Stable Preference: Redefining Training Paradigm of Human Preference Model for Text-to-image Synthesis

Hanting Li, Hongjing Niu and Feng Zhao*

University of Science and Technology of China, Hefei 230027, China

{ab828658, sasori}@mail.ustc.edu.cn, {fzhao956}@ustc.edu.cn

The 18th European Conference on Computer Vision ECCV 2024

Sun Sep 29th - Fri Oct 4th, 2024

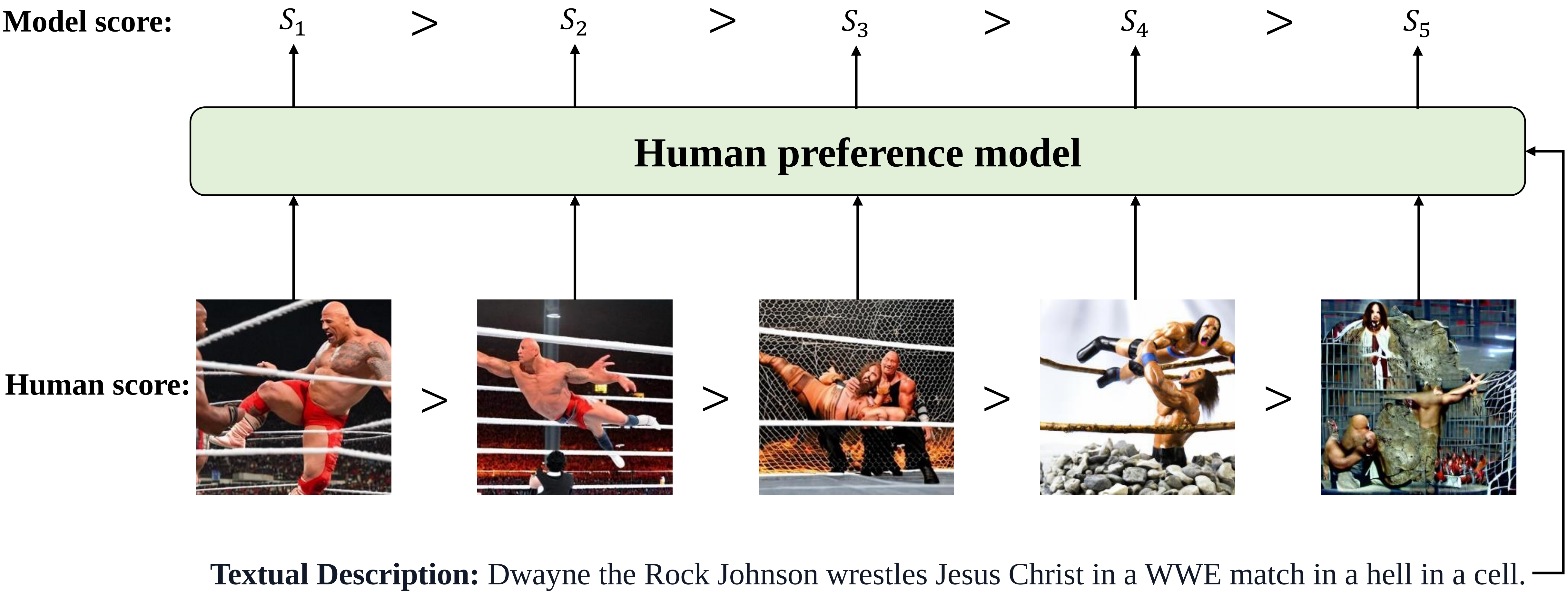
MiCo Milano, Milan, Italy

Outline

