

Up and Down Residual Blocks and Text to Image Algorithm

Hongzhi Liu

December 23rd, 2018

CONTENT



1. Review of Previous Work
 2. Additional Experiments
 3. Research about Text-to-image
 4. Further Research
-



1. Up and Down Residual Blocks

For the last 3 months, our team proposed up and down residual blocks in the paper submitted, dubbed upResBlock and downResBlock respectively.

In summary, the contribution of our method is three-fold as follows:

- 1) The upResBlocks and downResBlocks for convolutional GANs;**
- 2) Comprehensive experiments were carried out to evaluate its universality;**
- 3) Extensive experiments about text-to-image synthesis.**

2. The architecture of method

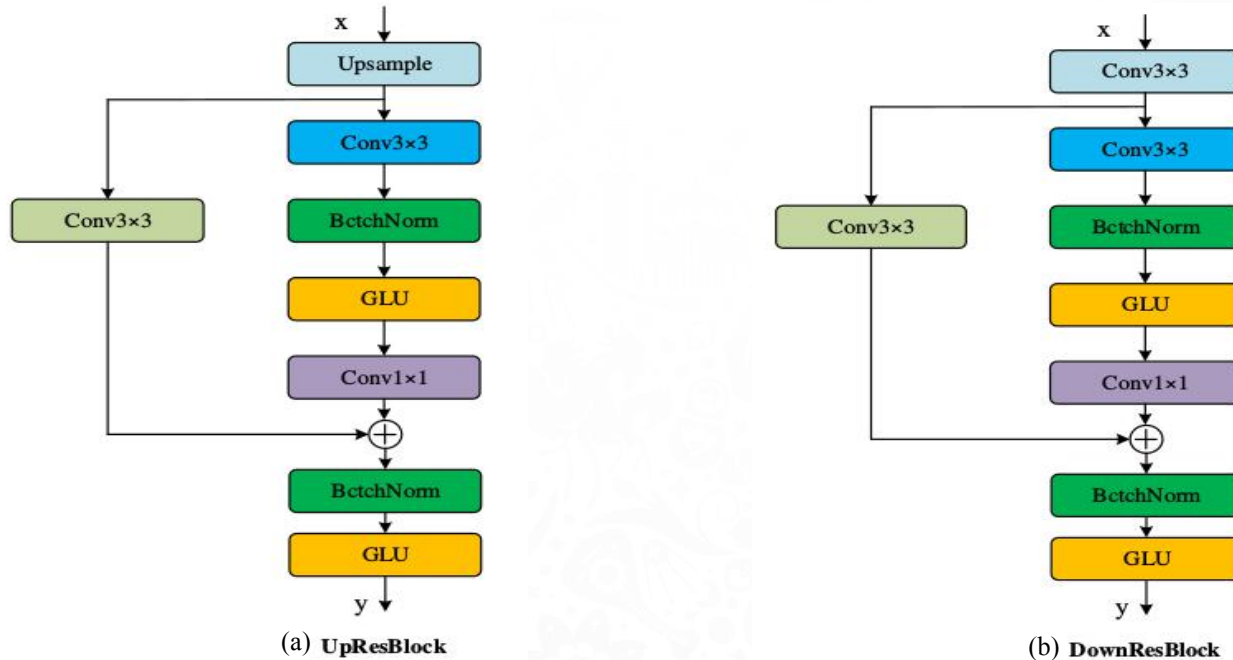


Figure 1. The architecture of upResBlock and downResBlock. (a) our upResBlock could make all information pass through and also make backward propagation of gradients more efficiently with the help of shortcut. (b) The downResblock could preserve the full features especially the usefull ones.

3. Experimental results

Network	Up and Down Residual Block		Inception Score
	upResBlock	downResBlock	CIFAR 10
GAN			2.55 \pm .18
	✓		2.69 \pm .17
		✓	2.65 \pm .14
	✓	✓	2.73 \pm .18
LSGAN			2.39 \pm .15
	✓		2.66 \pm .15
		✓	2.59 \pm .15
	✓	✓	2.94 \pm .23
DCGAN			6.57 \pm .06
	✓		6.72 \pm .09
		✓	6.84 \pm .09
	✓	✓	7.50 \pm .10
WGAN-GP			7.86 \pm .07
	✓		8.31 \pm .05
		✓	7.93 \pm .07
	✓	✓	8.40 \pm .06

Table 1. Inception score results of the GAN models without (blank) or with (✓) upResBlock/downResBlock for image generation on CIFAR10 dataset. The values in bold font indicate the best results.

1. Image-to-image translation—Pix2Pix

Network	Up and Down Residual Blocks		FCN scores (Mean pixel accuracy)
	upResBlocks	downResBlocks	
pix2pix			0.71
	✓		0.74
		✓	0.63
	✓	✓	0.65

Table 2. FCN score results of pix2pix model without (blank) or with (✓) upResBlock/downResBlock for image generation on cityscapes datasets. The values in bold font indicate the best results.

