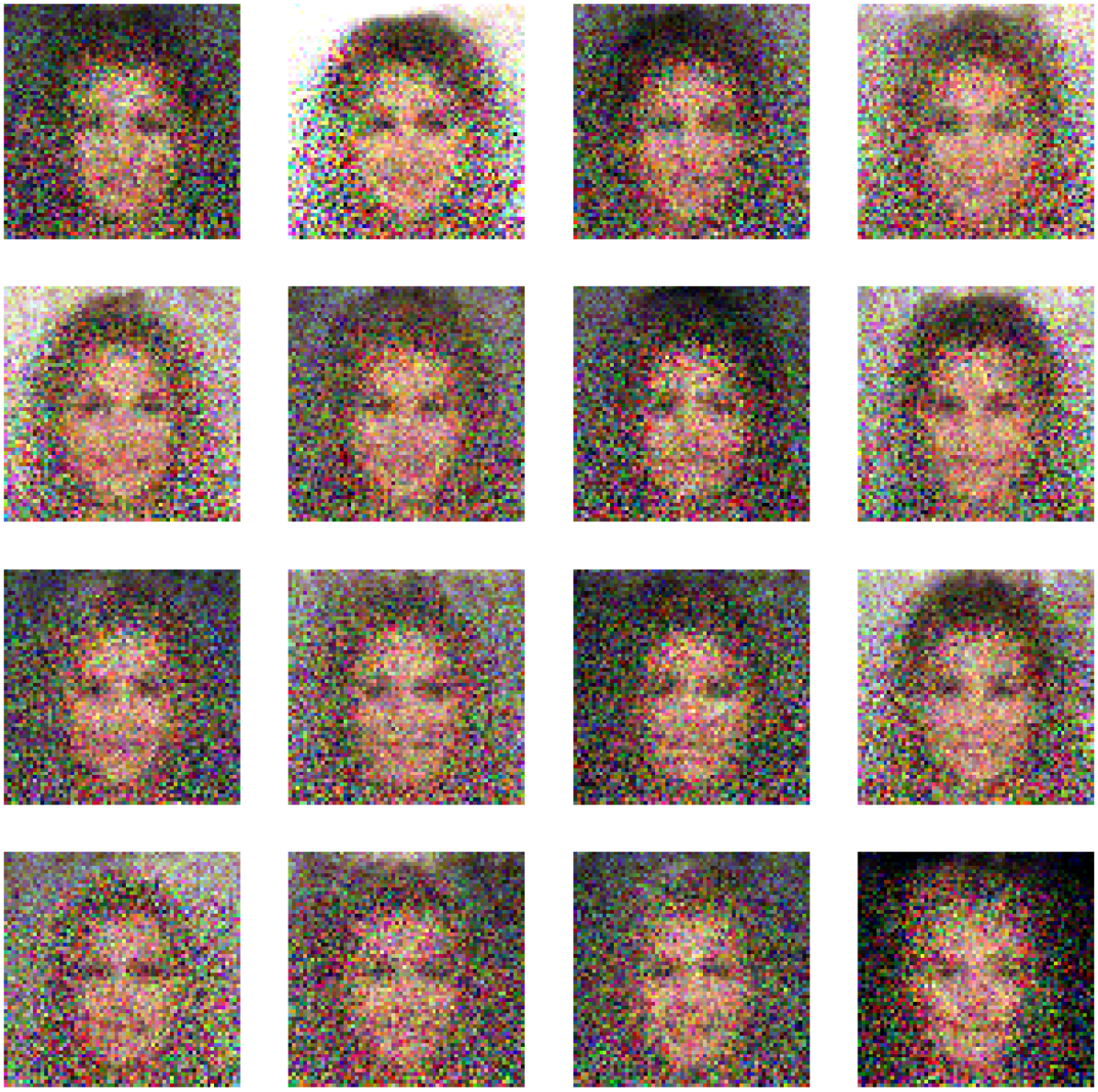
Report

NB: I was using LLMs to search for documentation and help in implementation

Normflow

Simple flow was built from composition of 5 MaskedAffineAutoregressiveTransform layers from nflows library.

