

Supplementary Material For Polynomial Implicit Neural Representations For Large Diverse Datasets

A. Training details

ImageNet: We train the Poly-INR model progressively with increasing resolution. The Poly-INR model is first trained on $200M$ images at 32×32 with 2048 batch size, followed by $72M$ images at 64×64 with 512 batch size, $21M$ images at 128×128 with 256 batch size, $10M$ images at 256×256 with 128 batch size and $2M$ images at 512×512 with 128 batch size. We use learning rate of $1e^{-4}$ for the generator and $2e^{-4}$ for the discriminator. We use Adam optimizer for both the generator and discriminator with $\beta = (0.0, .99)$ and $\epsilon = 1e^{-8}$ and the classifier guidance loss weight is set to 8.0 starting at 128×128 and higher resolution. We do not use style mixing regularization and path length regularization.

FFHQ: We also train the Poly-INR model progressively with increasing resolution on the FFHQ dataset. We first train our model with 64×64 on $60M$ images using a batch size of 2048, followed by $15M$ images at 128×128 with 256 batch size and $15M$ images at 256×256 with 256 batch size. The other training hyperparameters are same as the ImageNet experiments described above.

Table 1. Poly-INR performance on FFHQ-32x32 across various levels (lvl) and model size (number of parameters in Million).

	Lvl-2	Lvl-4	Lvl-7	Lvl-10	Lvl-14	Lvl-14
Feat. Dim.	512	512	512	512	512	1024
Params (M)	2.98	5.62	9.57	13.52	18.79	64.74
FID ↓	27.01	3.46	1.92	1.83	1.52	1.12
Precision ↑	0.85	0.68	0.67	0.68	0.68	0.70
Recall ↑	0.01	0.41	0.56	0.57	0.59	0.63

B. Ablation study on the number of levels and feature dimension

We present an ablation study in Table 1, demonstrating the Poly-INR performance on the FFHQ-32x32 dataset as levels increase. We observe that with increasing levels, the model’s performance improves. We utilize 10 levels in our experiments because of training stability and also achieve comparable performance compared to CNN-based models. In case of training with more than 10 levels, we can incrementally increase the number of levels by first training the model on a lower number, such as 10, and gradually add more levels as training progresses.

In Table 1, we increase the model capacity either by adding more levels (layers) or increasing the feature dimension on FFHQ-32x32. We observe that when the model capacity is very small, the recall score is also very poor, but as we increase the model parameters, the recall score gets much better.

C. Inference speeds across resolutions for our ImageNet model

Table 2. Inference speed (sec-per-image) of Poly-INR model trained on the ImageNet dataset across various resolutions.

32^2	64^2	128^2	256^2	512^2
0.007	0.013	0.047	0.179	0.720

Table 2 shows the inference speed (sec-per-image) across various resolutions of the Poly-INR model trained on the ImageNet dataset on a Nvidia-RTX-6000 GPU. The Poly-INR model synthesizes each pixel independently and performs all computations at the same resolution, resulting in slower inference time at higher resolutions.

D. Affine parameters mixing

An advantage of representing an image in the polynomial form is that it inherently breaks the image into shape and style. For example, the lower polynomial orders represent the object’s shape, whereas the higher orders represent finer details like the style of the image. In our Poly-INR model, manipulating the lower levels’ affine parameters changes the object’s shape, and manipulating higher levels’ affine parameters changes the style. Fig. 1 shows examples of style mixing from source A to source B images. In the figure, copying the affine parameters of source A to source B at higher levels (8 and 9) brings fine change in the style, whereas middle levels (5, 6, and 7) bring coarse style change. Fig. 2 shows affine parameters mixing at initial levels (0-5). In the figure, we observe that copying the affine parameters at these levels changes the shape of the source B image to the source A image.

E. Interpolation

Fig. 3 shows linear interpolation between samples of different classes in the affine parameters space. The Poly-INR model provides smooth interpolation between different classes. *Please see the attached video for better interpolation visualization.*

F. Qualitative comparison with StyleGAN-XL:

We also compare the quality of images generated by Poly-INR model against state-of-the-art CNN-based StyleGAN-XL model for different classes. Fig. 4, 5, and 6 show examples of images generated from different classes for the models trained on ImageNet at 256×256 . The Poly-INR generates samples qualitatively similar to the StyleGAN-XL model but without using any convolution or self-attention layers.

