Supplementary Material Outline

Section 1 lists all categories we used in training our models. Section 2 compares the performance of MRU to some other models on CIFAR-10. Section 3 shows samples of generated images from all 50 categories.

1. Category list

Here are the 50 categories we use for training and testing our models: airplane, ant, apple, banana, bear, bee, bell, bench, bicycle, candle, cannon, car, castle, cat, chair, church, couch, cow, cup, dog, elephant, geyser, giraffe, hammer, hedgehog, horse, hotdog, hourglass, jellyfish, knife, lion, motorcycle, mushroom, pig, pineapple, pizza, pretzel, rifle, scissors, scorpion, sheep, snail, spoon, starfish, strawberry, tank, teapot, tiger, volcano, zebra.

2. Evaluation of MRU on CIFAR-10

We introduce the Masked Residual Unit (MRU) to improve generative deep networks by giving repeated access to the conditioning signal (in our case, a sketch). But this network building block is also quite useful for classification tasks. We compare the performance of the MRU and other recent architectures on CIFAR-10 and show that the MRU performance is on par with ResNet. Accuracy numbers for other models are obtained from their corresponding papers. For convenience, we call the improved ResNet "ResNet-v2" in the table. In "MRU-108, LeakyReLU gate", we substitute the sigmoid activations in our MRU units with LeakyReLU [4], and normalize obtained masks to the range of $[0, 1]$.

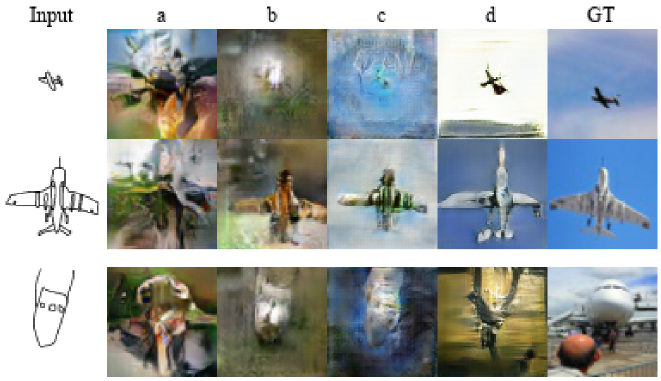
Model	error (%)
NIN [3]	8.81
Highway [5]	7.72
ResNet-110 [1]	6.61
ResNet-1202 [1]	7.93
ResNet-v2-164 [2]	5.46
MRU-108	6.34
MRU-108, LeakyReLU gate	**5.83**

Table 1: Comparison of error rates on CIFAR-10. Lower is better.

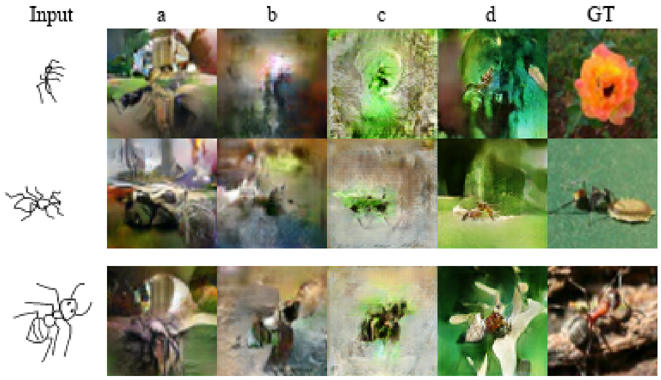
3. Samples from all 50 categories

Here we present samples from all 50 categories from pix2pix variants and our methods for comparison. Each category contains three input samples, among which the third sample is a failure case for our method. The six columns in each figure are: (Input) input sketch, (a) pix2pix on Sketchy, (b) pix2pix on Augmented Sketchy, (c) Label-supervised pix2pix on Augmented Sketchy, (d) our method, (GT) ground truth image.