Example of generation:



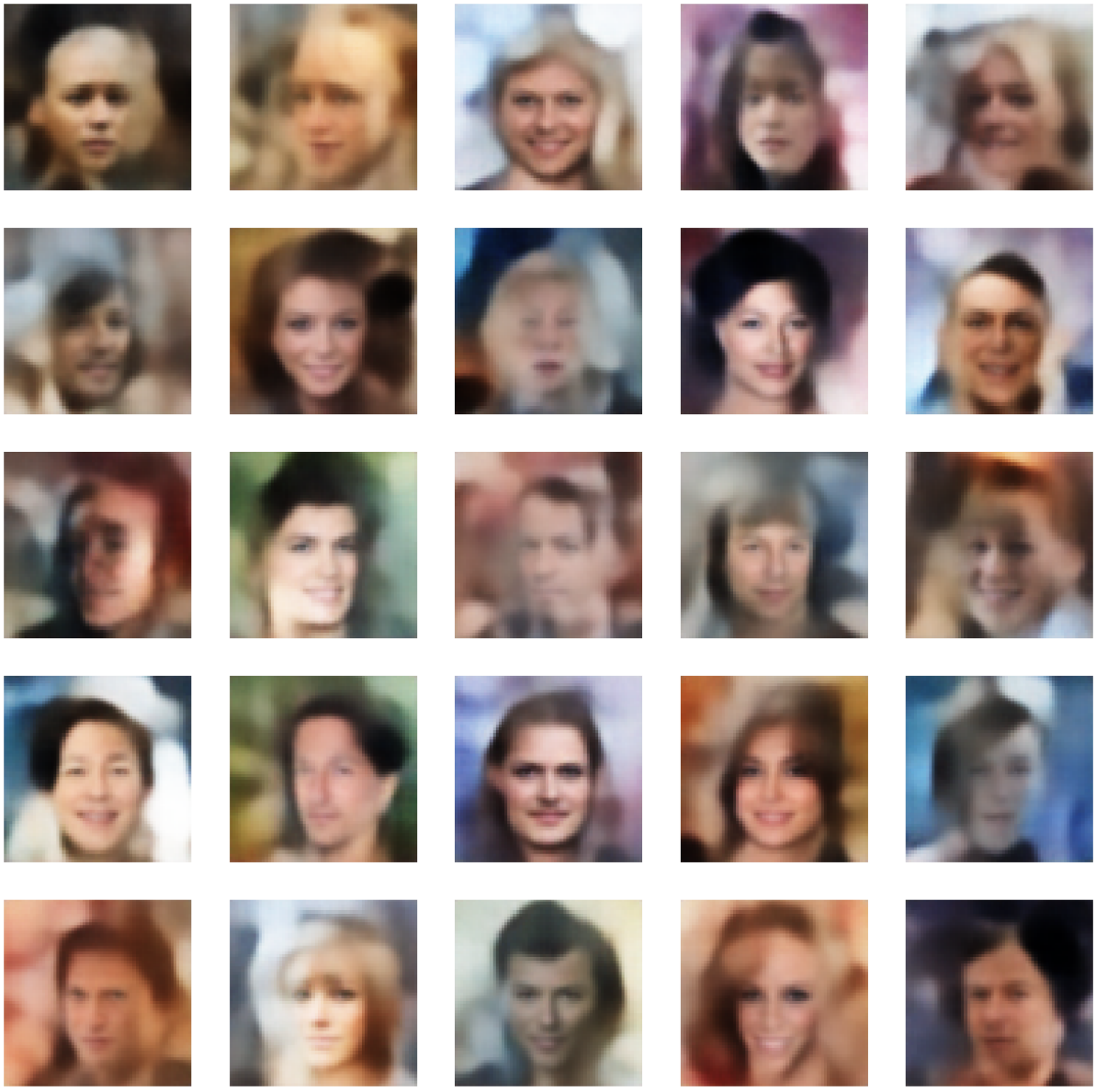
There is lot of noise, though faces features are distinguishable

On 1000 samples FID: 1247.963

VAE

Classic implementation with loss calculated from MSE and KL divergence. Dimension of latent space is 256

Example of generation:

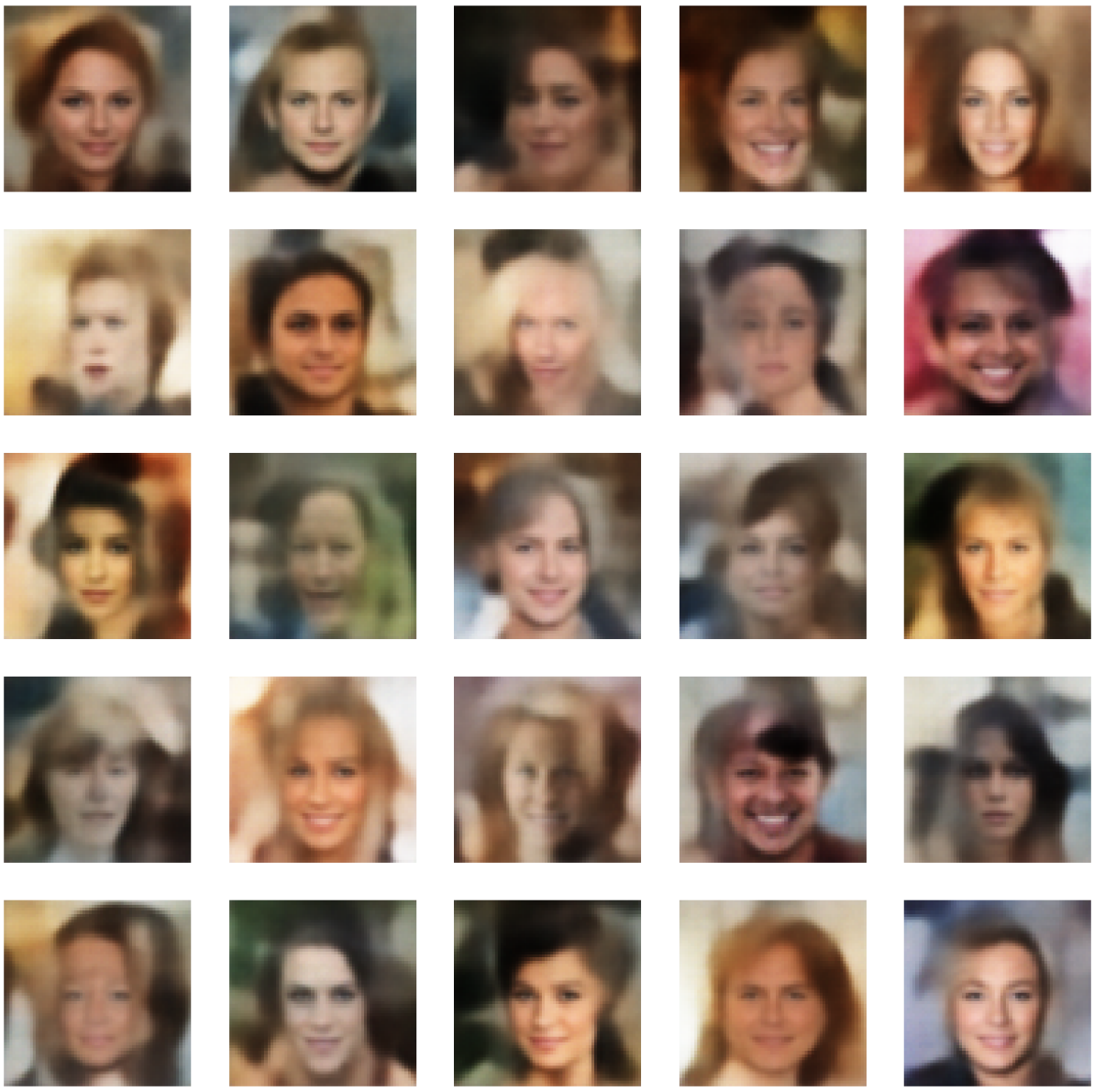


FID: 863.444

VAPNEV

Latent space dimension is also 256, loss Is calculated as sum of MSE and KL divergence but taking into account normalizing flow presented by log determinant.

Example of generation:



