UC San Diego
JACOBS SCHOOL OF ENGINEERING

Qualcomm

NeurIPS

Diffusion Meta-Prompting and Steering for Generalizable Foundation Model Adaptation

Deepak Sridhar[1], Yi Li[2], Kartikeya Bhardwaj[2], Shuangjun Liu[2], Taotao Jing[2]
Yuan Li[2], Shuai Zhang[2], Jiancheng Lyu[2], Dashan Gao[2], Nuno Vasconcelos[1]

[1]University of California, San Diego [2]Qualcomm AI Research

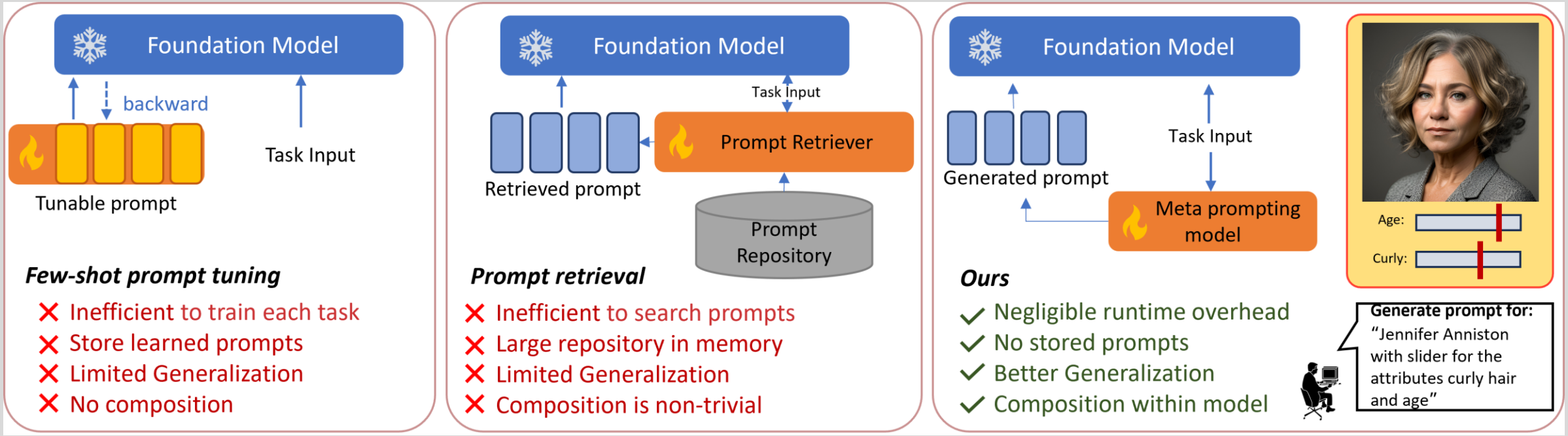
NeurIPS 2026

Code

SVCL UCSD

Prompt Learning: Limitations

- Prompt learning cheaply adapts foundation models (CLIP, Stable Diffusion), but prompts are **task-specific**
  - Re-tuned for every label set, concept or attribute; generalize poorly to new classes and compositions
- Prompt repositories (HuggingFace, Civitai) must be **stored, searched and manually composed**



Key Idea: Learn a Distribution over Prompts

- Treat prompts as **samples from a structured distribution**, not isolated optimization outcomes
- Learned prompts admit smooth **interpolation, composition and negation**
- Learn from **existing prompt repositories**: no double-loop meta-learning over task data

Can we generate prompts for any task directly from natural language?