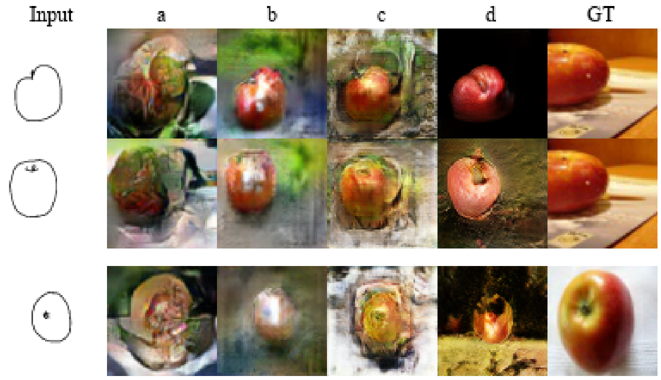
airplane

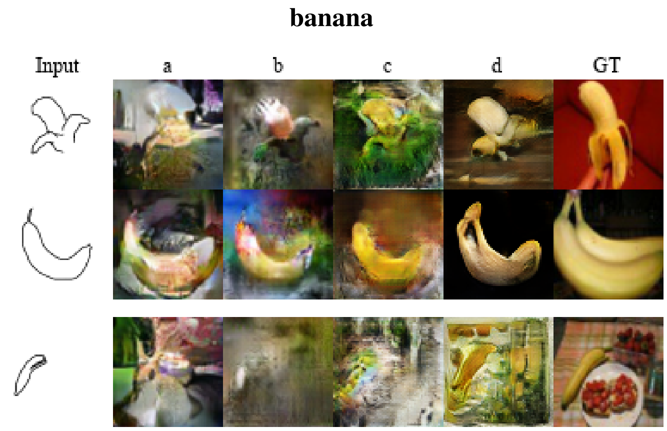


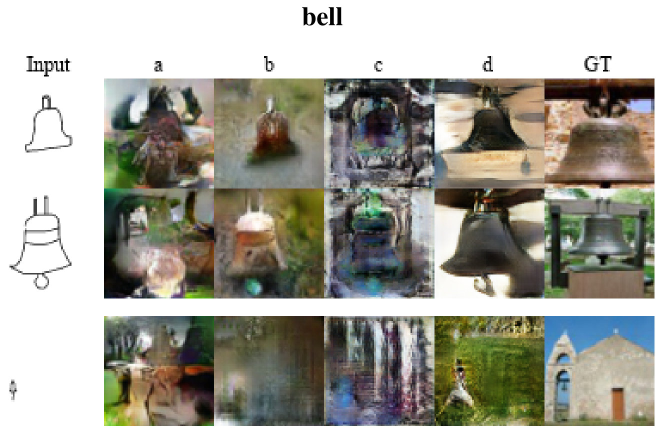
ant

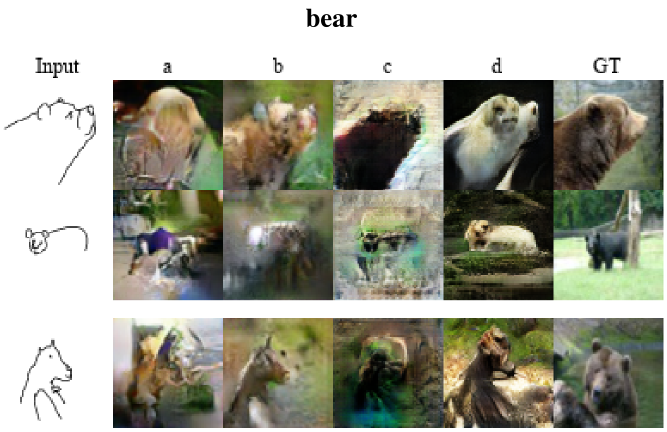


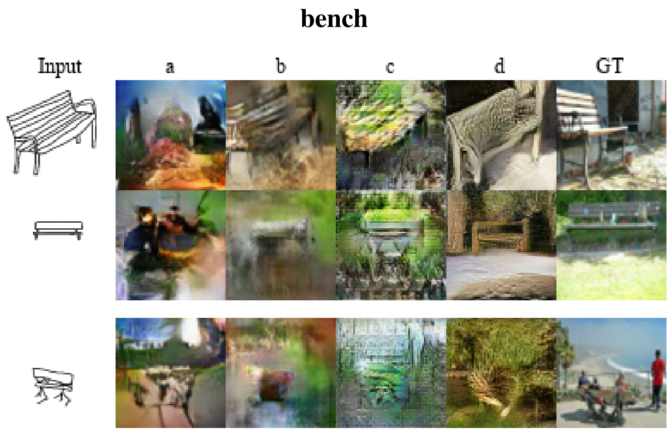
apple

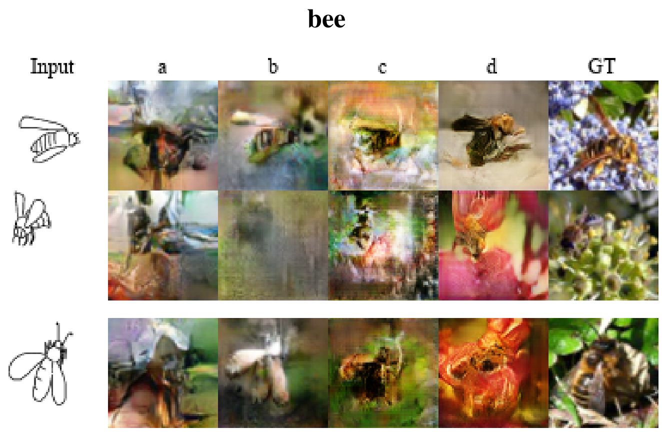


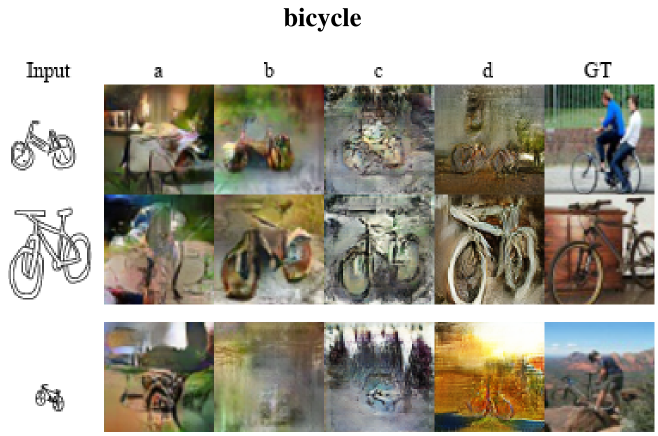

banana
Input
a
