Self-Supervised GANs via Auxiliary Rotation Loss (Supplementary Material)

Ting Chen*
University of California, Los Angeles
tingchen@cs.ucla.edu

Xiaohua Zhai
Google Brain
xzhai@google.com

Marvin Ritter
Google Brain
marvinritter@google.com

Mario Lucic
Google Brain
lucic@google.com

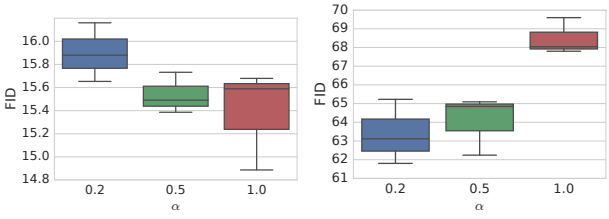
Neil Houlsby
Google Brain
neilhoulsby@google.com

1. FID Metric Details

We compute the FID score using the protocol as decribed in [1]. The image embeddings are extracted from an Inception V1 network provided by the TF library [2], We use the layer "pool_3". We fit the multivariate Gaussians used to compute the metric to real samples from the test sets and fake samples. We use 3000 samples for CELEBA-HQ and 10000 for the other datasets.

2. SS-GAN Hyper-parameters

We compare different choices of α, while fixing $\beta = 1$ for simplicity. A reasonable value of α helps aqa the generator to train using the self-supervision task, however, an inappropriate value of α could bias the convergence point of the generator. Table 1 shows the effectiveness of α. In the values compared, the optimal α is 1 for CIFAR10, and 0.2 for IMAGENET. In our main experiments, we used $\alpha = 0.2$ for all datasets.



(a) CIFAR10 (b) IMAGENET

Figure 1: Performance under different α values.

*Work done at Google.

3. Representation Quality

3.1. Implementation Details

We train the linear evaluation models with batch size 128 and learning rate of $0.1 \times \frac{\text{batch\_size}}{256}$ following the linear scaling rule [3], for 50 epochs. The learning rate is decayed by a factor of 10 after epoch 30 and epoch 40. For data augmentation we resize the smaller dimension of the image to 146 and preserve the aspect ratio. After that we crop the image to 128×128. We apply a random crop for training and a central crop for testing. The model is trained on a single NVIDIA Tesla P100 GPU.

3.2. Additional Results

Table 1 shows the top-1 accuracy with on CIFAR10 with standard deviations. The results are stable on CIFAR10 as all the standard deviation is within 0.01. Table 2 shows the top-1 accuracy with on IMAGENET with standard deviations. Uncond-GAN representation quality shows large variance as we observe that the unconditional GAN collapses in some cases.

Figure 2 shows the representation quality on all 4 blocks on the CIFAR10 dataset. SS-GAN consistently outperforms other models on all 4 blocks. Figure 3 shows the representation quality on all 6 blocks on the IMAGENET dataset. We observe that all methods perform similarly before 500k steps on block0, which contains low level features. While going from block0 to block6, the conditional GAN and SS-GAN achieve much better representation results. The conditional GAN benefits from the supervised lables in layers closer to the classification head. However, the unconditional GAN attains worse result at the last layer and the rotation only model gets decreasing quality with more training steps. When combining the self-supervised loss and the adversarial loss, SS-GAN representation quality becomes stable and outperforms the other models.

Figure 4 and Figure 5 show the correlation between top-1 accuracy and FID score. We report the FID and top-1 accuracy from training steps 10k to 100k on CIFAR10, and 100k to 1M on IMAGENET. We evaluate 10×3 models in total, where 10 is the number of training steps at which we evaluate and 3 is the number of random seeds for each run. The collapsed models with FID score larger than 100 are removed from the plot. Overall, the representation quality and the FID score is correlated for all methods on the CIFAR10 dataset. On IMAGENET, only SS-GAN gets better representation quality with better sample quality on block4 and block5.

References

[1] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, Günter Klambauer, and Sepp Hochreiter. GANs trained by a two time-scale update rule converge to a Nash equilibrium. In *Advances in Neural Information Processing Systems (NIPS)*, 2017.

[2] TFGAN: A Lightweight Library for Generative Adversarial Networks, 2017. URL https://ai.googleblog.com/2017/12/tfgan-lightweight-library-for.html.

[3] Priya Goyal, Piotr Dollár, Ross Girshick, Pieter Noordhuis, Lukasz Wesolowski, Aapo Kyrola, Andrew Tulloch, Yangqing Jia, and Kaiming He. Accurate, large minibatch SGD: Training Imagenet in 1 hour. *arXiv preprint arXiv:1706.02677*, 2017.

