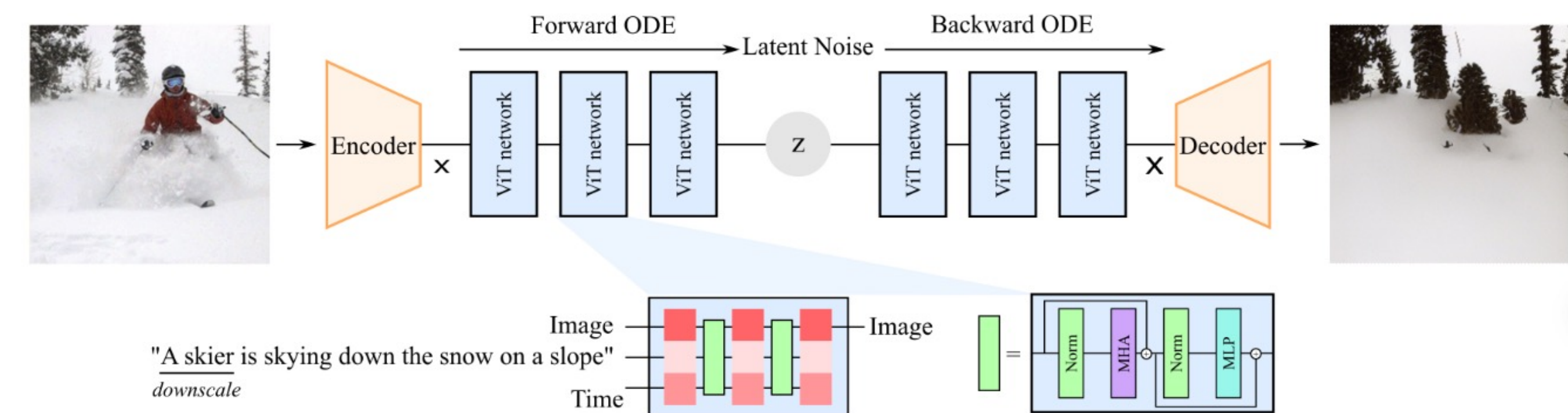## Transformer-based Flow Matching

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 θ, fix step ODE solver ODE_f with step N, adaptive ODE solver ODE_a, target attribute k.

Output: An edited image $\mathbf{x}_1$

$\mathbf{x}_0 \sim \mathcal{N}(0, I)$ a unit Gaussian random variable

$m = 0$

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

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

    $\mathbf{u}^i_{t_{j-1}}, \mathbf{x}_{t_{j-1}} = \text{ODE}_f(\mathbf{x}_{t_j}; \theta)$; where $t_j = \frac{j}{N}$

  end for

end for

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

  Calculate $\mathbf{s}^k_{t_j}$ by Equation (5)

end for

while $t_m \leq 1$ **do**

  if $0 < t < t_{\text{edit}}$ **then**

    Interpolate semantic direction $\mathbf{s}^k_t$ by Equation (8)

    $\mathbf{x}_{t_{m+1}}, t_{m+1} = \text{ODE}_a(\mathbf{x}_{t_m}, \mathbf{u}_{t_m} + w \cdot \mathbf{s}^k_{t_m}, t_m; \theta)$

  else

    $\mathbf{x}_{t_{m+1}}, t_{m+1} = \text{ODE}_a(\mathbf{x}_{t_m}, \mathbf{u}_{t_m}, t_m; \theta)$

  end if

end while

Return $\mathbf{x}_1$

Global Editing by Manipulating Direction

Guidance before K steps In 100 steps — K = 5, K = 10, K = 20, K = 30, K = 50, K = 100

Guidance every K steps in 100 steps — K = 20, K = 10, K = 5, K = 3, K = 2, K = 1

Editing should be performed in early timestep.