b
c
d
GT


bell
Input
a
b
c
d
GT

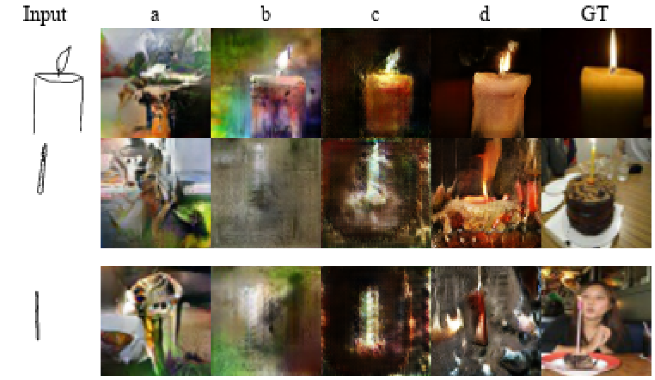

bear
Input
a
b
c
d
GT


bench
Input
a
b
c
d
GT

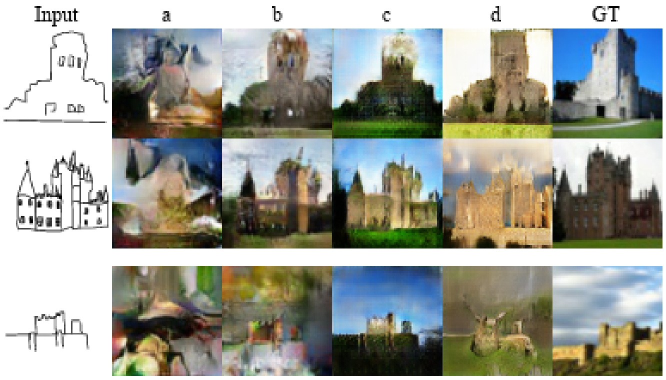

bee
Input
a
b
c
d
GT


bicycle
Input
a
b
c
d
GT

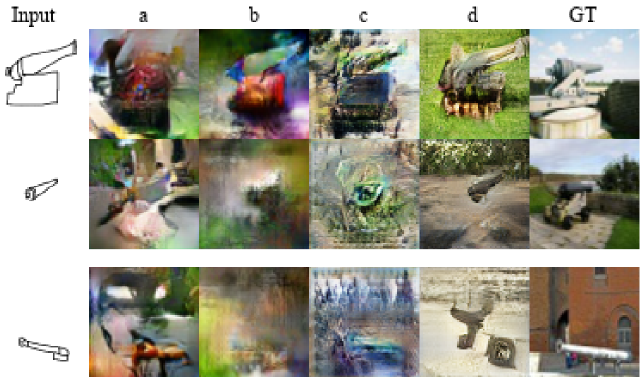
candle



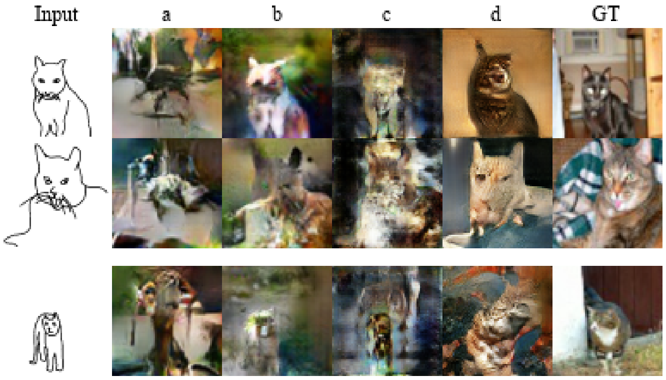
castle