Additional Experiments



1. Image-to-image translation—Pix2Pix



(a)



(b)



(c)

Figure 2. Example results of pix2pix on cityscapes datasets at 256x256 resolution. (a) the image generated by pix2pix with upResBlocks. (b) the image generated by pix2pix with downResBlocks. (c) the image generated by pix2pix with up and down ResBlocks.

2. Image-to-image translation—CycleGAN

Network	Up and Down Residual Blocks		FCN scores (Mean pixel accuracy)
	upResBlocks	downResBlocks	
CycleGAN			0.52
	✓		0.54
		✓	0.60
	✓	✓	0.65

Table 3. FCN score results of the CycleGAN model without (blank) or with (✓) upResBlock/downResBlock for image generation on cityscapes datasets. The values in bold font indicate the best results.

Additional Experiments



2. Image-to-image translation—CycleGAN



(a)



(b)



(c)

Figure 3. Example results of CycleGAN on Cityscapes datasets at 256x256 resolution. (a) the image generated by CycleGAN with upResBlocks. (b) the image generated by CycleGAN with downResBlocks. (c) the image generated by CycleGAN with up and down ResBlocks.

2. Image-to-image translation—CycleGAN

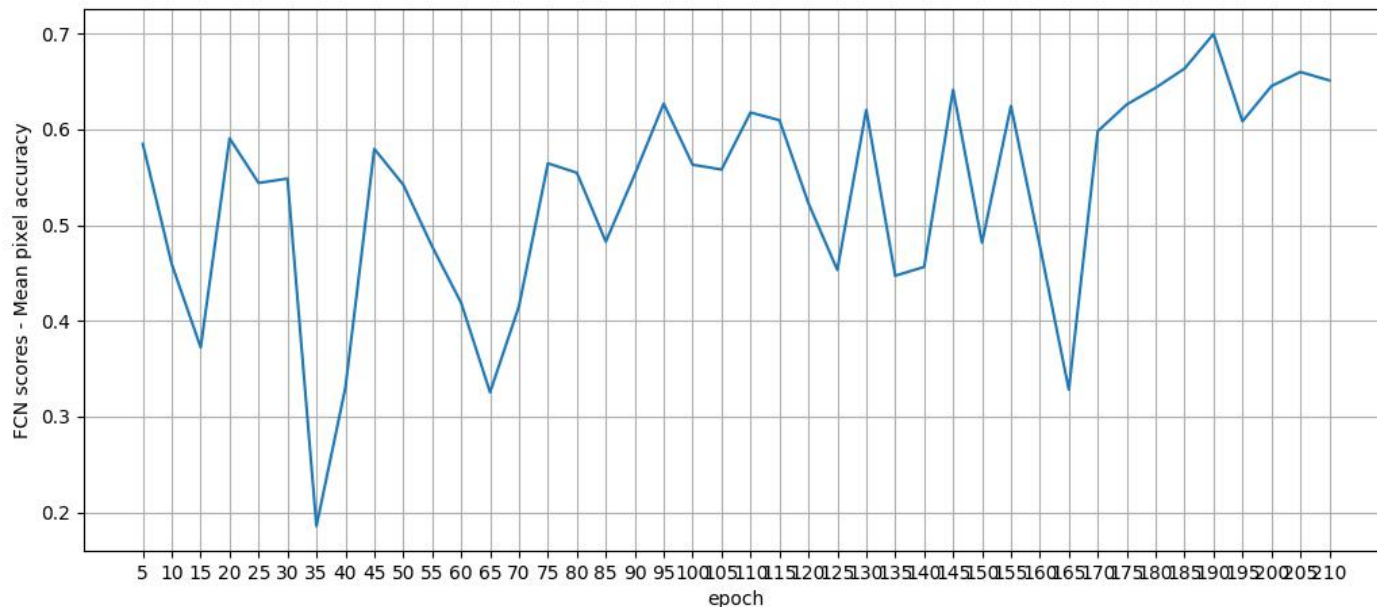


Figure 4. The curve shows mean pixel accuracy changes from 5 epoch to 210 epoch in CycleGAN experiment with up and down Resblocks. There is one peak at 190 epoch and the lowest value appears at 35 epoch.