G. Qualitative comparison with CIPS and INR-GAN on FFHQ dataset

We also provide qualitative comparison of Poly-INR model against previously proposed INR-based generative models like CIPS and INR-GAN. Fig. 7 shows samples generated by the three models trained on the FFHQ dataset at 256×256 . Our Poly-INR model generates qualitatively better samples than the CIPS and INR-GAN using significantly fewer parameters.



Figure 1. Source A and B images generated from random latent codes, and the remaining images are generated by copying the affine parameters of source A to source B at different levels. Copying the higher levels' (8 and 9) affine parameters leads to finer style changes, whereas copying the middle levels' (7, 6, and 5) leads to coarse style changes.

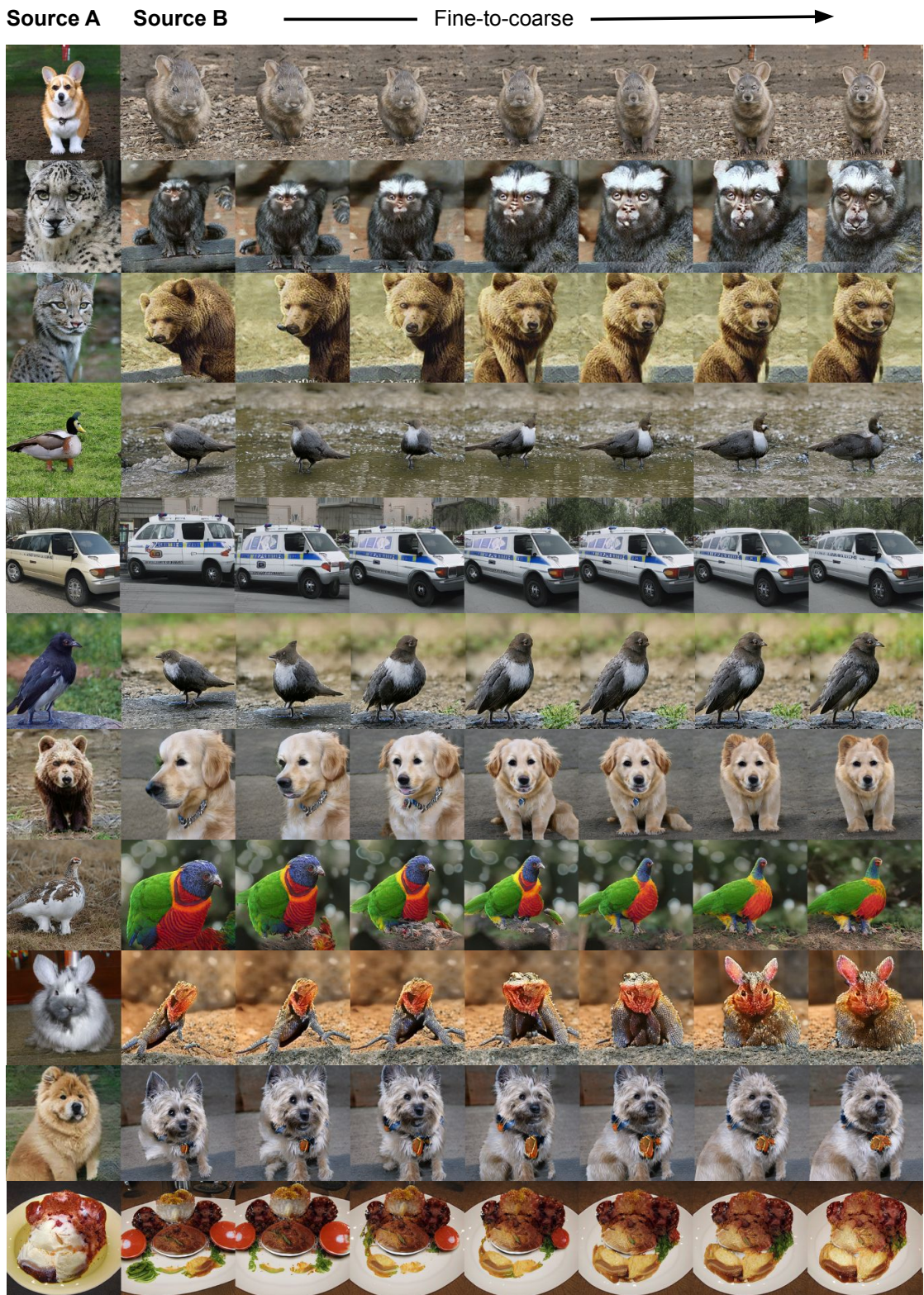


Figure 2. Source A and B images are generated from random latent codes, and remaining images are generated by copying the affine parameters of source A to source B at different levels. Copying the initial levels' (0, 1, and 2) affine parameters leads to finer shape changes, whereas copying slightly higher levels' (3, 4, and 5) leads to coarse shape changes.



Figure 3. The poly-INR model generates smooth interpolations between samples of different classes.

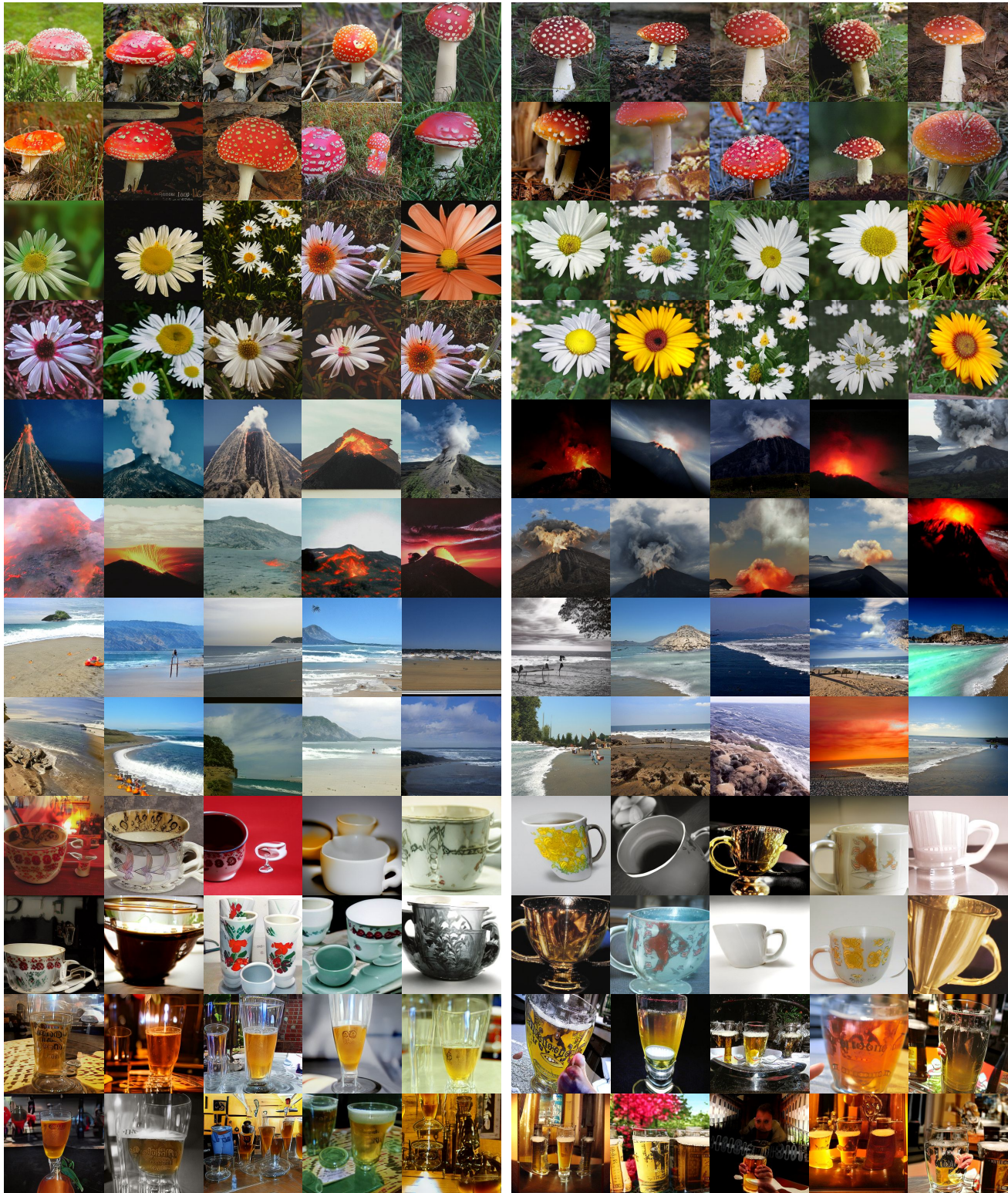


Figure 4. Qualitative comparison between StyleGAN-XL (left column) and Poly-INR (right column). Classes from top to bottom: agaric, daisy, volcano, seashore, cup, and beer glass.

