Drag Your GAN: Interactive Point-based Manipulation on the Generative Image Manifold

Августёнок Алина, БПМИ211

Plan:

- problem definition
- reminder of StyleGAN
- DragGAN
- experiments
- limitations
- images!

Problem

ideal controllable image synthesis approach:

- flexible
- precise
- general

Existing solutions

- (un)conditional GANs
- 3D-aware GANs
- diffusion models

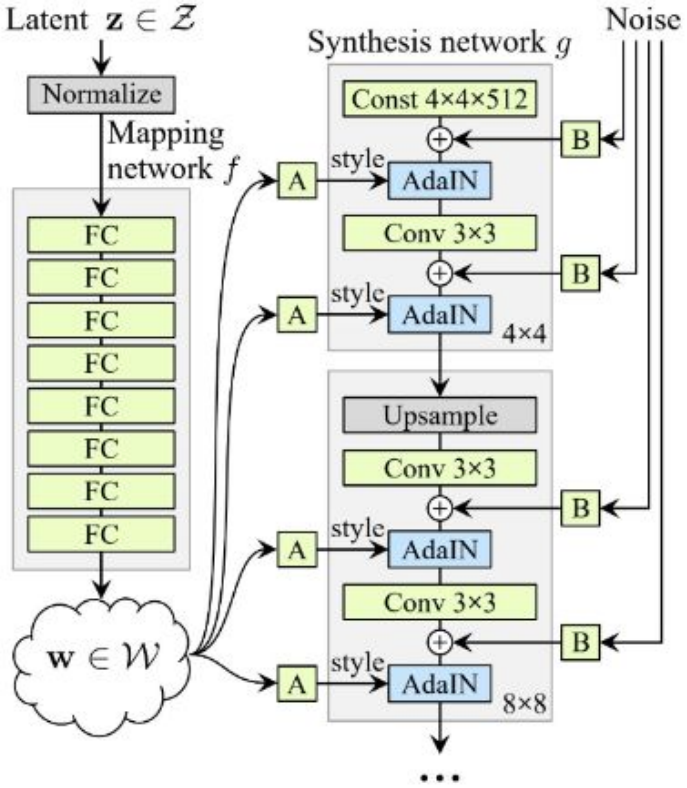
StyleGAN

$$\mathbf{z} \in \mathcal{N}(0, I)$$

