A. Training settings

The architectural and training settings used in Sections 3, 4, and 5 are presented in Tables 4, 5, and 6, respectively. For the FID comparison on CIFAR-10 and CelebA in Section 5, we used the same architectures as Table 6 but with a batch size of 64 on both datasets, and ran for 78K iterations on CIFAR-10 and 125K iterations on CelebA.

B. Sigmoid initial value

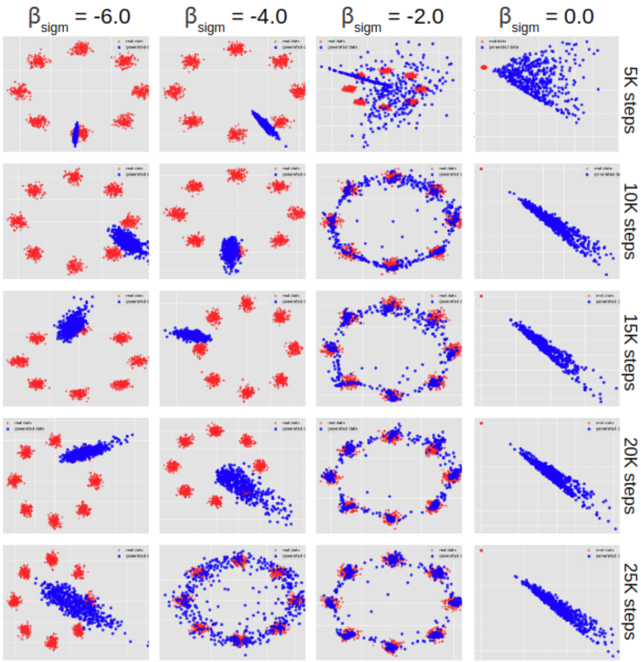
In Figure 8, we show and discuss the effects of using different β_{sigm} on α_{sigm} on the toy dataset, giving more insights regarding the choice of $\beta_{sigm} = -1.8$ mentioned in Section 3.

C. Toy dataset comparisons

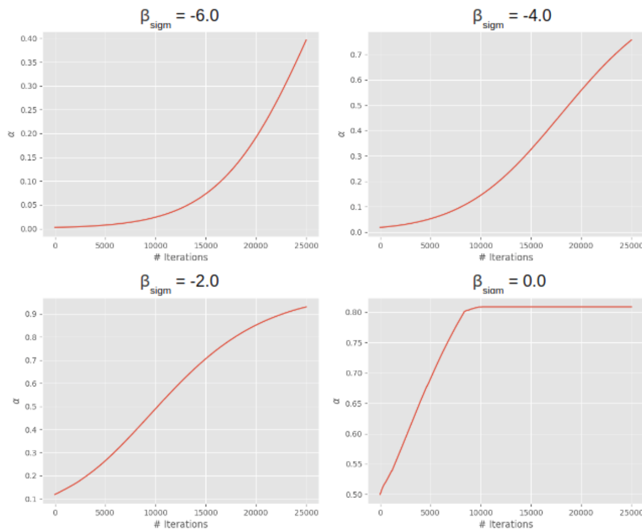
Figure 9 shows how different methods compare using the above mentioned toy dataset. We compared microbatchGAN's results (K = 8, α_{sigm}) to the standard GAN ([12]), UnrollledGAN ([26]), D2GAN ([30]), and MGAN ([15]). We observe bigger sample diversity with our method, while still approximating the real data distribution.

D. Extended Results

Additional results for CIFAR-10, STL-10, and ImageNet are presented bellow.



(a) Generated samples.



(b) α evolution.

Figure 8. Analysis of self-learning α_{sigm} with different initial values of β. The generated samples in (a) show that using lower β_{sigm} values lead the model to mode collapse, since only low α values are used throughout the whole training. On the other hand, using higher values, *e.g.*, $\beta_{sigm} = 0.0$, leads to a steeper increase of α values, inducing the model to only generate varied, but not realistic, samples. We empirically found that using $-2.0 \leq \beta_{sigm} \leq -1.8$ led to diverse plus realistic looking samples from early iterations due to the mild, yet meaningful, increase of α throughout training. The evolution of α's values are presented in (b).