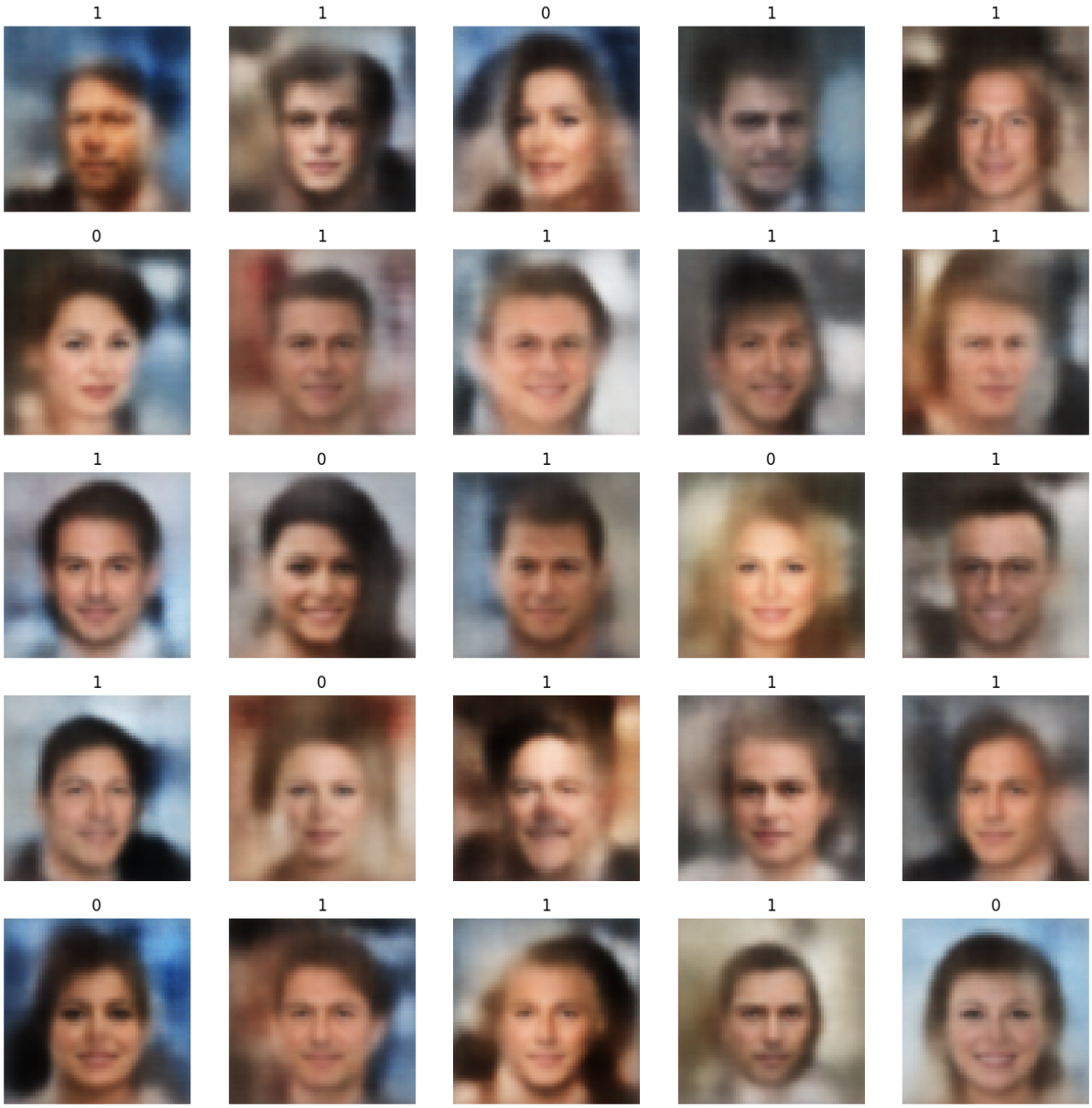
FID: 907.805

FID appeared to be relatively same as for regular VAE, but visually images seem more natural. To calculate FID, I was using only 1000 samples, maybe that's the reason

Conditional VAPNEV

Now when labels are added into model, for better interpretability I took two classes based on gender.

Example of generation (1 – male, 0 – female):



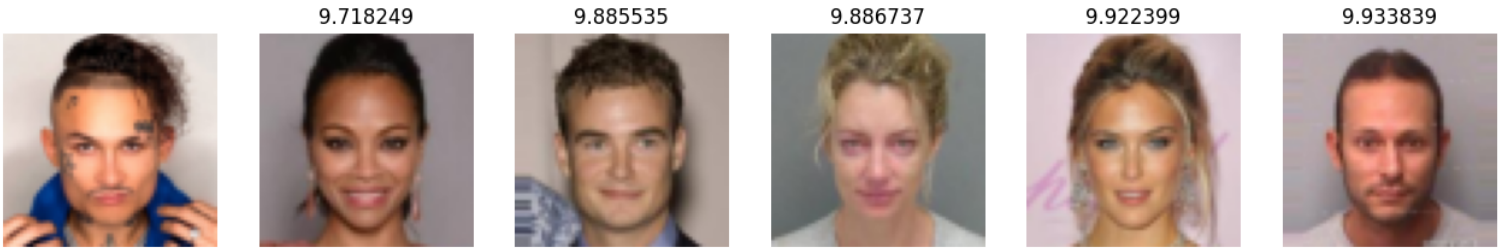
Visually features of corresponding genders are quite distinguishable

FID on females: 546.958

FID on males: 1221.81

Nearest to external image

By encoder from VAPNEV, measured by distance in latent space



Most funny generation