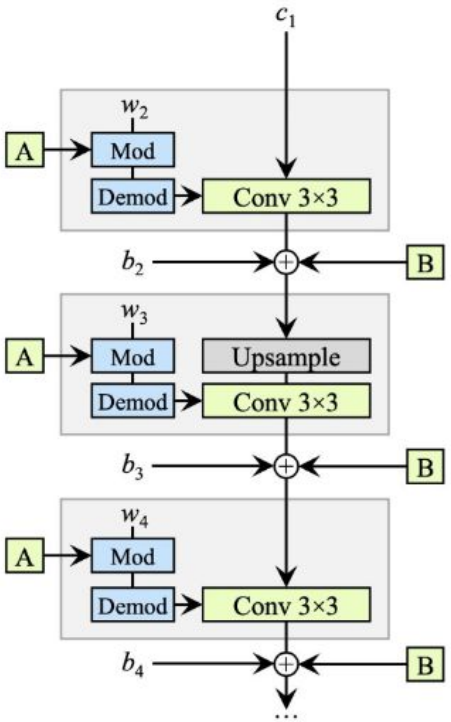
$$\mathbf{w} \in \mathbb{R}^{512}$$

generator G

$$Image = G(\mathbf{w})$$

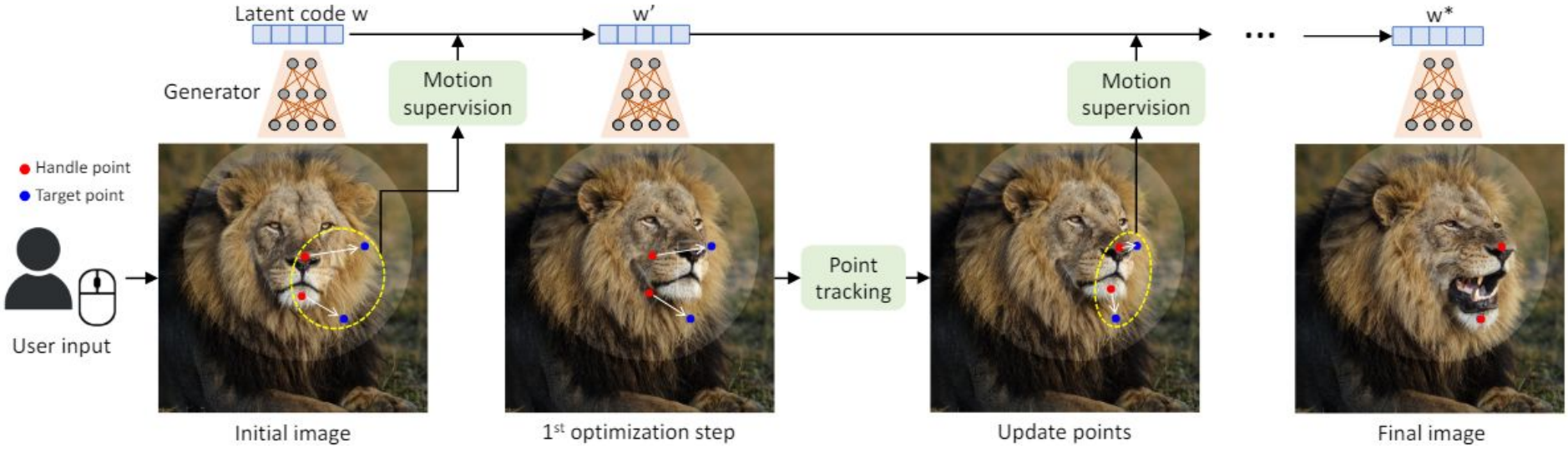


(b) Style-based generator

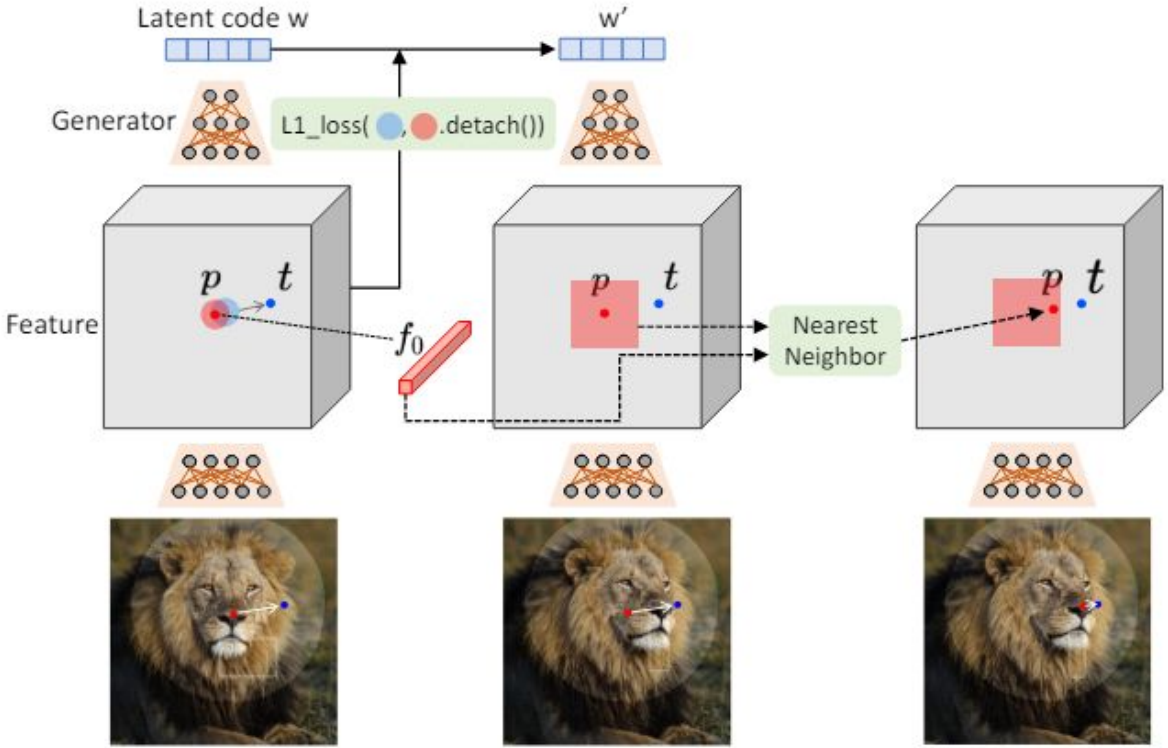


(d) Weight demodulation

DragGAN


Latent code w
w'
w*
...
Generator
Motion supervision
Motion supervision
Handle point
Target point
User input
Point tracking
Initial image
1st optimization step
Update points
Final image

