Guide-and-Rescale

Self-guidance Mechanism for Effective Tunning-Free Real Image Editing

Vadim Titov, Madina Khalmatova, Alexandra Ivanova, Dmitriy Vetrov, Aibek Alanov

European Conference on Computer Vision, October, 2024

EUROPEAN CONFERENCE ON COMPUTER VISION

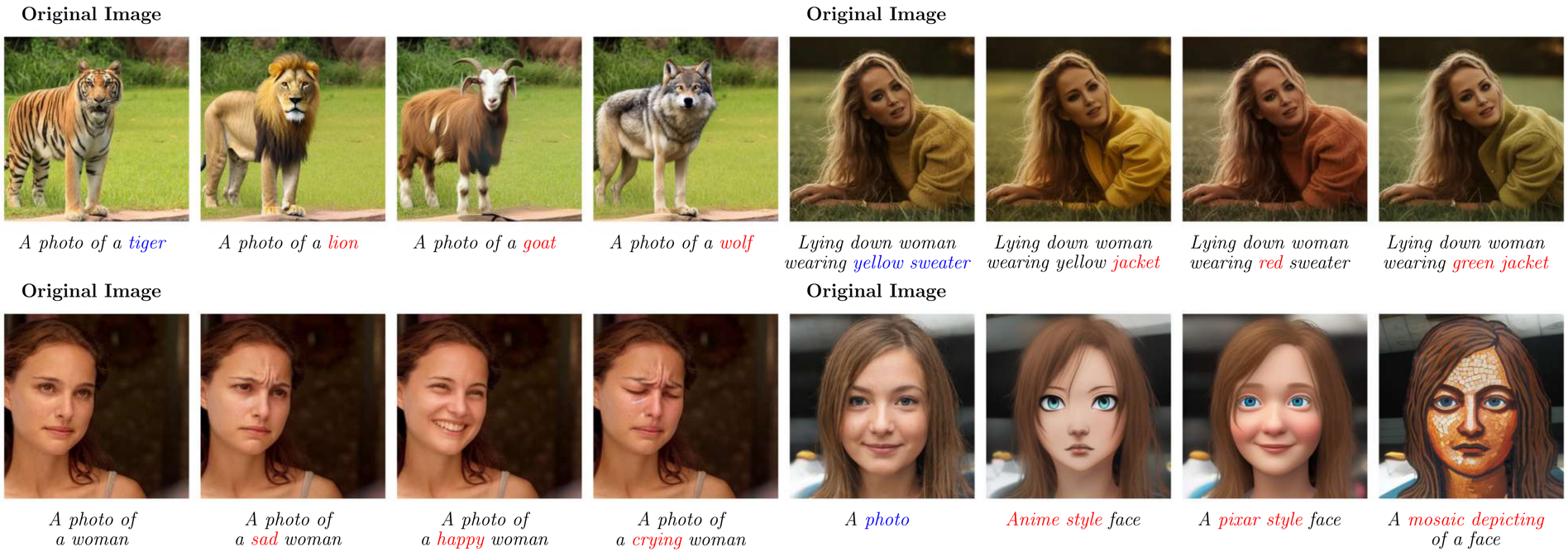
MILANO

2024

Paper summary

Introduction, Problem Statement

- Challenge
  - Manipulating real images with t2im models while preserving original structure



Paper summary
Introduction, Baselines

- Existing approaches
  - Fine-tuning (Memory Intensive + Time Consuming)
  - Internal representation exchange (Limited versatility)
  - High-quality reconstruction (Time consuming)

Paper summary

Introduction, Key contributions

- Proposed Framework: Guide-and-Rescale
  - Guidance technique to preserve original structure
  - No-fine tuning or exact inversion required
- Key benefits:
  - Controllable structure preservation via special guidance
  - Automative Noise Rescaling technique
  - Best balance between edit quality and original image preservation

Technical Details

Guidance recap

- Classifier guidance & classifier-free guidance

$$\hat{\epsilon}_t = \epsilon_\theta(z_t; t, y) - s\sigma_t \nabla_{z_t} \log p(y|z_t)$$

$$\hat{\epsilon}_t = (1+s)\epsilon_\theta(z_t; t, y) - s\epsilon_\theta(z_t; t, \emptyset)$$

- Combined guidance approach

$$\hat{\epsilon}_t = (1+s)\epsilon_\theta(z_t; t, y) - s\epsilon_\theta(z_t; t, \emptyset) + v\sigma_t \nabla_{z_t} g(z_t; t, y)$$

Technical Details

Proposed guidance scheme

$$\hat{\varepsilon}_\theta(z_t, t, y) = \mathrm{CFG}(z_t, t, y_{\mathrm{trg}}, 7.5) + v \cdot \nabla_{z_t} g(z_t, z_t^*, t, y_{\mathrm{src}}, \mathcal{I}^*, \bar{\mathcal{I}})$$

