Comparison between Wasserstein Auto-Encoders and Introspective Variational Auto-encoders

Francesco Bongini

January 2021

Introduction

Generative models are a particular type of model that allow us to generate new data, such as images, sounds and so on.

Figure 1: generating Obama's young face

WAE and IntroVAE

There exist a lot of variants, most of the time combining the VAEs and the GANs properties like:

1. **WAE** (Wasserstein Auto-Encoders)

2. **IntroVAE** (Introspective Variational Autoencoders)

WAE objective

WAE uses the Optimal Transportation (OT) to compare P_X and P_G (the original and generated image), through their distance's distributions.

OT provides useful gradients for training, for instance low dimensional manifolds in the input space X, in which most of the f-divergences fail.

OT requires the regularization.

WAE objective

However we don't know P_x, while P_G is from a neural network.
If we are using the KL-divergence, we can find the minimum through the **variational lower bound**.

With the OT we can compute $d(Q_z, P_z)$ instead of $d(P_x, P_G)$, in fact:

$$\inf_{\Gamma \in P(X \sim P_X, Y \sim P_G)} E_{(X,Y)\sim\Gamma}[c(X,Y)] = \inf_{Q: Q_Z = P_Z} E_{P_X} E_{Q(Z|X)}[c(X, G(Z))]$$

This makes the calculations easier.

The WAE's regularization

VAE forces $Q(Z|X = x)$ to match P_z, while WAE forces the **continuous mixture** $Q_Z = \int Q(Z|X)dP_x$ to match P_z (usually the standard Normal).