Motion Supervision



Motion Supervision: Loss

$$\mathcal{L} = \sum_{i=0}^{n} \sum_{q_i \in \Omega_1(p_i, r_1)} \|\mathbf{F}(q_i) - \mathbf{F}(q_i + d_i)\|_1 + \lambda \|(\mathbf{F} - \mathbf{F}_0) \cdot (1 - \mathbf{M})\|_1$$

$\Omega_1(p_i, r_1)$ – r_1 - neighborhood of p_i

F – feature maps, $\mathbf{F}(q)$ – feature values at pixel q, $\mathbf{F}_0$ – feature maps of initial image

$d_i = \frac{t_i - p_i}{\|t_i - p_i\|_2}$ – normalized vector from p_i (handle point) to t_i (target point)

$\mathbf{M}$ – binary mask

$\lambda = 20, r_1 = 3,$ – hyperparameters

Point Tracking

$$p_i := \underset{q_i \in \Omega_2(p_i, r_2)}{\arg\min} \|\mathbf{F}'(q_i) - f_i\|_1.$$

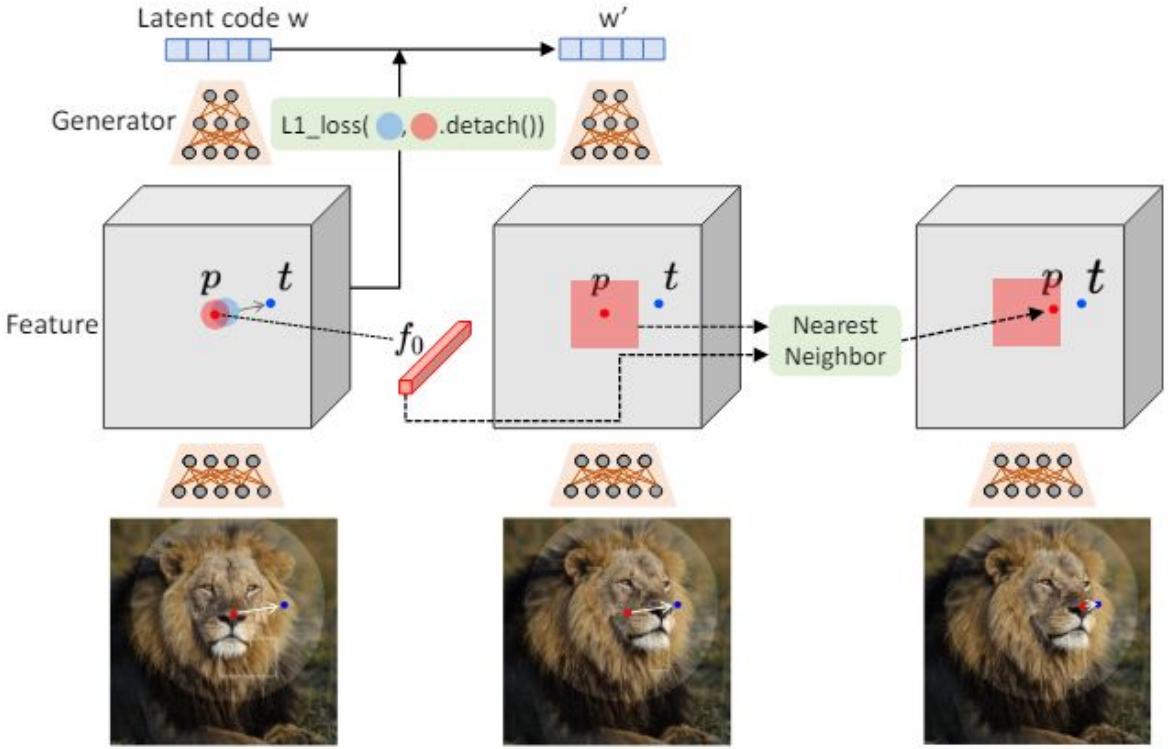
$\Omega_2(p_i, r_2) = \{(x, y) \mid |x - x_{p,i}| < r_2, |y - y_{p,i}| < r_2\}$ – r_2 - neighborhood of p_i