Method	Uncond-GAN	Cond-GAN	Rot-only	SS-GAN (sBN)
Block0	0.719 ± 0.002	0.719 ± 0.003	0.710 ± 0.002	$\mathbf{0.721 \pm 0.002}$
Block1	0.762 ± 0.001	0.759 ± 0.003	0.749 ± 0.003	$\mathbf{0.774 \pm 0.003}$
Block2	0.778 ± 0.001	0.776 ± 0.005	0.762 ± 0.003	$\mathbf{0.796 \pm 0.005}$
Block3	0.776 ± 0.005	0.780 ± 0.006	0.752 ± 0.006	$\mathbf{0.799 \pm 0.003}$
Best	0.778 ± 0.001	0.780 ± 0.006	0.762 ± 0.003	$\mathbf{0.799 \pm 0.003}$

Table 1: Top-1 accuracy on CIFAR10 with standard variations.

Method	Uncond-GAN	Cond-GAN	Rot-only	SS-GAN (sBN)
Block0	0.074 ± 0.074	0.156 ± 0.002	0.147 ± 0.001	$\mathbf{0.158 \pm 0.001}$
Block1	0.063 ± 0.103	0.187 ± 0.010	0.134 ± 0.003	$\mathbf{0.222 \pm 0.001}$
Block2	0.073 ± 0.124	0.217 ± 0.007	0.158 ± 0.003	$\mathbf{0.250 \pm 0.001}$
Block3	0.083 ± 0.142	0.272 ± 0.014	0.202 ± 0.005	$\mathbf{0.327 \pm 0.001}$
Block4	0.077 ± 0.132	0.253 ± 0.040	0.196 ± 0.001	$\mathbf{0.358 \pm 0.005}$
Block5	0.074 ± 0.126	0.337 ± 0.010	0.195 ± 0.029	$\mathbf{0.383 \pm 0.007}$
Best	0.083 ± 0.142	0.337 ± 0.010	0.202 ± 0.005	$\mathbf{0.383 \pm 0.007}$

Table 2: Top-1 accuracy on IMAGENET with standard variations.

