ColorPeel: Color Prompt Learning with Diffusion Models via Color and Shape Disentanglement

Muhammad Atif Butt[1,2], Kai Wang[1], Javier Vazquez-Corral[1,2], and Joost Van De Weijer[1,2]

[1] Computer Vision Center, Spain. [2] Department of Computer Science, Universitat Autònoma de Barcelona, Spain.

mabutt@cvc.uab.es, kwang@cvc.uab.es, jvazquez@cvc.uab.es, joost@cvc.uab.es

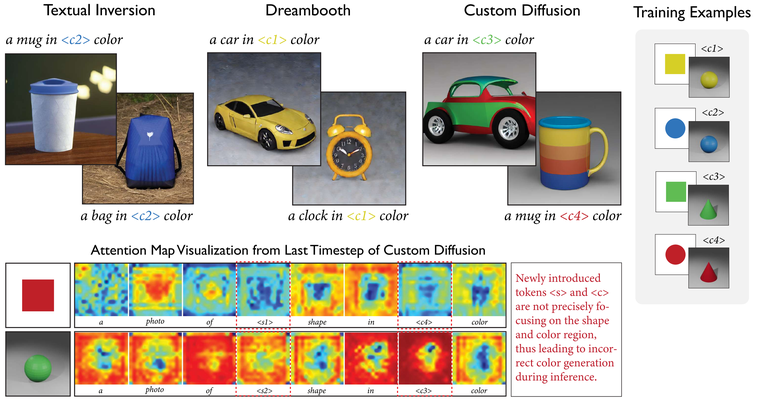
INTRODUCTION

Problem: Text-to-Image models allow users to specify object colors using linguistic color names. However, these labels encompass broad color ranges, making it challenging to achieve precise color generation.

Stable Diffusion v1.4

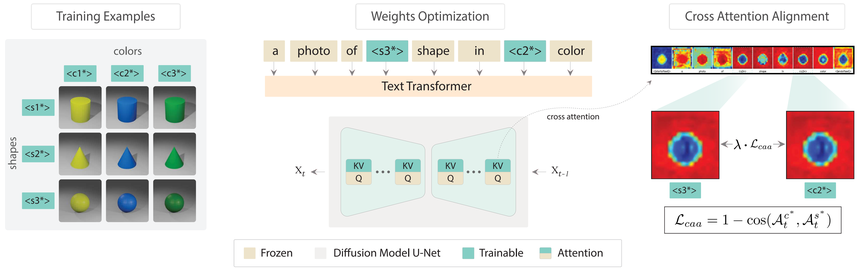
T2I Personalization Methods

- Single-concept learning methods — Textual Inversion and Dreambooth can generate text-guided objects in specified colors, however fail to generate consistent colors.
- Custom Diffusion inter-mixes the colors while also reducing sample variation.



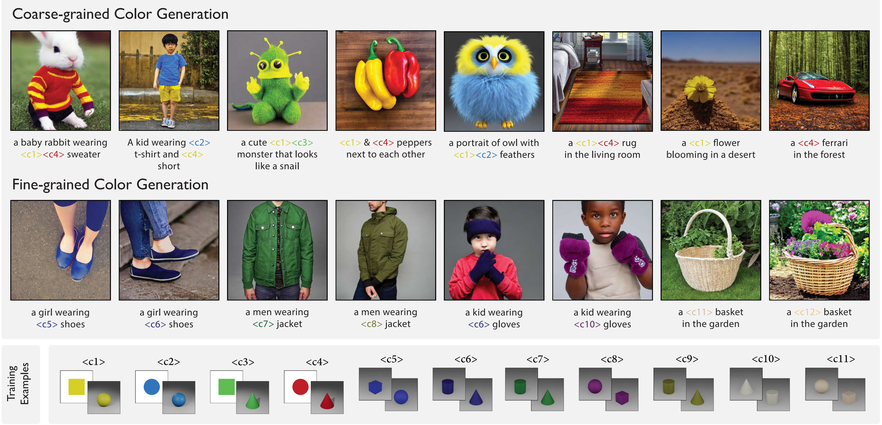
PROPOSED METHOD

TL,DR: We propose to learn specific color prompts tailored to user provided color.



- Instance images and training templates (text-prompts) are generated given user-provided RGB triplets.
- New modifier tokens <s*> and <c*> are introduced to learn shapes and colors, respectively.
- The key and value projection matrices in DM cross-attention layers are optimized with modifier tokens.
- Cross Attention Alignment (CAA) Loss is introduced to ensure disentanglement of color from shape.

QUALITATIVE RESULTS



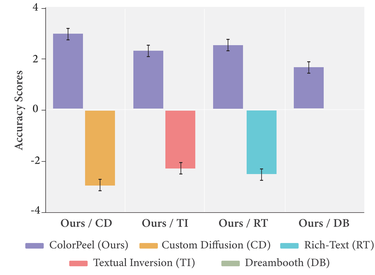
COMPARISONS

(a) Quantitative Results

Quantitative comparison with baselines over various evaluation metrics

Method	ΔE	ΔE_{ch}	MAE (rgb) 10%	MAE (rgb) 50%	MAE (rgb) 100%	MAE (Hue) 10%	MAE (Hue) 50%	MAE (Hue) 100%	Time (min)
SD [41]	47.45	41.55	12.89	20.04	26.93	30.17	54.14	86.38	-
Rich-Text [14]	36.62	32.48	9.91	13.29	18.53	50.55	72.77	93.51	-
TI [12]	48.98	44.29	15.22	19.51	23.90	52.66	69.35	90.88	118
DB [43]	50.71	46.29	14.75	19.30	23.70	47.12	67.13	88.72	56
CD [27]	48.47	42.23	13.43	17.93	22.43	31.63	55.07	78.43	24
ColorPeel (3D)	21.39	16.51	**4.36**	**7.76**	**12.08**	**2.63**	**6.47**	**21.35**	**19**
ColorPeel (2D)	**20.45**	**15.29**	4.83	7.88	12.13	3.18	7.43	21.46	

(b) User Preference Study (15 participants)



GENERALIZABILITY