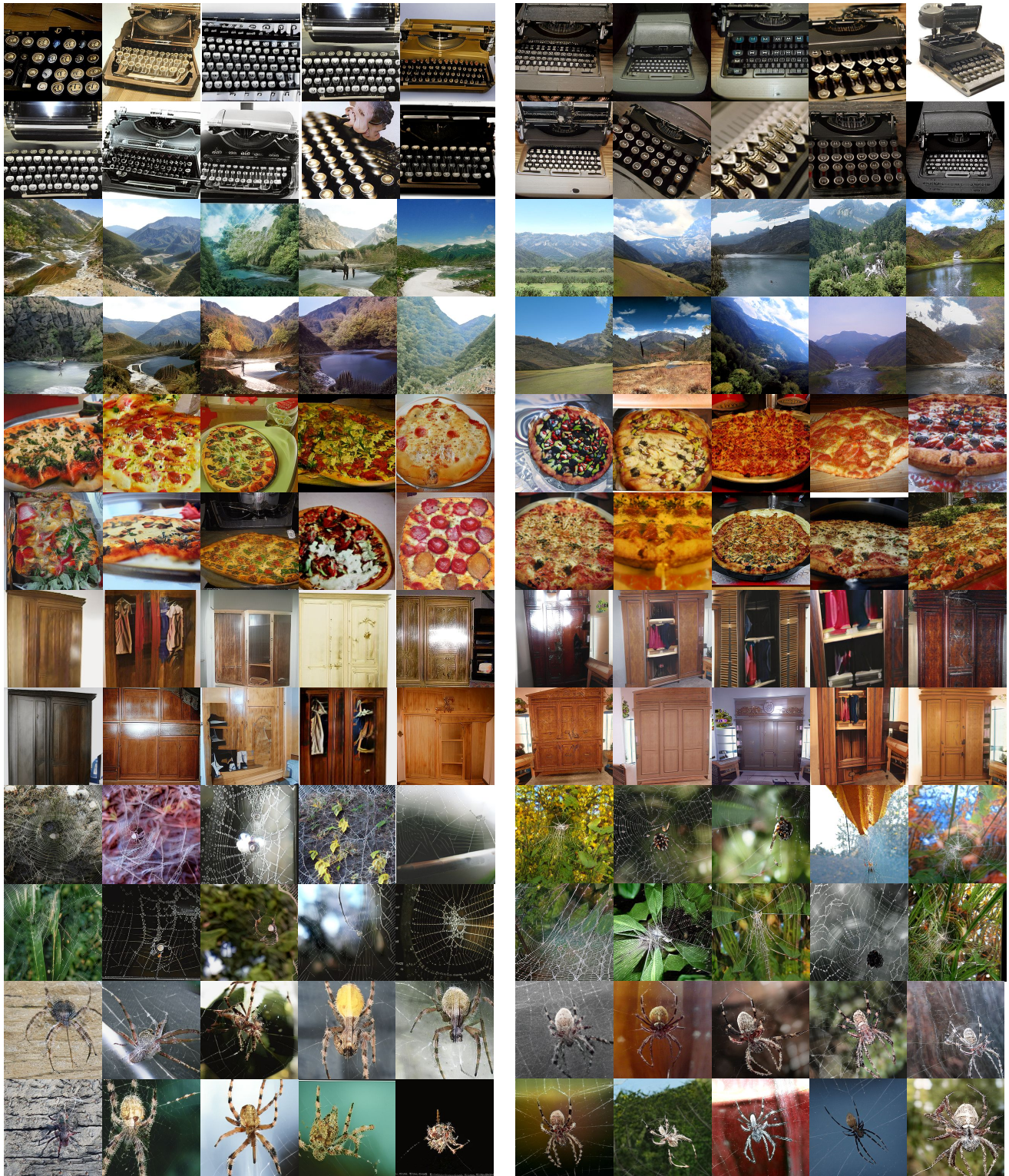


Figure 5. Qualitative comparison between StyleGAN-XL (left column) and Poly-INR (right column). Classes from top to bottom: type writer, valley, pizza, wardrobe, spider web, barn spider.