$f_i = \mathbf{F}_0(p_i)$ – feature of the initial handle point

$\mathbf{F}'(q_i)$ – feature values after motion supervision

$r_2 = 12$ – hyperparameter

Point Tracking



Experiments

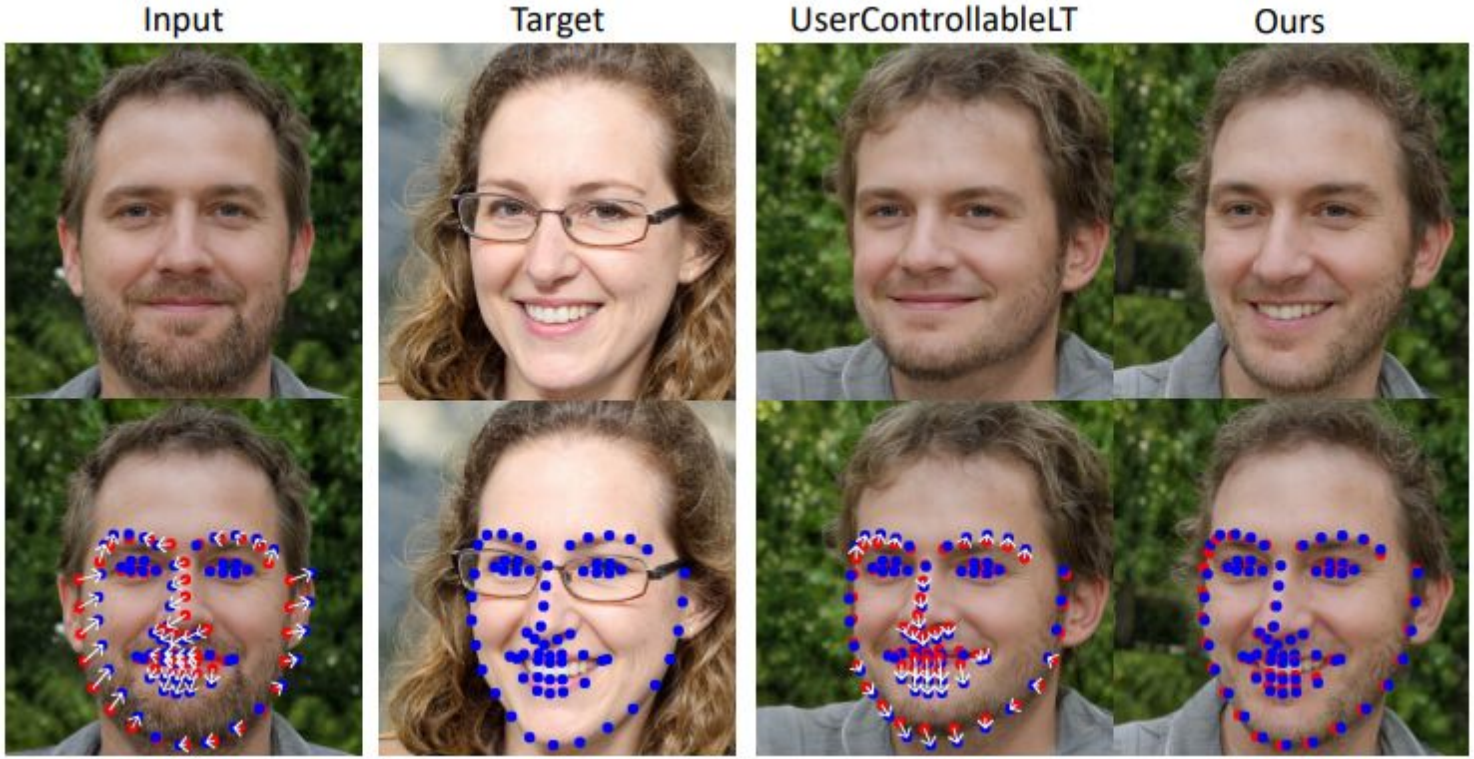
face landmark manipulation:

Method	1 point	5 points	68 points	FID	Time (s)
No edit	12.93	11.66	16.02	-	-
UserControllableLT	11.64	10.41	10.15	25.32	0.03
Ours w. RAFT tracking	13.43	13.59	15.92	51.37	15.4
Ours w. PIPs tracking	2.98	4.83	5.30	31.87	6.6
Ours	**2.44**	**3.18**	**4.73**	**9.28**	2.0