- Introduction
- Methodologies
- Experimental results
- Conclusion

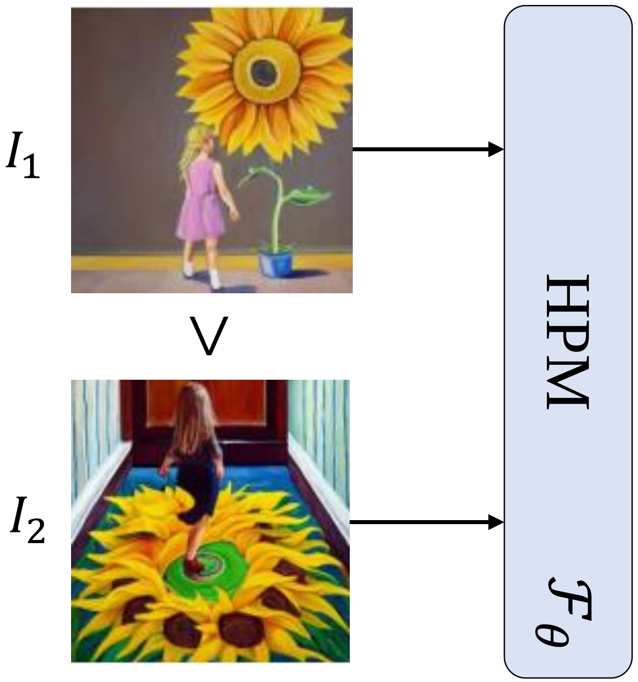
Introduction

- **Human preference models (HPMs) for text-to-image synthesis**



Introduction

● Current training paradigm of HPMs



$$S_1 = \mathcal{F}_\theta(I_1, T)$$

$$S_2 = \mathcal{F}_\theta(I_2, T)$$

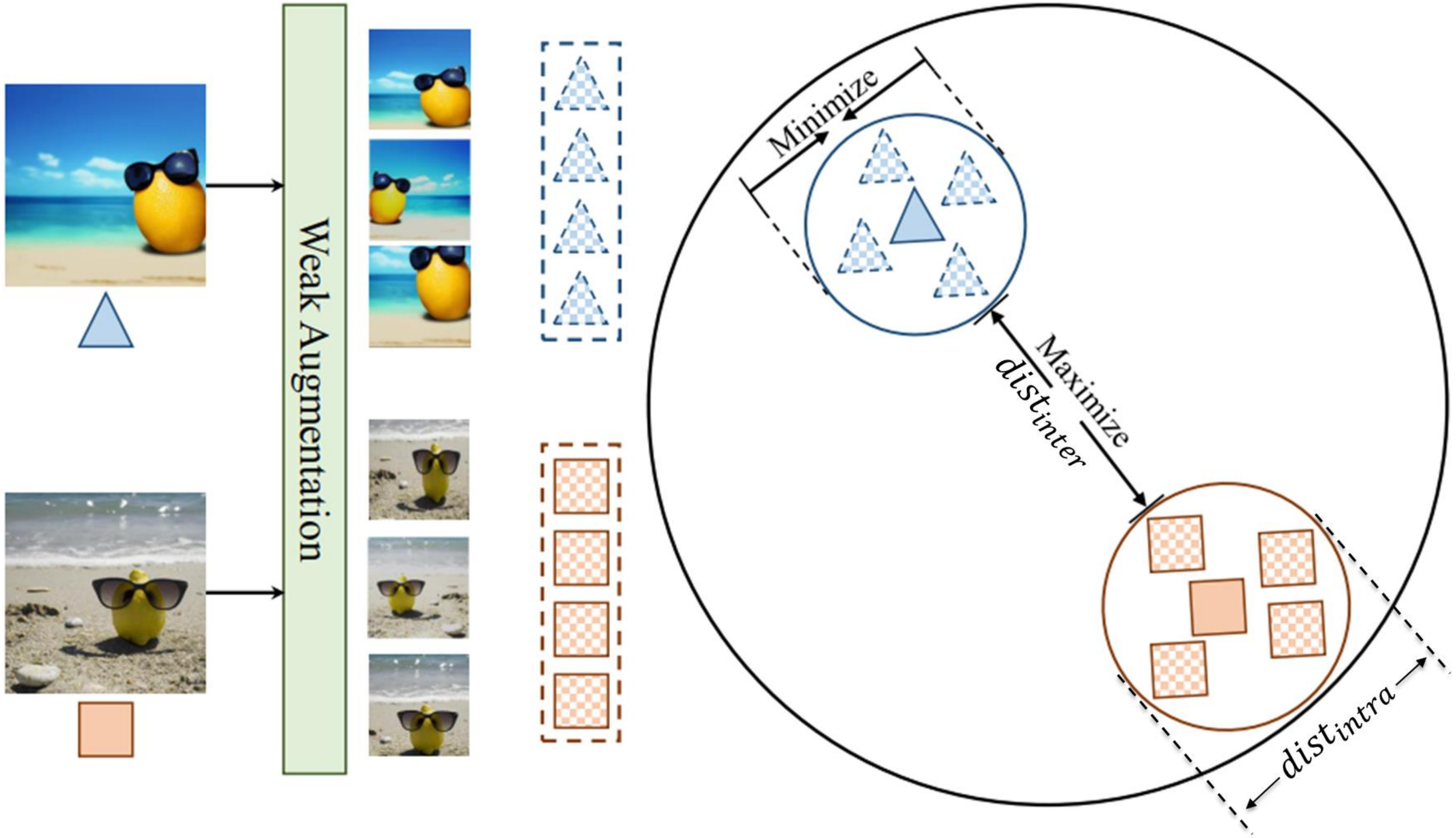
Training Loss: $$\mathcal{L}_{pref} = \sum_{i=1}^{2} y_i \log \hat{y}_i,$$