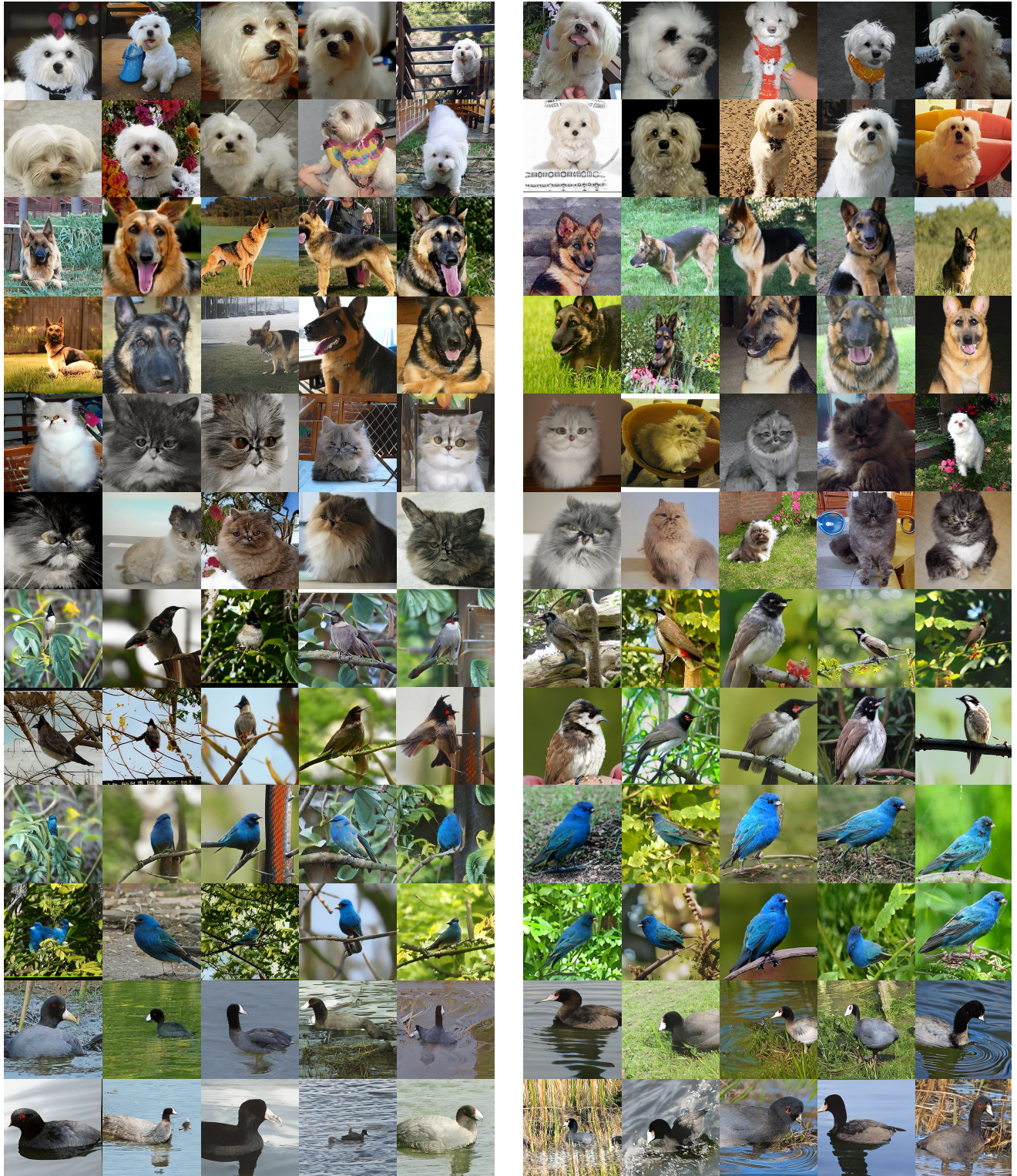


Figure 6. Qualitative comparison between StyleGAN-XL (left column) and Poly-INR (right column). Classes from top to bottom: maltese dog, german shepherd, persian cat, bulbul, robin, american coot.