Table 4. Training settings for the toy dataset.

	Feature maps	Nonlinearity
$G(z) : z \sim Normal(0, I)$	256	
Fully connected	128	ReLu
Fully connected	128	ReLu
Fully connected	2	Linear
$D(x)$	2	
Fully connected	128	ReLu
Fully connected	1	Softplus
Number of discriminators	8	
α (static)	$\{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$	
α (self-learned)	$\{\alpha_{sigm}, \alpha_{soft}, \alpha_{tanh}, \alpha_{ident}\}$	
Batch size	512	
Iterations	25K	
Optimizer	Adam ($lr = 0.0002, \beta_1 = 0.5$)	

Table 5. Training settings for MNIST, CIFAR-10, and CelebA.

	KERNEL	STRIDES	FEATURE MAPS	BATCH NORM.	NONLINEARITY
$G(z) : z \sim Uniform[-1, 1]$	-	-	100	-	-
TRANSPOSED CONVOLUTION	3×3	4×4	128	YES	RELU
TRANSPOSED CONVOLUTION	5×5	2×2	64	YES	RELU
TRANSPOSED CONVOLUTION	5×5	2×2	32	YES	RELU
TRANSPOSED CONVOLUTION	5×5	2×2	1/3	NO	TANH
$D(x)$	-	-	$32 \times 32 \times 1/3$	-	-
CONVOLUTION	3×3	2×2	32	YES	LEAKY RELU (0.2)
CONVOLUTION	3×3	2×2	64	YES	LEAKY RELU (0.2)
CONVOLUTION	3×3	2×2	128	YES	LEAKY RELU (0.2)
FULLY CONNECTED	-	-	1	NO	SIGMOID
NUMBER OF DISCRIMINATORS	$\{1, 2, 5, 10\}$				
α (STATIC)	$\{0\}$				
α (SELF-LEARNED)	$\{\alpha_{sigm}, \alpha_{soft}, \alpha_{tanh}, \alpha_{ident}\}$				
BATCH SIZE	100				
ITERATIONS	50K				
OPTIMIZER	ADAM ($lr = 0.0002, \beta_1 = 0.5$)				

Table 6. Training settings for CIFAR-10, STL-10, and ImageNet.

	KERNEL	STRIDES	FEATURE MAPS	BATCH NORM.	NONLINEARITY
$G(z) : z \sim Uniform[-1, 1]$	-	-	100	-	-
TRANSPOSED CONVOLUTION	3×3	4×4	256	YES	RELU
TRANSPOSED CONVOLUTION	5×5	2×2	128	YES	RELU
TRANSPOSED CONVOLUTION	5×5	2×2	64	YES	RELU
TRANSPOSED CONVOLUTION	5×5	2×2	1/3	NO	TANH
$D(x)$	-	-	$32 \times 32 \times 1$	-	-
CONVOLUTION	3×3	2×2	64	YES	LEAKY RELU (0.2)
CONVOLUTION	3×3	2×2	128	YES	LEAKY RELU (0.2)
CONVOLUTION	3×3	2×2	256	YES	LEAKY RELU (0.2)
FULLY CONNECTED	-	-	1	NO	SIGMOID
NUMBER OF DISCRIMINATORS	$\{2\}$				
α (SELF-LEARNED)	$\{\alpha_{sigm}, \alpha_{soft}, \alpha_{tanh}\}$				
BATCH SIZE	100				
ITERATIONS	200K, 400K, 1M				
OPTIMIZER	ADAM ($lr = 0.0002, \beta_1 = 0.5$)				

