Improving Compositional Generation with Diffusion Models Using Lift Scores

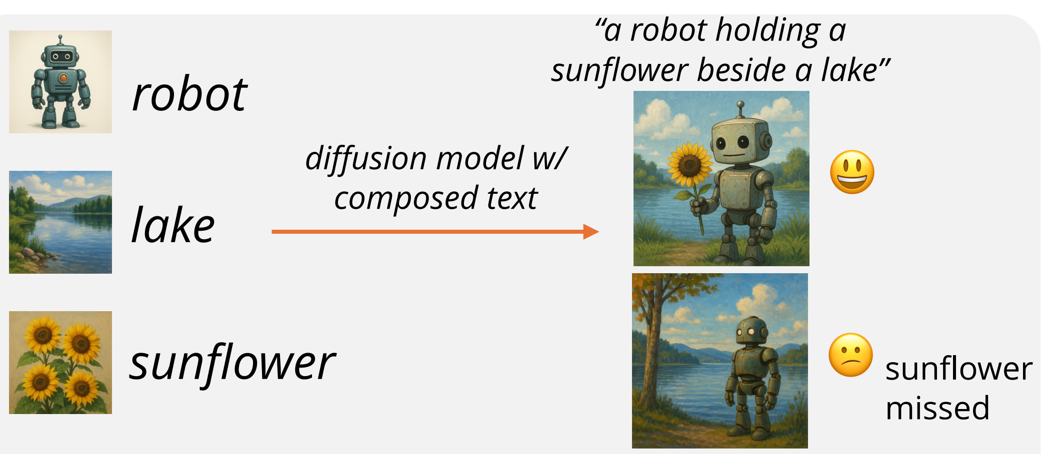
Chenning Yu and Sicun Gao, UC San Diego

ICML
International Conference On Machine Learning

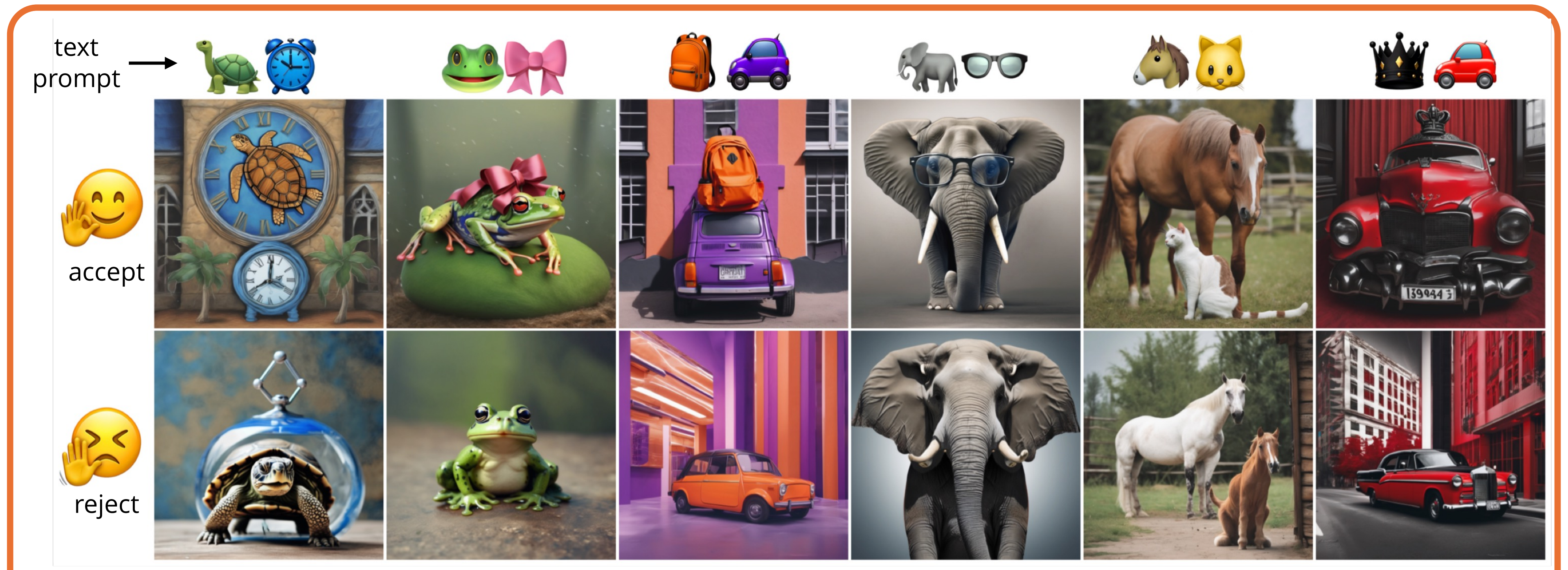
Blog | Paper | Code

We use diffusion model itself to reject bad images, in compositional way

Background



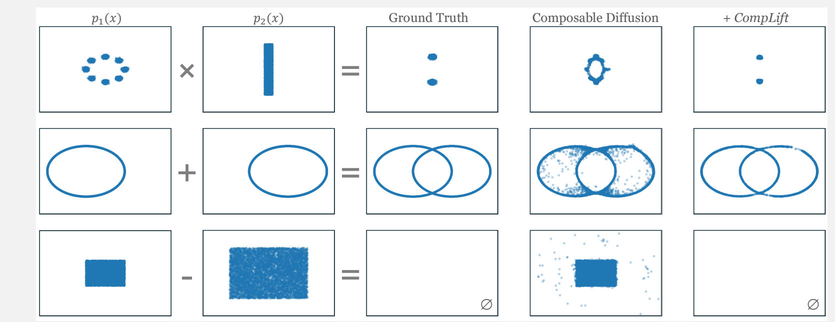
diffusion models sometimes miss components in text prompt



accepted and rejected Stable Diffusion XL images using lift scores

Key Highlights

- Training-free, use diffusion model itself
- Simple concept – compare with 0 to accept/reject
- Boosts base models by 106% (2D generation) and 16% (Image generation with SD XL)



Lift Score

An existing concept in data mining (Brin et al., 1997)

$$lift(\boldsymbol{x}|\boldsymbol{c}) := \log \frac{p(\boldsymbol{x}|\boldsymbol{c})}{p(\boldsymbol{x})}.$$

conditional distribution

unconditional distribution

sample (e.g., an image)

constraint (e.g., a text component)

Diffusion Models can Easily Estimate Lift Score

$$\log \frac{p(\boldsymbol{x}|\boldsymbol{c})}{p(\boldsymbol{x})} \begin{array}{l} \approx \exp(-\mathbb{E}_{t,\epsilon}\|\epsilon - \epsilon_\theta(\boldsymbol{x}_t, \boldsymbol{c})\|^2 + C), \\ \approx \exp(-\mathbb{E}_{t,\epsilon}\|\epsilon - \epsilon_\theta(\boldsymbol{x}_t, \varnothing)\|^2 + C). \end{array}$$