+Smile

+Smile +Black hair

+Smile +Black hair +Mustache

less — Male — More

less — Mustache — More
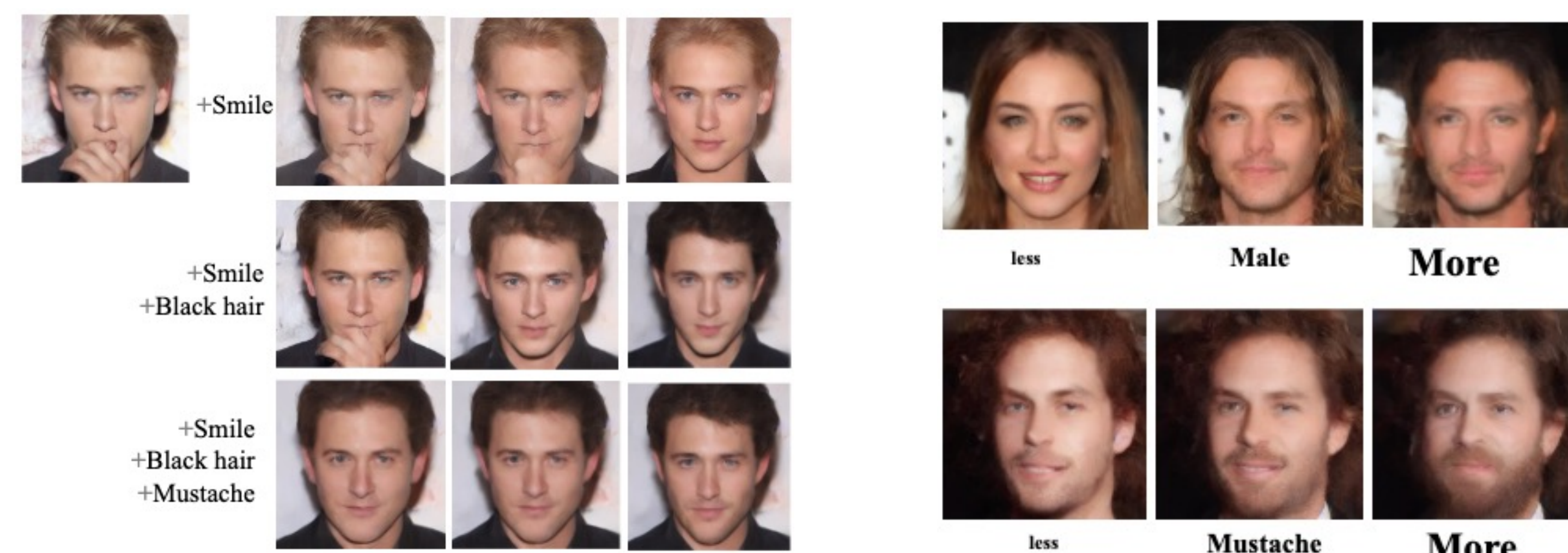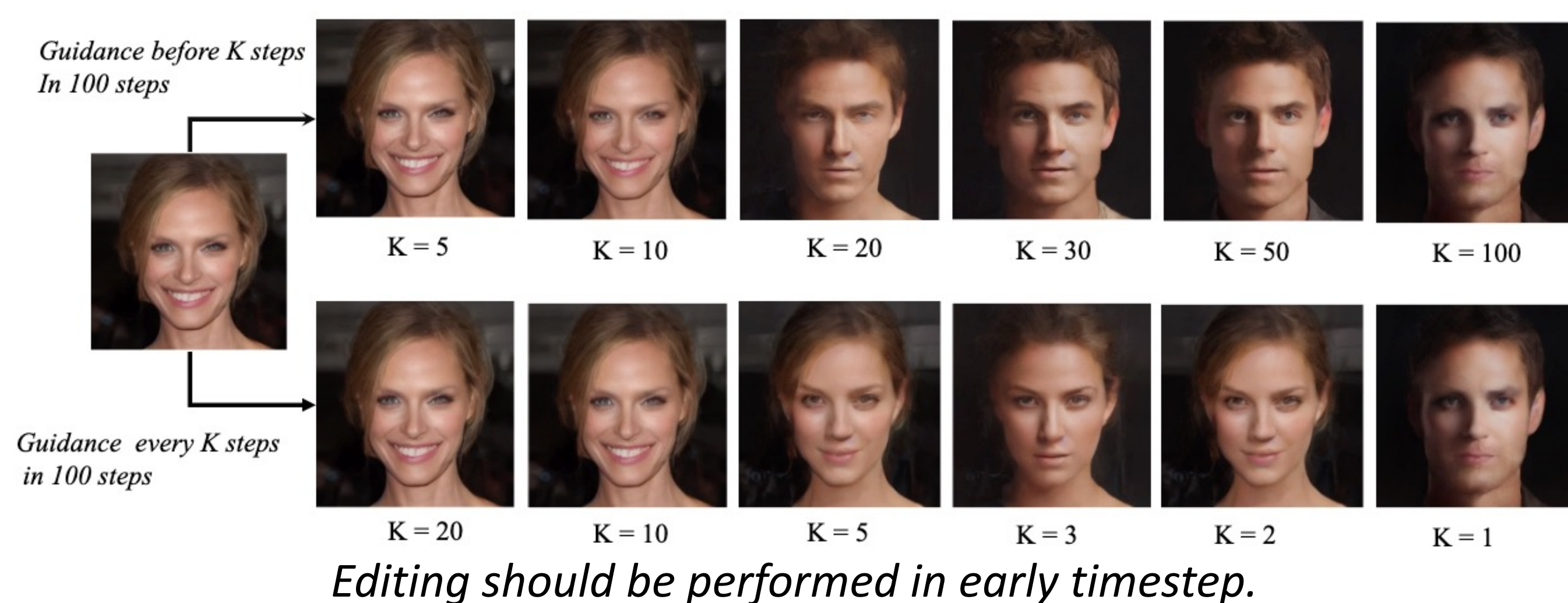
Editing can be conducted in a controllable, accumulative and composable manner.

