# Semantic Image Editing

Adéla Šubrtová

# Introduction

- 2nd year PhD student
- Visual Recognition Group at FEE CTU
- Supervisor: Jan Čech, Supervisor specialist: Jiří Matas
- Research interests:
  super-resolution, image synthesis, image editing
- Personal interests:
  Bouldering , Reading, Psychology

# Outline

- Intro to semantic image editing
- Generative models
- GANs
- Editing using GANs
- Project: Hairstyle Transfer between Face Images
- Project: Chunky GAN: Real Image Inversion via Segments
  - Image Inversion Methods
  - ChunkyGAN method
- Conclusion

# What is Semantic Image Editing?



„Less Angry“

# Generative models

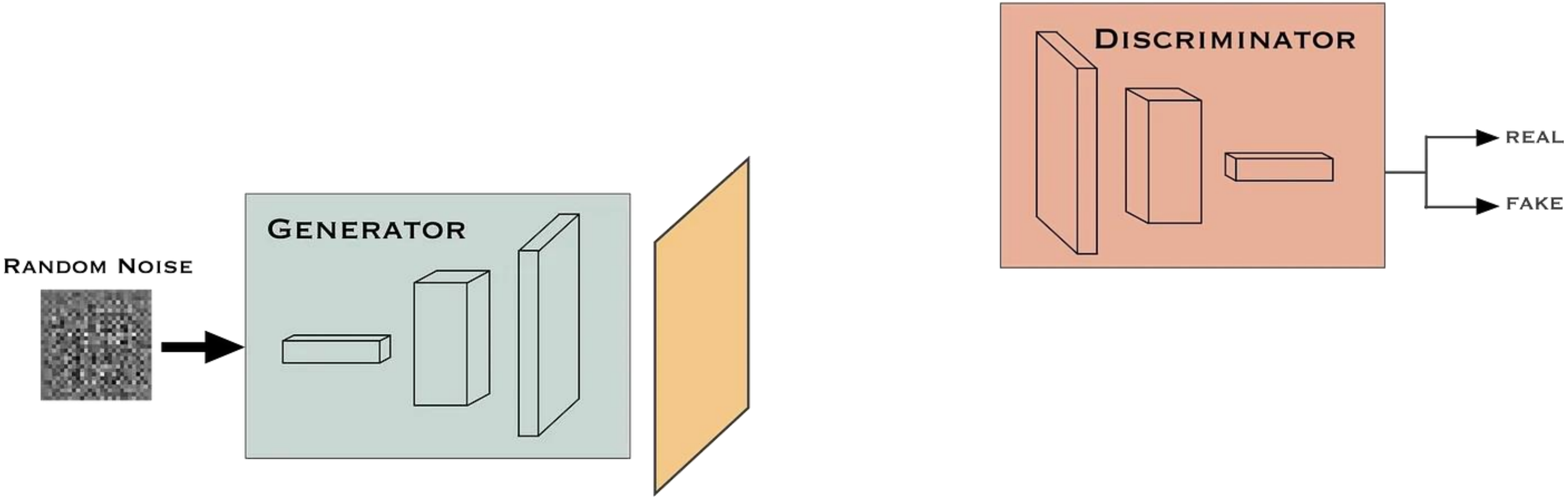
- Aim to model our data distribution p(x)
- Transform distribution that we can sample from to a complex one
- Examples
  - Variational Autoencoders (VAE)
  - **Generative Adversarial Networks (GAN)** – StyleGAN, BigGAN
  - Diffusion models – DALLE 2, Imagen, Stable Diffusion

„teddy bears mixing sparkling chemicals as mad scientists in a steampunk style"

1024 x 1024 px

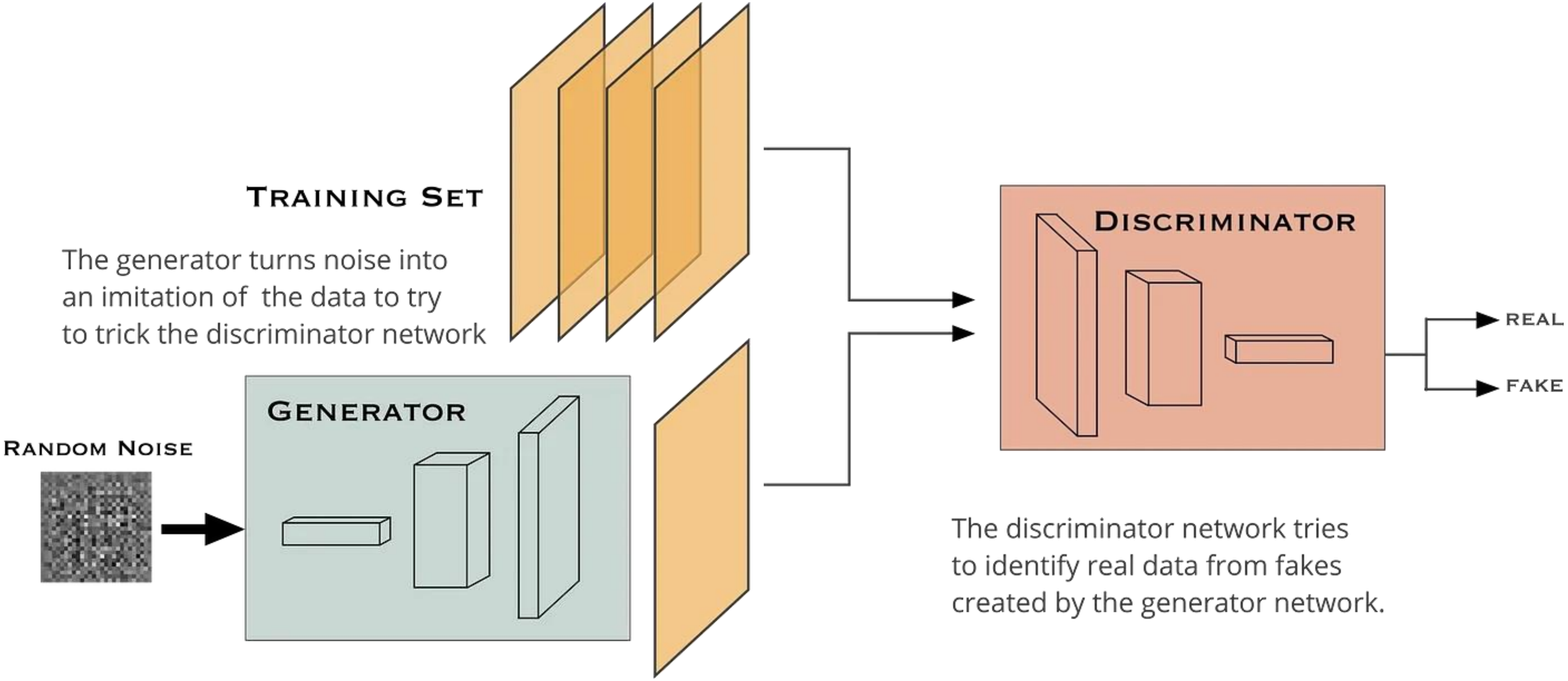
512 x 512 px

# Generative Adversarial Networks (GAN)




Goodfellow et al. "Generative Adversarial Nets“, Advances in Neural Information Processing Systems, 2014

# GAN Training




Source: https://towardsdatascience.com/image-generation-in-10-minutes-with-generative-adversarial-networks-c2afc56bfa3b

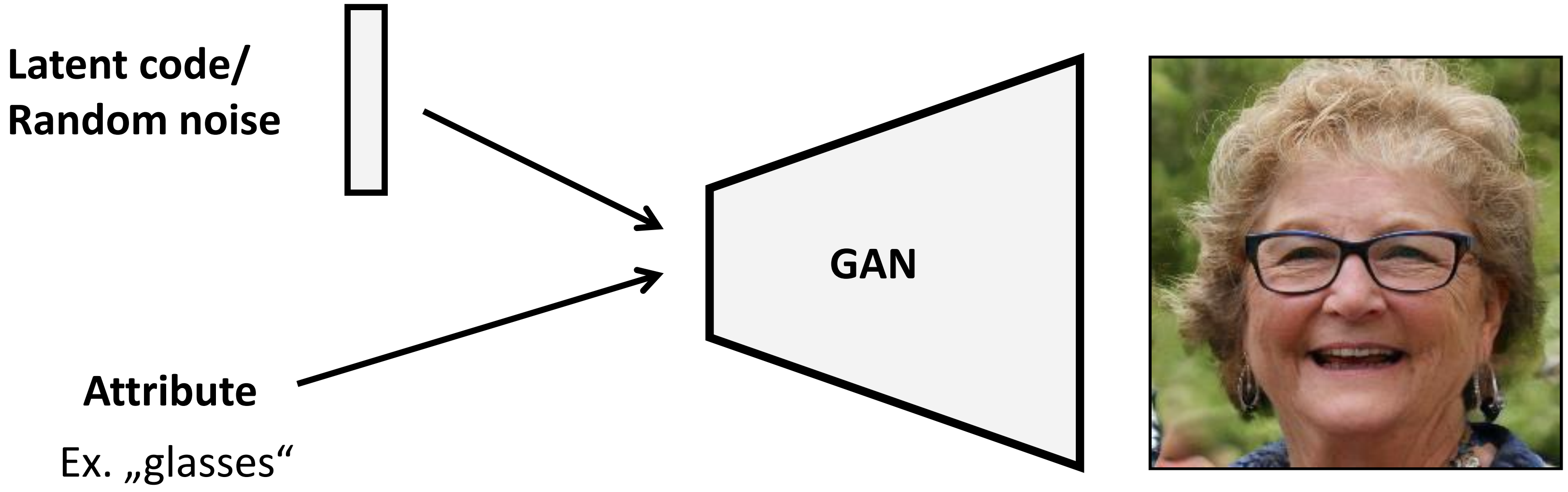
# GAN Training

$$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))]$$

- Alternating optimization of G and D
- Training is unstable – mode collapse, vanishing gradient
- Many works improve/stabilize the training