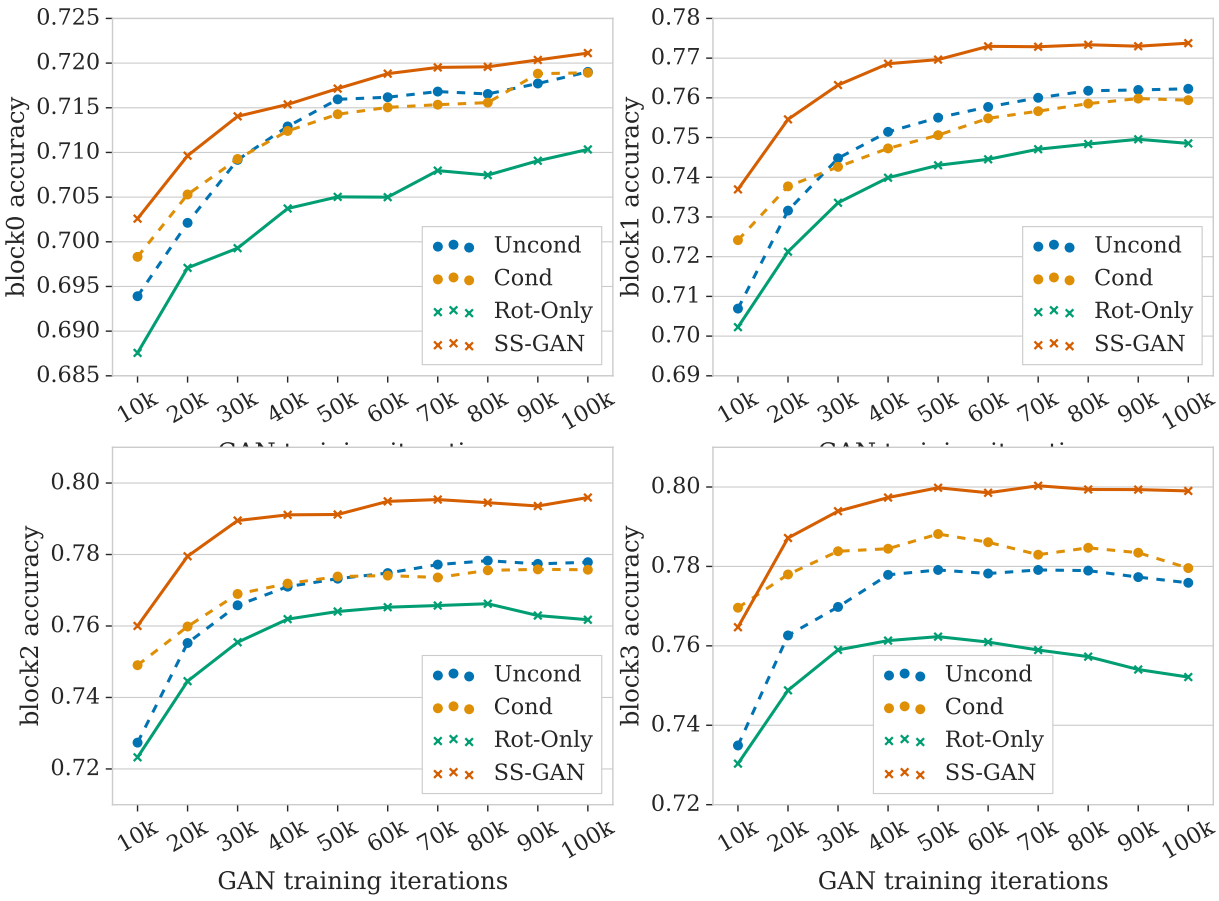


Figure 2: Top 1 accuracy on CIFAR10 with training steps from 10k to 100k.