Local Editing by Local-Prompt

The man has brown hair, He has a mustache, he is smiling, he has bushy eyebrows, he has pale skin.
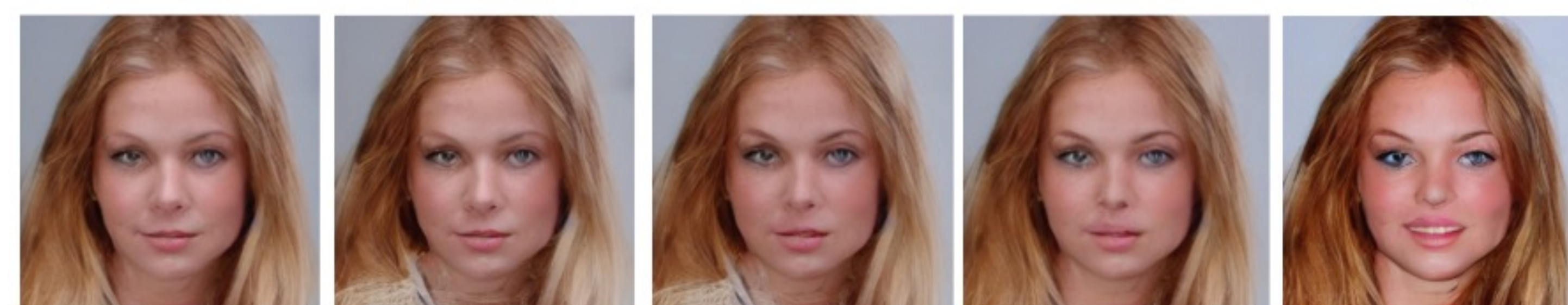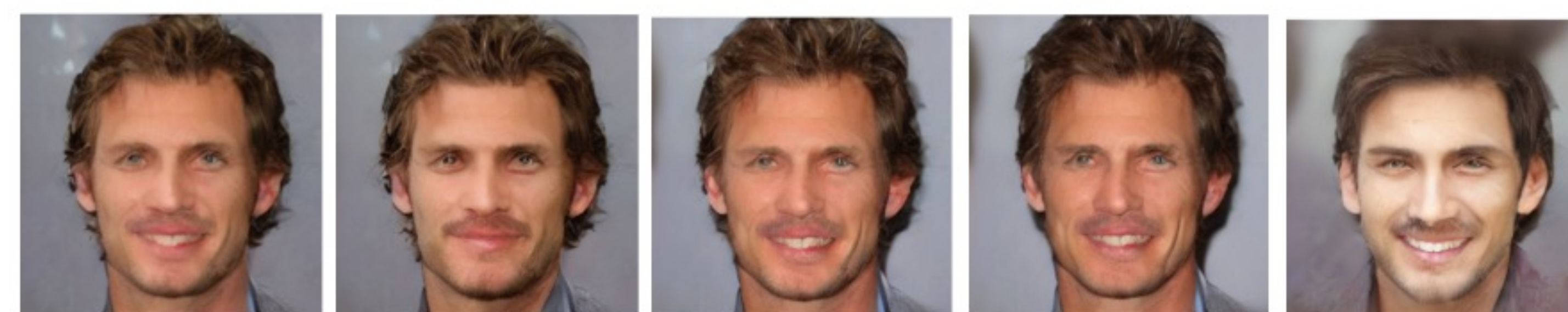