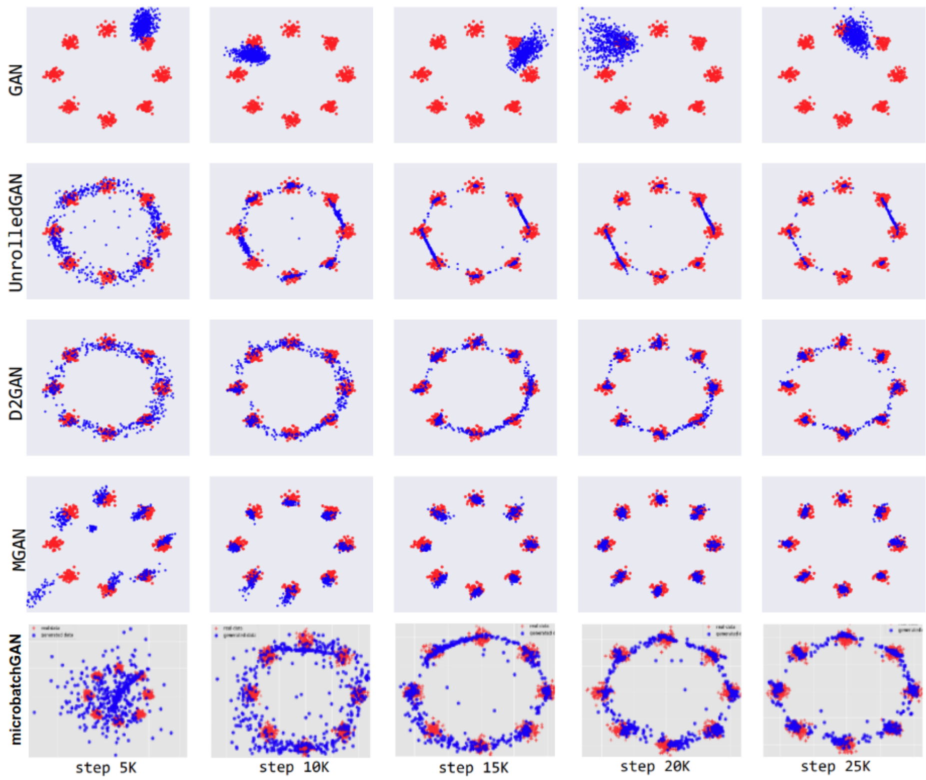


Figure 9. Method comparisons on the toy dataset.

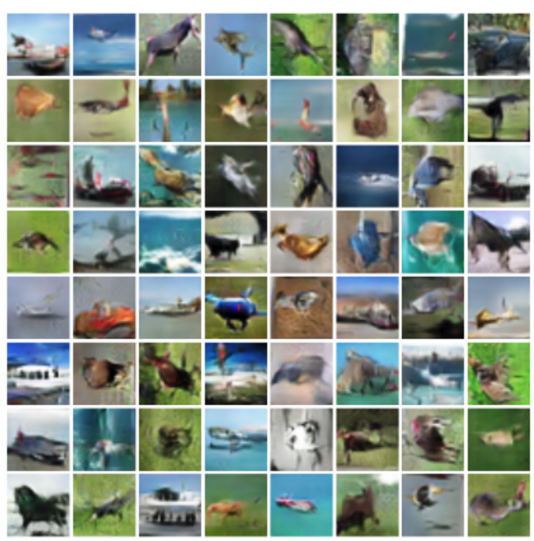


Figure 10. CIFAR-10 extended results using K = 2 and α_{sigm}.

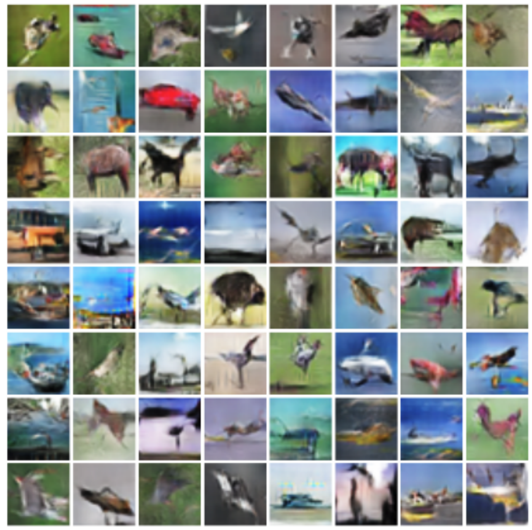


Figure 11. CIFAR-10 extended results using K = 2 and α_{soft}.