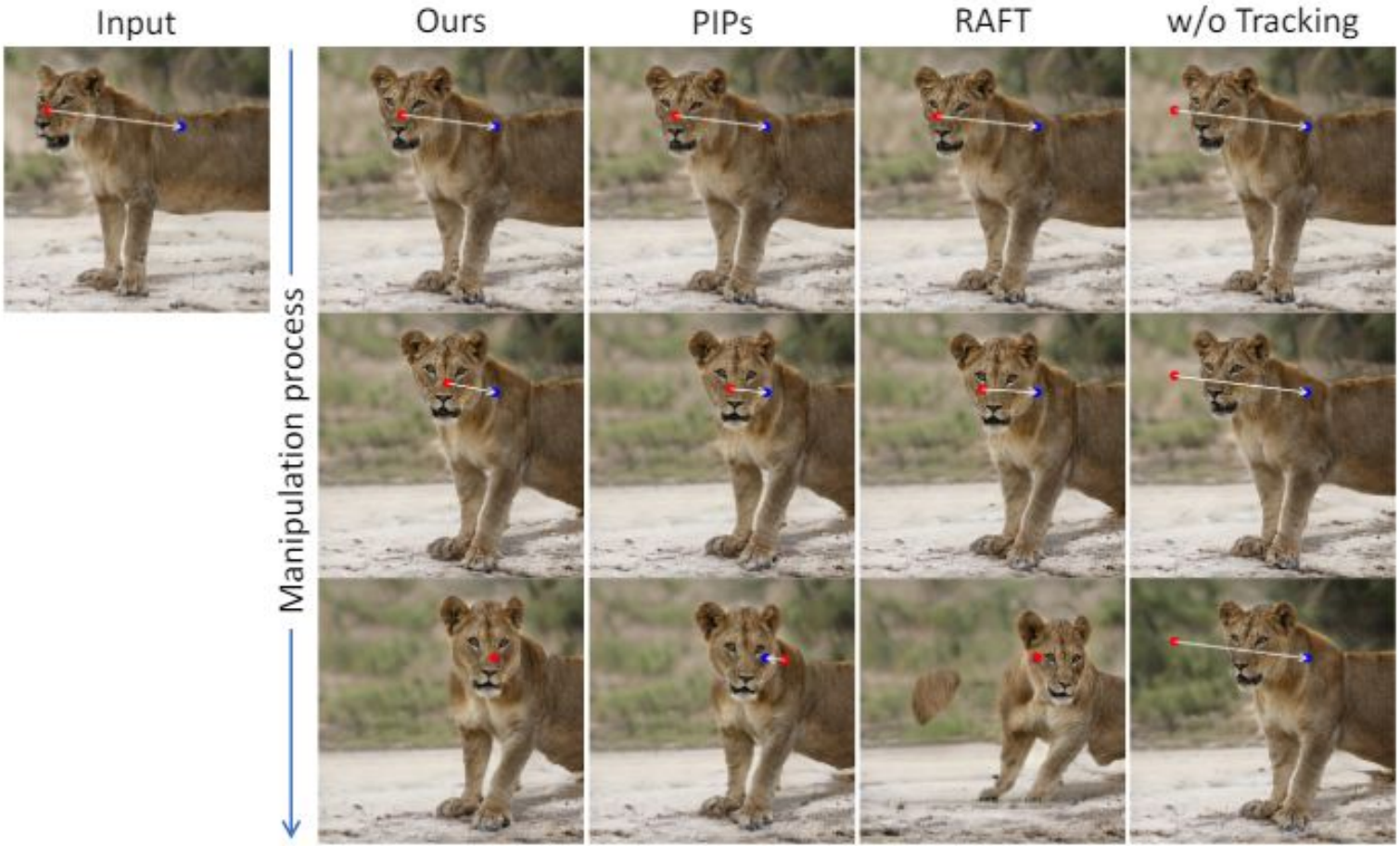
paired image reconstruction:

Dataset	Lion		LSUN Cat		Dog		LSUN Car	
Metric	MSE	LPIPS	MSE	LPIPS	MSE	LPIPS	MSE	LPIPS
UserControllableLT	1.82	1.14	1.25	0.87	1.23	0.92	1.98	0.85
Ours w. RAFT tracking	1.09	0.99	1.84	1.15	0.91	0.76	2.37	0.94
Ours w. PIPs tracking	0.80	0.82	1.11	0.85	0.78	0.63	1.81	0.79
Ours	**0.66**	**0.72**	**1.04**	**0.82**	**0.48**	**0.44**	**1.67**	**0.74**