$$\hat{y}_i = \frac{\exp(\mathcal{F}_\theta(I_i, T))}{\sum_{j=1}^{2} \exp(\mathcal{F}_\theta(I_j, T))}$$

1. Current HPMs displays sensitivity towards small visual perturbations
2. The image selection process of human is not strictly dichotomous

Methodologies

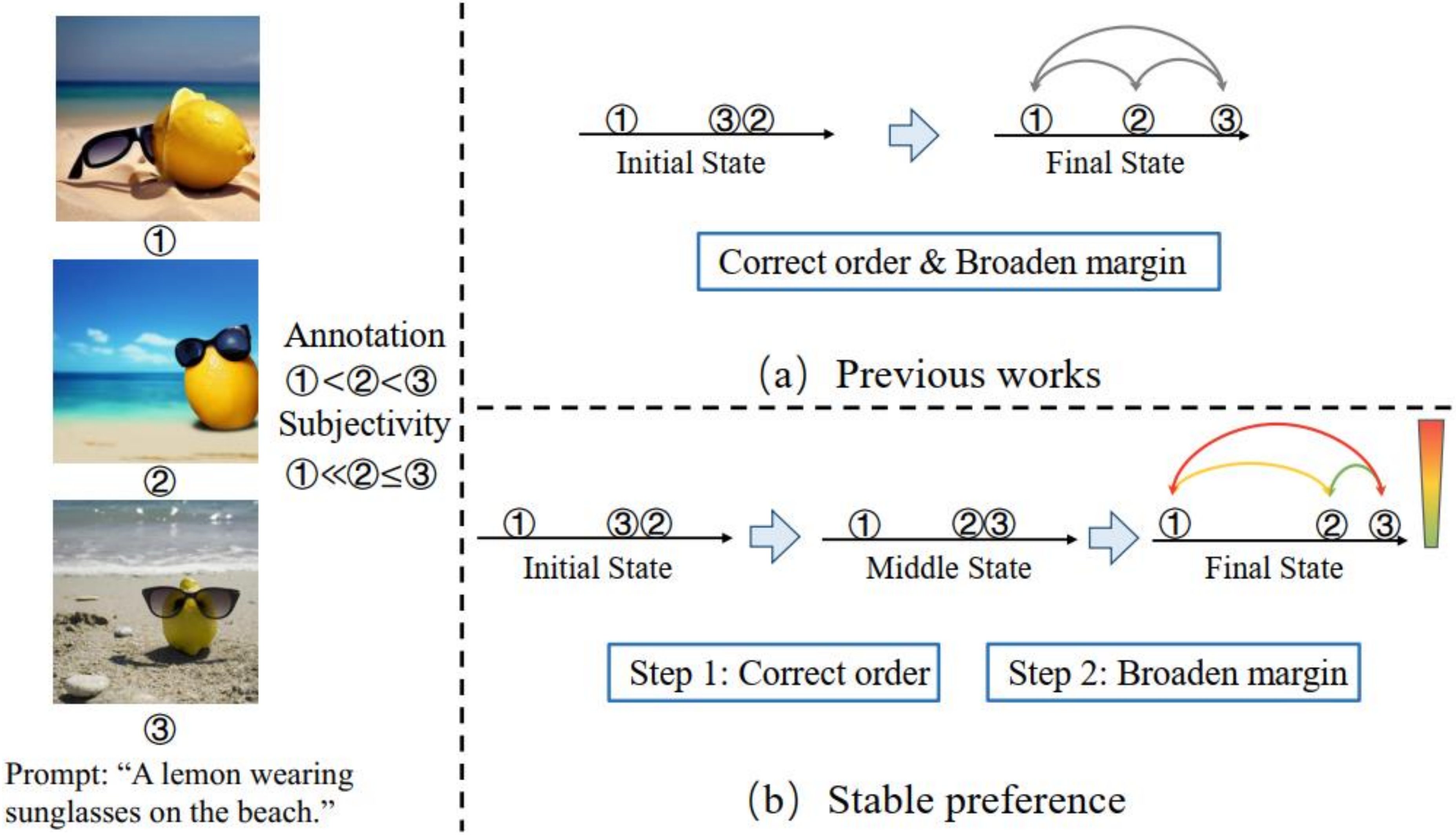
● **Anti-interference loss**



$$\mathcal{L}_{ai} = -\log \frac{e^{dist_{inter}}}{e^{dist_{inter}} + e^{dist_{intra}}}$$

Methodologies

- **Stable preference**