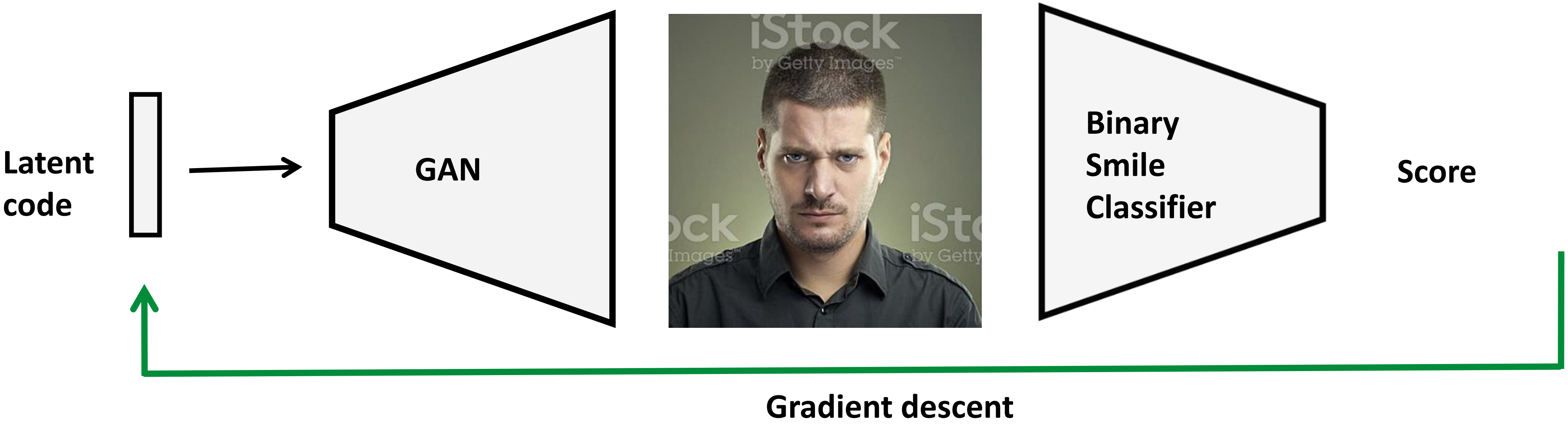
# GAN-based image editing

1. Conditioned GANs on semantic segmentation labels or other attributes



# GAN-based image editing

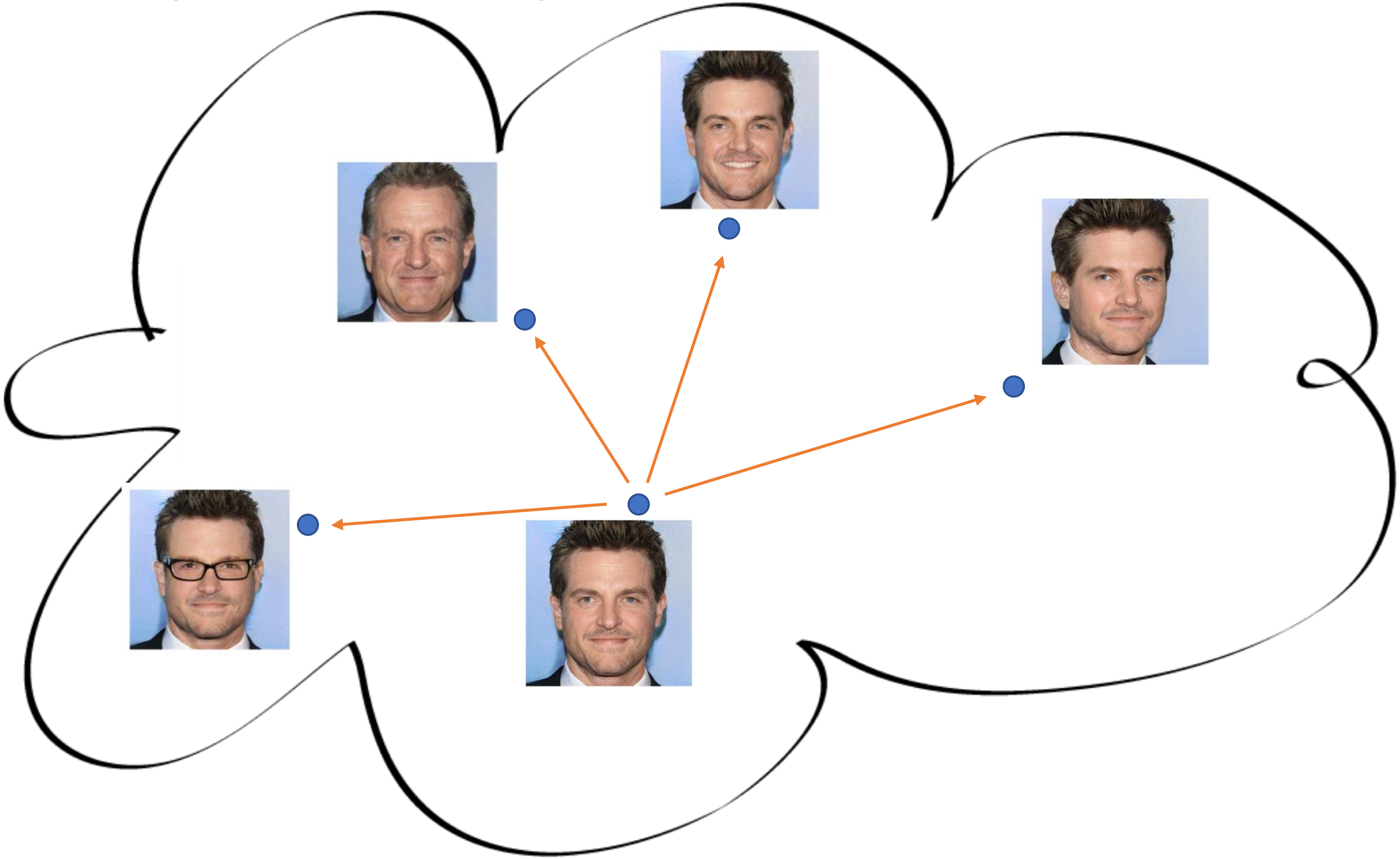
2. Attribute-classifier-guided manipulation



# GAN-based image editing

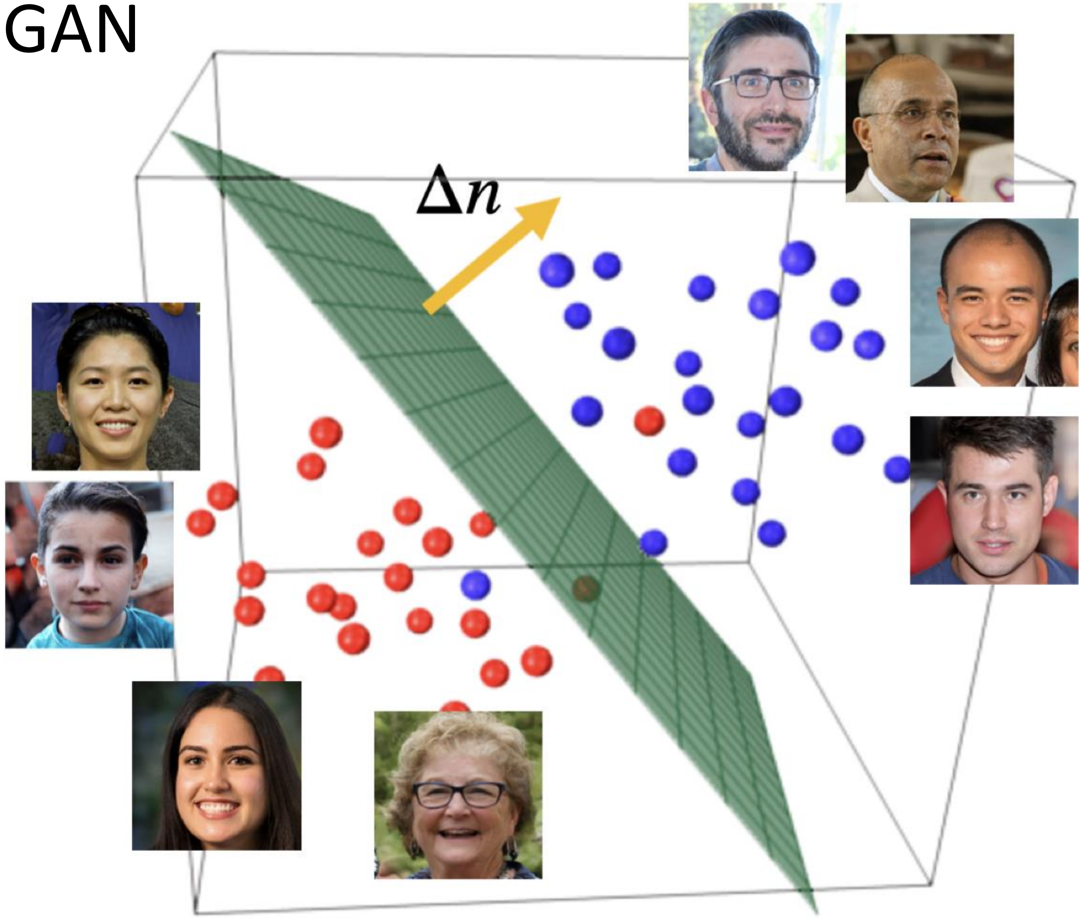
3. Latent space dissection – finding interpretable disentangled directions in the latent space of a pre-trained GAN

# Latent space manipulation



# GAN-base image editing

- Latent space dissection – finding interpretable disentangled directions in the latent space of a pre-trained GAN




**Supervised**

- Binary classification / regression in the latent space

**Unsupervised**

- PCA in the latent space
  + interpretation of the directions

# Hairstyle Transfer between Face Images

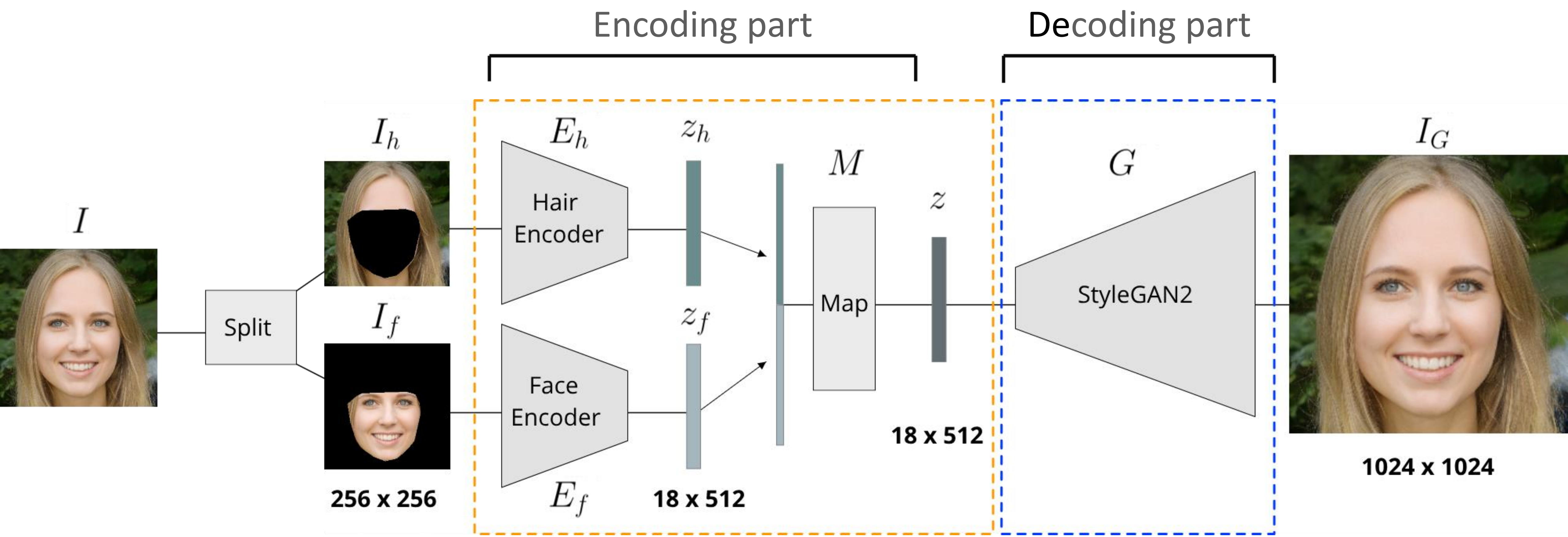
A. Šubrtová, J. Čech, V. Franc, Hairstyle Transfer between Face Images,
In Proc. IEEE Automatic Face and Gesture Recognition, 2021

# Are you thinking about getting a new haircut?

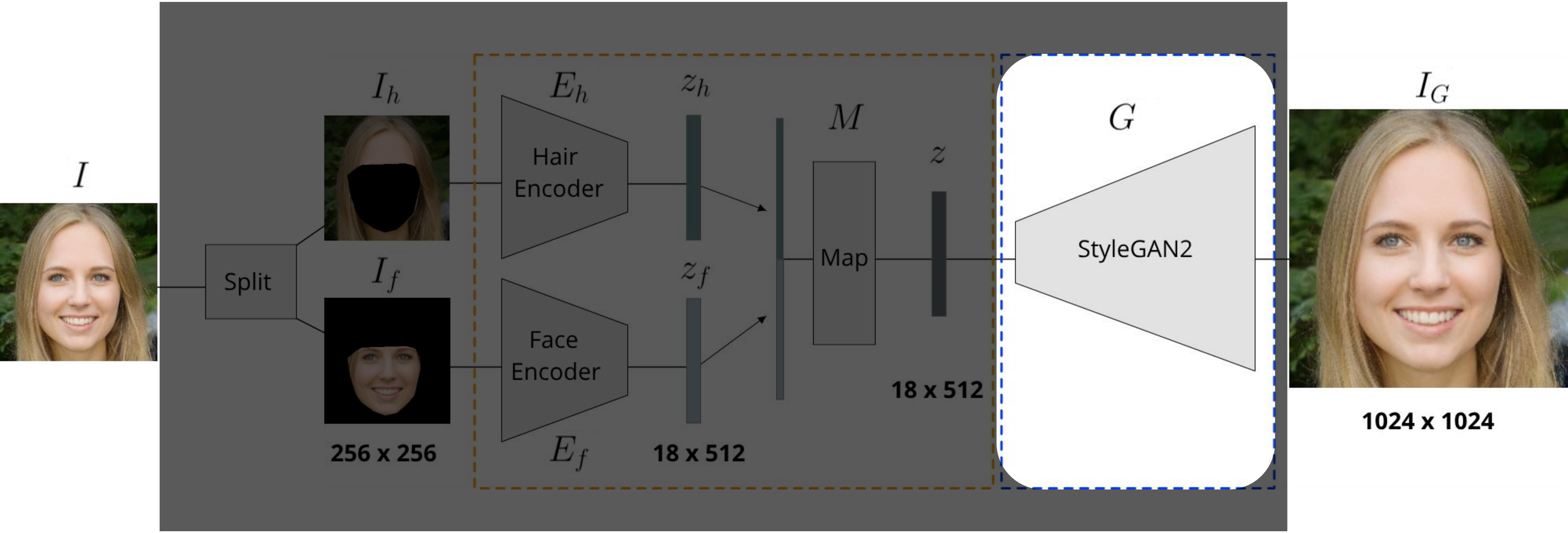
# Hairstyle Transfer between Face Images

- Fully automatic, including hair and face region separation
- Handles different pose and illumination