Face Landmark Manipulation



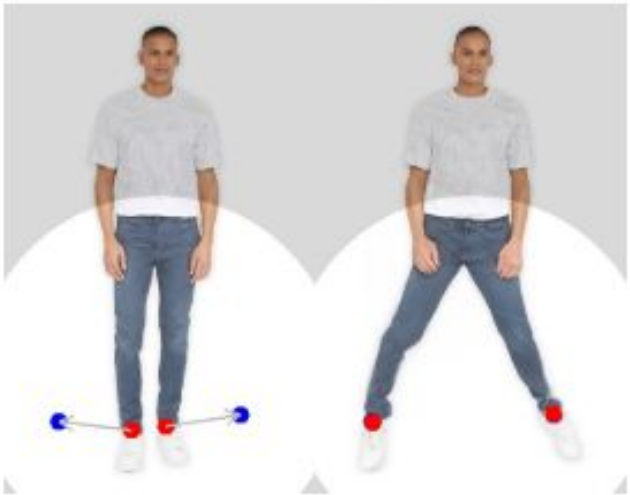
Experiments

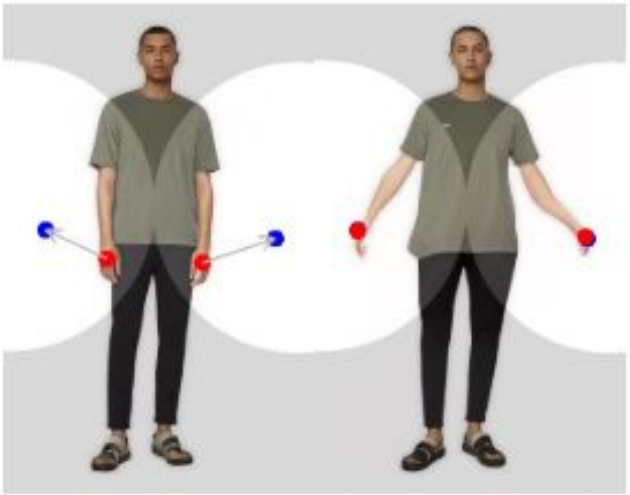


Effects of the mask

Limitations

- diversity of training data





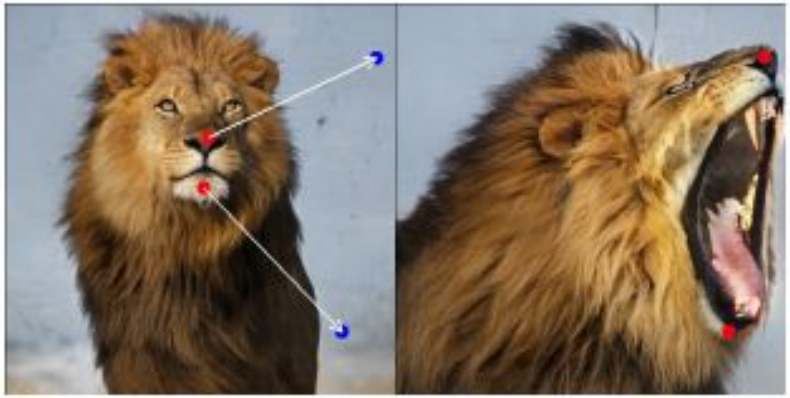