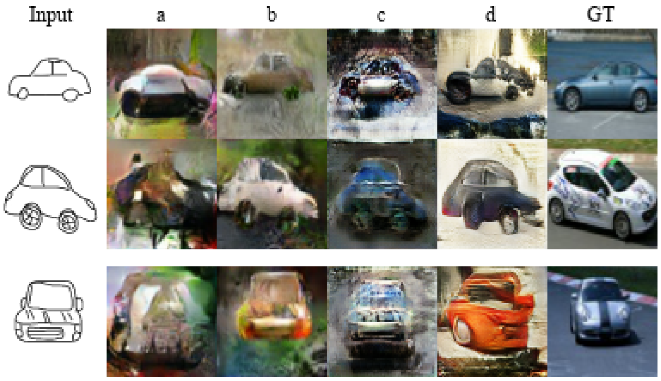
cannon



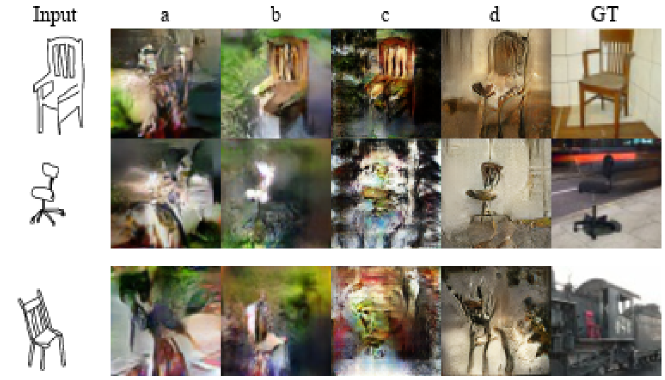
cat



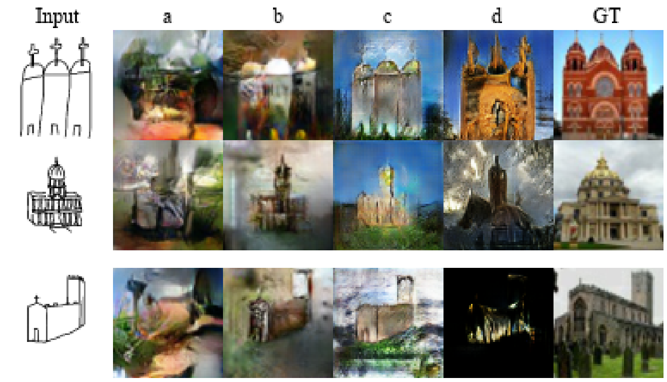
car



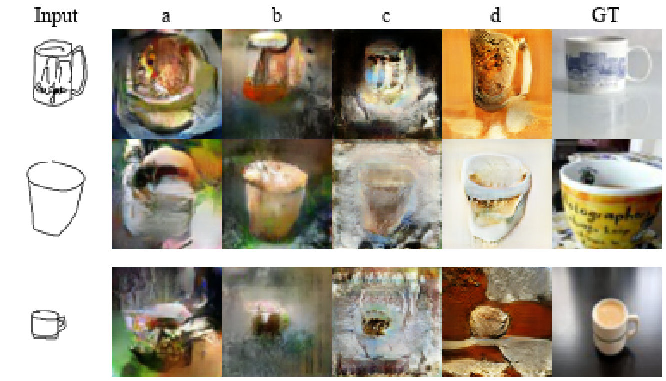
chair



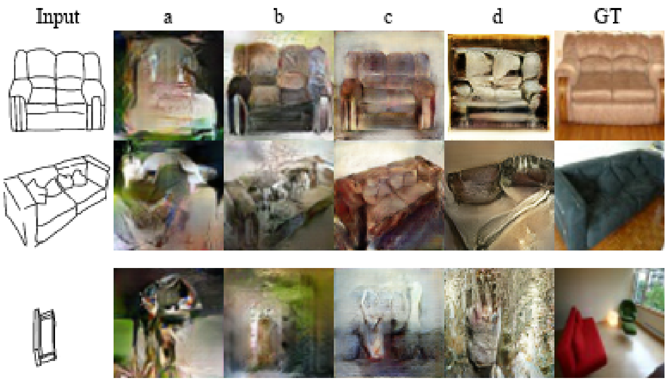
church



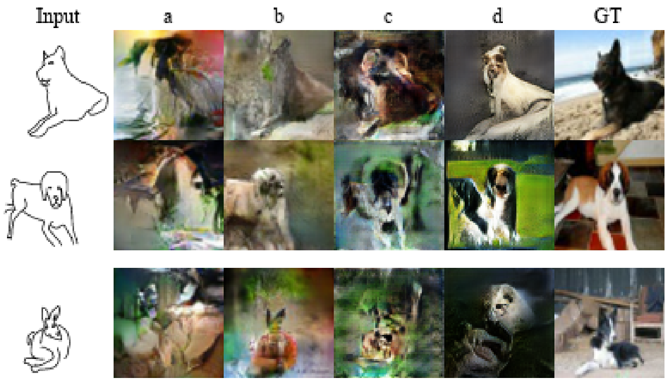
cup



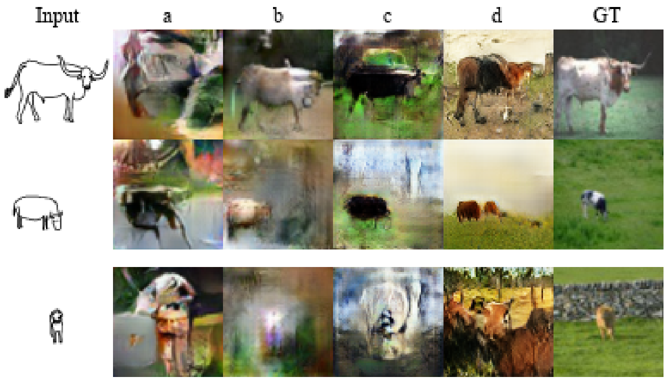
couch