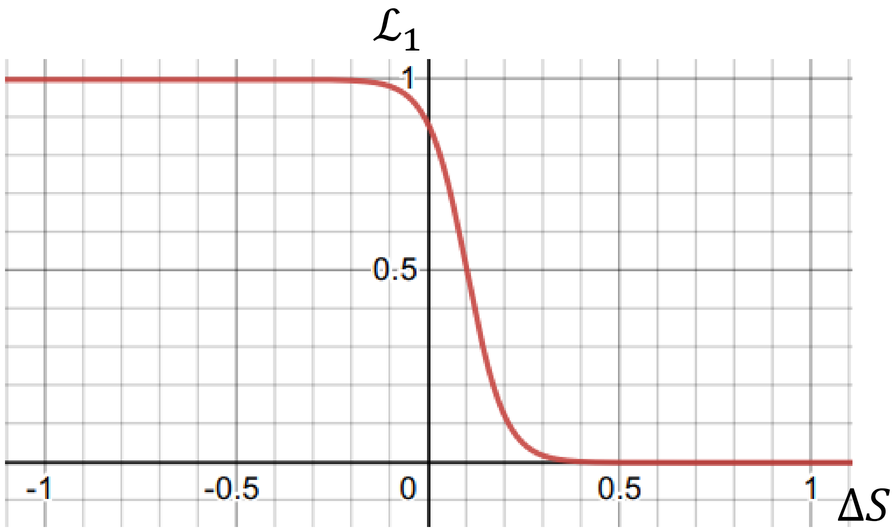
Methodologies

● Stable preference

- Step 1: Correct the preference order

Training Loss: $\mathcal{L}_1 = \dfrac{\mathcal{L}_{pref} + \mathcal{L}_{ai}}{1 + e^{(\Delta S - b)/\tau}}$ if I_1 is better than I_2, then $\Delta S = S_1 - S_2$

e.g., $\dfrac{1}{1+e^{(\Delta S - 0.1)/0.05}}$



- Step 2: Broaden the margin

Training Loss: $$\mathcal{L}_2 = \frac{e^{\Delta S_j}}{\sum_{i=1}^{N} e^{\Delta S_i}} (\mathcal{L}_{pref} + \mathcal{L}_{ai})$$