1 **No task data or task losses**

Trained only on a repository of previously learned prompts

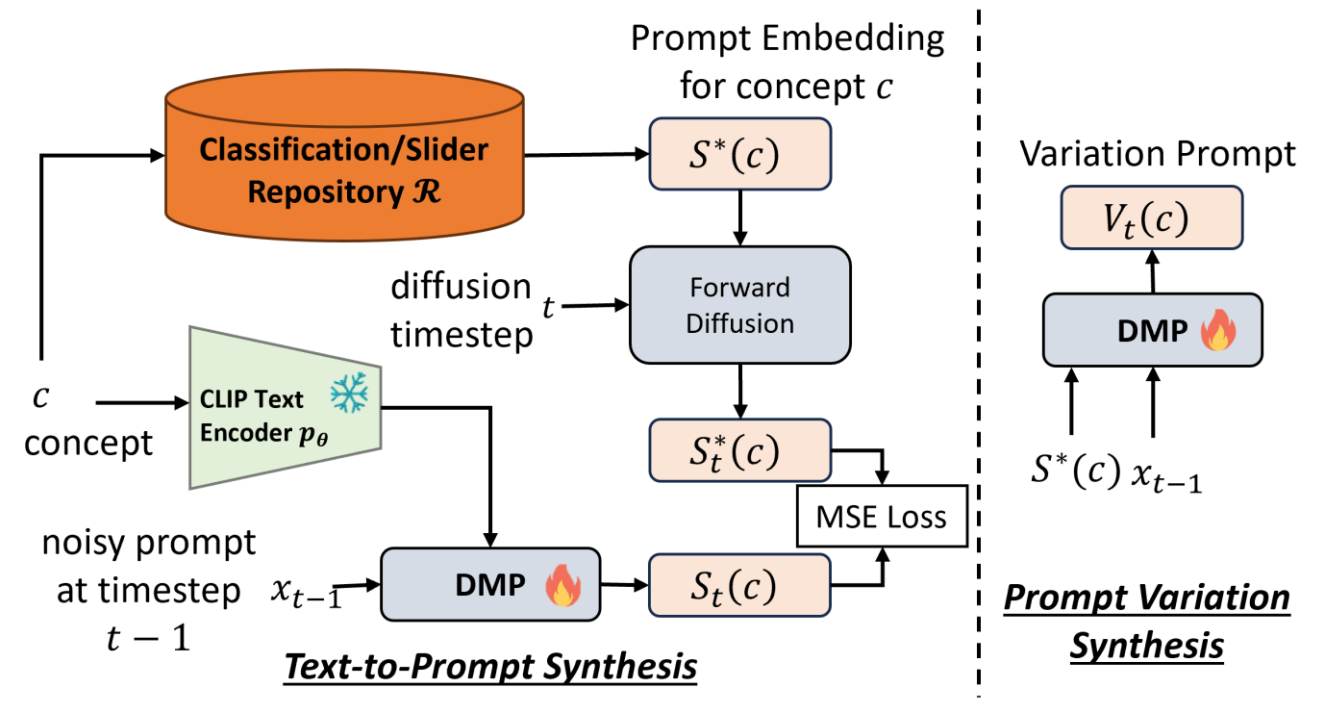
2 **One model, many tasks**

Classification, personalization, sliders, and retrieval embeddings

3 **Composition for free**

Concept composition and negative prompting without explicit training

Diffusion Meta-Prompting (DMP)



- Repository of learned prompts x(c) with task texts y(c)
- **1D Stable-Diffusion U-Net** over prompt embeddings, conditioned on CLIP text
- Standard denoising loss, classifier-free guidance
- Small 1D VAE (< 1M params) for long prompts
- **One-shot synthesis**: 50 DDIM steps ≈ 1 s

$$\mathcal{L}_{denoise}(\theta) = \mathbb{E}_{t,\, x_0,\, y,\, \epsilon_t} \left\| \epsilon_t - \epsilon_\theta(x_t, \tau(y), t) \right\|^2$$

Guarantee and Composition

Theoretical guarantee

If the denoising loss ≤ ε on n repository prompts, then with probability ≥ 1 − δ:

$$\mathbb{E}_{p_\theta(x|y)}[\mathcal{L}_{task}] - \mathbb{E}_{\hat{p}_{data}(x|y)}[\mathcal{L}_{task}] \leq L_{max}\sqrt{2C\epsilon} + L_{max}\sqrt{\frac{\log(2/\delta)}{2n}} + o(1)$$

→ DMP matches prompt learning for a large repository and a small denoising error

Concept composition in prompt space

$$\hat{\epsilon}(x_t, t) = \epsilon_\theta(x_t, t) + \sum_{i=1}^{n} w_i \left(\epsilon_\theta(x_t, \tau(y^{(c_i)}), t) - \epsilon_\theta(x_t, t)\right)$$