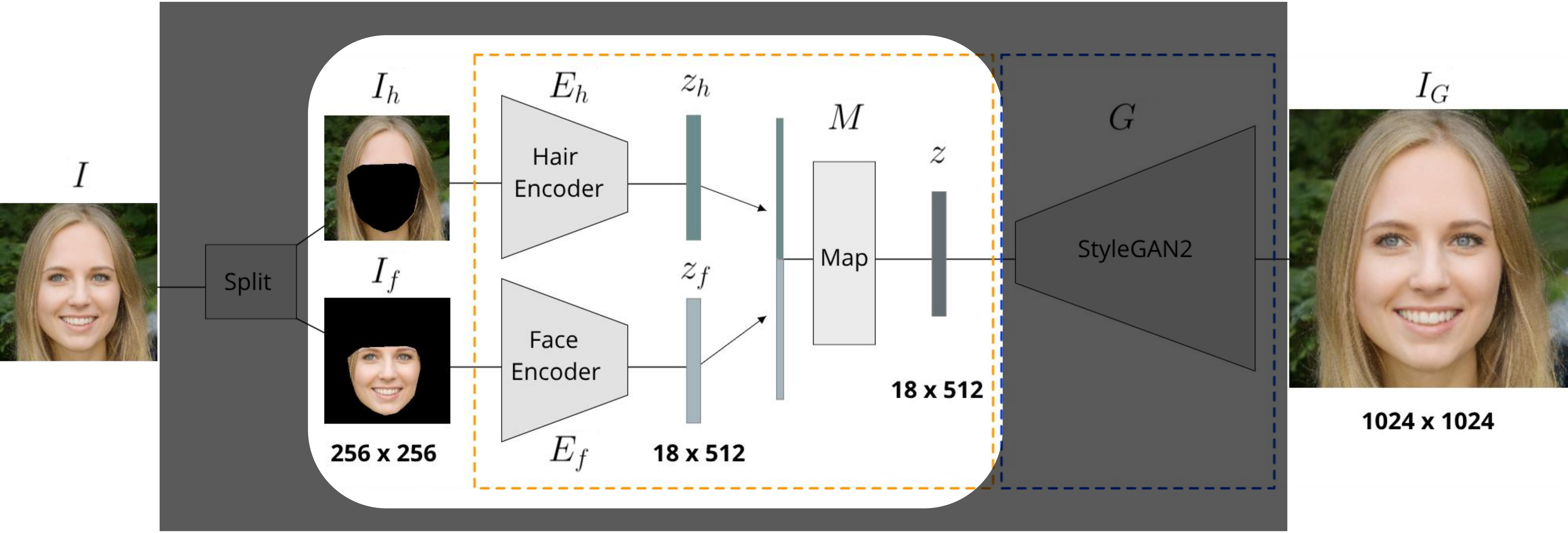
# Architecture



# Architecture



# Architecture



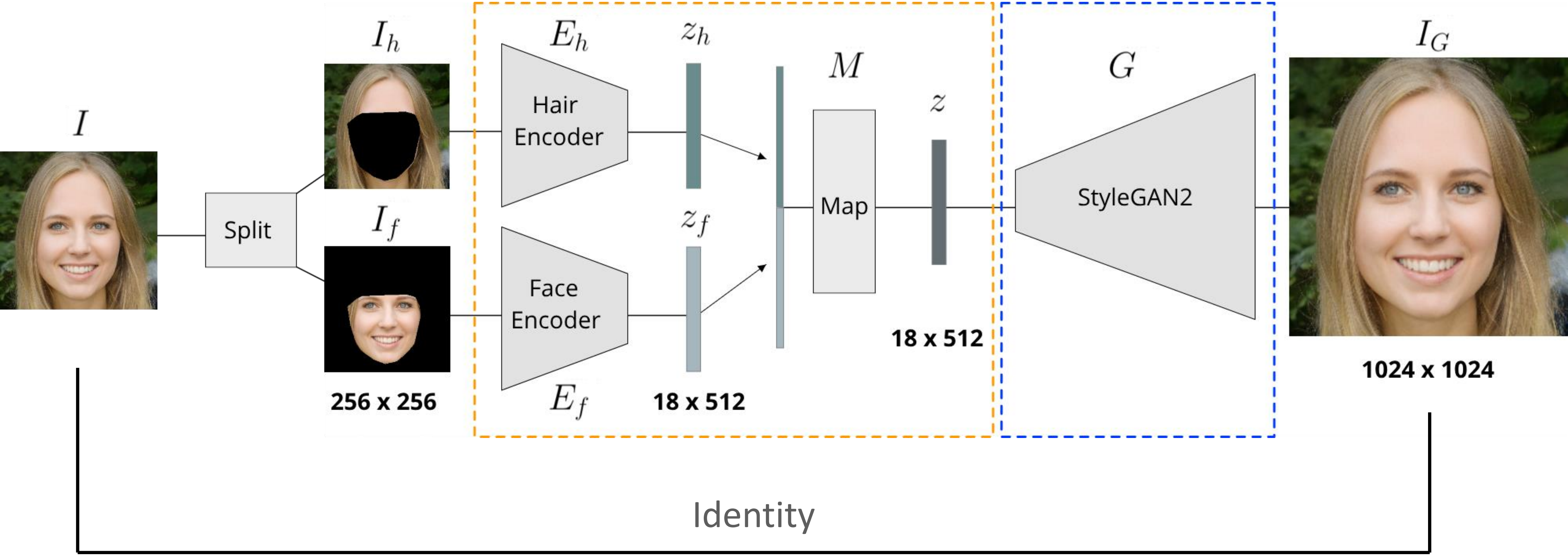
# Synthetic Dataset

- Generatived by pre-trained GAN (StyleGAN2)
- Photorealistic face images with high resolution (1024 x 1024 px)
- Potentially unlimited source of training data

Randomly generated images

Karras et. al, Analyzing and Improving the Image Quality of StyleGAN, CVPR, 2020

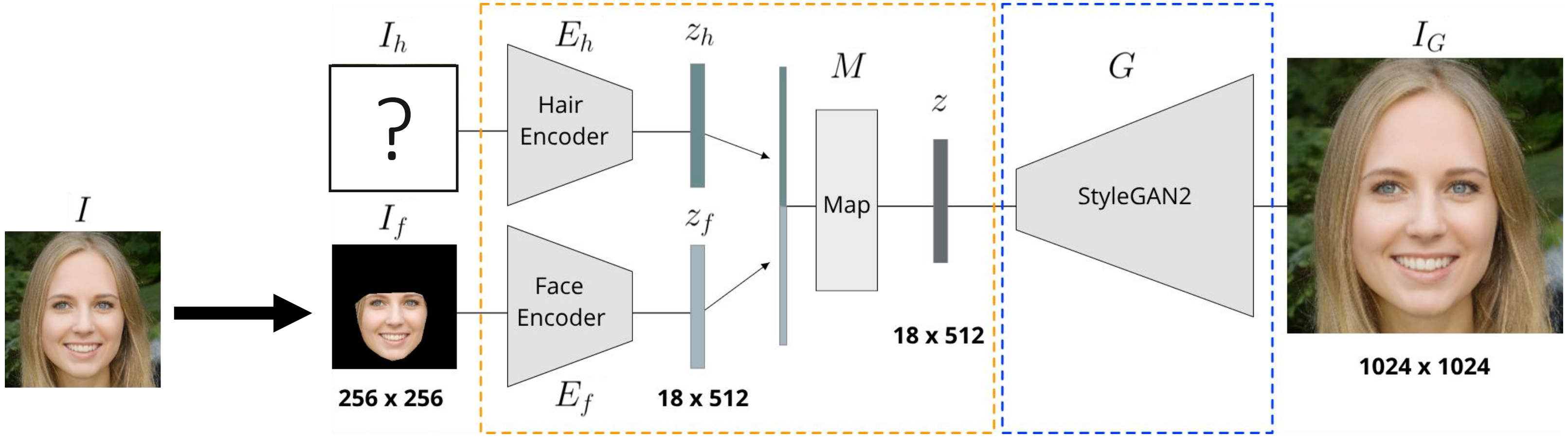
# Training - baseline



# Baseline Training – problems with alignment

# Path to Hairstyle Transfer

- Goal: Hairstyle transfer for **unaligned** pair of images
- Hair encoder should ignore the geometry
- Train from images with different pose

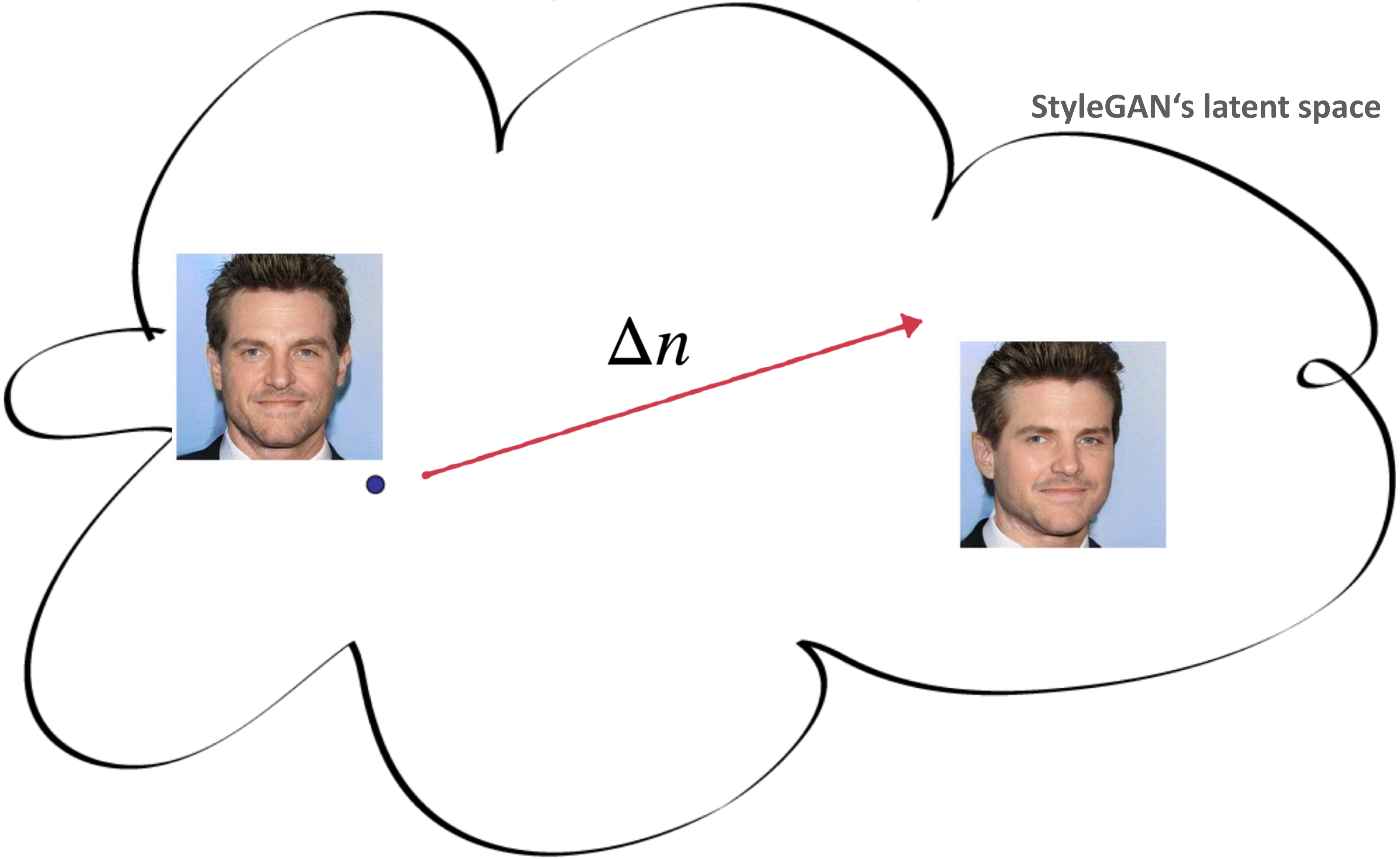


# Improving the Dataset

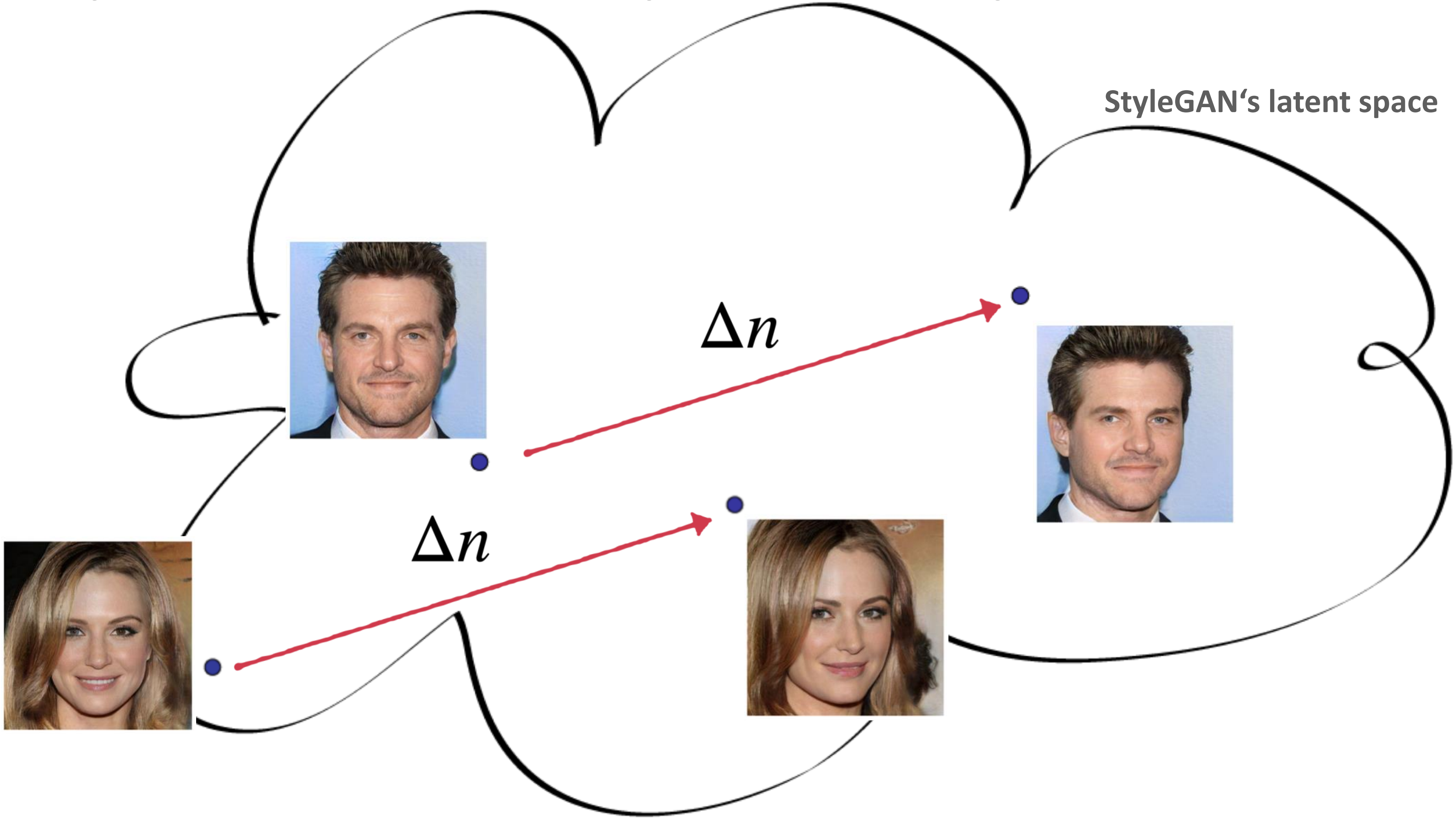
- Where to get misaligned pairs of images with the same hair?