Experimental results

Datasets and implementation details

- Datasets: ImageReward, Human Preference and DrawBench Datasets
- Evaluation Metric: Accuracy (of preferred image selection)
- Input sizes: all images are resized to 224×224
- Optimizer: AdamW optimizer with a learning rate initialized to 2×10^{-6}
- Training process: stage 1 for 3,000 steps and stage 2 for 27,000 steps
- Model: CLIP-H and CLIP-L

Experimental results

- **Comparison of human preference models sensitivity to small visual perturbations on HPD v2 and ImageReward datasets. “ORG” represents the baseline result on original test split. “HP” and “CC” stand for horizontal flip and center crop, respectively. Numbers in brackets represent the side length ratio of the center crop. SP represents our stable preference training paradigm.**

Method	Dataset	ORG	HP&CC (0.97)	HP&CC (0.95)	HP&CC (0.93)	HP&CC (0.90)
HPS v2	HPD v2	83.3	82.2 (-1.1)	82.2 (-1.1)	81.8 (-1.5)	81.7 (-1.6)
ImageReward		74.2	73.7 (-0.5)	73.6 (-0.6)	73.6 (-0.6)	74.0 (-0.2)
SP (CLIP-L)		77.2	77.3 (+0.1)	77.0 (-0.2)	76.9 (-0.3)	77.0 (-0.2)
SP (CLIP-H)		80.7	81.4 (+0.7)	80.3 (+0.4)	80.4 (+0.3)	80.7 (+0.0)
HPS v2	ImageReward	65.7	64.8 (-0.9)	63.8 (-1.9)	64.2 (-1.5)	63.9 (-1.8)
ImageReward		65.2	64.5 (-0.7)	64.8 (-0.4)	64.8 (-0.4)	65.3 (+0.1)
SP (CLIP-L)		66.3	65.7 (-0.6)	65.6 (-0.7)	65.9 (-0.4)	66.0 (-0.3)
SP (CLIP-H)		66.8	67.4 (+0.6)	66.4 (-0.4)	66.5 (-0.3)	66.7 (-0.1)

Experimental results

- **Comparison with state-of-the art methods on test split of ImageReward dataset. † CLIP-H is initialized with the HPS v2 checkpoint.**

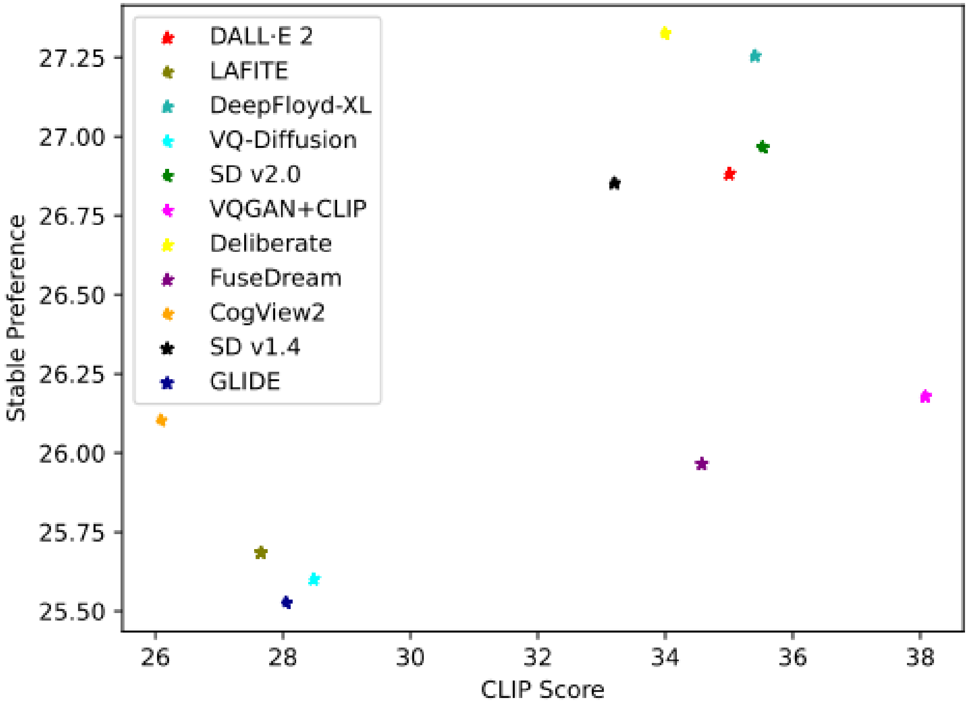
Method	ImageReward
CLIP-L [11, 25]	54.8
CLIP-H [11, 25]	57.1
Aesthetic Score Predictor [35]	57.4
HPS v1 [38]	61.2
PickScore [13]	62.9
ImageReward [39]	65.1
HPS v2 [37]	65.7
Single Human vs. Single Human	65.3
Single Human vs. Averaged Human	53.9
Stable Preference (CLIP-L)	66.3
Stable Preference (CLIP-H)	66.8
Stable Preference (CLIP-H†)	**68.0**