Limitations

- handle points in texture-less regions sometimes suffer

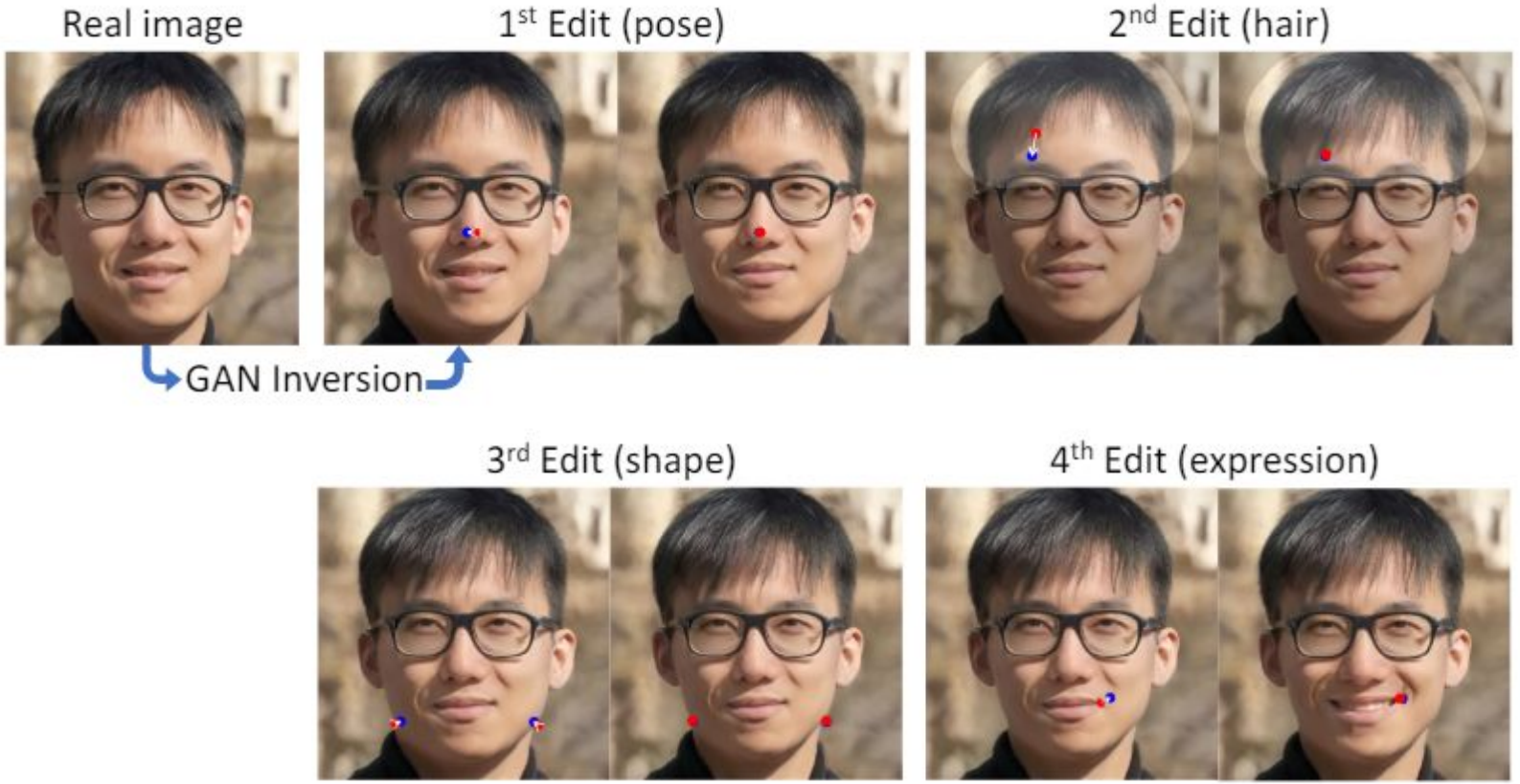
(b) Texture-less handle point

(c) Texture-rich handle point

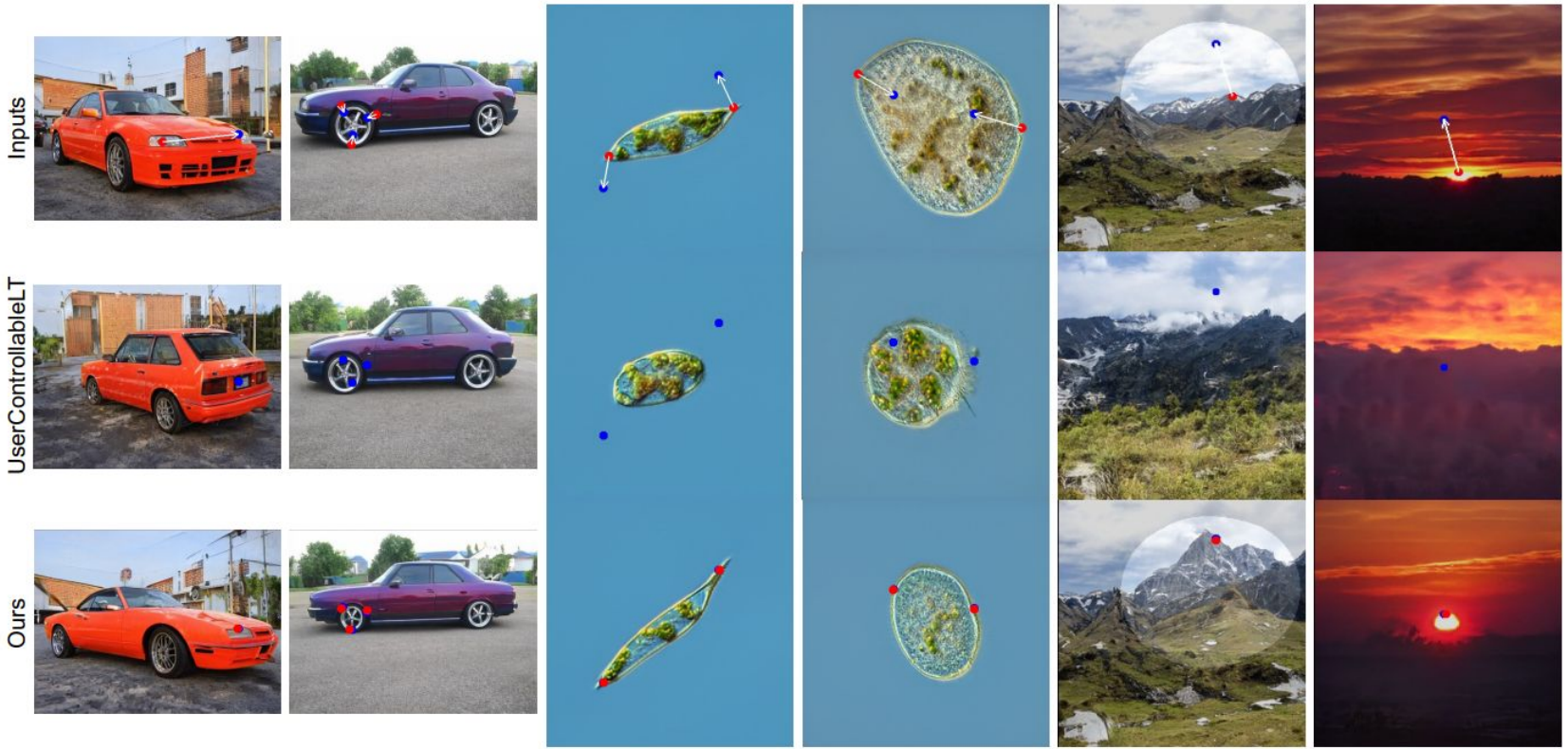