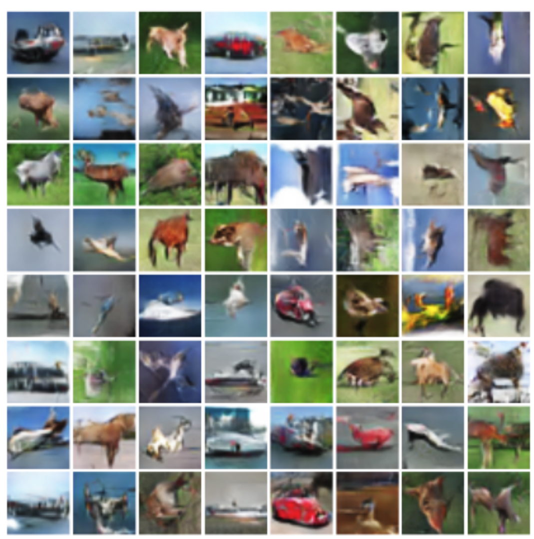


Figure 12. CIFAR-10 extended results using K = 2 and α_{tanh}.

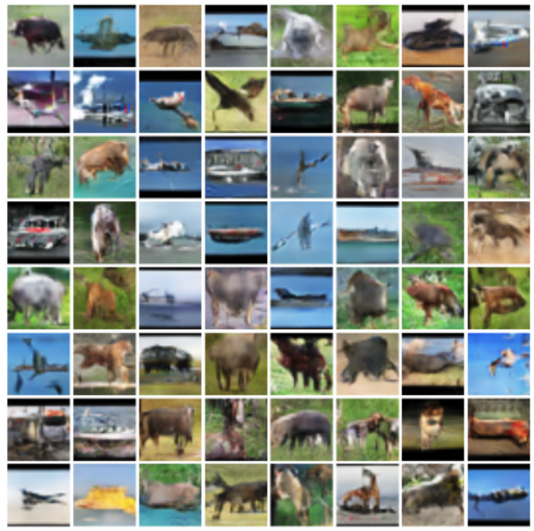


Figure 13. STL-10 extended results using K = 2 and α_{sigm}.

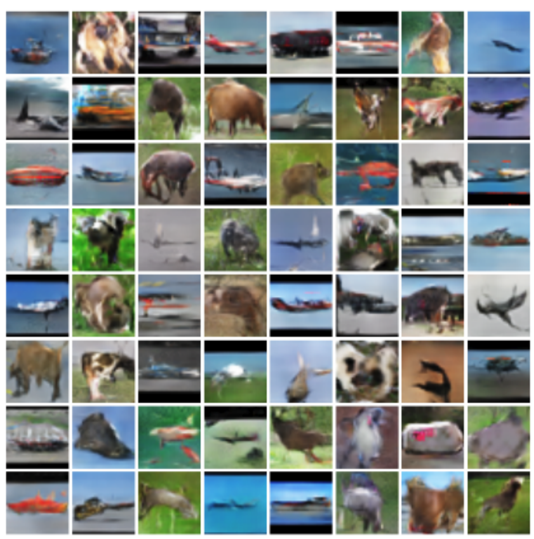


Figure 14. STL-10 extended results using K = 2 and α_{soft}.

