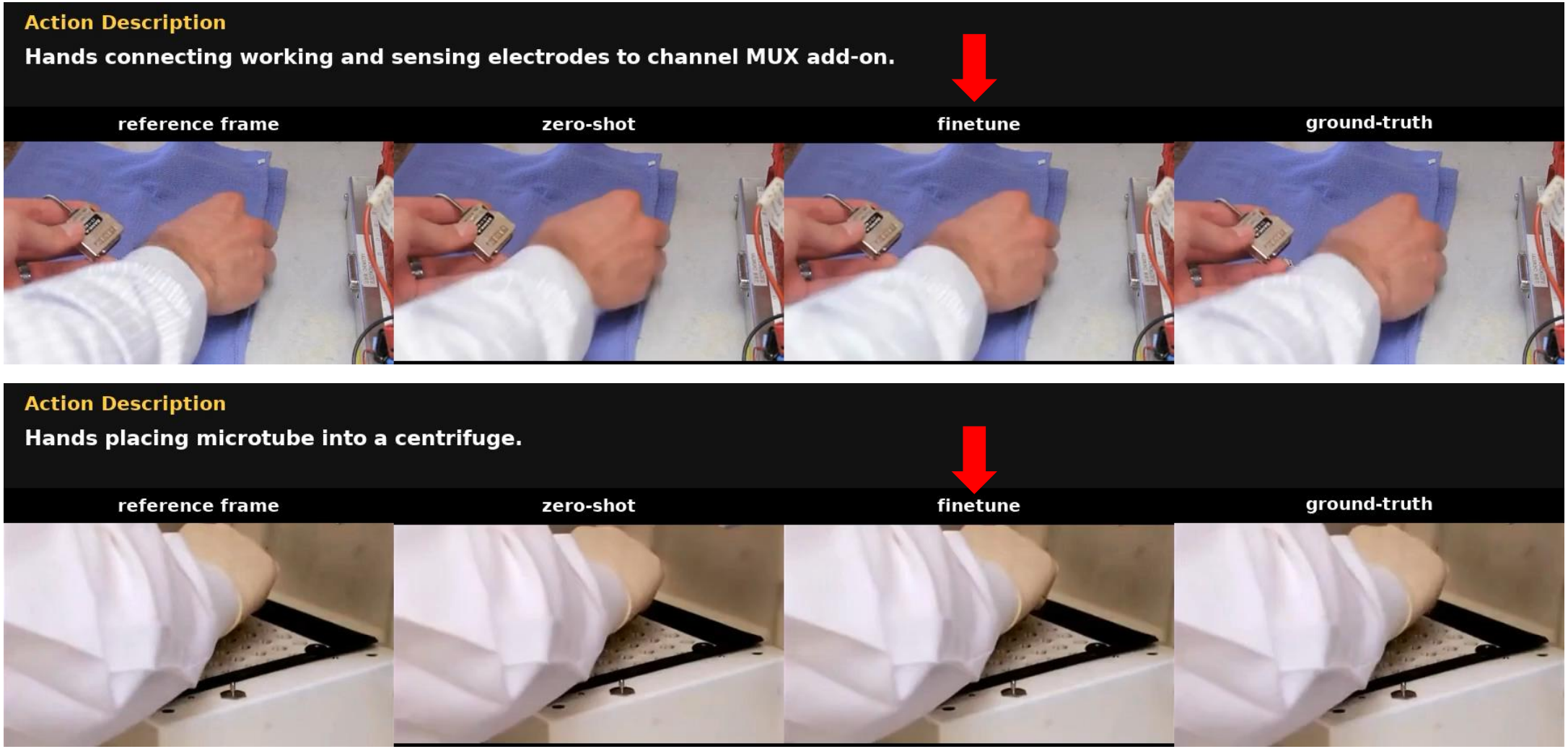


Dilute the Bradford reagent (Bio-Rad) 2.5-fold in deionized water.