- Standard composition runs the downstream diffusion model **n times per step**
- DMP composes inside the small prompt model: **one prompt** for all concepts; negative weights negate

Test-Time Steering

Different denoising trajectories give prompts with different generalization. **Steer DDIM toward an anchor prompt:**

$$\tilde{x}_0 = (1-\lambda)\,\hat{x}_0(x_t, y^{(i)}, t) + \lambda\, x_i^*$$

$$x_i^* = \arg\max_m \; \mathrm{Acc}_{train}(x_{i,m})$$

- Anchor chosen once from **training accuracy**: no test classes, no retraining
- Composite tasks: average of the K dataset anchors; λ = 0.2
- Robust to anchor perturbations; a random or resampled anchor hurts

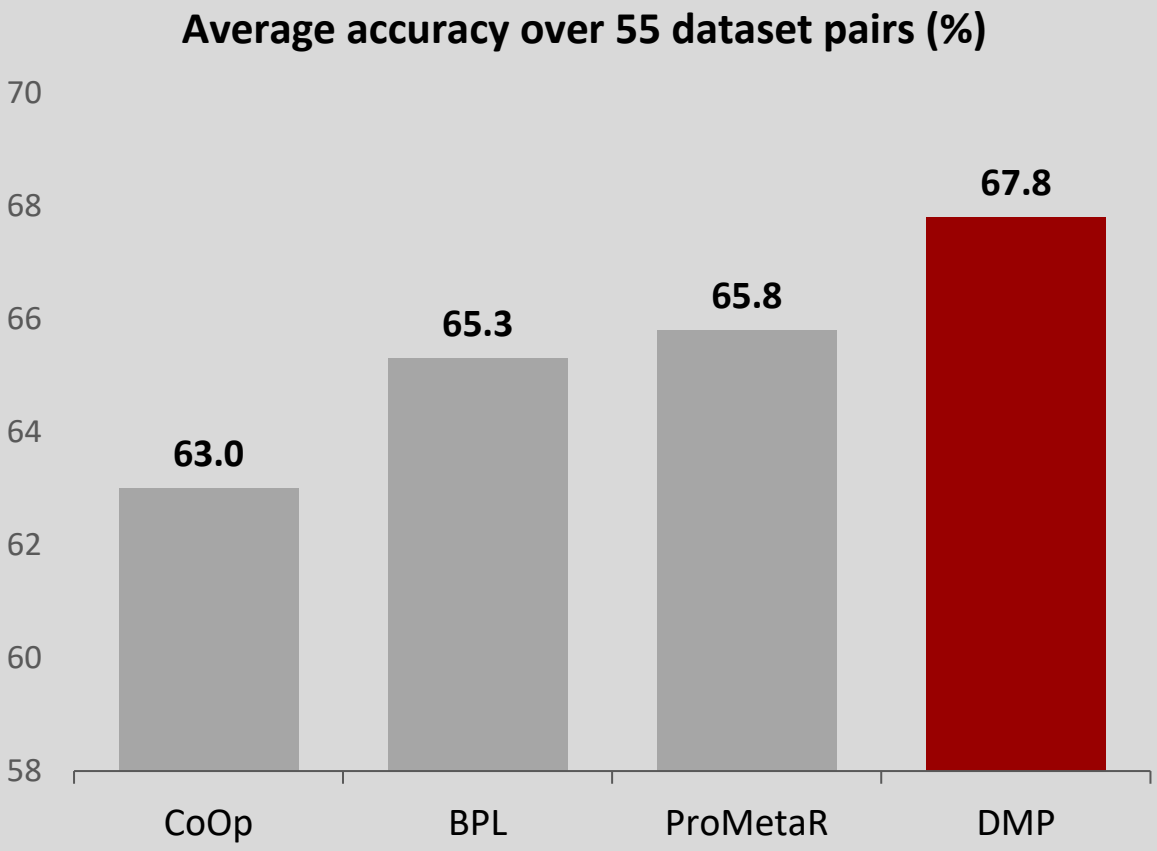
Composite classification, 55 dataset pairs