**Latent Space Manipulation…**

# StyleGAN: Latent space manipulation

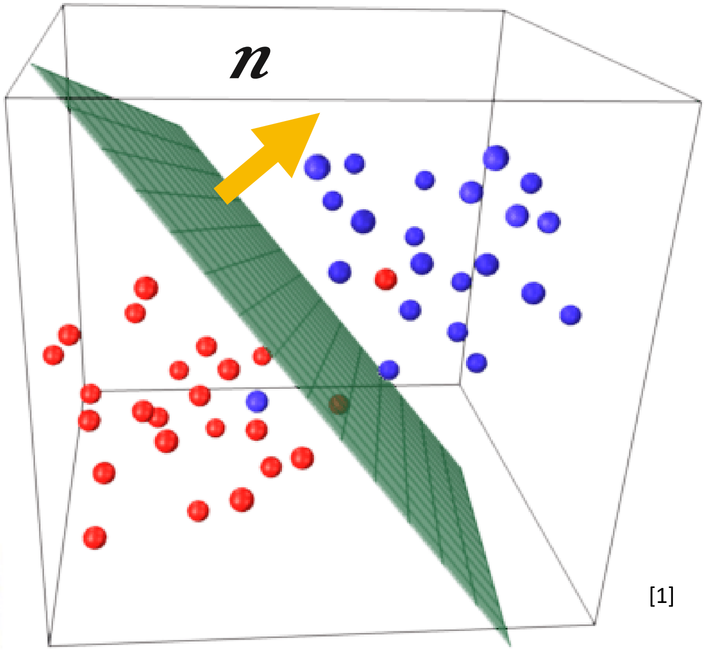


# StyleGAN: Latent space manipulation



# Finding yaw direction in face space of StyleGAN

- Direction estimation ~ classification in the latent space
- 2 classes: faces looking to the left and right
- $z$ – code of a face image

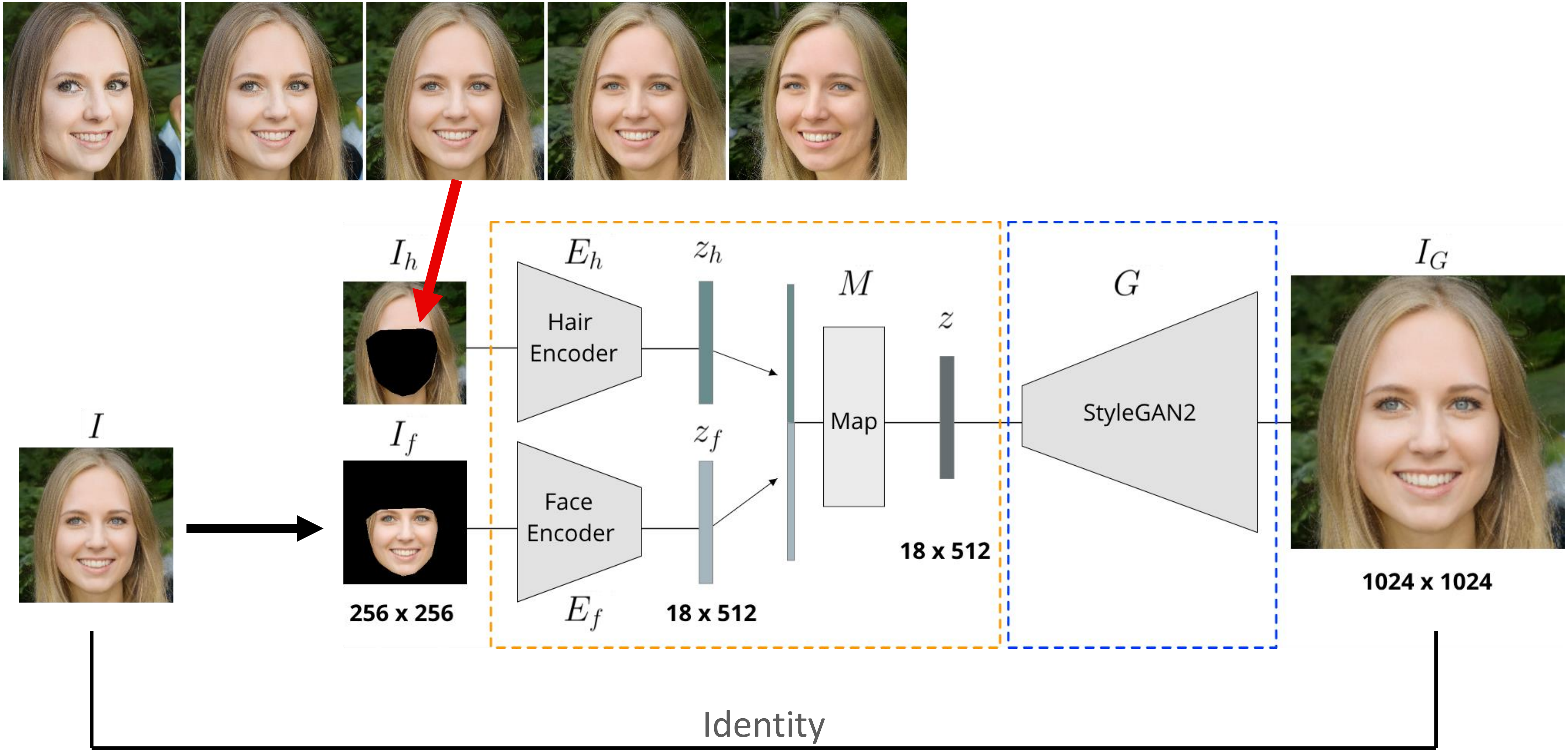



[1]

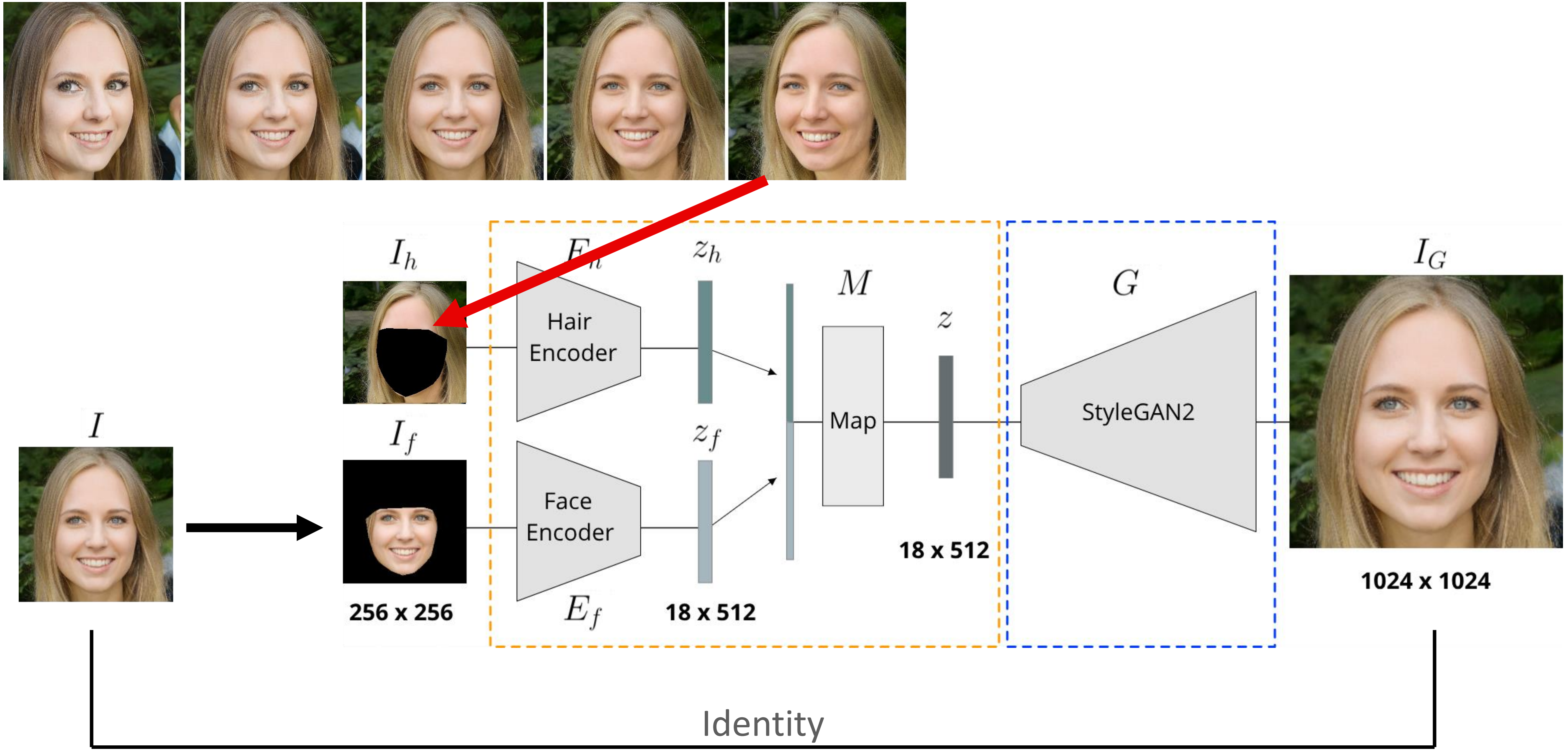
$z - 2k \cdot n$ $z - k \cdot n$ $z$ $z + k \cdot n$ $z + 2k \cdot n$