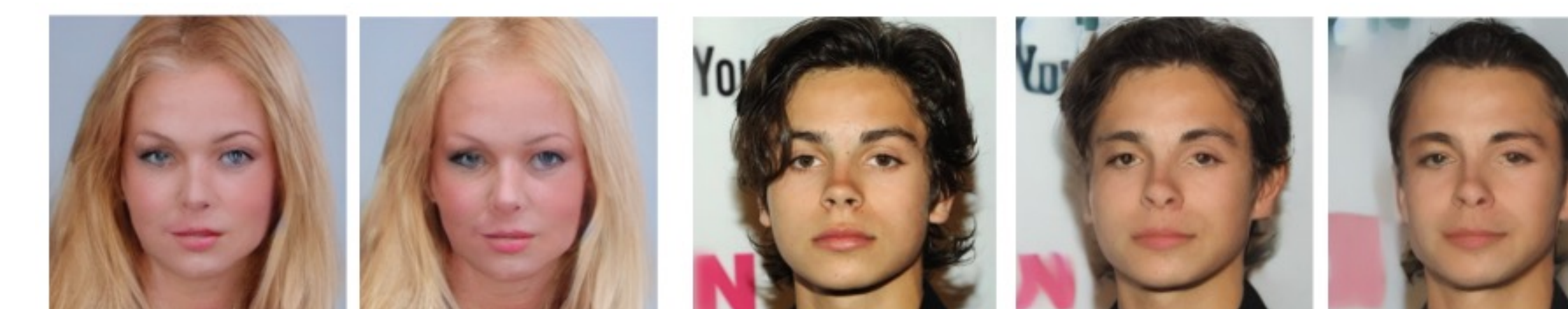
She has brown hair, she is young, she has mouth slightly open, she has a pointy nose, she has rosy cheeks.
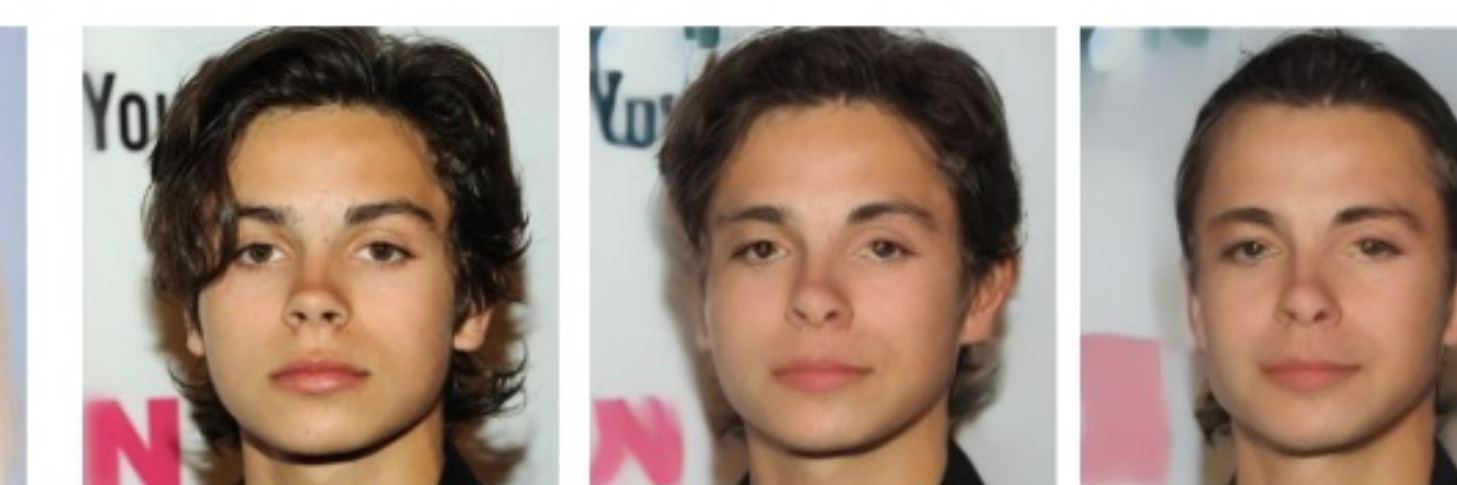