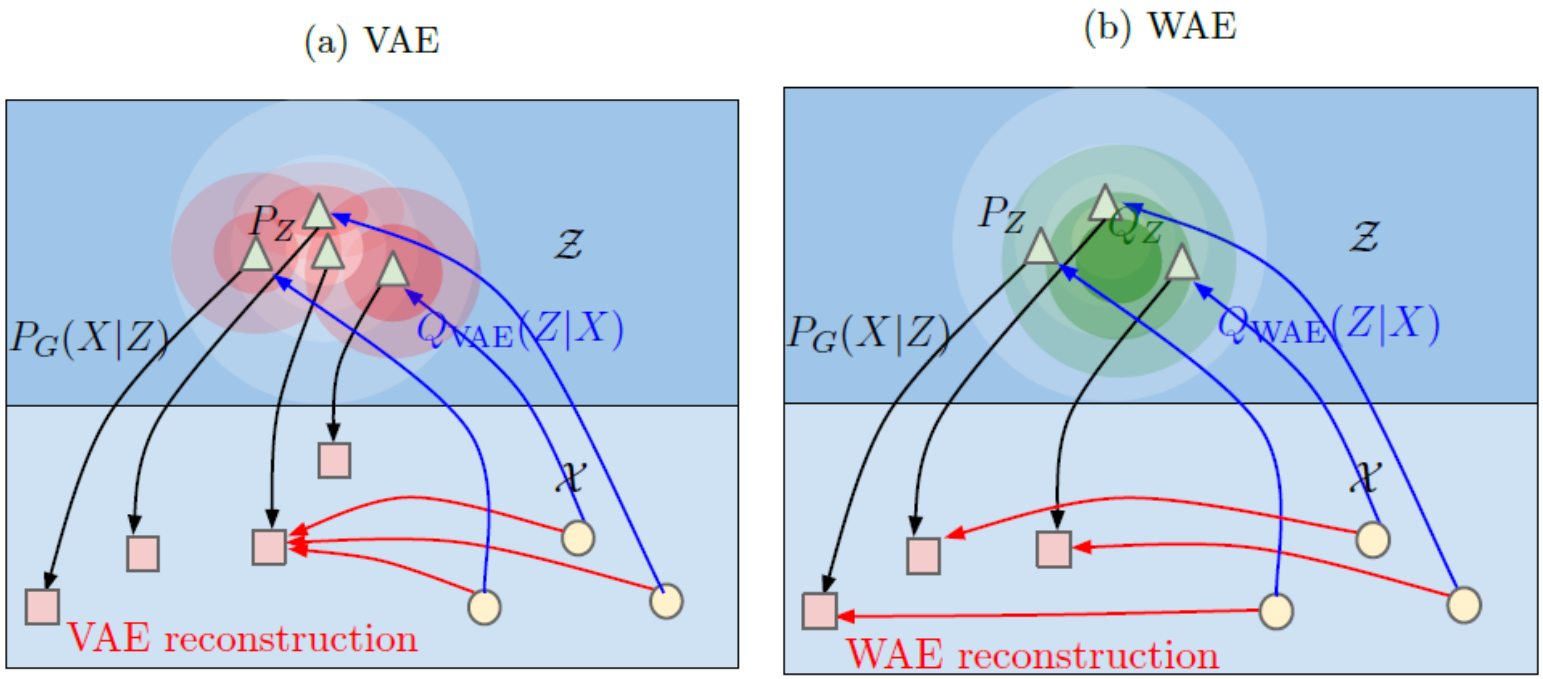


Figure 2: Difference between VAE and WAE

Regularization

The paper presents two penalities $D_z(Q_z, P_z)$:

1. WAE-MMD (maximum mean discrepancy):

$$MMD_k(P_z, Q_z) = || \int_Z k(z, .) dP_Z(z) - \int_Z k(z, .) dQ_Z(z) ||_{H_k}$$

2. WAE-GAN:

$$\sum_{i=1}^{n} \log D_\gamma(z_i) + \log(D_\gamma(\hat{z}_i))$$

$$\sum_{i=1}^{n} c(x_i, G_\theta(\hat{z}_i)) - \log(D_\gamma(\hat{z}_i))$$

riguardare WAE-GAN, prob da togliere c(x,G)

Further improvement

Given:

- gen= decode the codes
- ae_gen = encode and decode gen

The idea is to minimize the loss between gen and ae_gen to select a better sample from P_z.

IntroVAE

The idea is to consider the inference model (*encoder*) as discriminator and the generative model (*decoder*) as generator of the GANs.

The inference and generative models are defined as:

$$L_E(x, z) = E(x) + [m - E(G(z))]^+$$

$$L_G(z) = E(G(z))$$

where $E(x) = D_{KL}(q_\theta(z|x)p(z))$, $[.]^+ = max(0, .)$
and $m = positive\ mangin$.
Minimizing both models leads to the *minimax adversarial training*.

Introspective Variational Inference

However, like with GANs, finding the optimal E* and G* may cause some problems like the *mode collapse* and the *training instability*.
We therefore consider L_{AE} to make sure the generated images match the true ones.

The IntroVAE's objective is:

$$L_E(x, z) = E(x) + [m - E(G(z))]^+ + L_{AE}$$
$$L_G(z) = E(G(z)) + L_{AE}$$

Where $L_{AE} = -E_{q_\theta(z|x)} \log p_\theta(x|z)$

IntroVAE training

Finally, the objective of IntrovAE is:

$$L_E = L_{REG}(Enc(x)) + \alpha \sum_{s=r,p} [m - L_{REG}(Enc(ng(x_s)))]^+ + \beta L_{AE}(x, x_r)$$

$$L_g = \alpha \sum_{s=r,p} L_{REG}(Enc(x_s)) + \beta L_{AE}(x, x_r)$$

α, β and m are three parameters to make either the reconstruction or the regularization more or less important.

Comparing WAE and IntroVAE

The two models are compared considering:

1. The recostruction of the train and test sets
2. The good latent manifold
3. The sample diversity

Data Set

The *CelebA* data set contains 202,599 pictures of celebrities. We split the data set in:

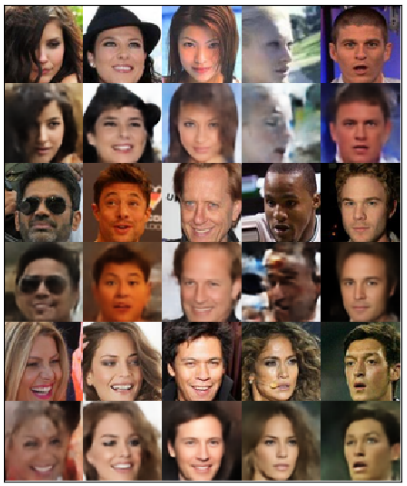
- Training set = 182,559 celebrities
- Test set = 20,000 celebrities