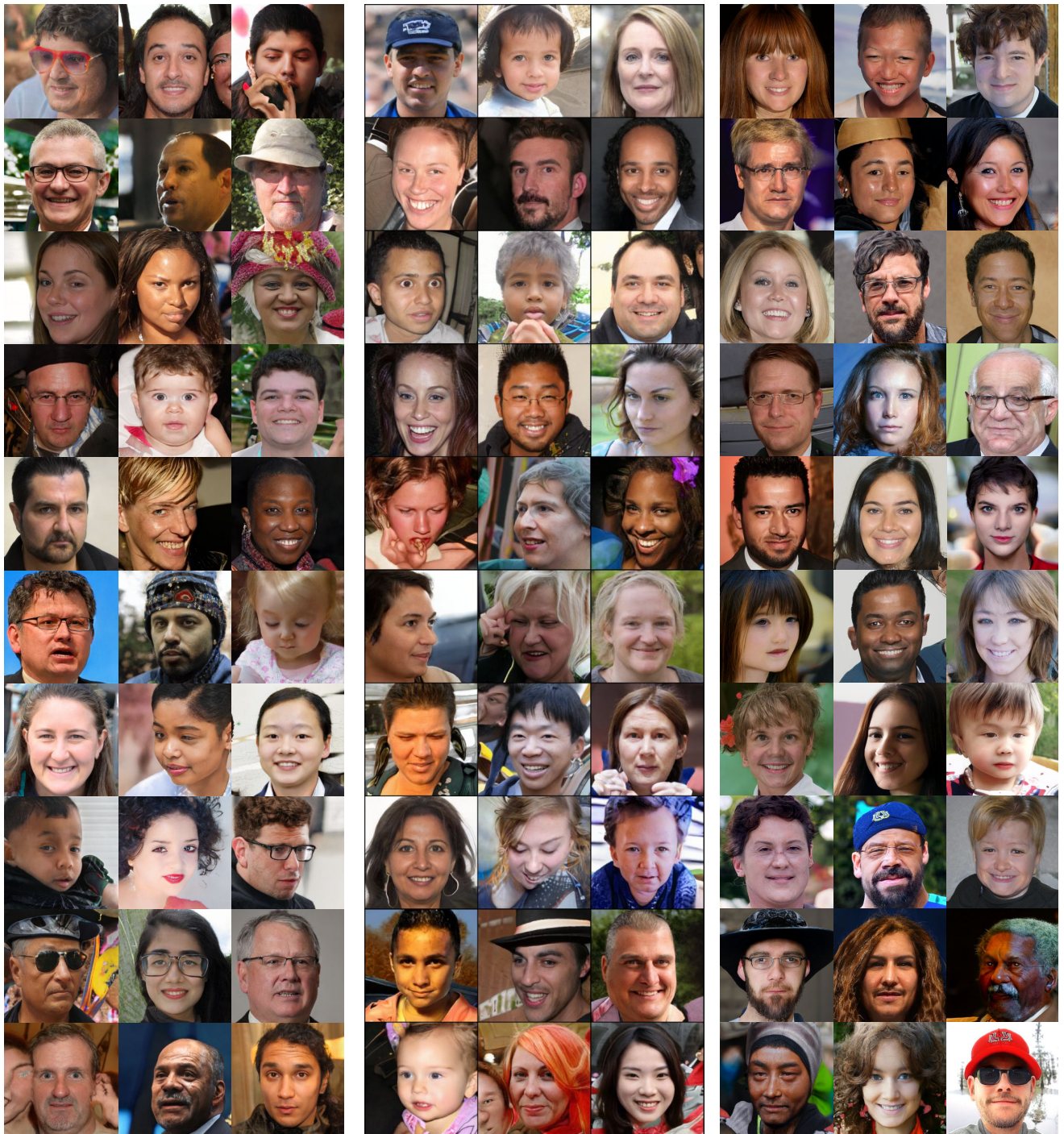


Figure 7. Qualitative comparison between INR-GAN (left column), CIPS (middle column), and Poly-INR (right column) on FFHQ dataset at 256×256 .