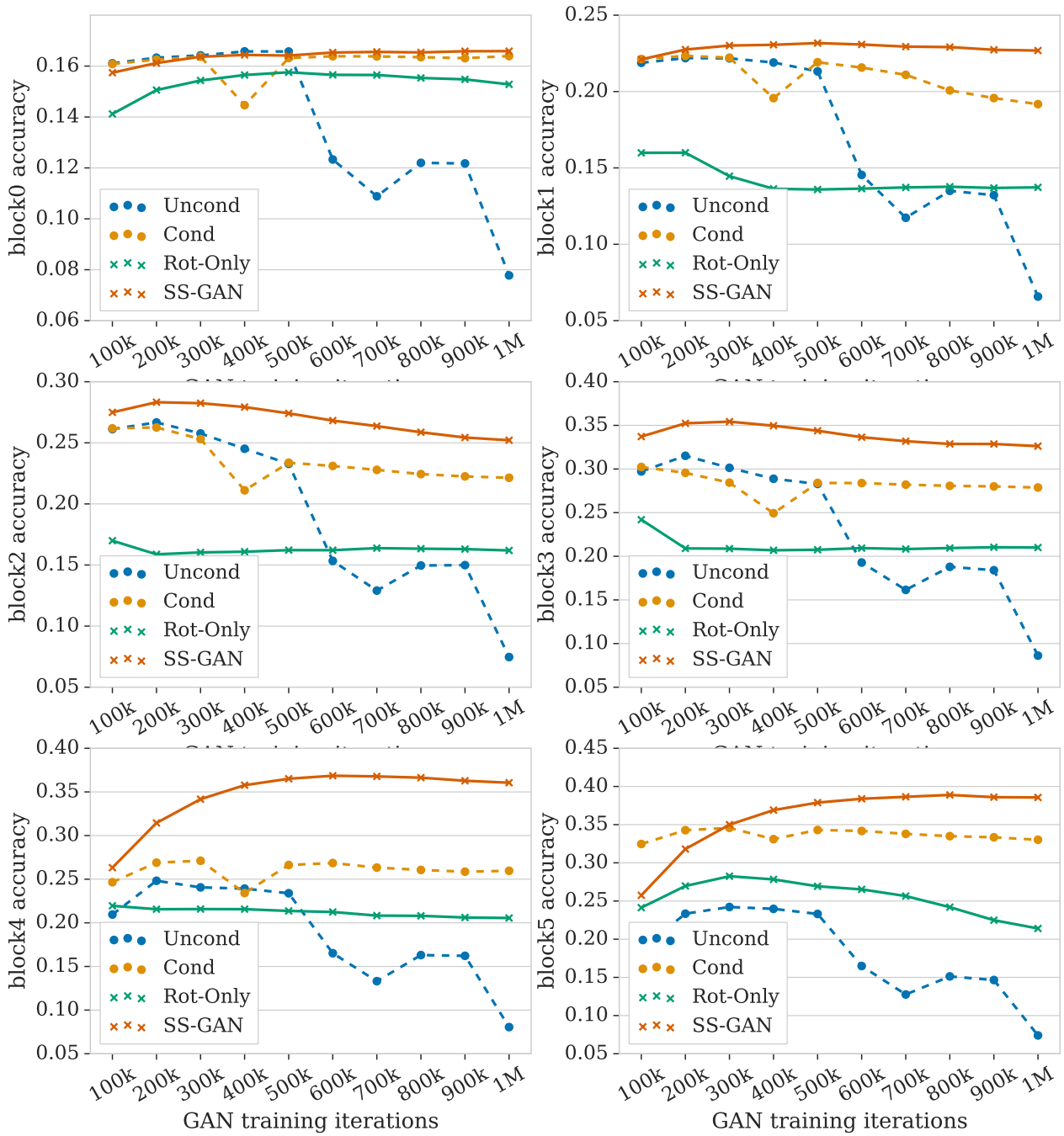


Figure 3: Top 1 accuracy on IMAGENET validation set with training steps from 10k to 1M.

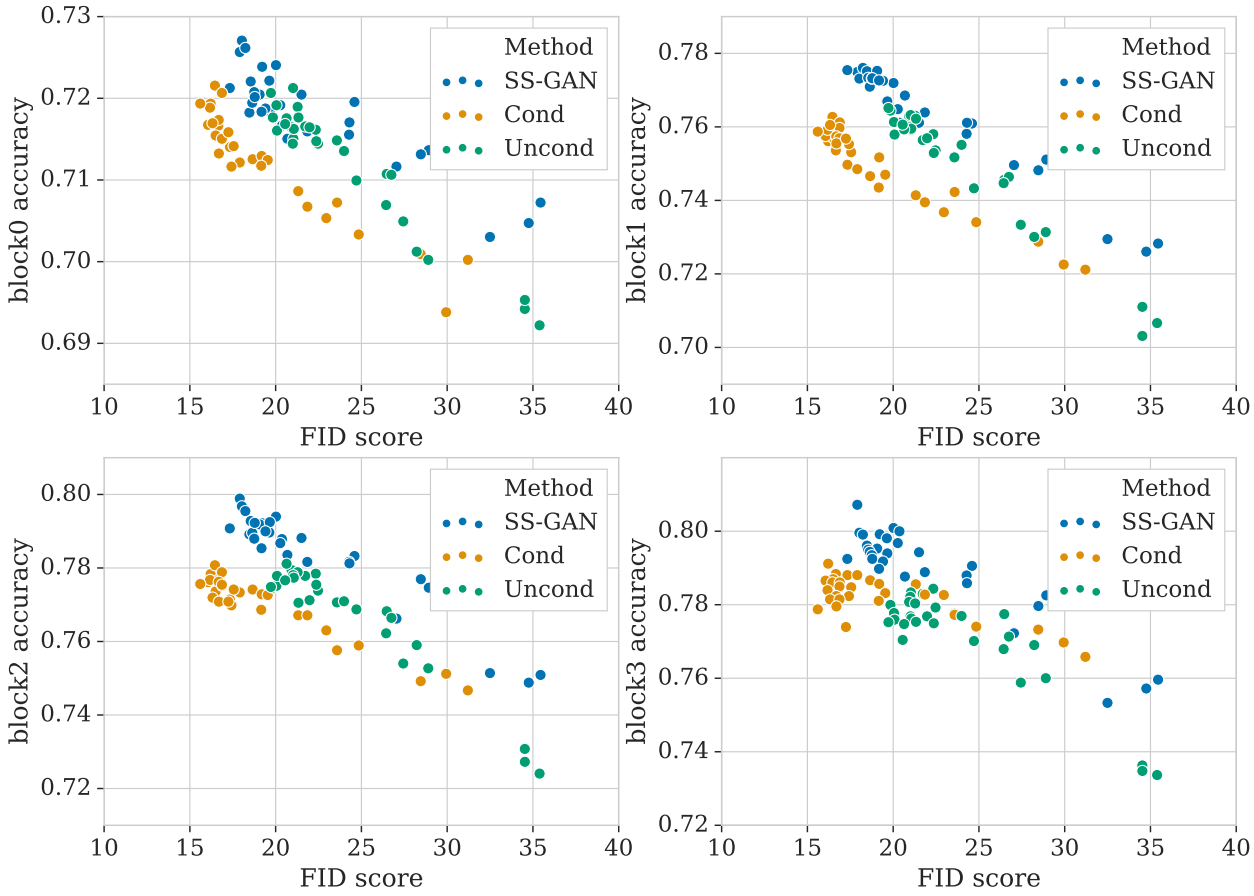


Figure 4: Correlation between top-1 accuracy and FID score for different numbers of GAN training steps from 10k to 100k on CIFAR10. Overall, the representation quality and the FID score is correlated for all methods. The representation quality varies up to 4% with the same FID score.