Out-of-distribution Manipulations



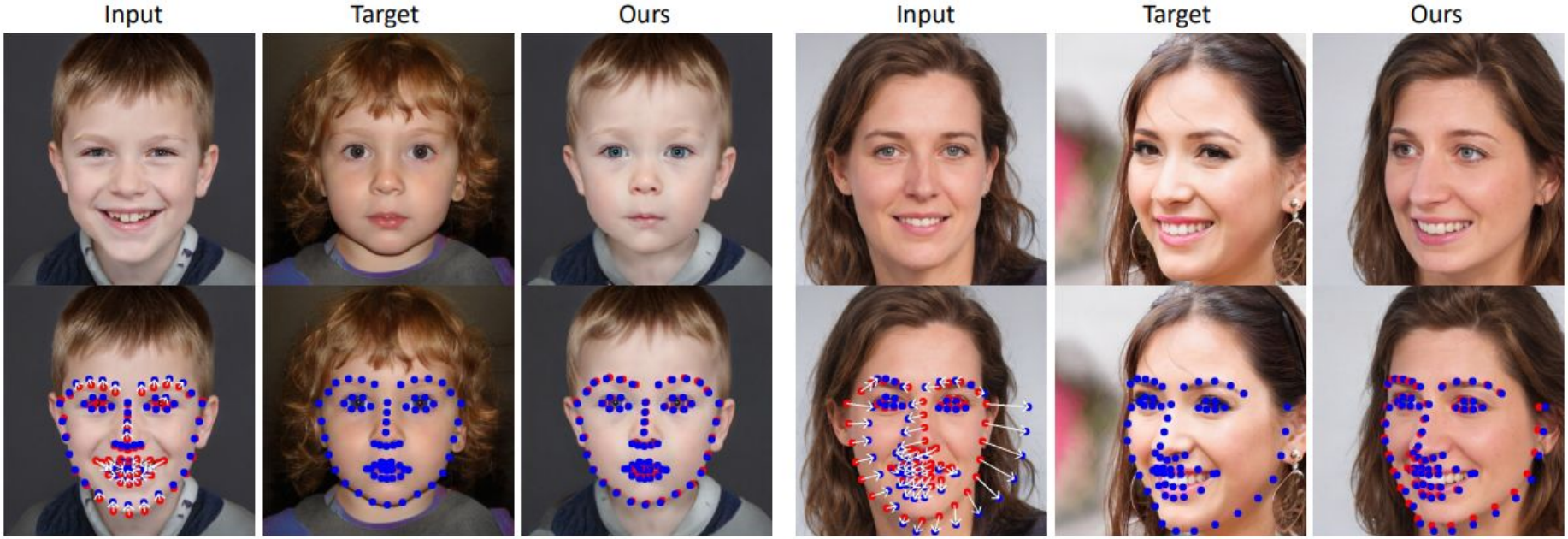
Real Image Manipulation



More Examples



More Examples



More Examples

Sources

- Drag Your GAN project
- Paper
- Web Demo