[1] https://appliedmachinelearning.files.wordpress.com/2017/03/svm_logo1.png

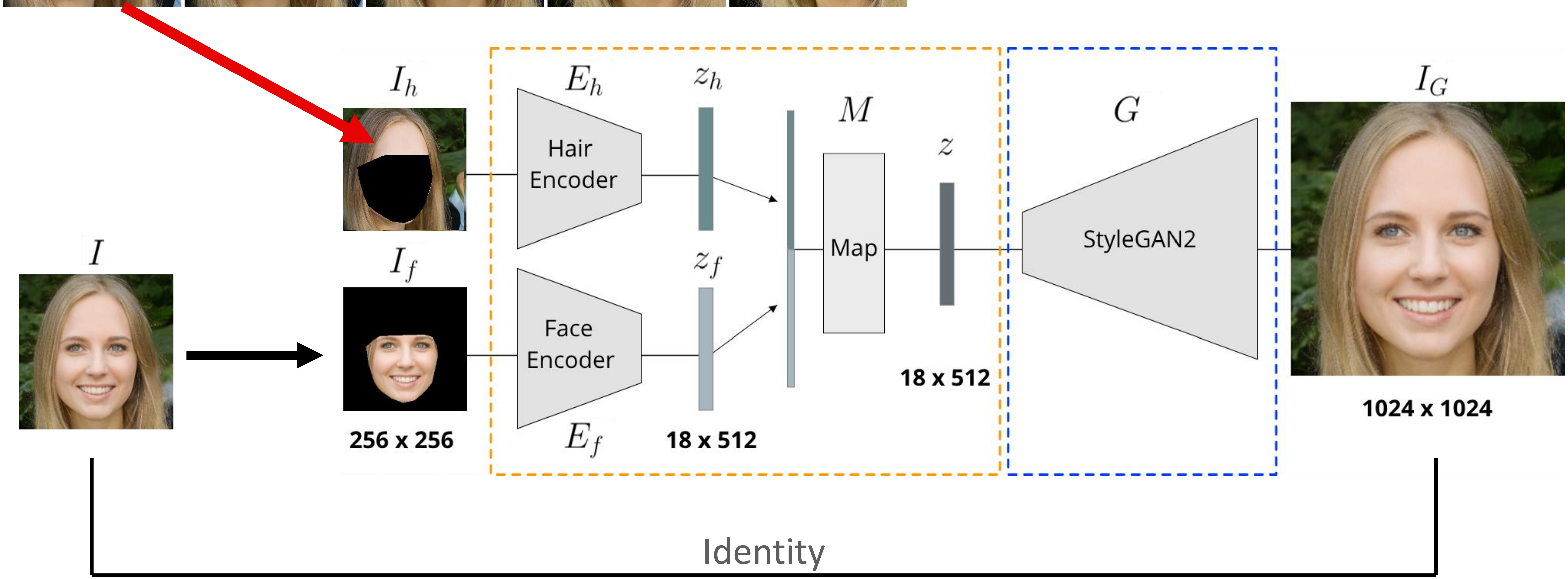
# Improved Training



# Improved Training

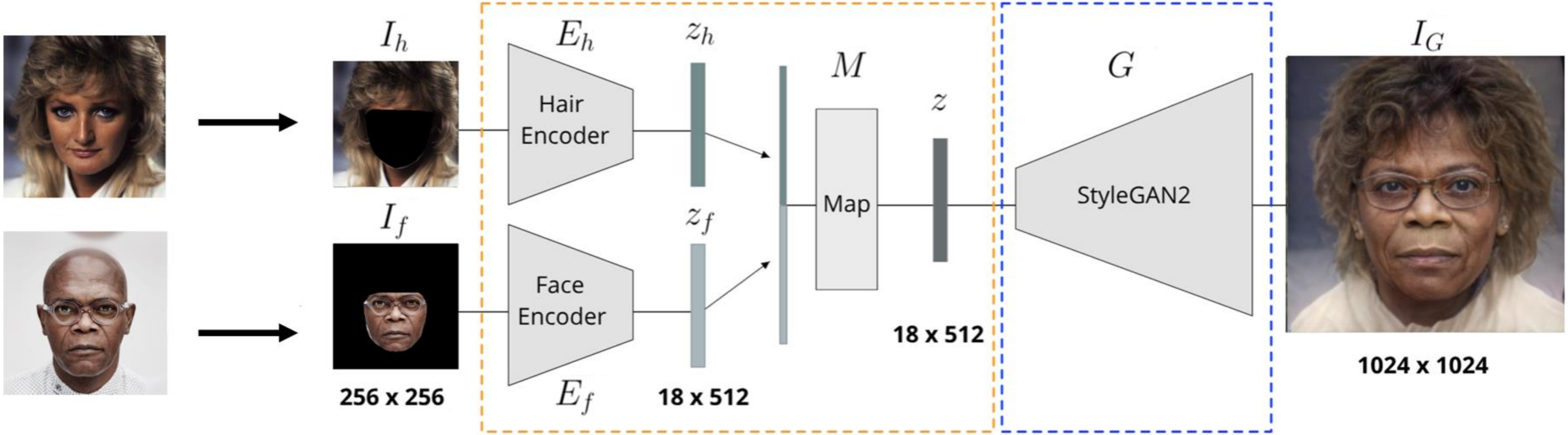


# Improved Training



# Experiments

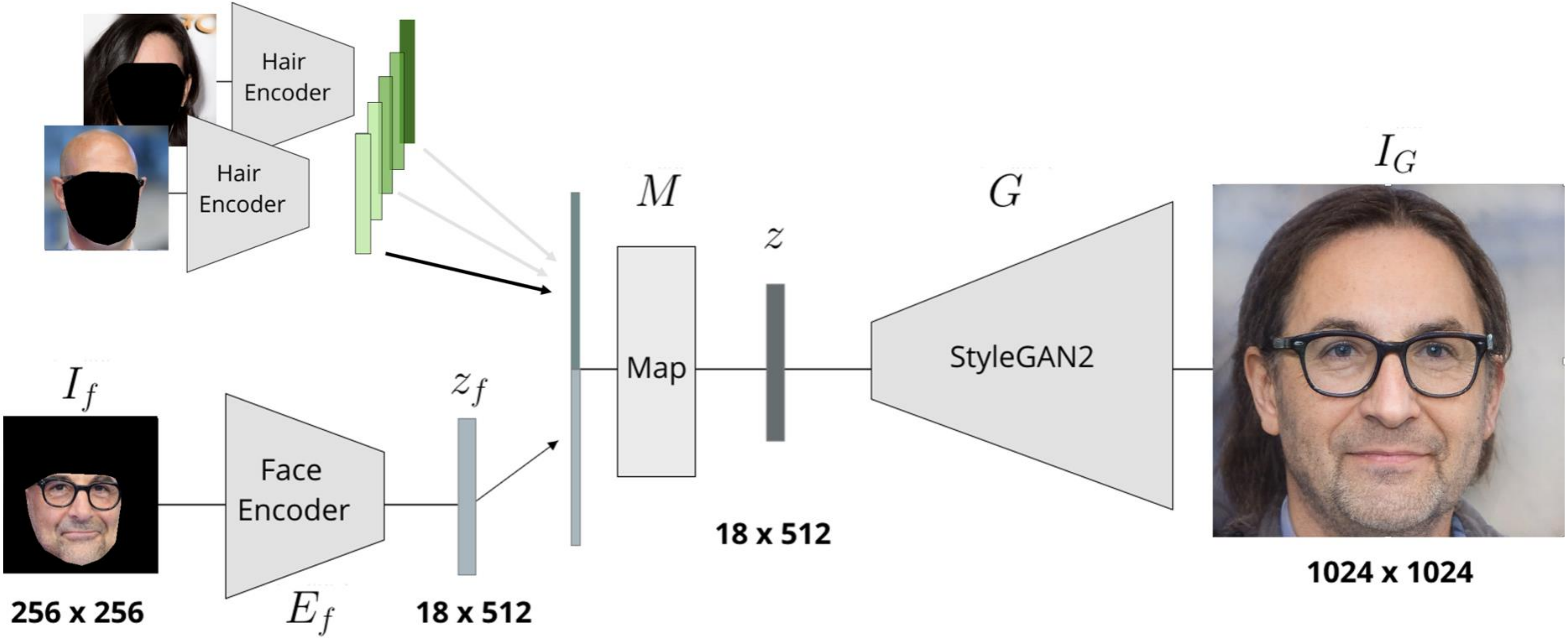
# Hairstyle Transfer



# Hairstyle Transfer

# Hairstyle Transfer

# Interpolation

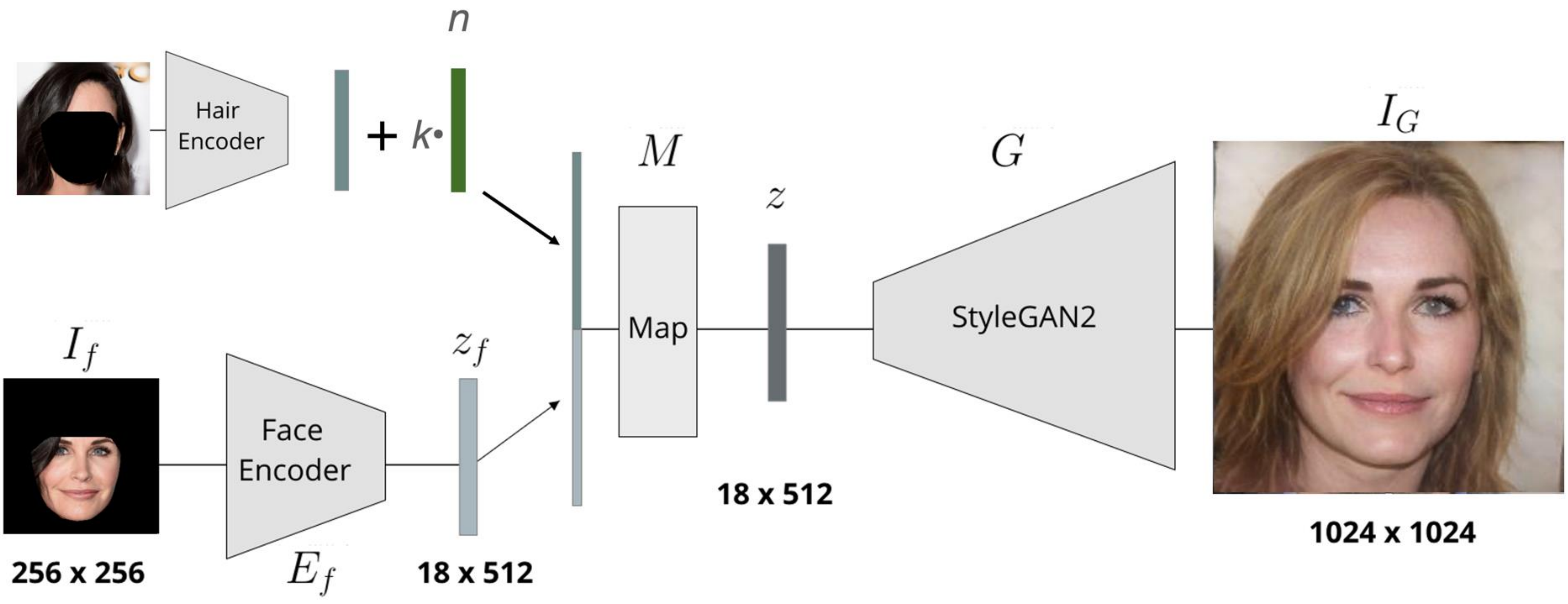


# Interpolation




# Hair Manipulation

- Same principle as in face rotation



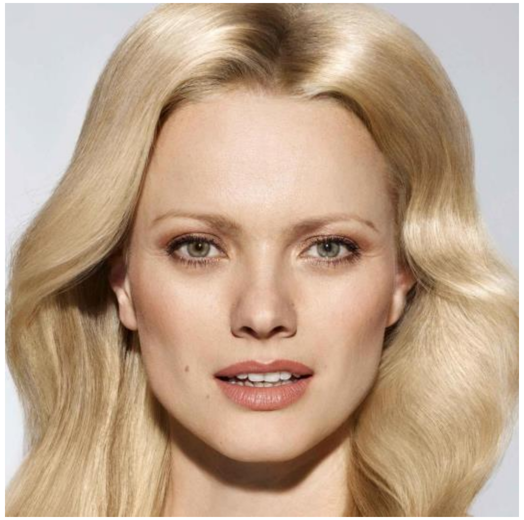
# Experiments - Quantitative

**Identity preservation assesment**

- 100k random images generated from StyleGAN
- Hairstyle transfer on random subset of 1k images
- Image retrieval on the full 100k dataset with the hairstyle-transfered images as a queries
- Ranking based on cosine similarity of ArcFace descriptors
- In 98.3% the first ranked was correct.
- Average rank was 1.143 .

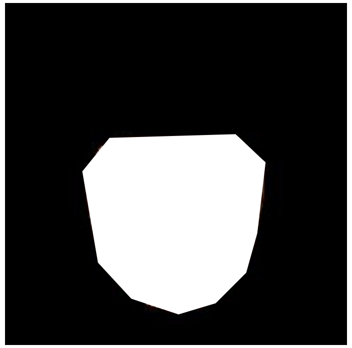
# Limitations – Identity preservation on real images