- Self-attention guider

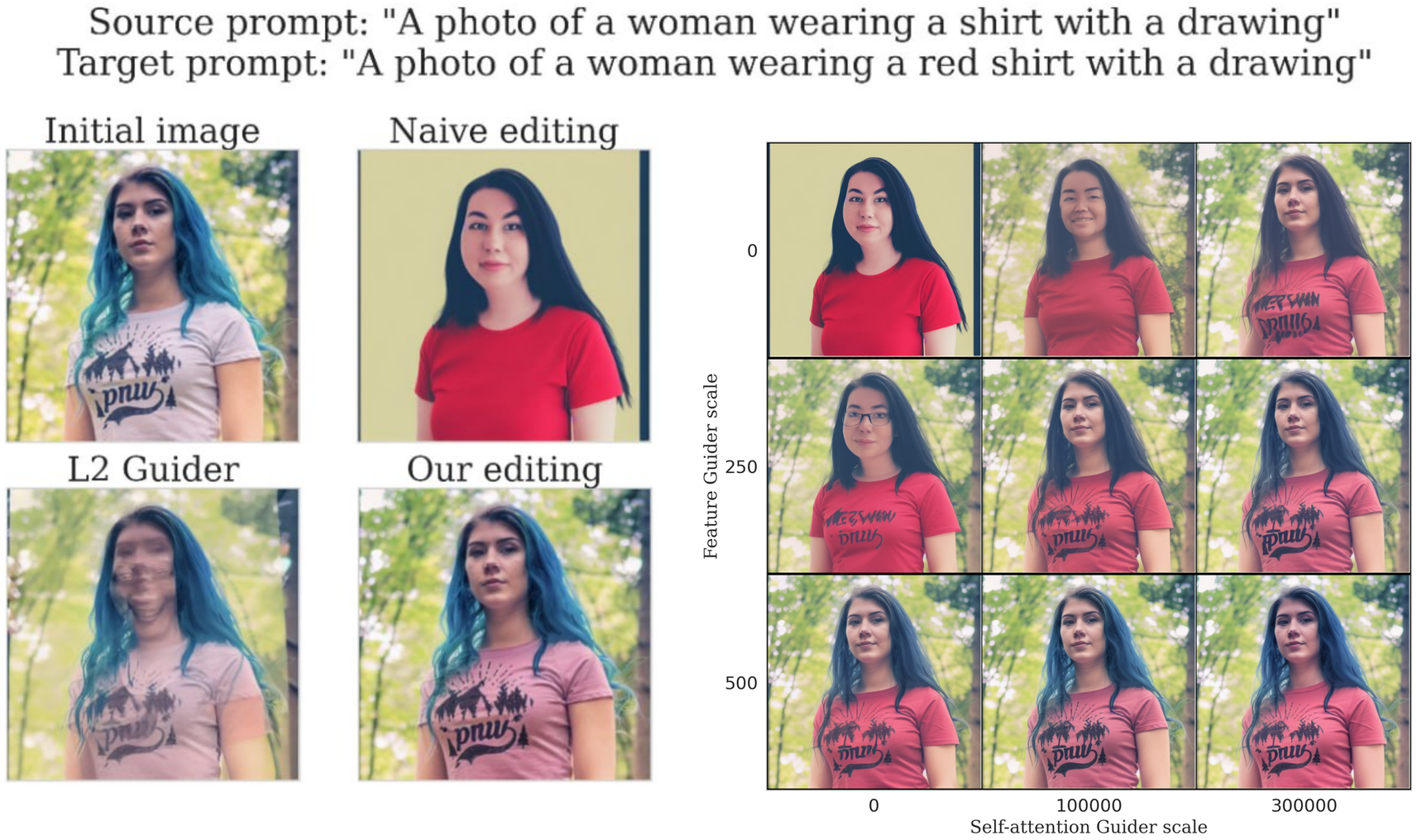
$$g(z_t, z_t^*, t, y_{\mathrm{src}}, \{\mathcal{A}_i^{*\mathrm{self}}\}, \{\bar{\mathcal{A}}_i^{\mathrm{self}}\}) = \sum_{i=1}^{L} \mathrm{mean} \|\mathcal{A}_i^{*\mathrm{self}} - \bar{\mathcal{A}}_i^{\mathrm{self}}\|_2^2$$

- Feature guider

$$g(z_t, z_t^*, t, y_{\mathrm{src}}, \Phi^*, \bar{\Phi}) = \mathrm{mean} \|\Phi^* - \bar{\Phi}\|_2^2$$

Technical Details

Proposed guidance scheme, motivation



Technical Details

Rescaling technique

- Overall guidance scheme

$$\epsilon_t = \varepsilon_\theta(z_t, t, \varnothing) + w\big(\varepsilon_\theta(z_t, t, y) - \varepsilon_\theta(z_t, t, \varnothing)\big) + \\ + \gamma \sum_i v_i \cdot \nabla_{z_t} g_i(z_t, z_t^*, t, y_{\text{src}}, \mathcal{I}^*, \overline{\mathcal{I}}),$$