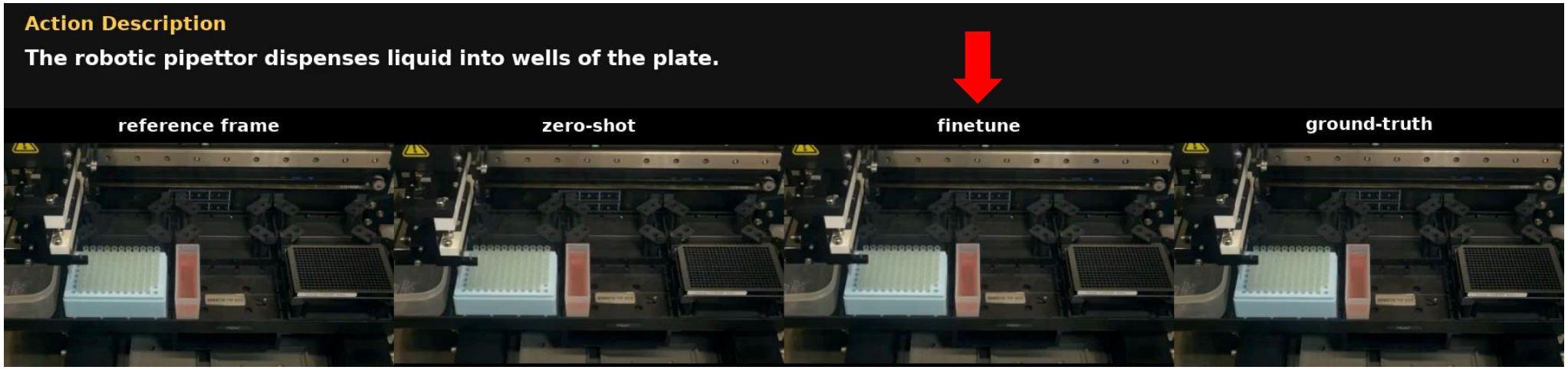
… …

Deng, Andong, Taojiannan Yang, Shoubin Yu, Lincoln Spencer, Mohit Bansal, Chen Chen, Serena Yeung-Levy, and Xiaohan Wang. “Scivideobench: Benchmarking scientific video reasoning in large multimodal models.” ICML 2026

Action Conditioned I2V



Action Conditioned I2V





Future Directions: Benefits and Risks

-

RISKS

Deepfakes

Enable misinformation and impersonation

Trust

Make authentic and generated media harder to distinguish

Ethics

Raise concerns about consent, bias, privacy, and misuse

References

- Li, Ming, et al. “ControlNet++: Improving conditional controls with efficient consistency feedback” ECCV, 2024
  - Code: https://liming-ai.github.io/ControlNet_Plus_Plus/
- Lyu, Zonglin, et al. “CPO: Condition Preference Optimization for Controllable Image Generation.” NeurIPS, 2025
  - Code: https://zonglinl.github.io/CPO_page/
- Liu, Xinxin, et al. “Learning from Noisy Preferences: A Semi-Supervised Learning Approach to Direct Preference Optimization.” ICLR, 2026
  - Code: https://github.com/L-CodingSpace/semi-dpo
- Li, Ming, et al. "ViPO: Visual Preference Optimization at Scale." ICLR, 2026
  - Project: https://liming-ai.github.io/ViPO/

Acknowledgements

ControlNet++

ECCV 2024



Ming Li

CPO

NeurIPS 2025



Zonglin Lyu

Semi-DPO

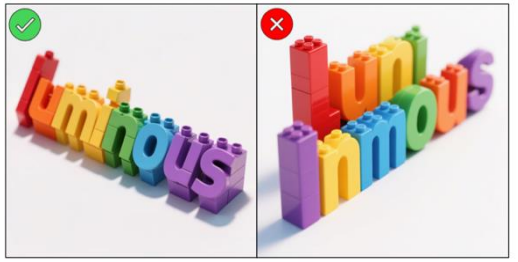
ICLR 2026



Xinxin Liu

ViPO

ICLR 2026



Ming Li

Thank you to all co-authors and collaborators.
Thanks for the support from USDA-NIFA (CPO work).

Thank you!