Hair



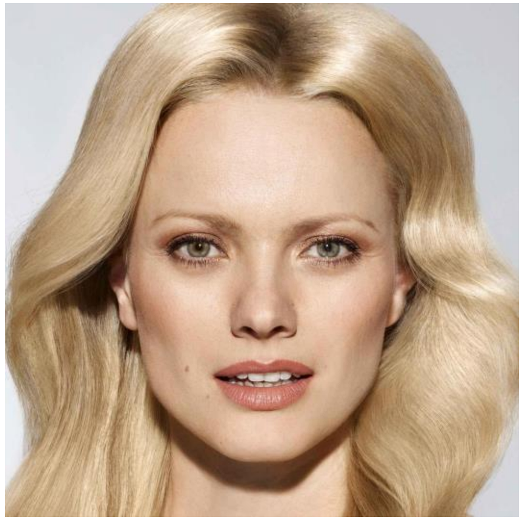
Identity

Generated



# Limitations – Identity preservation on real images

Hair



Identity

Generated

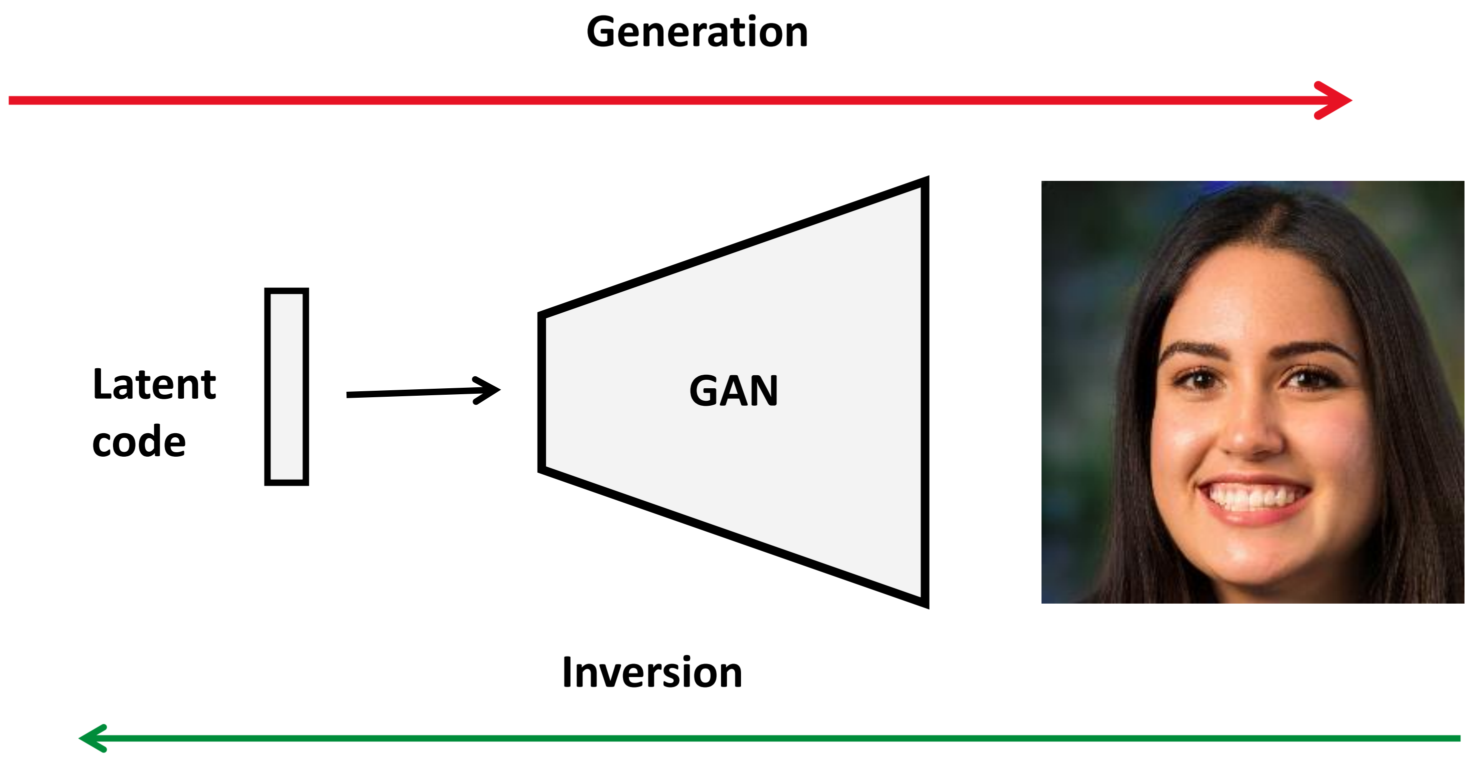
Alpha-blended

# Project page with demo

# ChunkyGAN: Real Image Inversion via Segments

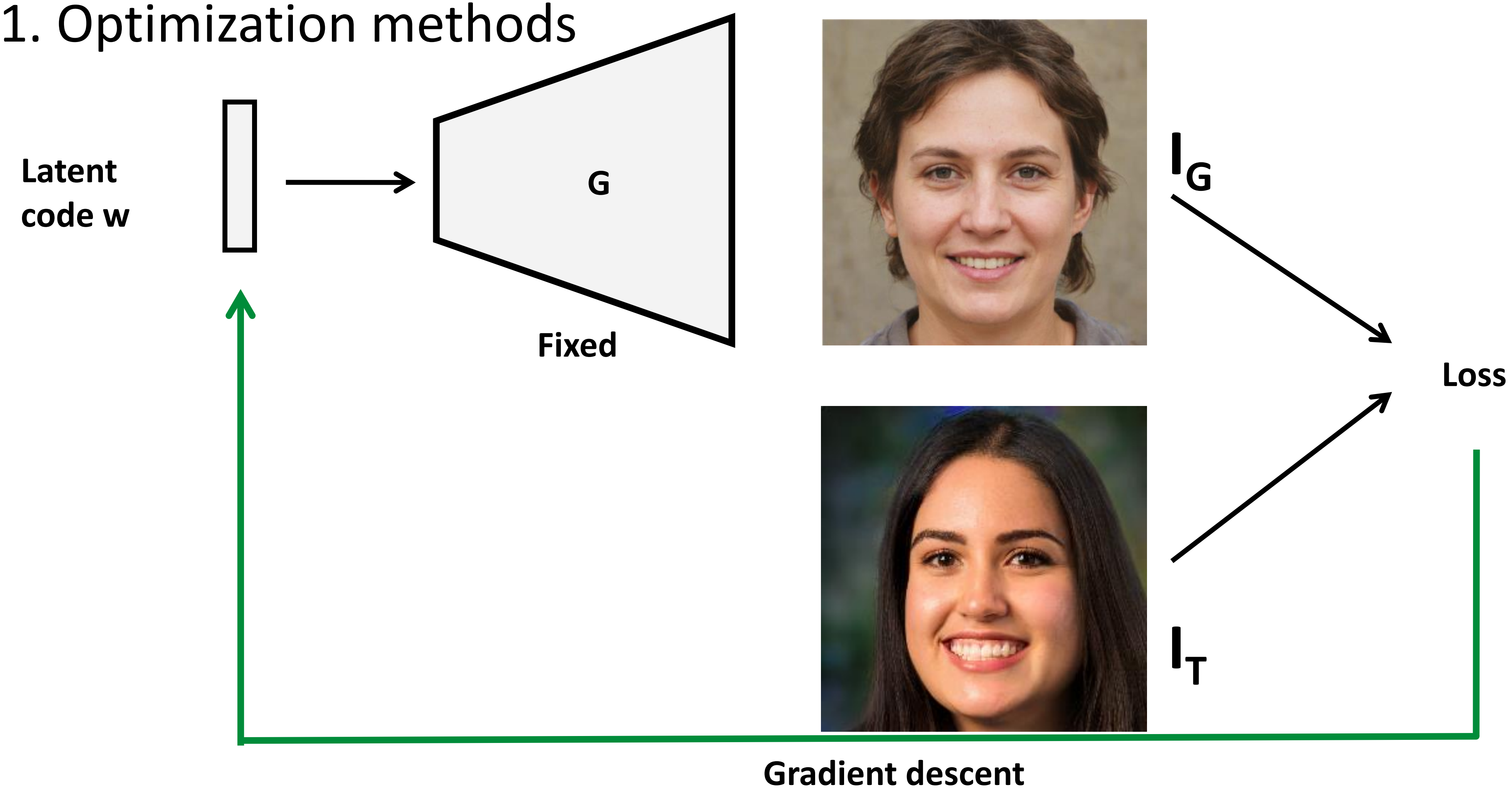
A. Šubrtová, D. Futschik, J. Čech, M. Lukáč, E. Shechtman, D. Sýkora, ChunkyGAN: Real Image Inversion via Segments, In Proc. ECCV, 2022

# Editing Real Images using GANs

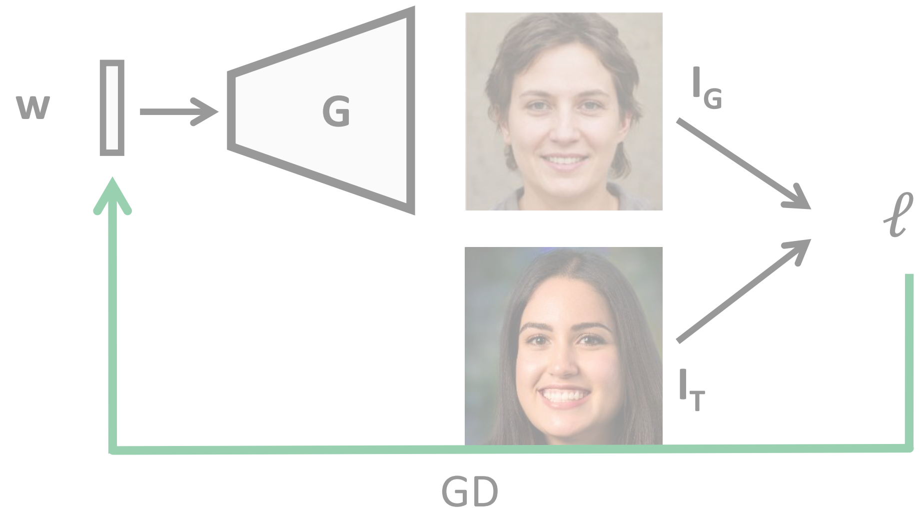


# Image Inversion Methods

- 1. Optimization methods



# Image Inversion Methods




- 1. Optimization methods

$$\hat{w} = \underset{\boldsymbol{w}}{\operatorname{argmin}}\, \ell(IT, G(w))$$

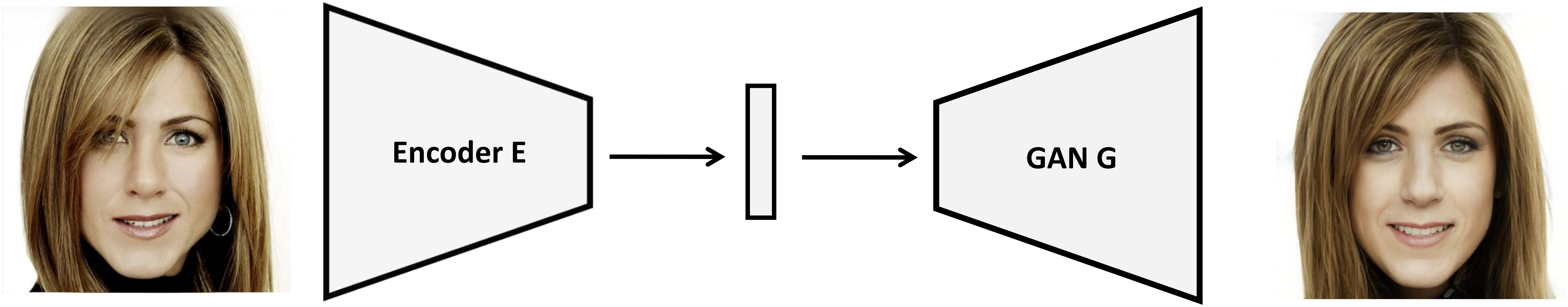
$$\ell = \text{combination of perceptual loss [1] and identity loss [2]}$$

[1] Zhang et. al., The Unreasonable Effectiveness of Deep Features as a Perceptual Metric, CVPR, 2018

[2] Deng et. al., ArcFace: Additive Angular Margin Loss for Deep Face Recognition, CVPR, 2019

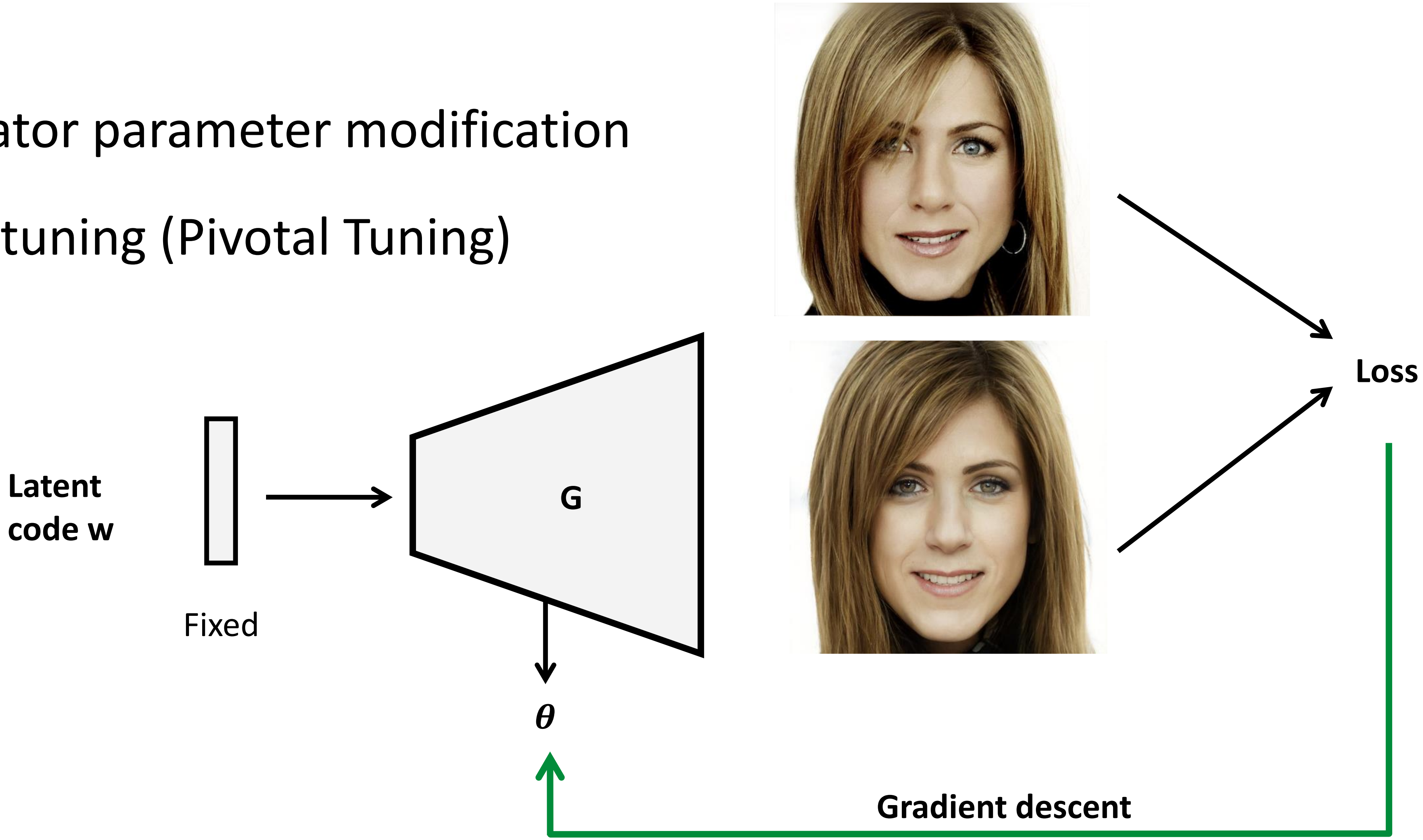
# Image Inversion Methods

- 2. Encoder-based methods



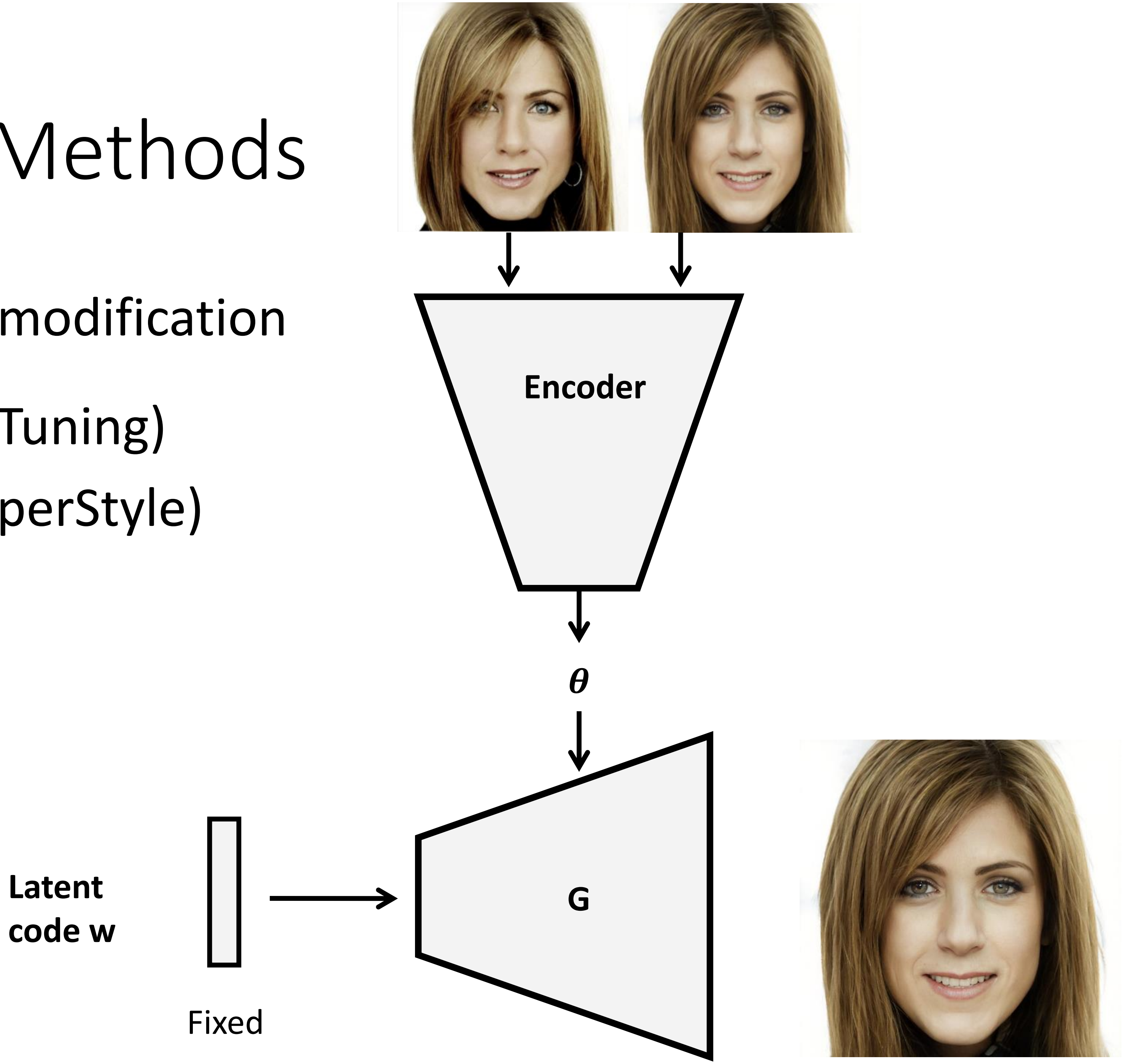
# Image Inversion Methods

- 3. Generator parameter modification
    - Finetuning (Pivotal Tuning)



