Transformer-based Flow Matching

Global Editing by Manipulating Direction

Contribution 1: We find an editing space named $u$-space that can perform controllable, accumulative and composable editing.

Contribution 2: A local-prompt for replacing, adding, removing, and rescaling prompts.

Algorithm 1 Semantic direction manipulation in $u$-space.

Input: parameter of pretrained network $\theta$, fix step ODE solver $\text{ODE}_f$ with step $N$, adaptive ODE solver $\text{ODE}_a$, target attribute $k$.

Output: An edited image $x_t$

$x_0 \sim \mathcal{N}(0, I)$ a unit Gaussian random variable

$m = 0$

for $i = 1, 2, \ldots, M$

for $j = N, N-1, \ldots, 1$

$u_{t-1, j} = \text{ODE}_a(t_j, \theta)$; where $t_j = \frac{1}{N}$

end for

for $j = N, N-1, \ldots, 0$

Calculate $s^k_j$ by Equation (5)

end for

while $t_m \leq 1$

if $0 < t_m < t_f$ then

Interpolate semantic direction $s^k_j$ by Equation (8)

else

$X_{t-1, j} = \text{ODE}_a(t_j, s^k_j, t_m, \theta)$

end if

end while

Return $x_t$

Local Editing by Local-Prompt