	All	Base	New	Avg.
w/o steering	61.4	67.3	68.6	65.8
w/ steering	**63.1**	**69.4**	**70.8**	**67.8**
Gain	+1.7	+2.1	+2.2	+2.0

Anchor ablation (All split)

Anchor	Acc.	Δ
Training-selected	**63.10**	**–**
Perturbed (L2 = 1.0)	63.09	−0.01
Fixed random prompt	62.36	−0.74
Resampled every step	61.27	−1.83

Results: Composite Classification

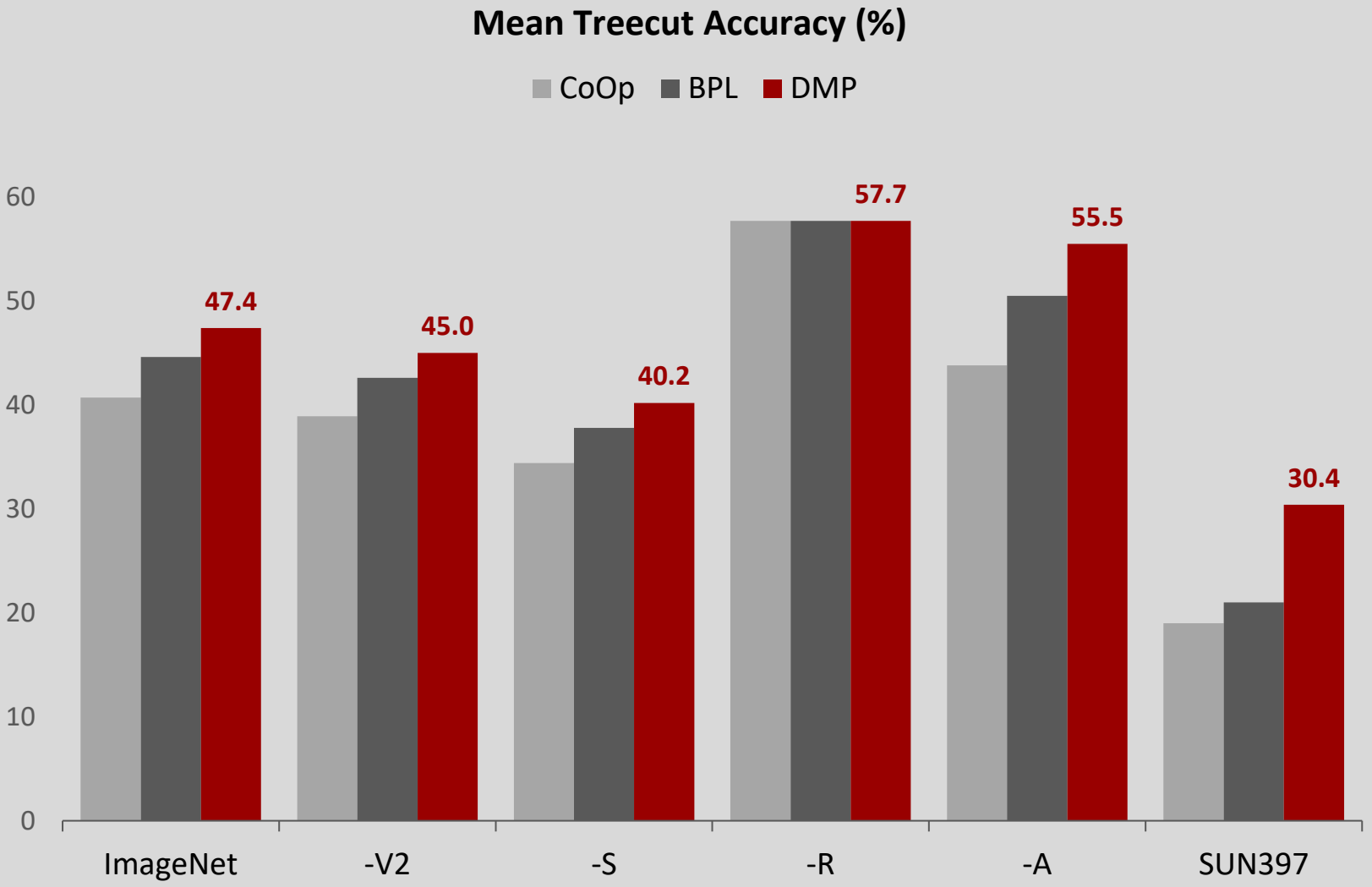


EuroSAT + Flowers

Method	All	Base	New
CoOp (zero-shot)	35.5	50.2	59.4
BPL (zero-shot)	52.6	61.3	73.4
ProMetaR (zero-shot)	48.0	59.9	67.1
BPL (trained on pair)	56.2	76.1	75.1
DMP (zero-shot)	**61.1**	**74.5**	**77.8**