# Image Inversion Methods

- 3. Generator parameter modification
    - Finetuning (Pivotal Tuning)
    - Hypernetworks (HyperStyle)



# Overview

**Optimization**

+ Accurate*
- Slower
- Poor editability

**Encoders**

+ Fast
+ Good editability
- Identity not well preserved

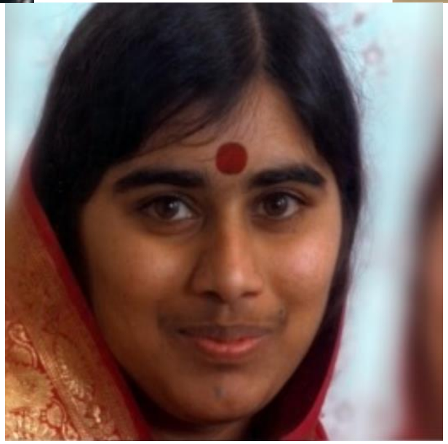
**Generator modification**

+ Accurate*
+ Good editability
- Need to store the weights

* For in-domain images

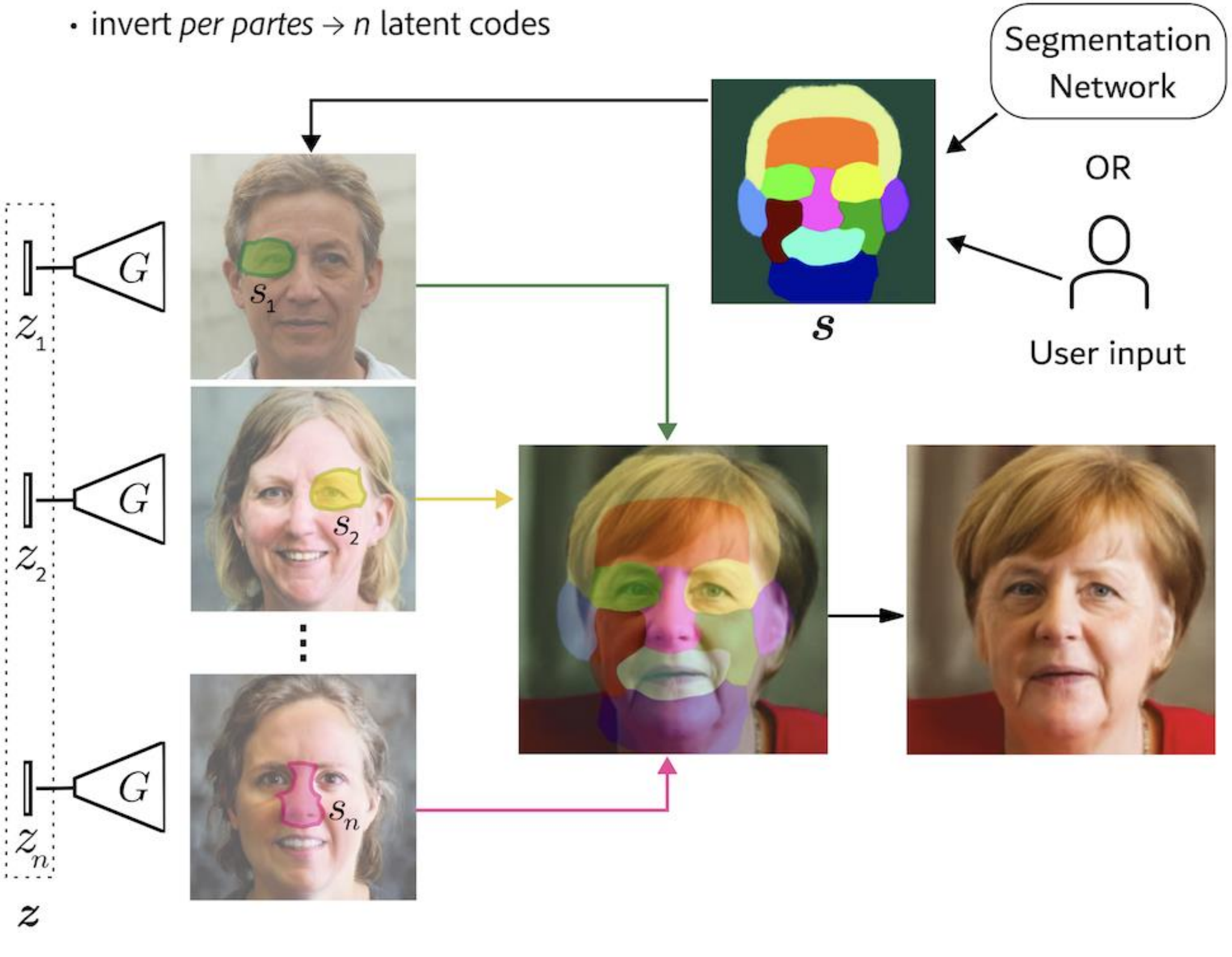
# Shortcomings

- Hard/Impossible to reconstruct unusual features ( face mask, bindi,…)



# ChunkyGAN - Method

- invert *per partes* → *n* latent codes




$$O(\boldsymbol{z}, \boldsymbol{s}) = \sum_{i=1}^{n} G(z_i) \cdot s_i$$

# Optimalization

- Optimization problem:

$$\min_{\boldsymbol{z},\boldsymbol{s}} \mathcal{L}_{LPIPS}\big(I,\ O(\boldsymbol{z},\boldsymbol{s})\big) + \lambda_{reg}\sum_{i=1}^{n} \parallel z_i - z_\mu \parallel^2$$

- Regularization → latent codes lie close to the natural image manifold

# Projection

# Results – inversion

# Inversion – difficult examples

# Quantitative evaluation of projection

| | Projection | LPIPS | Identity | $L_2$ |
|---|---|---|---|---|
| Optimization | $\mathcal{W}$ | $0.4190 \pm 0.0363$ | $0.1745 \pm 0.1328$ | $0.0725 \pm 0.0699$ |
| | Ours in $\mathcal{W}$ | $0.3697 \pm 0.0396$ | $0.1384 \pm 0.1117$ | $0.0481 \pm 0.0289$ |
| | $\mathcal{W}^+$ | $0.3675 \pm 0.0387$ | $0.1195 \pm 0.1047$ | $0.0436 \pm 0.0623$ |
| | Ours in $\mathcal{W}^+$ | $\mathbf{0.3194 \pm 0.0365}$ | $0.0937 \pm 0.0855$ | $0.0207 \pm 0.0151$ |
| | Ours in $\mathcal{W}^+$ reg. | $0.3330 \pm 0.0350$ | $\mathbf{0.0894 \pm 0.074}$ | $0.0217 \pm 0.0130$ |
| | $\mathcal{S}$ | $0.3577 \pm 0.0397$ | $0.1070 \pm 0.0965$ | $0.0328 \pm 0.0188$ |
| | Ours in $\mathcal{S}$ | $0.3572 \pm 0.0401$ | $0.1053 \pm 0.0928$ | $0.0319 \pm 0.0187$ |
| Encoders | e4e [9] | $0.4444 \pm 0.0418$ | $0.1912 \pm 0.1343$ | $0.0468 \pm 0.0165$ |
| | pSp [7] | $0.4433 \pm 0.0418$ | $0.1706 \pm 0.1182$ | $0.0351 \pm 0.0135$ |
| | ReStyle [2] - 5 iters | $0.4444 \pm 0.0430$ | $0.1900 \pm 0.1318$ | $0.0433 \pm 0.0162$ |
| Generator parameters modification | Pivotal Tuning [8] | $0.3332 \pm 0.0353$ | $0.0936 \pm 0.0616$ | $\mathbf{0.0135 \pm 0.0071}$ |
| | HyperStyle [3] - 5 iters | $0.4297 \pm 0.0404$ | $0.1420 \pm 0.1003$ | $0.0247 \pm 0.0115$ |

# Regularization

- Optimization problem:

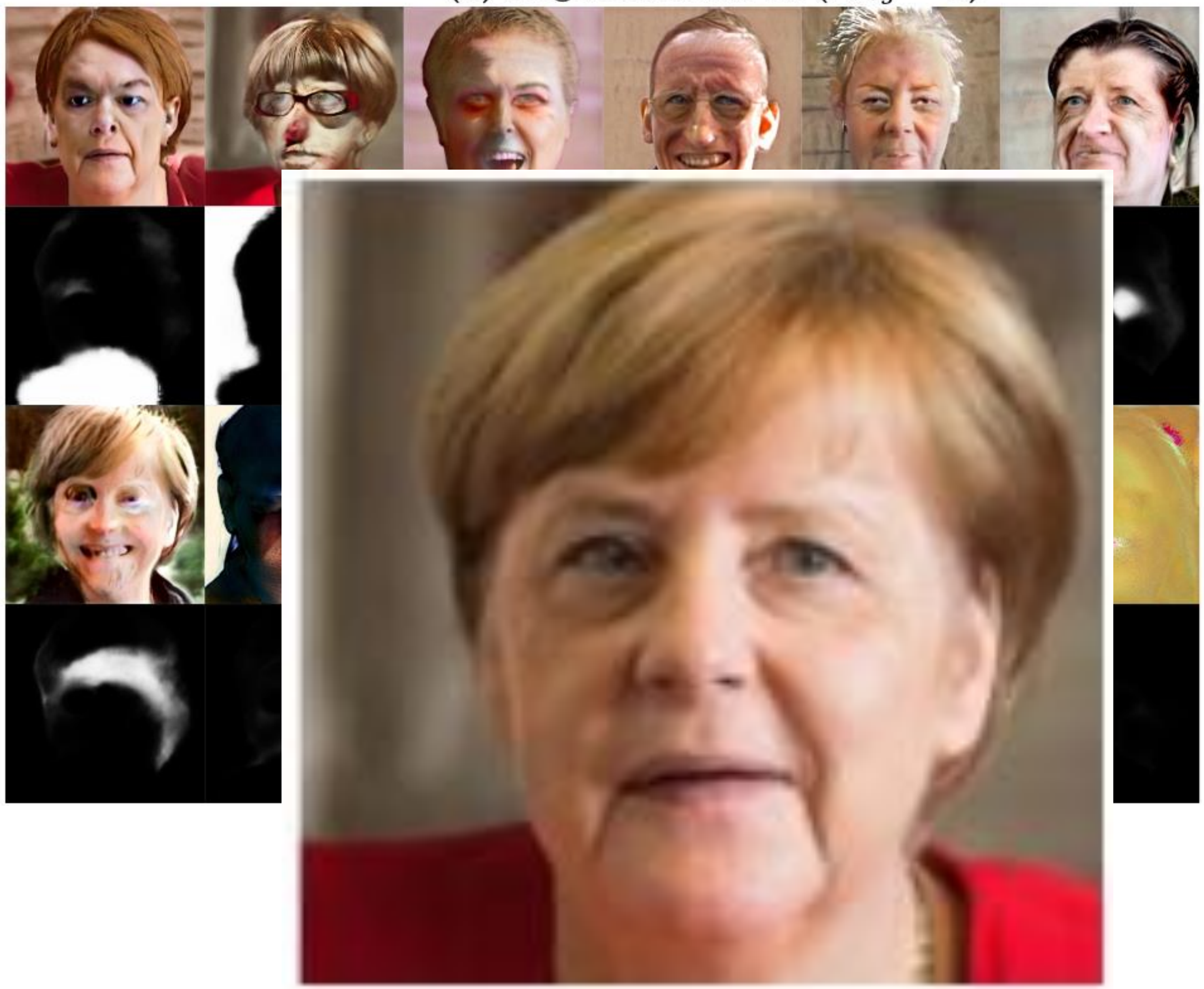
$$\min_{\boldsymbol{z},\boldsymbol{s}} \mathcal{L}_{LPIPS}\big(I,\ O(\boldsymbol{z},\boldsymbol{s})\big) + \boxed{\lambda_{reg} \sum_{i=1}^{n} \| \ z_i - z_\mu \ \|^2}$$