- **Comparison of cross-domain performance. All the models are trained on the training set of ImageReward and tested on the test split of HPD v2. † CLIP-H is initialized with the HPS v2 checkpoint.**

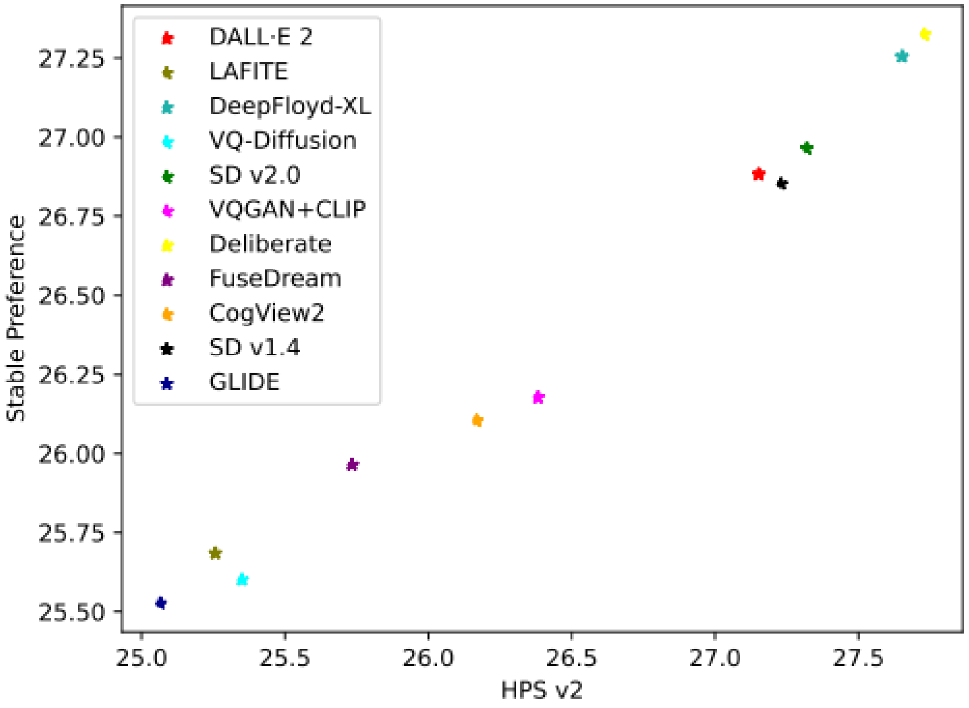
Method	HPD v2
CLIP-L [11, 25]	72.8
CLIP-H [11, 25]	74.8
BLIP [15]	74.2
Single Human vs. Single Human	78.1
Single Human vs. Averaged Human	85.0
Stable Preference (CLIP-L)	77.2
Stable Preference (CLIP-H)	80.7
Stable Preference (CLIP-H†)	**82.5**

Experimental results

- **Correlation between stable preference and other human preference models. The model score is calculated by the average score of all images in DrawBench**