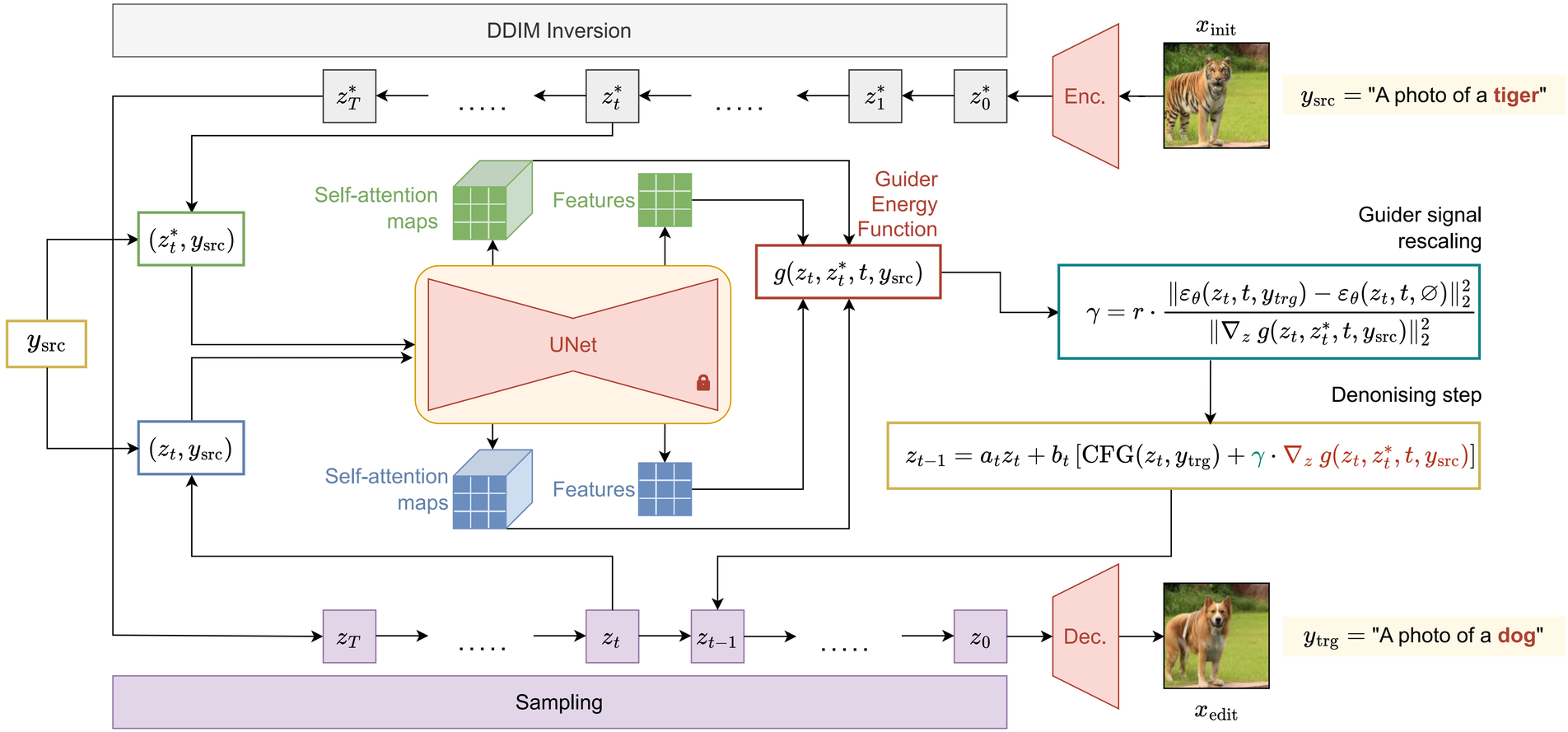
- Automative rescaling scheduling

$$r_{\text{cur}}(t) = \frac{\| w\big(\varepsilon_\theta(z_t, t, y) - \varepsilon_\theta(z_t, t, \varnothing)\big) \|_2^2}{\| \sum_i v_i \cdot \nabla_{z_t} g_i(z_t, z_t^*, t, y_{\text{src}}) \|_2^2},$$

$$\gamma = r \cdot r_{\text{cur}}(t).$$

Technical Details

Overall pipeline



Results

Qualitative

Results

Quantitative

Method	LPIPS ↓	CLIP ↑	FID ↓	Time (s) ↓
ProxMasaCtrl [5]	0.267	0.215	94.53	12.94
MasaCtrl [2]	0.306	0.223	100.62	13.73
EDICT [27]	<u>0.221</u>	0.229	47.13	68.13
P2P [6] + NTI [15]	0.279	0.233	42.46	66.77
P2P [6] + NPI [14] Prox [5]	**0.170**	0.233	43.16	8.59
P2P [6] + NPI [14]	0.251	0.234	44.05	8.54
PnP [24]	0.366	**0.256**	<u>39.55</u>	197.0
Guide-and-Rescale (ours)	0.228	<u>0.243</u>	**39.07**	24.26

Contacts

HF Demo:

Code:

Project page:

Vadim Titov

Madina Khalmatova

Alexandra Ivanova

Dmitry Vetrov

Aibek Alanov