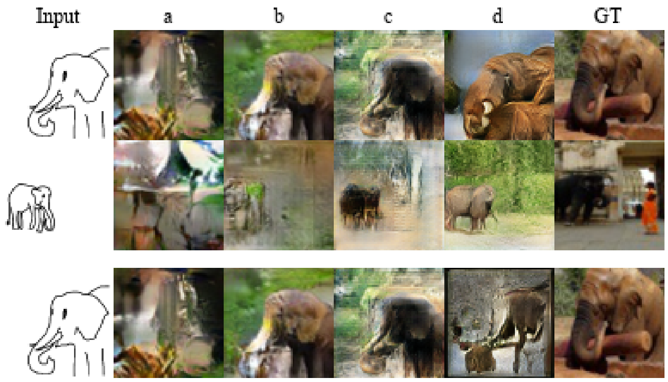
dog



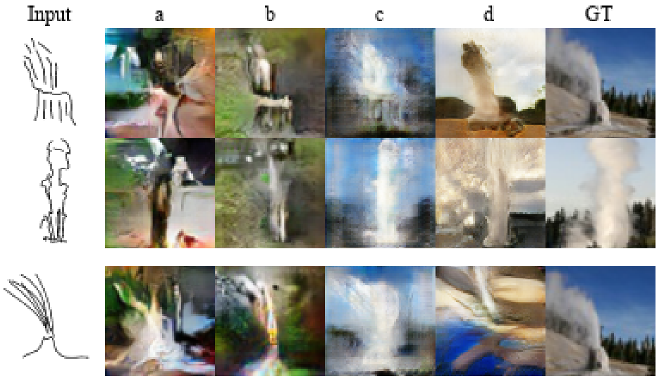
cow



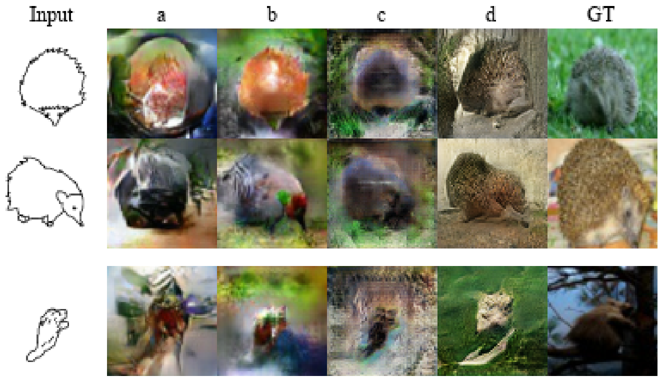
elephant



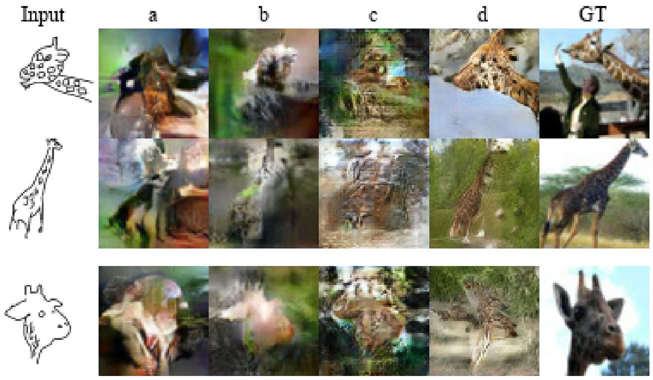
geyser



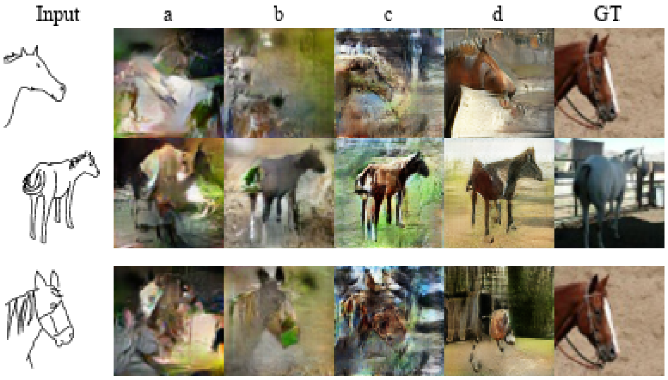
hedgehog



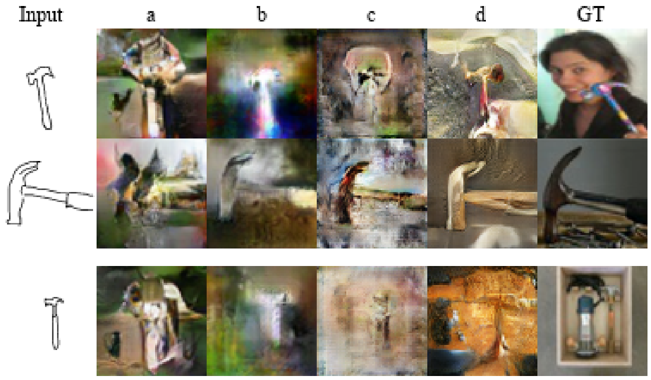
giraffe



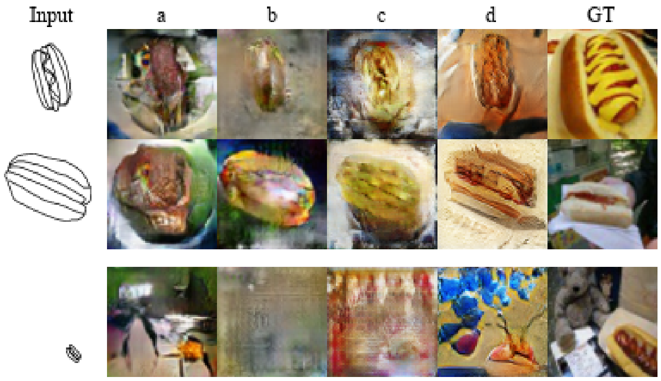