Figure 3: Example of celebA data set

Results of WAE-MMD

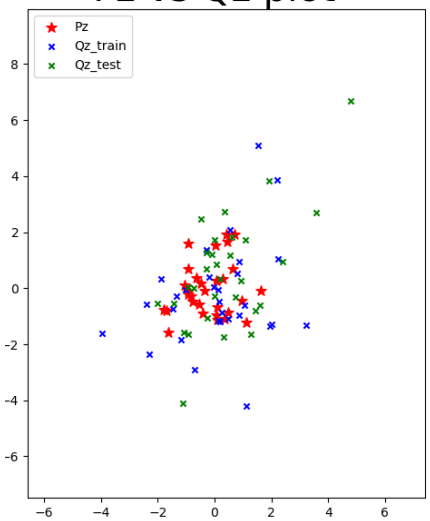
train reconstruction

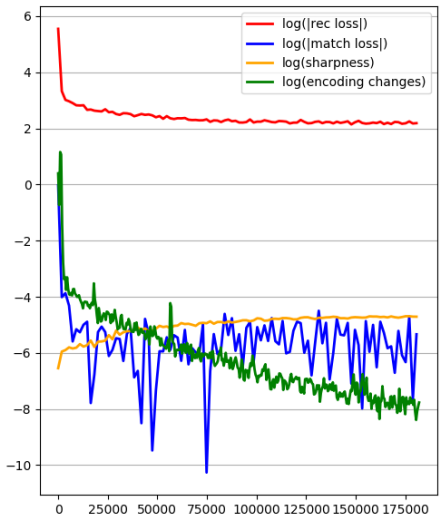
test reconstruction



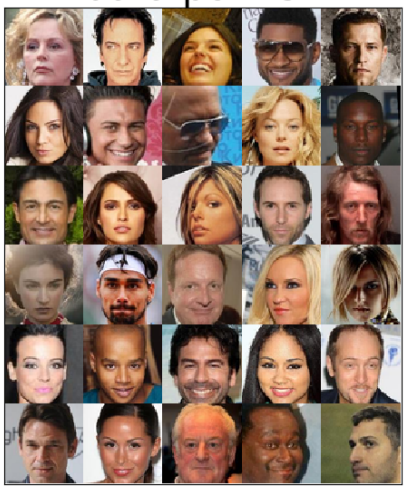
generated samples

Pz vs Qz plot





data points



1) RMSE of the reconstruction

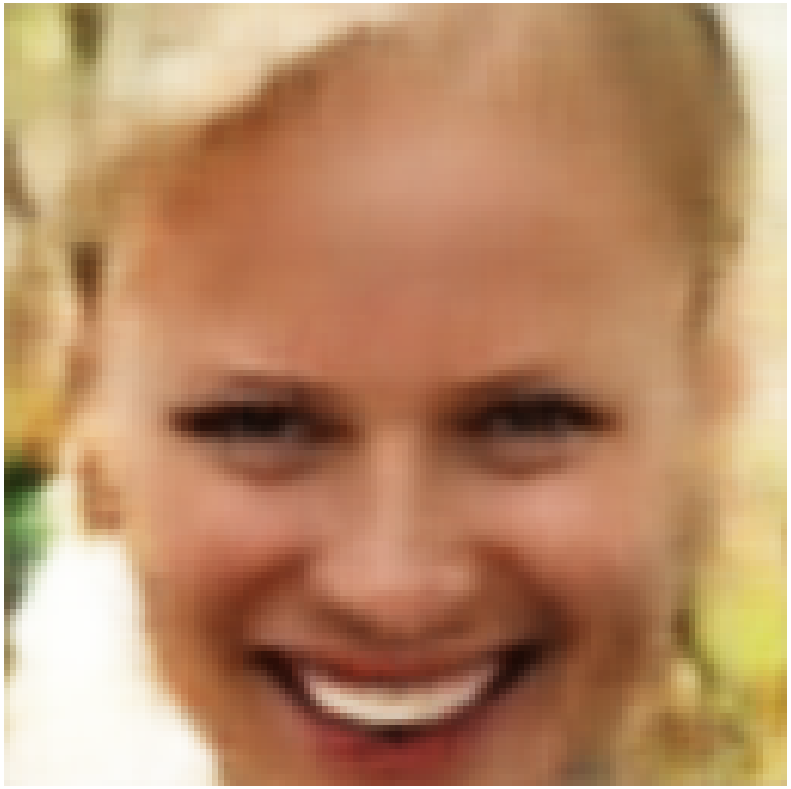
The performance of the reconstruction is calculated as the average of the RMSE between real and fake images. Figure 4 shows a test set example and their reconstruction.