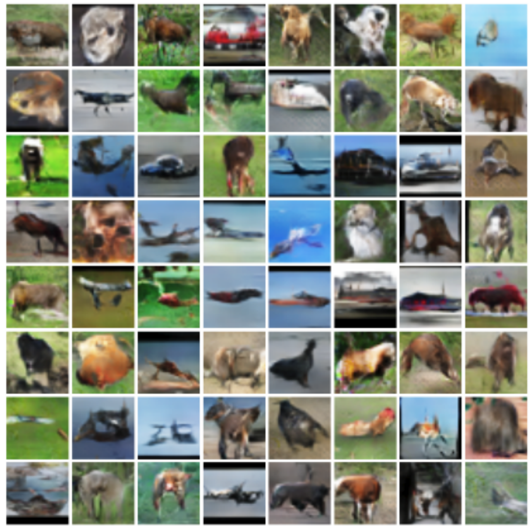


Figure 15. STL-10 extended results using K = 2 and α_{tanh}.

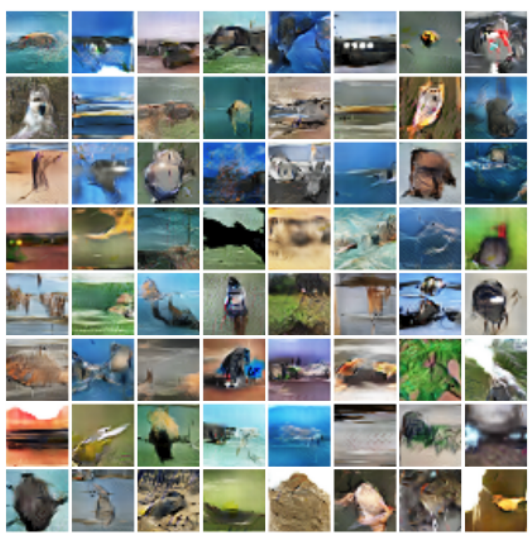


Figure 16. ImageNet extended results using K = 2 and α_{sigm}.

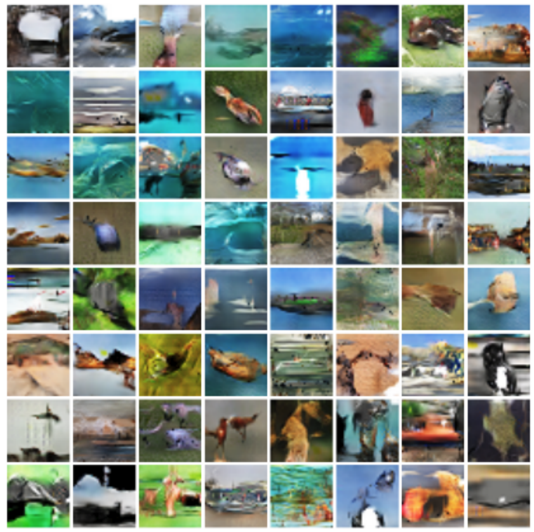


Figure 17. ImageNet extended results using K = 2 and α_{sigm}.

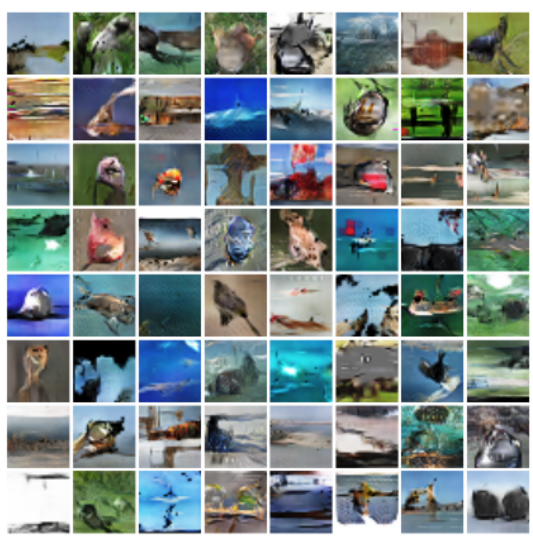


Figure 18. ImageNet extended results using K = 2 and α_{sigm}.