horse



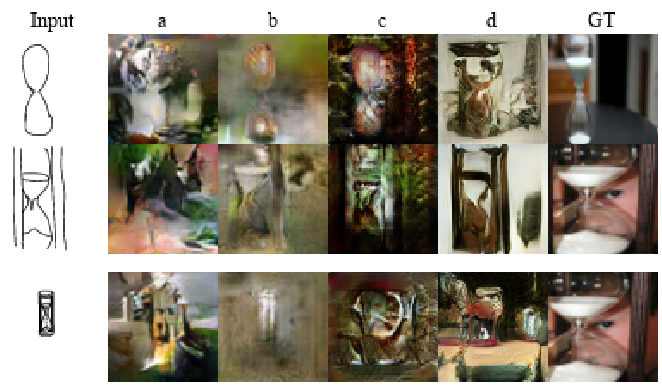
hammer



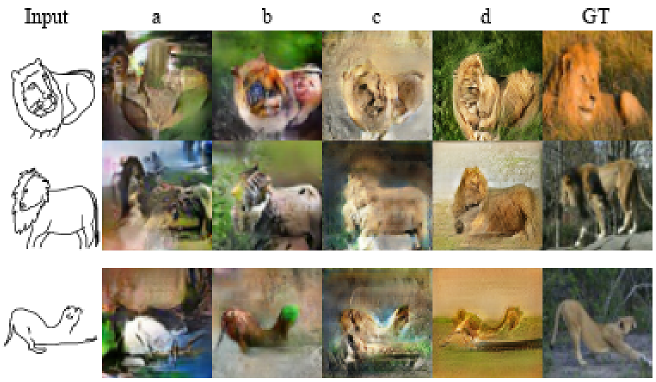
hotdog



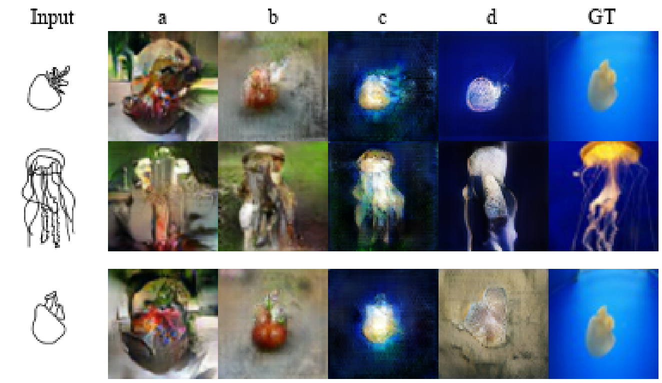
hourglass



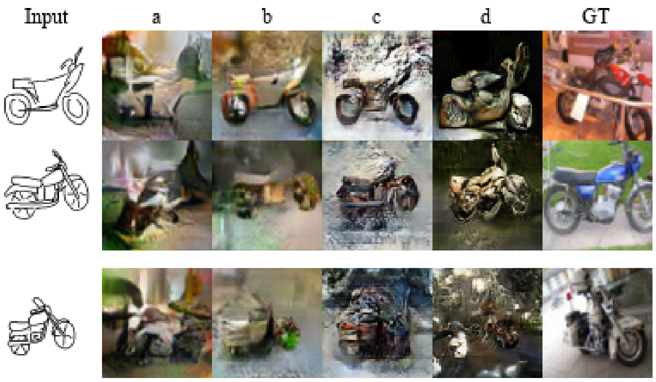
lion



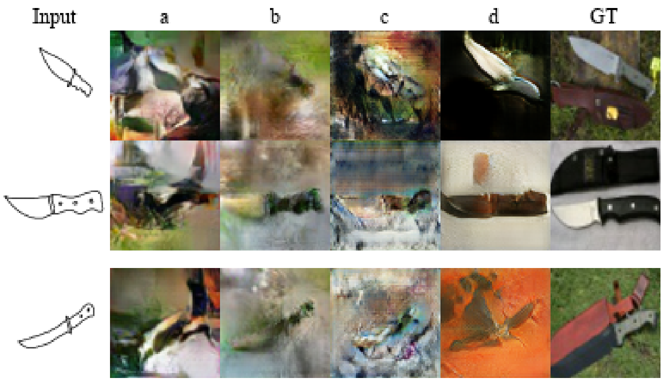
jellyfish



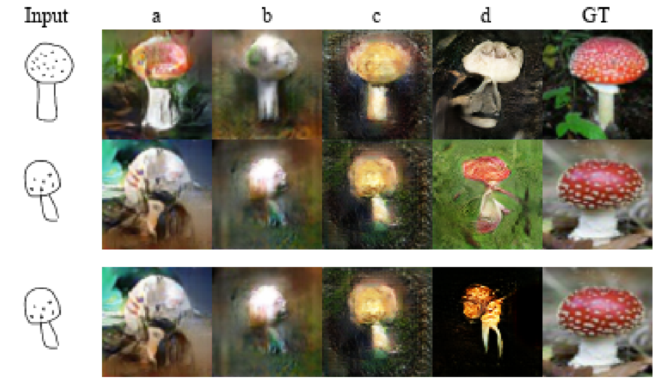
motorcycle