random sample timestep/noise

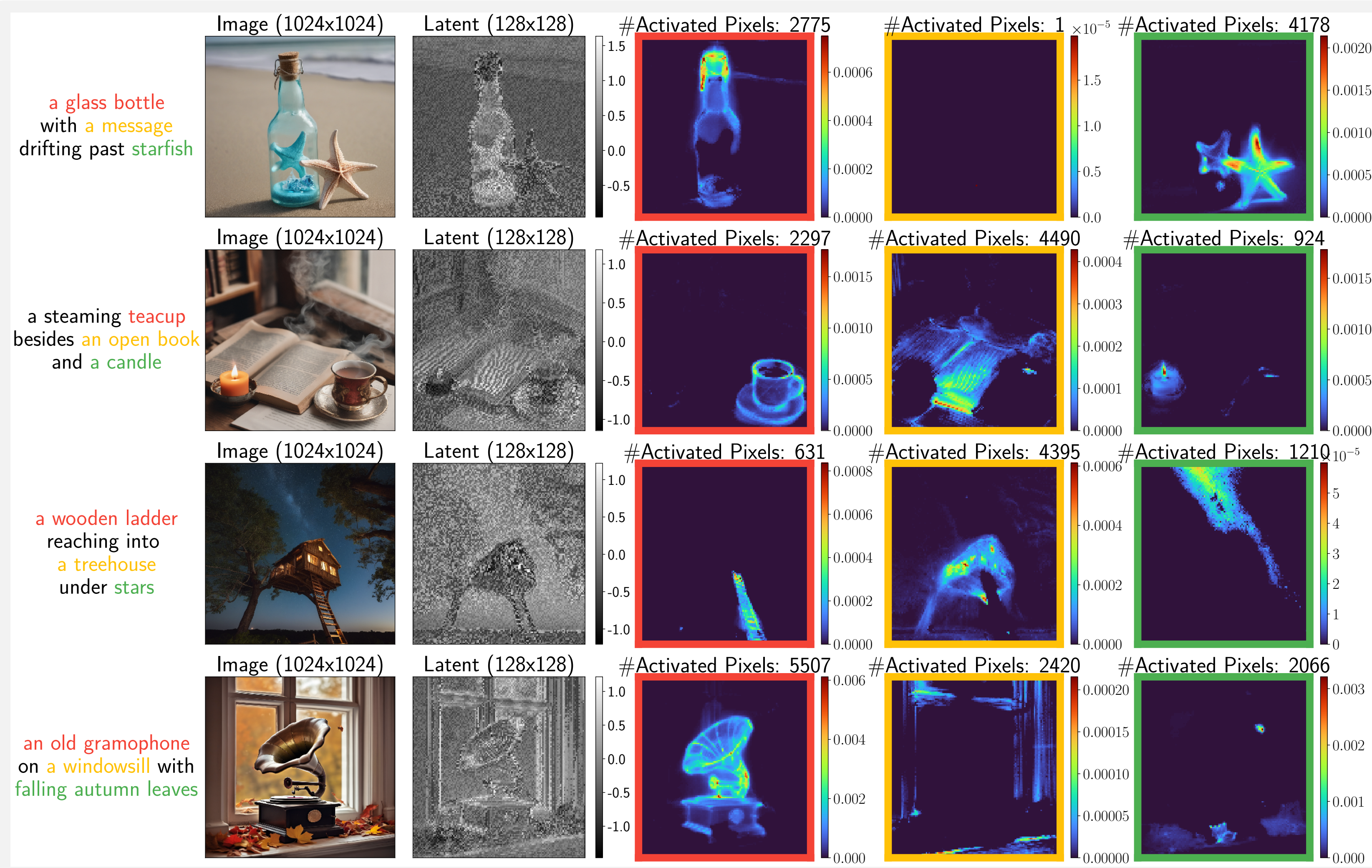
diffusion model

We use lift score as rejection criterion to check alignment of each individual component; then compose the criterion to accept/reject

Type	Algebra	Acceptance Criterion
Product	$\boldsymbol{c}_1 \wedge \boldsymbol{c}_2$	$\min_{i\in[1,2]} lift(\boldsymbol{x}\|\boldsymbol{c}_i) > 0$
Mixture	$\boldsymbol{c}_1 \vee \boldsymbol{c}_2$	$\max_{i\in[1,2]} lift(\boldsymbol{x}\|\boldsymbol{c}_i) > 0$
Negation	$\neg \boldsymbol{c}_1$	$lift(\boldsymbol{x}\|\boldsymbol{c}_1) \leq 0$