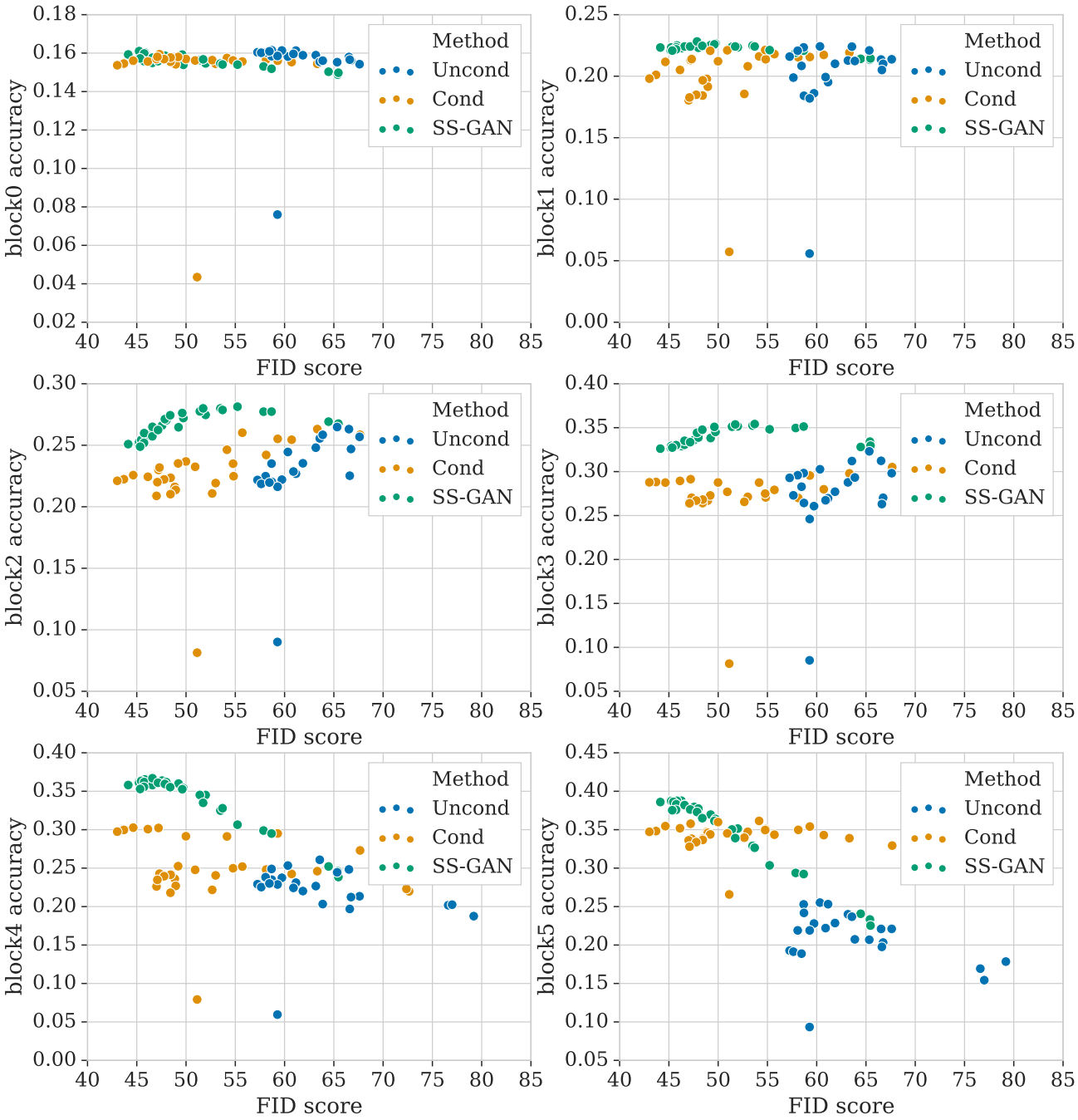


Figure 5: Correlation between top-1 accuracy and FID score for different numbers of GAN training steps from 100k to 1M on IMAGENET. Representation quality and FID score are not correlated on any of block0 to block4. This indicates that low level features are being extracted, which perform similarly on the IMAGENET dataset. Starting from block4, SS-GAN attains better representation as the FID score improves.