1. Semantic Image Manipulation

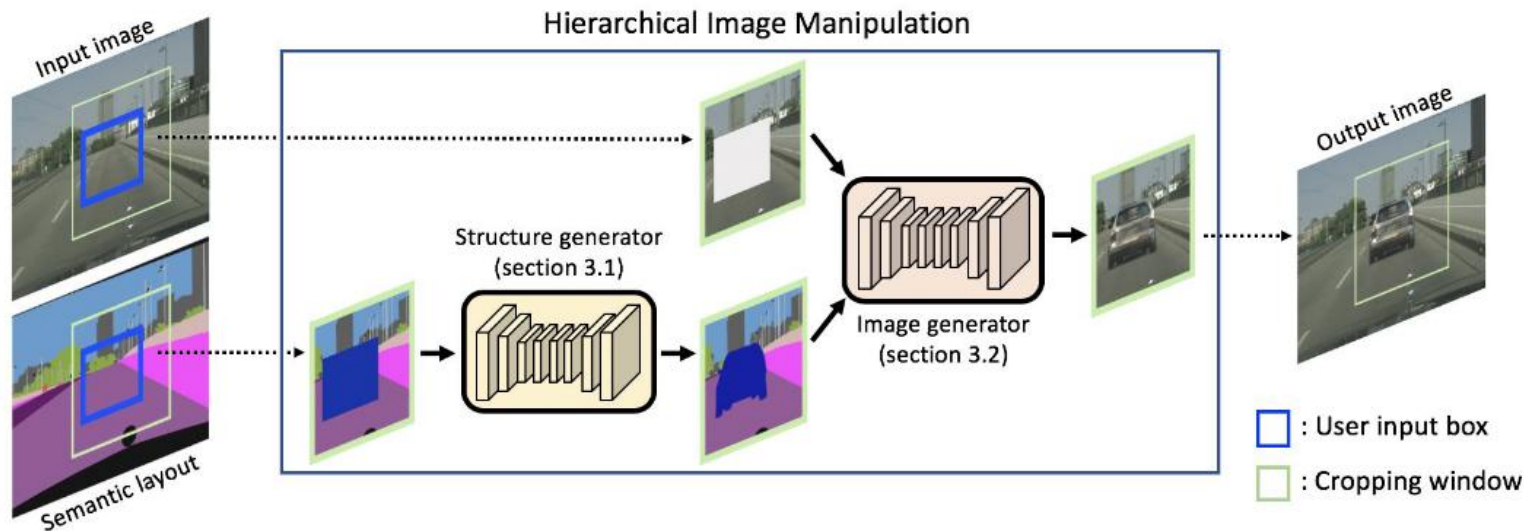


Figure 5. Overall pipeline of the proposed semantic manipulation framework.

2. Hierarchical Text-to-Image Synthesis

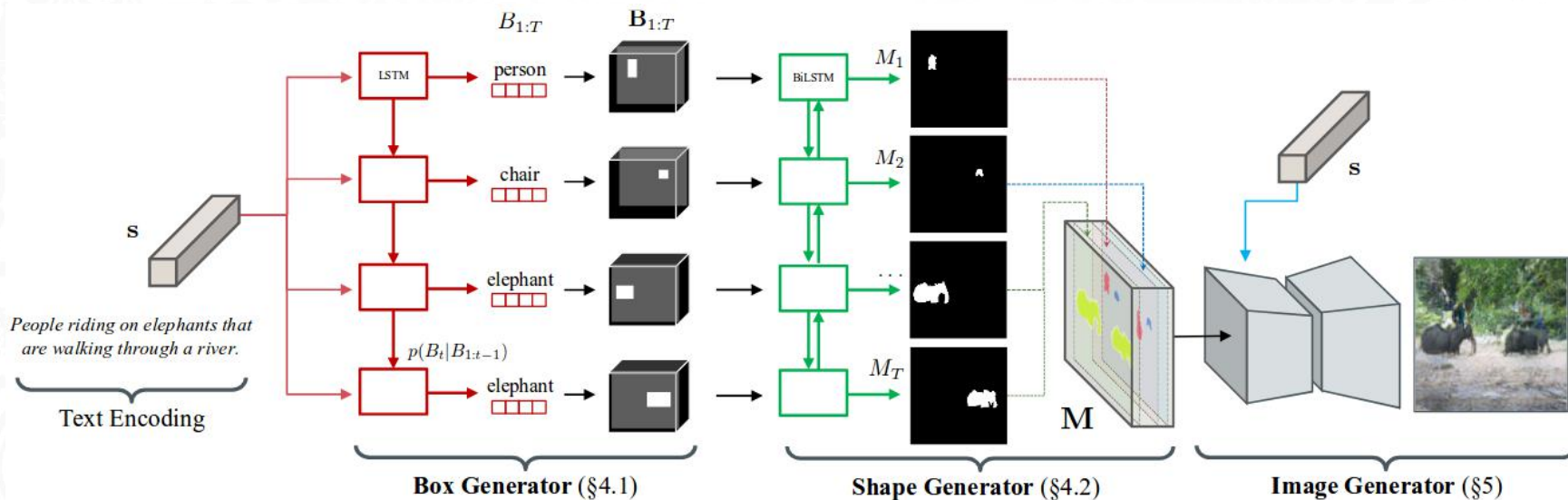


Figure 6. Given a text embedding, the algorithm first generates a coarse layout of the image by placing a set of object bounding boxes using the box generator, and further refines the object shape inside each box using the shape generator. Combining outputs from the box and the shape generator leads to a semantic label map defining semantic structure of the scene. Conditioned on the inferred semantic layout and the text, a pixel-wise image is finally generated by the image generator.



1. Additional Experiments for up and down ResBlocks
2. AttnGAN with semantic segmentation image datasets
3. Looking for better advice and methods

Q & A