- Classify over the union of class sets from two datasets: **out-of-domain composition** never seen in training
- **+2.0%** over the best meta-prompting baseline; up to **+8.5%** on a single pair

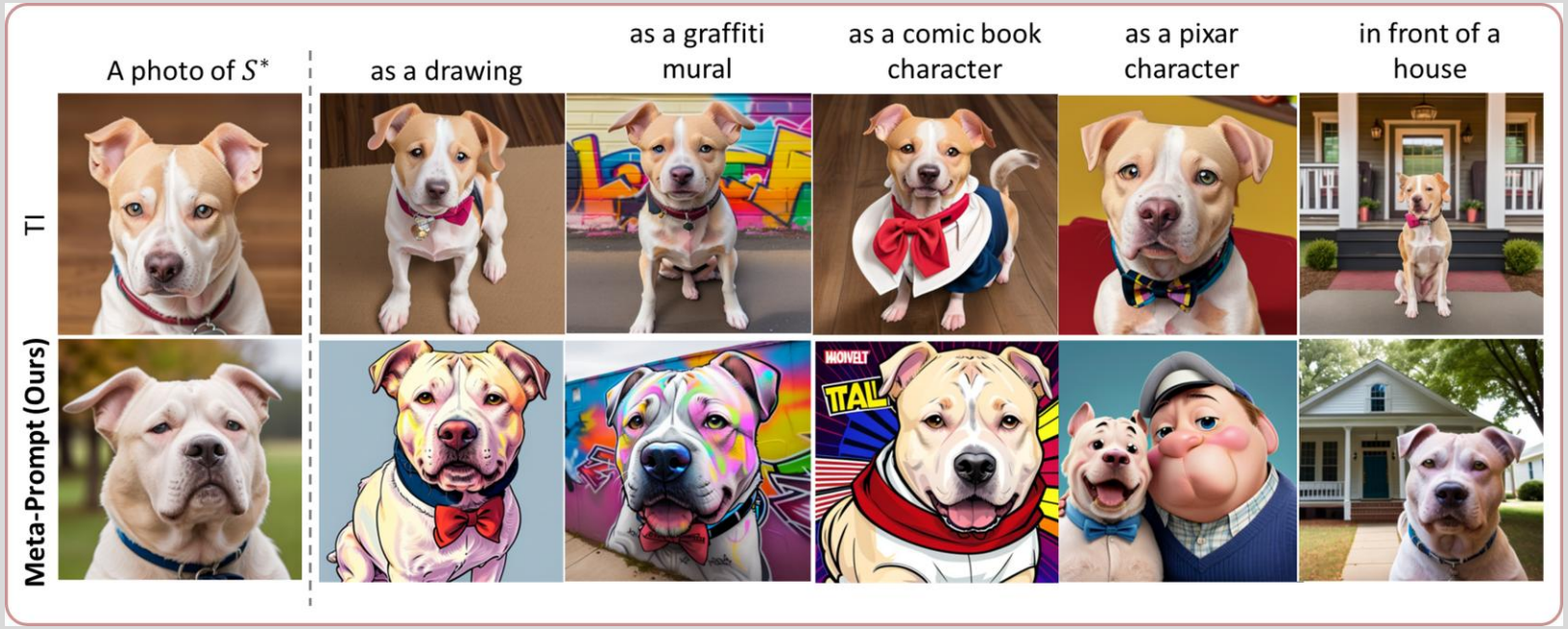
Results: Taxonomic Classification



- Classify at **any level** of the class hierarchy (25 treecuts)
- DMP never trained for hierarchical classification
- **+2–9%** over BPL, up to **+11.7%** over CoOp

Results: Generalizable Personalization

DMP-Variation, trained on Textual Inversion (TI) prompts of **3,000 CelebA identities**, follows new contexts and **generalizes to dogs**.



