Table 1. Compression Ratios and Image Quality Metrics for MNIST, CIFAR-10, Celeb-A. The respective metrics (*IS*, Inception Score and *FID*, Frechet Inception Distance) are shown for both the student GAN (Stu.) compared to a regularly trained GAN (Reg.) of corresponding size. For MNIST, CIFAR-10, and Celeb-A, the ratio shown is with respect to a teacher generator size of depth $d = 256$ (47,324,929 parameters, *IS* = 7.02), $d = 64$ (3,573,697 parameters, *FID* = 7.42), and $d = 128$ (12,652,417 parameters, *FID* = 4.39) respectively. All teacher GANs were trained with a discriminator of corresponding teacher depth.

		GAN Size (d)							
		2	4	8	16	32	48	64	128
No. of Parameters		28,351	62,077	145,657	377,329	109,8721	216,4177	3,573,697	12,652,417
MNIST	Ratio	1669:1	762:1	325:1	125:1	43:1	—	13:1	4:1
	IS (Stu.)	5.80	6.41	6.60	6.83	6.87	—	6.93	6.97
	IS (Reg.)	1.86	3.63	4.73	5.07	6.08	—	6.51	6.63
CIFAR-10	Ratio	126:1	58:1	25:1	9:1	3:1	2:1	—	—
	FID (Stu.)	11.76	11.00	9.57	8.39	7.80	7.58	—	—
	FID (Reg.)	38.72	14.28	11.85	9.90	7.86	7.64	—	—
Celeb-A	Ratio	446:1	204:1	87:1	34:1	12:1	6:1	4:1	—
	FID (Stu.)	12.15	10.97	8.78	6.29	4.84	—	4.54	—
	FID (Reg.)	45.49	18.72	11.06	9.14	5.05	—	4.62	—

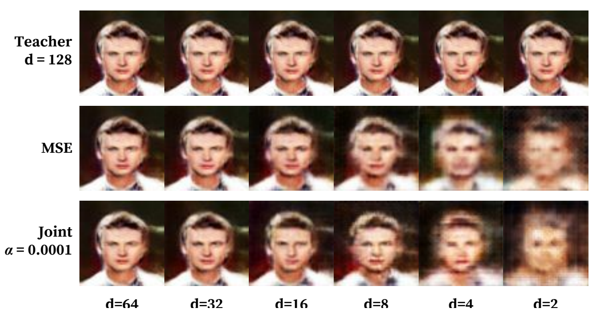

Figure 11. Compression artifacts on the Celeb-A dataset from images generated from a teacher GAN ($d = 128$), student GAN trained using MSE loss and student GAN trained using joint loss at $\alpha = 0.0001$.

smallest compressed GANs before significant observable degradation is present in the generated images. This observation suggests a potential limit to compression depending on the complexity of the data set.

6.2. Our Contributions

Our work contributes to the topic of GAN compression. To summarize, we have made the following contributions in this paper:

- We have developed two compression schemes for GANs using a student-teacher learning architecture (Figures 3, 4).
- We have evaluated the proposed compression methods over MNIST, CIFAR-10, and Celeb-A datasets. Our results show that the quality of generated imagery is maintained at high compression rates (1669:1, 58:1, 87:1 respectively) as measured by the Inception Score and Frechet Inception Distance metrics.
- We show that training a GAN of the same size without knowledge distillation produces comparatively diminished results, supporting the conjecture that overparameterization is both helpful and necessary for neural networks to find a good function for GANs.
- We observe a qualitative limit to GAN's compression for all the aforementioned datasets. We conjecture that there exists a fundamental compression limit of GANs similar to Shannon's compression theory (MacKay, 2002).

7. Conclusion

Overall, we have demonstrated that applying the knowledge distillation method to GAN training can produce compressed generators without loss of quality or generalization. More specifically, we demonstrated that the student generators are able to outperform a traditionally trained GAN of the same size and approximate the underlying function of the teacher generator for the whole latent space. This further supports the necessity for over-parameterization when training an effective generator prior to distillation. Further, a qualitative limit to GAN compression has been observed for MNIST, CIFAR-10 and Celeb-A datasets.

References

Allen-Zhu, Z., Li, Y., and Liang, Y. Learning and generalization in overparameterized neural networks, going