The man has brown hair, He has a mustache, he is smiling, he has bushy eyebrows, he has pale skin.
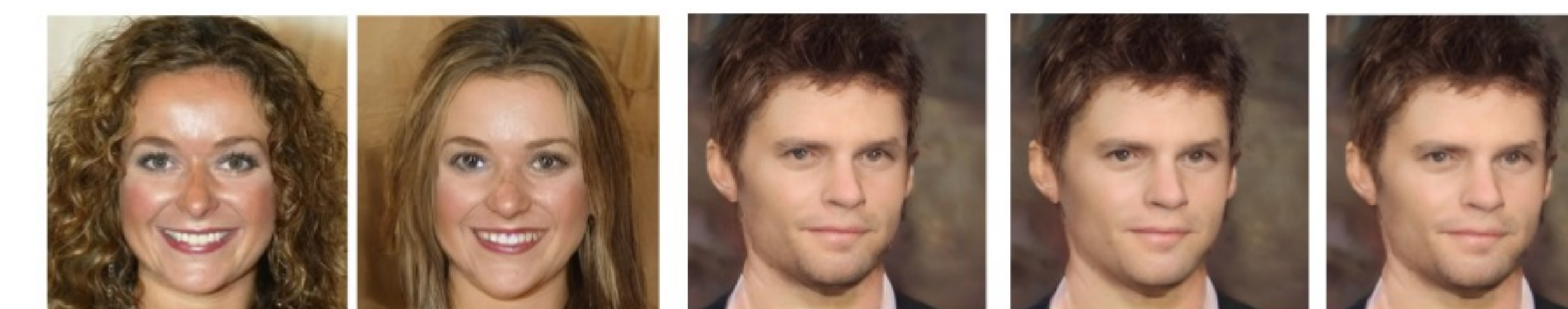
He is smiling, He has a mustache, The man has brown hair, he has bushy eyebrows, he has pale skin.

She has **white**(🔄) hair

This person is young and has wavy **hair**(🔄), bushy eyebrows

This person has **wavy**(long ✏️) hair, rosy cheeks, high cheekbones.

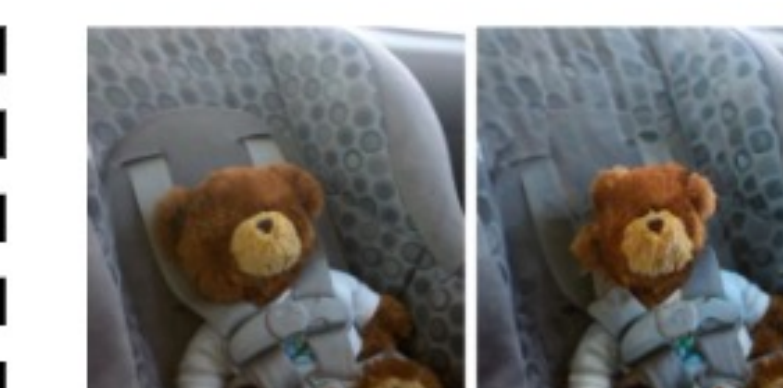A man with a wavy hair, he has a **beard**(❌).

A man with a wavy hair, **he has**(❌) a beard

A **stop sign**(🔄) is sitting on the side of a rural road.

A **dog**(🔄) that is in the air with a frisbee.

Stuffed(🔄) teddy bear strapped into child safety seat.

A **large pile**(🔄) of oranges lying next to some apples.

People walking(🔄) under a large clock and people sitting under a large umbrella

A **train**(🔄) coming out of a bridge next to a river.