- Regularization → latent codes lie close to the natural image manifold

# Effect of Regularization

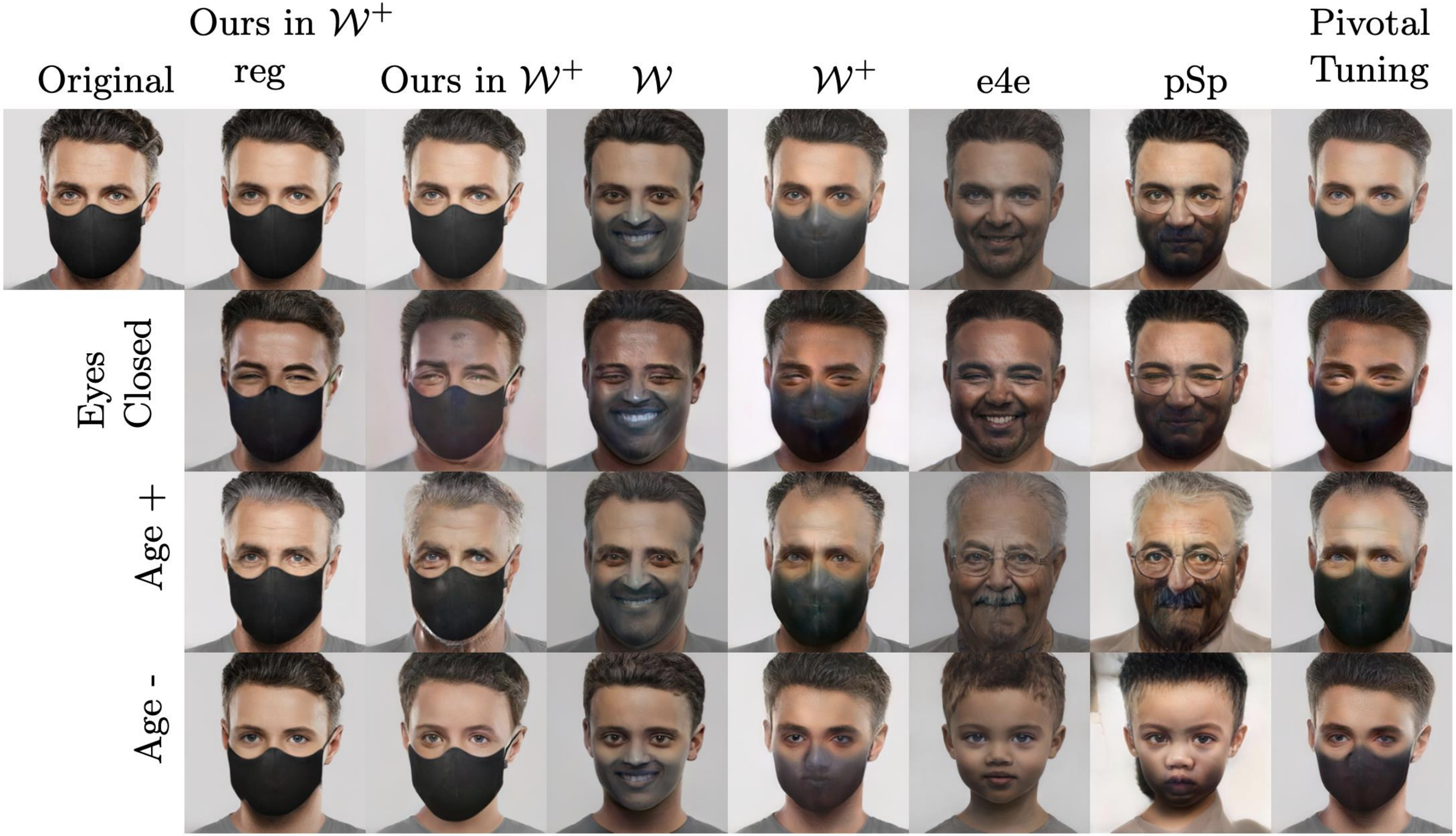
(a) Regularization on ($\lambda_{reg} = 1$)

(b) Regularization off ($\lambda_{reg} = 0$)

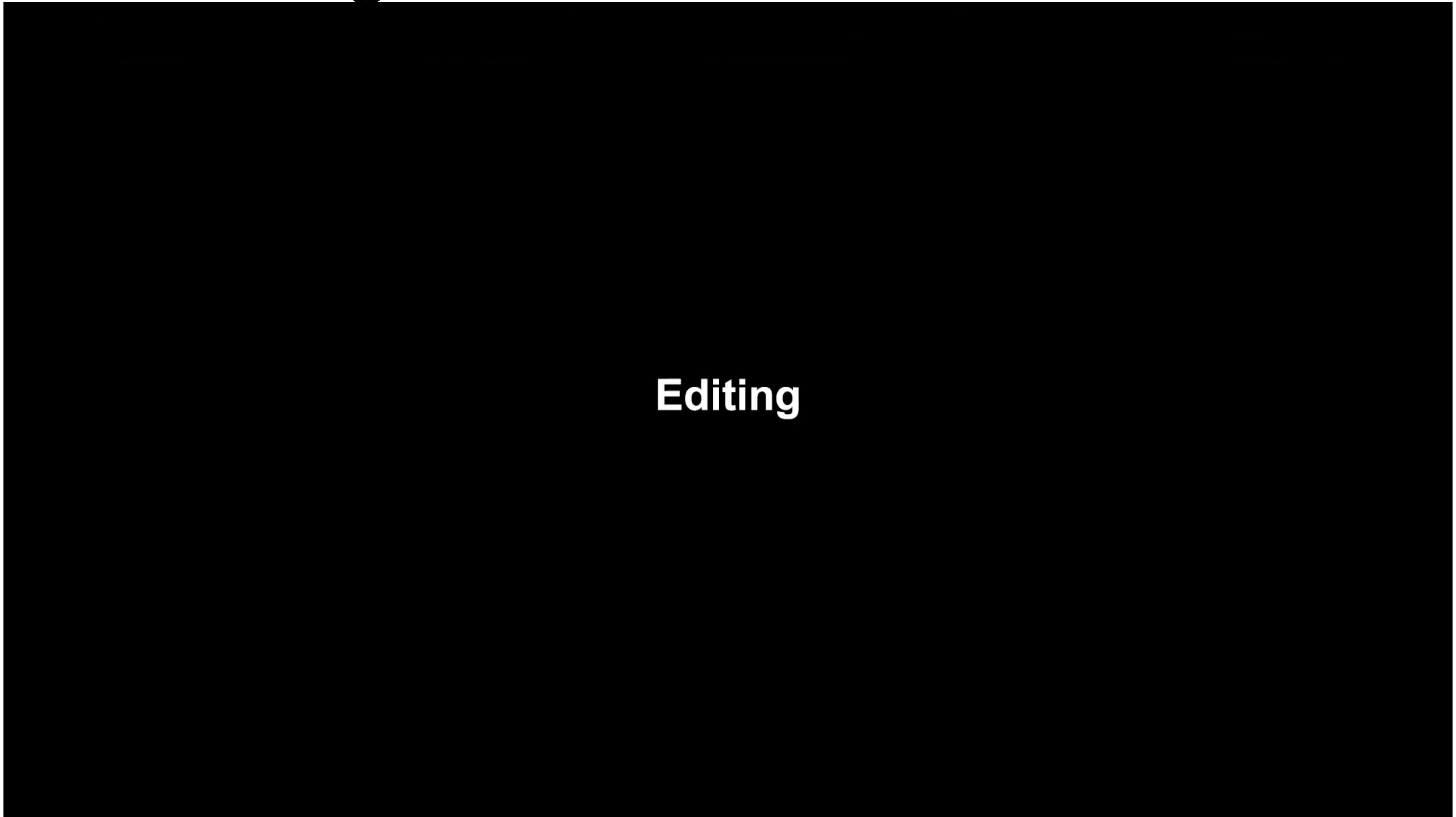


# Editing

# Results - Editing



# Local editing – artist workflow

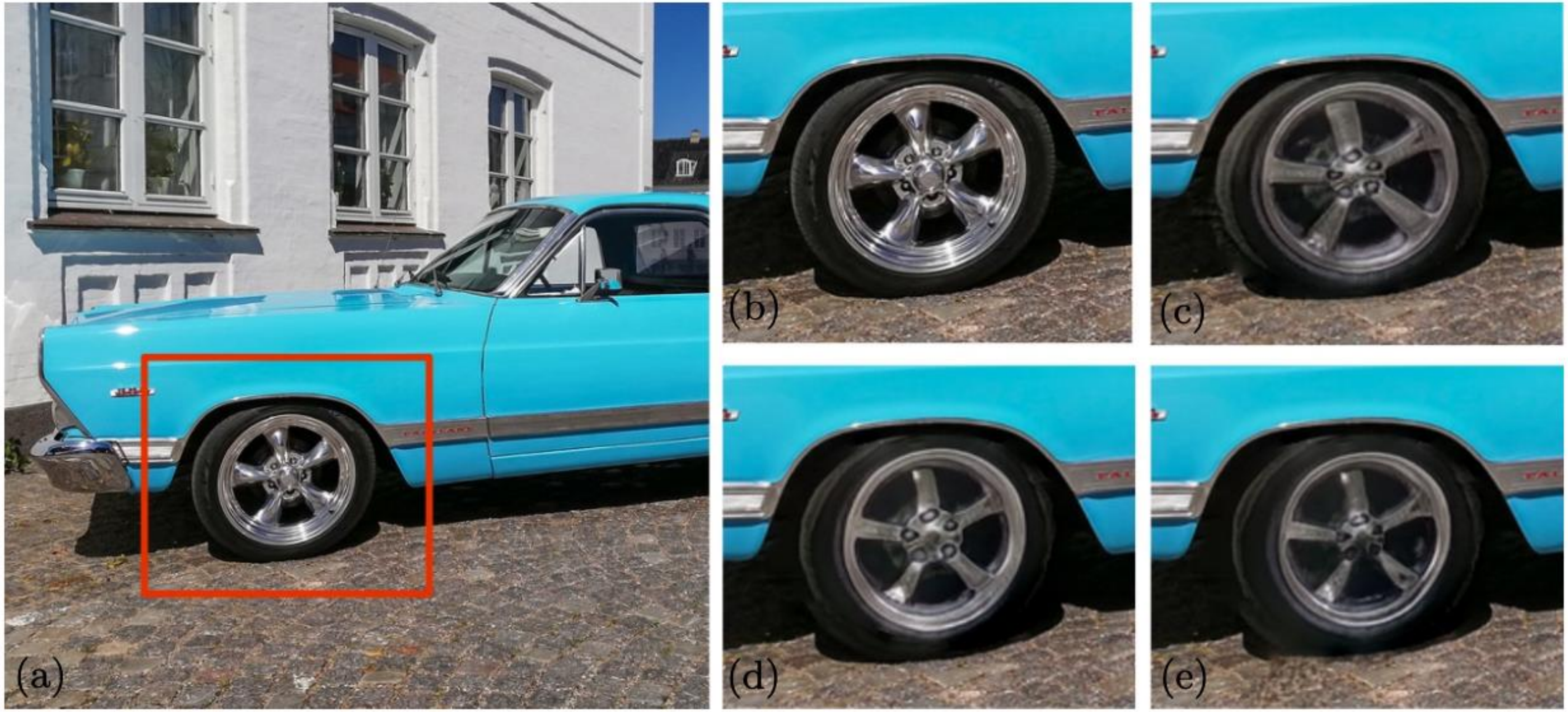


# Identity Preservation after Edits

| | | (a) | | | | (b) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | gender | smile | age | beard | gender | smile | age | beard |
| Optimization | $\mathcal{W}$ | 0.169 | 0.022 | 0.07 | 0.279 | 0.249 | 0.18 | 0.191 | 0.328 |
| | $\mathcal{W}^+$ | 0.209 | 0.02 | 0.095 | 0.296 | 0.256 | 0.128 | 0.171 | 0.325 |
| | Ours in $\mathcal{W}^+$ | 0.298 | 0.049 | 0.151 | 0.312 | 0.325 | 0.125 | 0.203 | 0.333 |
| | Ours in $\mathcal{W}^+$ reg. | 0.126 | **0.018** | 0.069 | 0.091 | 0.169 | **0.099** | **0.129** | **0.144** |
| Encoders | e4e [9] | **0.088** | 0.024 | **0.054** | 0.239 | 0.26 | 0.242 | 0.245 | 0.351 |
| | pSp [7] | 0.153 | 0.026 | 0.126 | **0.074** | 0.282 | 0.223 | 0.258 | 0.248 |
| | ReStyle [2]- 5 iters | 0.097 | 0.030 | 0.081 | 0.213 | 0.417 | 0.409 | 0.399 | 0.453 |
| Generator parameters modification | Pivotal Tuning [8] | 0.135 | 0.037 | 0.089 | 0.329 | 0.237 | 0.176 | 0.200 | 0.388 |
| | HyperStyle [3]- 5 iters | 0.107 | 0.12 | 0.135 | 0.107 | **0.15** | 0.163 | 0.166 | 0.157 |

Reconstructed vs Edited — Input vs Edited

# Also works for other domains



# Limitations

# Conclusion

- Semantic image editing using GANs
- Hairstyle Transfer between Face Images
  - Fully automatic method
  - Handles different illumination and pose
- ChunkyGAN: Real Image Inversion via Segments
  - Method for image modelling and editing using GANs
  - Works for images with unusual features ( glasses, face masks, bindi, etc.)
- Future of GANs? -> Diffusion models

# Questions & Demo