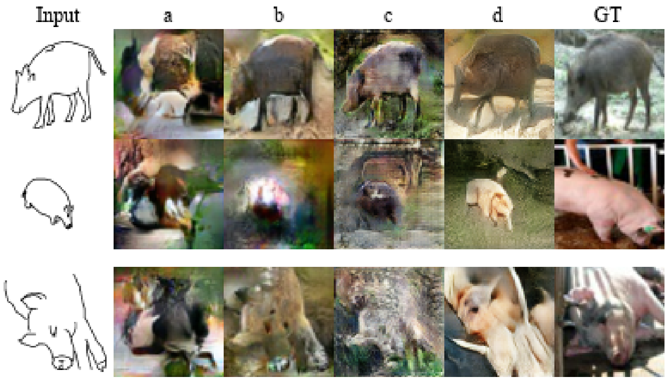
knife



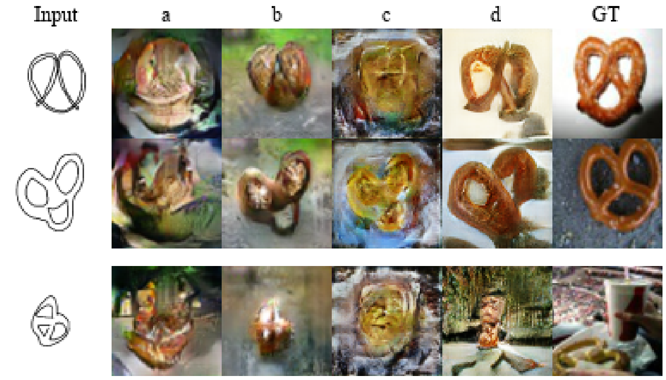
mushroom



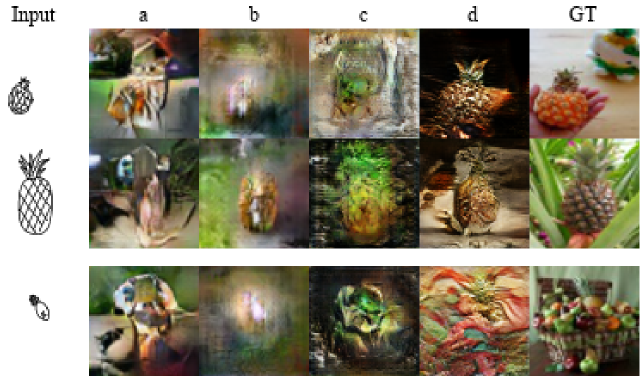
pig



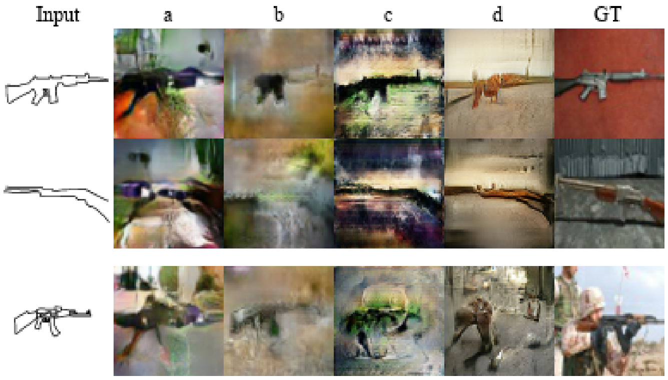
pretzel



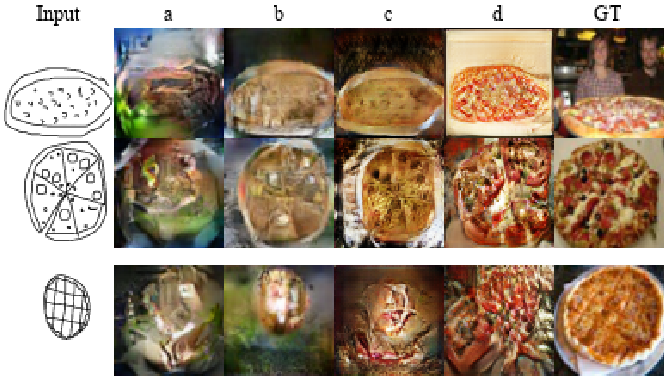
pineapple



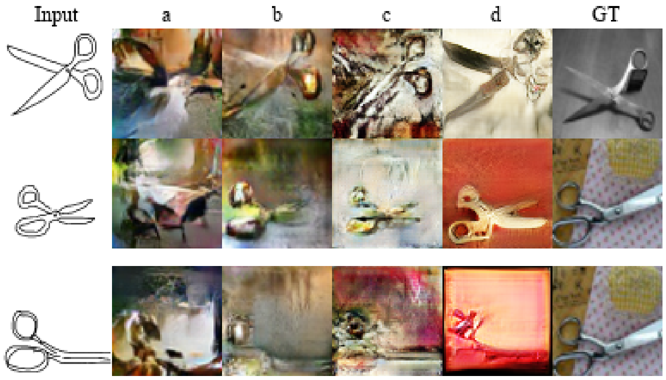
rifle



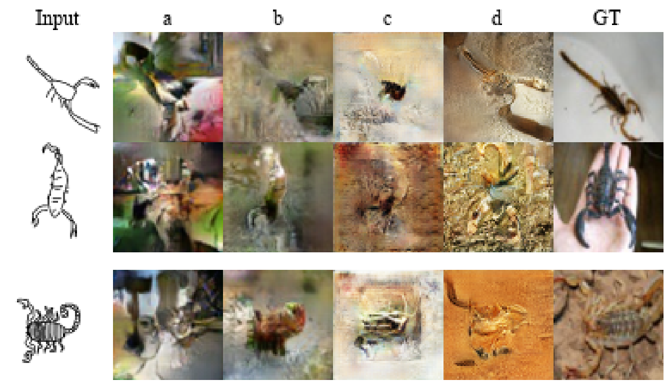