Metric	TI	DMP
Face-ID ↓	0.428	0.299
CLIP-T ↑	27.0	28.4
CSD ↑	50.4	61.8
HPSv2 ↑	26.6	27.3

Lower Face-ID: more diverse but faithful variations.

- TI prompts **overfit the subject** and ignore the context
- DMP prompts follow every instruction, though trained **only on faces**

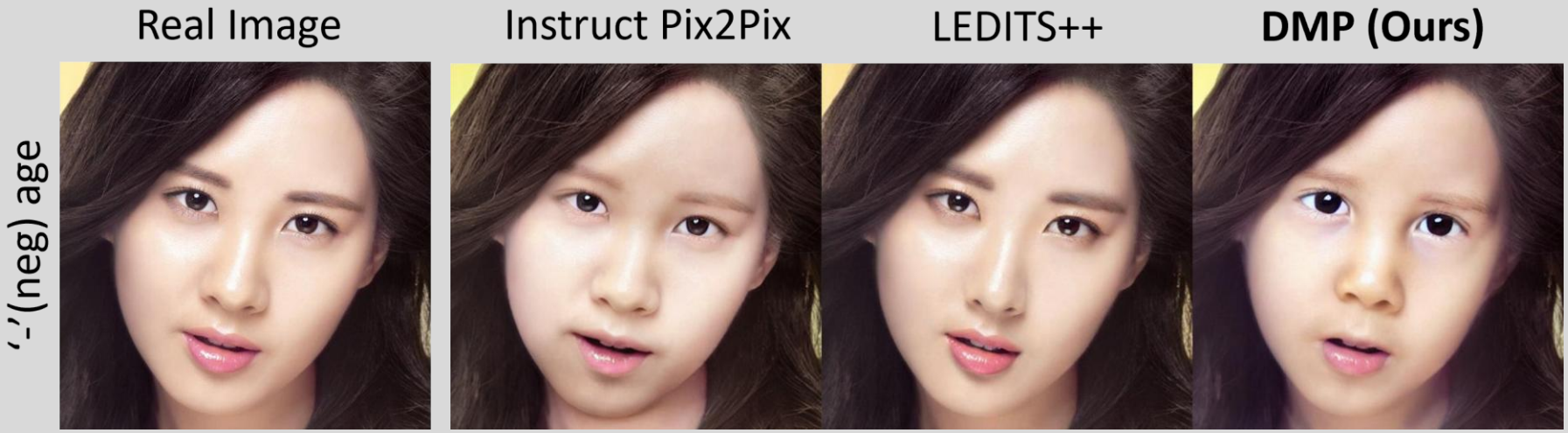
Composition and Negative Prompting

Identity composition: one DMP prompt, increasing composition scale



- Composition and negation **without explicit training**
- No extra passes through Stable Diffusion

Negative prompting with DMP



Major Takeaways

Compositional Scalability

1. Learns to combine multiple types of prompts
2. Does not require fine-tuning per composition

Efficiency in Storage

1. Reduces the need for large prompt repositories and prompt management by applications
2. Removes overhead of search and retrieval of task specific prompts

Generalization & Adaptability

1. Capable of generating prompts across diverse tasks and models
2. Enhanced model performance with steering on fine-grained or personalized tasks

Universal & Multi-task ability

1. Supports many parameter efficient training methods like Prompts or LoRAs
2. Prompts for different tasks (personalization, concept slider, etc.) can be synthesized by the same model

Summary

- **DMP** learns a distribution over prompts from existing repositories, with no task data or losses
- **Test-time steering** toward a training-selected anchor: up to +2.0% without retraining
- **Guarantee** bounding the task loss of DMP-sampled prompts
- Better generalization in **composite (+2.0%)** and **taxonomic (+2–9%)** classification
- Personalization, composition and negation; **> 90%** lower storage and inference cost

Thank You

desridha@ucsd.edu

github.com/DeepakSridhar/dmp