Figure 4: Harry Judd Mcfly picture (left) and his reconstruction (right).

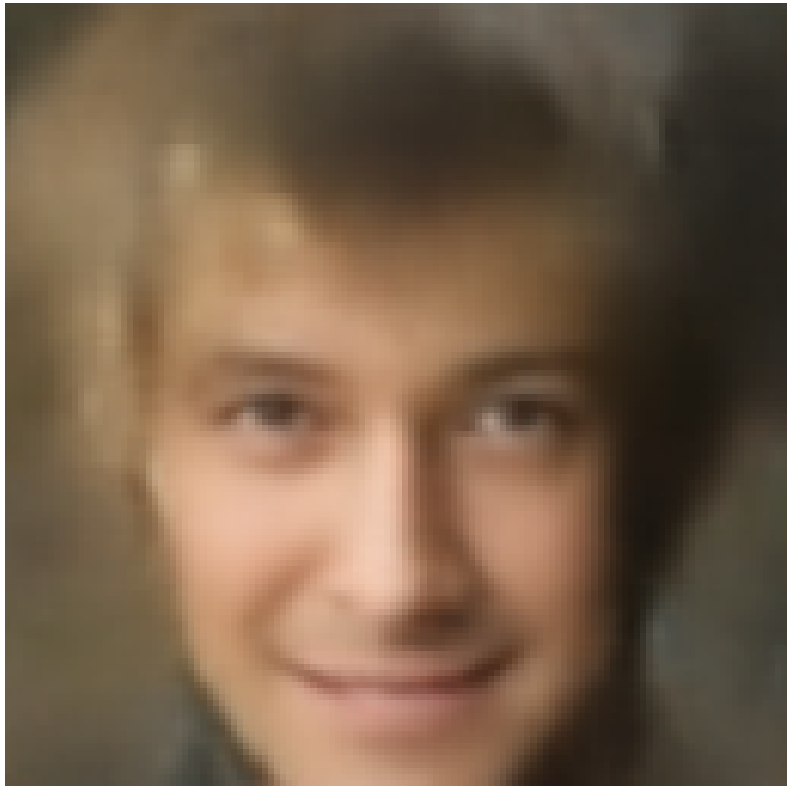
Model	RMSE training set	RMSE test set
IntroVAE	23.39	27.50
WAE-MMD	**11.70**	**15.44**
WAE-GAN	13.64	16.12

Table 1: RMSE results (lower is better)

2) WAE Latent manifold



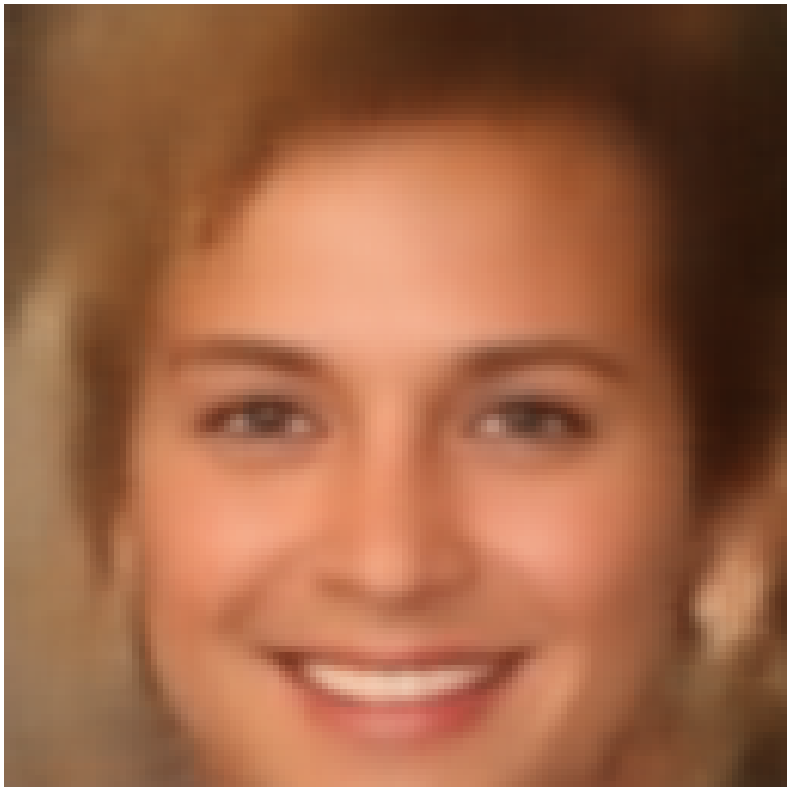
WAE-GAN



WAE-MMD

These manifold continuities verify that the proposed models generalize the image contents instead of simply memorizing them.

