Generating Multiple **Objects at Spatially Distinct Locations**



Tobias Hinz, Stefan Heinrich, Stefan Wermter

INTRO

- one **challenge** with Generative Adversarial Nets is that it is hard to control the layout of the generated scenes
- a popular approach for increased control is to use scene layouts as additional input, however, this requires a lot of labeling
- our approach only needs a bounding box (thereby specifying location and shape) and a class label for each of the foreground objects we want to be in the image

METHODS

- we add an **object pathway** to the generator and the discriminator of a normal GAN
- the **generator** applies the object pathway iteratively at each bounding box location, conditioned by the label
- the **discriminator's** object pathway checks each bounding box location iteratively and evaluates whether the specified object is actually recognizable

RESULTS

- experiments show that the architecture does lead to increased control over the image generation process
- experiments on the Multi-MNIST and CLEVR datasets show that the architecture can generalize to novel object characteristics and locations, as well as to different numbers of generated objects per scene
- experiments on the MS-COCO dataset show that the **object pathway learns features** for the individual foreground objects and can lead to an overall **higher quality** of the generated images (based on IS and FID)



Take a picture to get to our Github page



Ge	nerator	
A)	image c	apt
	dog standing	ma g in
b	ounding box	
la	abels I _{onehoti}	
	person	
	person —	
	dog —	



We extend the GAN architecture in order to control object locations and identities in generated images.

mages with differ	rent number of digits per image
5 9,7 Z	3,3,8 5
3,8,2 9,0,3,4 JJJ	9,6,1,5,9 6,1,8,5,2 6,1,1,8,0 6,1,1,1,8,0 6,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1
es with changed	size of bounding box of some digits
8,0 7 ,0,9 7	Z B B B Z Z D Z 2,7,6 8,9,6 2,4,8 Z 3,2,6 9,8,4
oottom half: 5-9	Failure case: training: botttom half empty
2 3 7,3,9 2 0,2,9	5,0,4 3,6,9 5,7 5,0,4 1 1,8,9 1 9,5,7 1,8,9 1 1,8,9 1 1,8,9 1 1,8,9 1 1,0,3
isabled	Object Pathway in Discriminator disabled
3,8,1 6,2,0 2 2 2 2 2 2 2 2 2 2 2 2 2	
rator disabled	All Object Pathways disabled
7,9 7 6,1,0	7777777777777

Gen	erated d	igits whe	n traineo	d with ima	ages conta	aining di	gits only	in the to	p half
720	446	455	809	59	1,0,7	3, 0, 1	9,4,5	5,9,6	6, 5, 4
							5	3 10	1)S

7,2,6	4,4,6	4, 5, 5	8,0,9	5,9,1	(L) <	3(2) (1	3 10	
4 ,8,1	4. 3,9	Z 3 Q 7,3,0	38 Z 3,8,2	9,6,6	9,4,1	8,4,6	1,2,1	6,9,9	6,3,5 II55
3,4 ,6	5,9,5	9 3 5 9,3,5	6 ,4,9	1 ,3,7	5,5,3	4,1,5	3,7,9	1,6,1	9,7,2
Z]]][] 2,7,8	03 0,3,7) 7 X 3,4,2	6,7,1	9, 0, 0	2,2,4	4,8,3	4,2,1	7,6,4	9,2,1

0 ⁸ 8		•, • •	•		▫▯▫▫ ▫	9 ^{* 0}	
		Gene	ralization -	- Color			
• .	•	- 0 ⁰ 0	10 .	·a 8·•	5 a B	••••••	•

Model	Resolution	$ $ IS \uparrow	$ $ FID \downarrow
GAN-INT-CLS	$ 64 \times 64$	7.88 ± 0.07	60.62
StackGAN-V2	256×256	8.30 ± 0.10	81.59
StackGAN	256×256	8.45 ± 0.03^{1}	74.05
PPGN	227×227	9.58 ± 0.21	
ChatPainter (StackGAN)	256×256	9.74 ± 0.02	
Semantic Layout	128×128	11.46 ± 0.09^2	
HDGan	256×256	11.86 ± 0.18	71.27 ± 0.12^{3}
AttnGAN	256×256	23.61 ± 0.21^4	33.10 ± 0.11^3
StackGAN + Object Pathways (Ours) ⁵ AttnGAN + Object Pathways (Ours)	$\begin{vmatrix} 256 \times 256 \\ 256 \times 256 \end{vmatrix}$	$\begin{array}{c} 12.12 \pm 0.31 \\ 24.76 \pm 0.43 \end{array}$	55.30 ± 1.78 33.35 ± 1.15

Recently updated to 10.62 ± 0.19 in its source code.

² When using the ground truth bounding boxes at test time (as we do) the IS increases to 11.94 ± 0.09 . ³ FID score was calculated with samples generated with the pretrained model provided by the authors. ⁴ The authors report a "best" value of 25.89 ± 0.47 , but when calculating the IS with the pretrained model provided by the authors we only obtain an IS of 23.61. Other researchers on the authors' Github website report a similar value for the pretrained model.

We use the updated source code (IS of 10.62) as our baseline model.

Table 1: Comparison of the Inception Score (IS) and Fréchet Inception Distance (FID) on the MS-COCO data set for different models. Note: the IS and FID values of our models are not necessarily directly comparable to the other models, since our model gets, in addition to the image caption, up to three bounding boxes and their respective object labels as input at test time.

Take a picture to get to our Github page