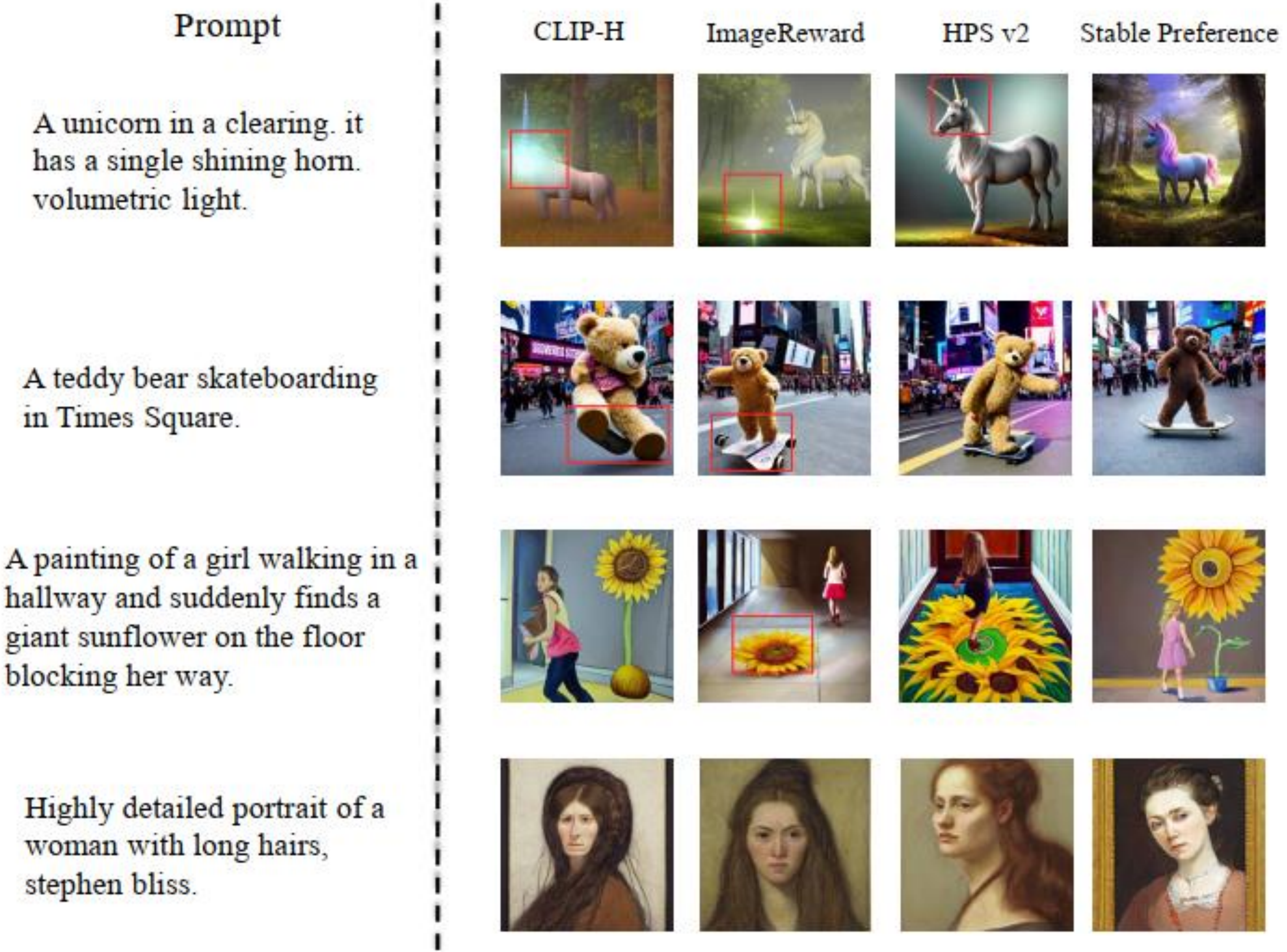
(a) Stable preference & CLIP Score



(b) Stable preference & HPS v2

Experimental results

- **Top-1 images out of 100 (Stable Diffusion v1.4) generations selected by stable preference and other HPMs.**



Conclusion

- We propose Stable Preference, a new training paradigm for human preference models. Training in the order of first aligning preference order and then mainly broaden the margin between images with significant difference effectively mitigates the risk of overfitting.

- We designed an anti-interference loss to reduce the sensitivity of preference model to small visual perturbations that do not affect human preferences

Thank you

Q & A