2) IntroVAE Latent manifold



IntroVAE

Model	RMSE
IntroVAE	
WAE-MMD	
WAE-GAN	

The table shows the mean of the RMSE values between the interpolated pictures and the training set of the three models,

3) Generating faces

To generating a new face, I generate a sample from the multidimensional P_z distribution that will then be decoded to an image. The video shows the training phase of the network and the relative sample generations.

https://www.youtube.com/watch?v=FKSkBO_Hms0&ab_channel=FrancescoBongini

Generating faces with WAE-GAN

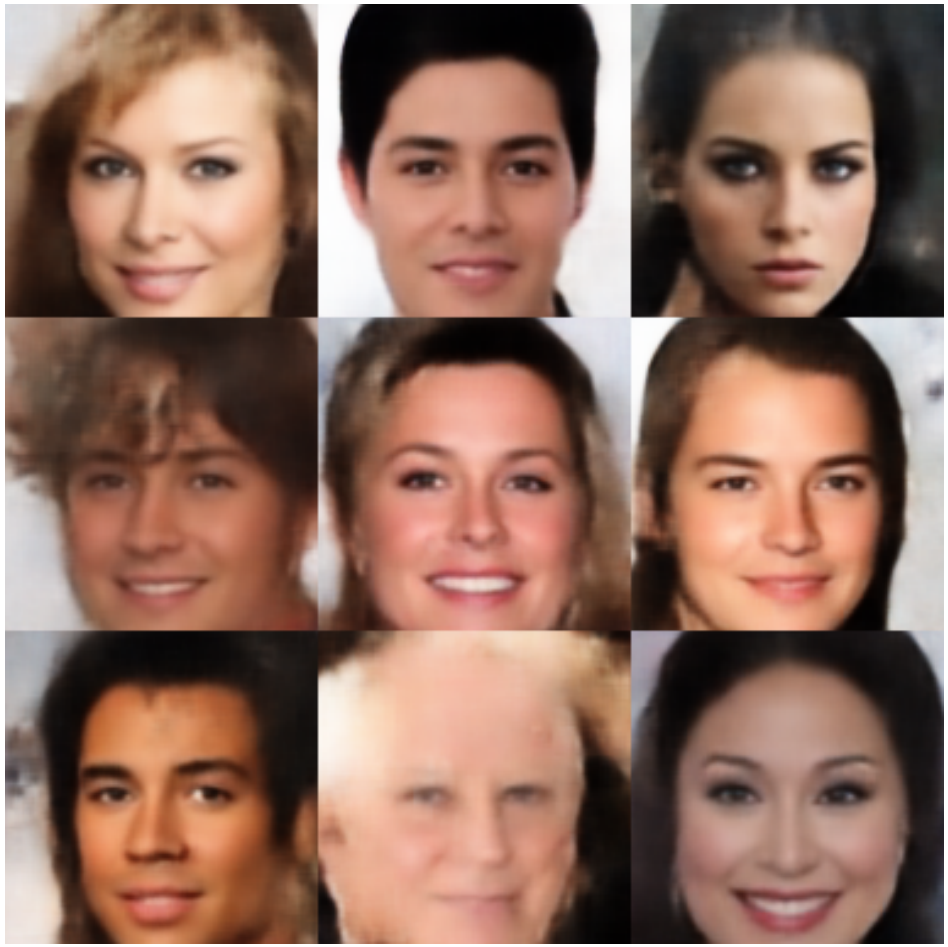


Figure 5: Improved WAE-GAN

3) MS-SSIM

To compare their capacity of generating different samples, the MS-SSIM scores are computed among an average of 10K pairs of synthesized images at 128x128 for CelebA.
The MS-SSIM index measures the similarity between two pictures. Here are the results (lower is better):

Model	MS-SSIM
IntroVAE paper result	**0.2989**
IntroVAE reimplemented	0.3042
Improved WAE-GAN	0.3159
Improved WAE-MMD	0.3719
WAE-GAN	0.4488
WAE-MMD	0.5178