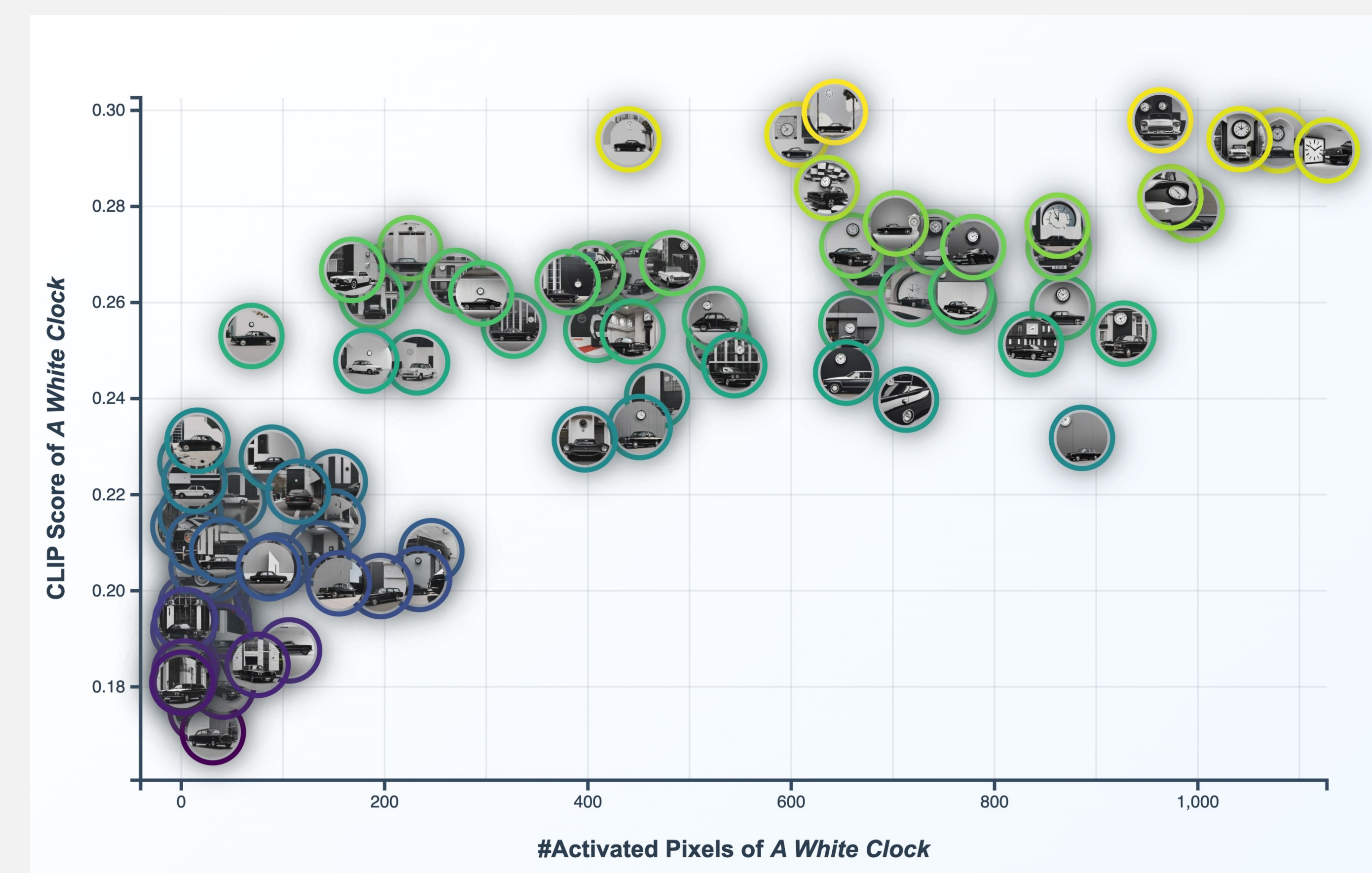
Table 1. **Examples of** `Compose` **for multiple conditions.**

Visualization in Latent Space



lift score shows strong correspondence to each text component; activated pixels are pixels with positive lift scores

Correlation with CLIP Score



images with missing components tend to have (1) lower CLIP score and (2) fewer activated pixels with positive lift scores
prompt: a black car and a white clock