Hyperparameter tuning

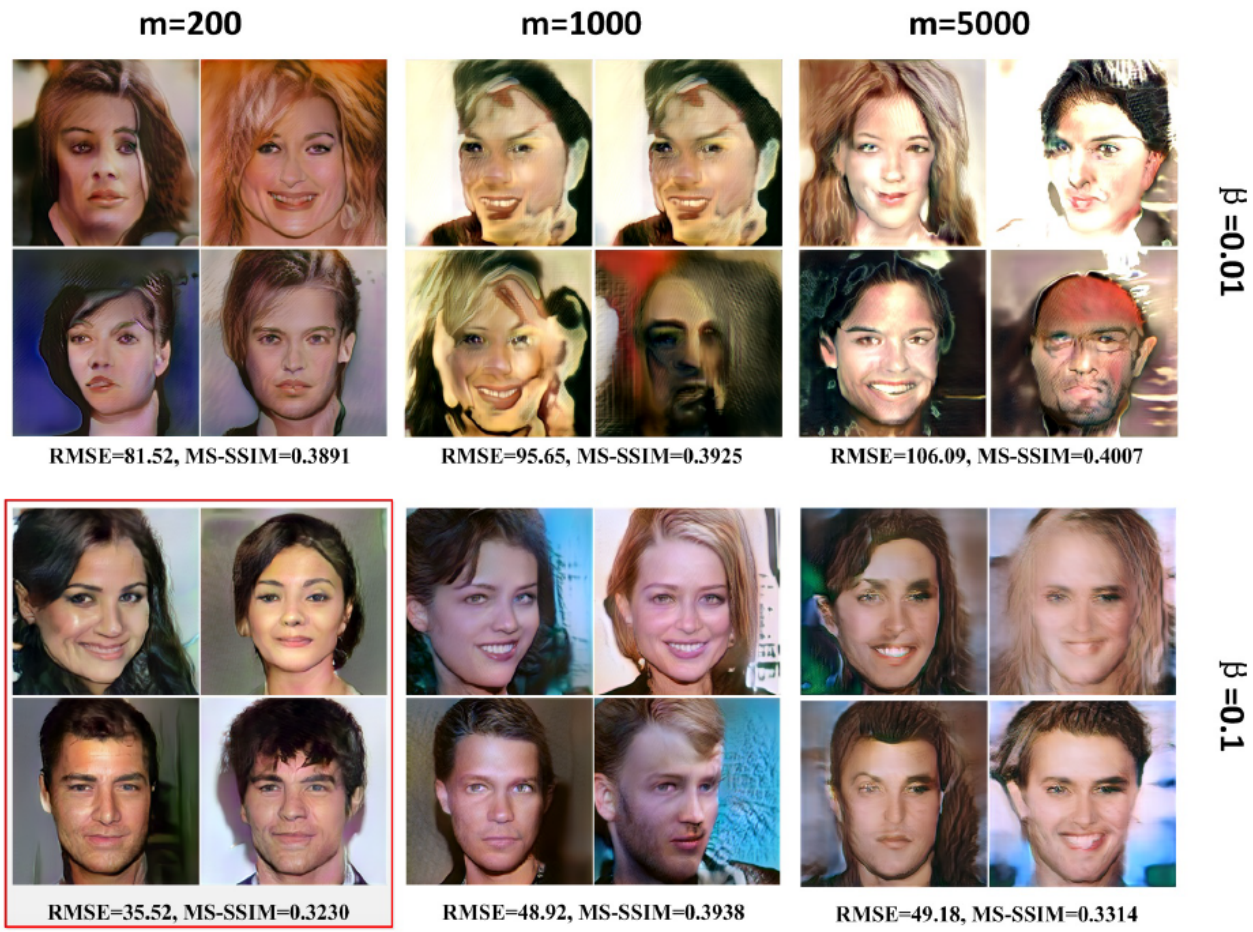


Figure 6: MS-SSIM and RMSE with multiple m and β values

Hyperparameter tuning/2

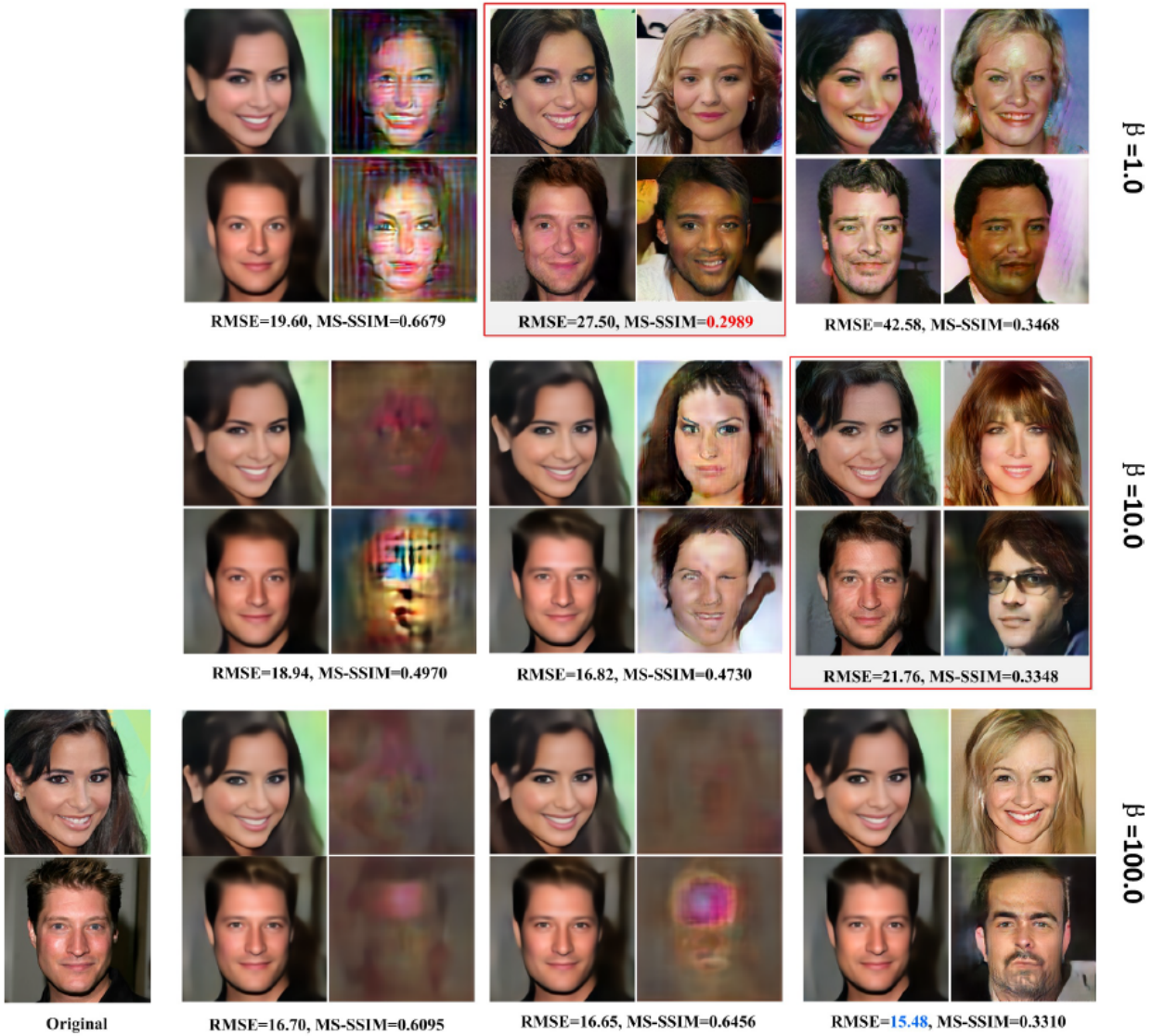


Figure 7: MS-SSIM and RMSE with multiple m and β values

Adversarial in WAE-GAN and IntroVAE

While the adversarial training in WAE-GAN is done in the latent space, IntroVAE does it at the end of the autoencoder, working on complex distributions.
This makes WAE-GAN more stable and faster to train than IntroVAE. However IntroVAE generates more realistic and different faces.

Conclusions

IntroVAE and WAE are both two valid generative models. We conclude that:

1. WAE-MMD achieves the best performances in the reconstruction of the train and test set.

2. WAE-GAN possess the good qualities of both the VAE and GAN, improving the samples generation of WAE-MMD.

3. Intro-VAE is less stable and slower to train than WAE models. It generates samples that are more realistic and diverse than WAE.