pizza



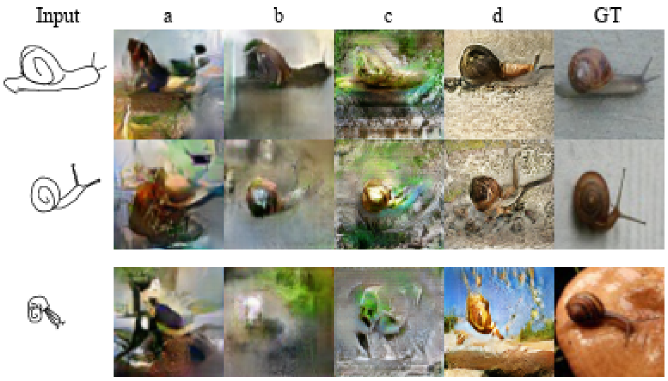
scissors



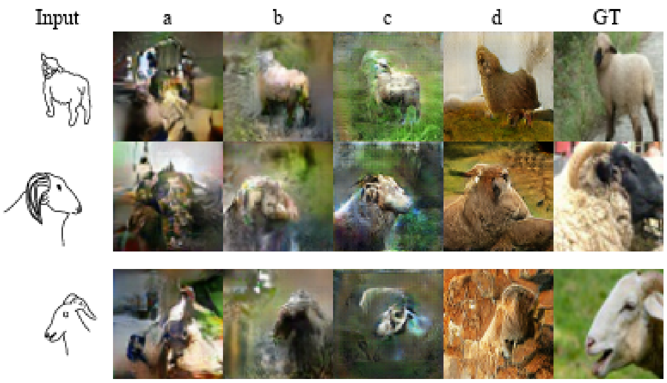
scorpion



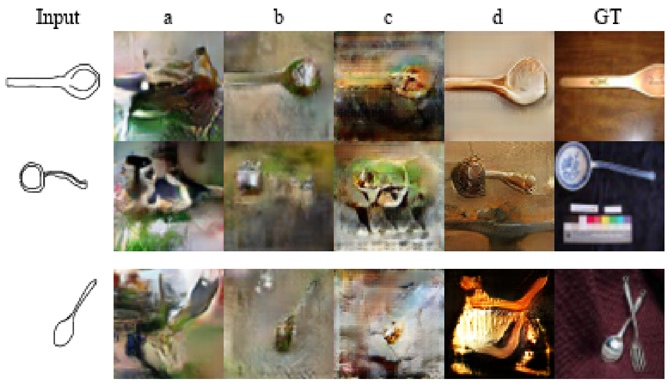
snail



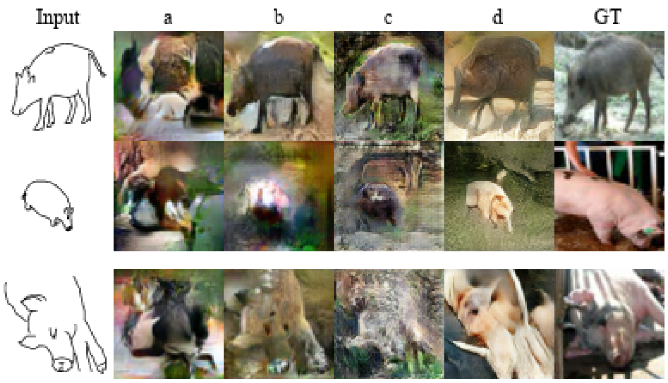
sheep



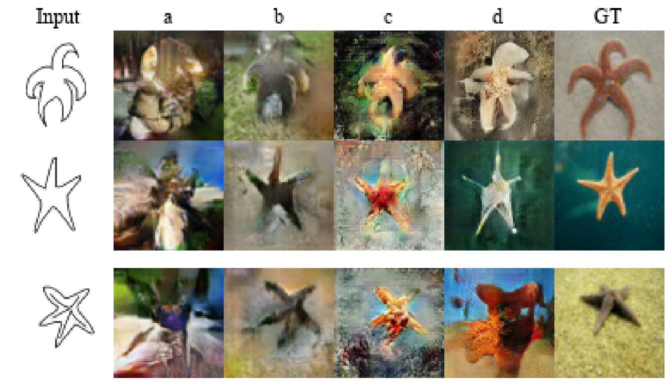
spoon



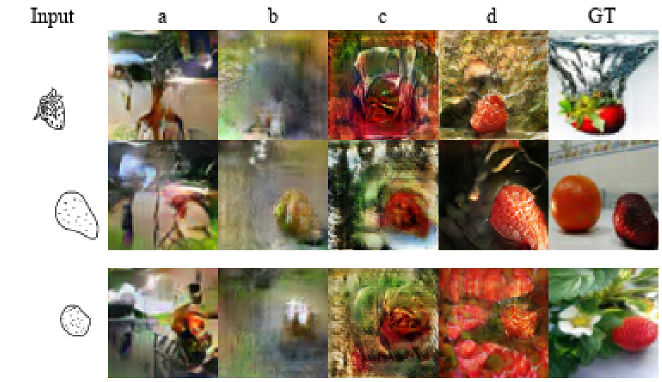
pig



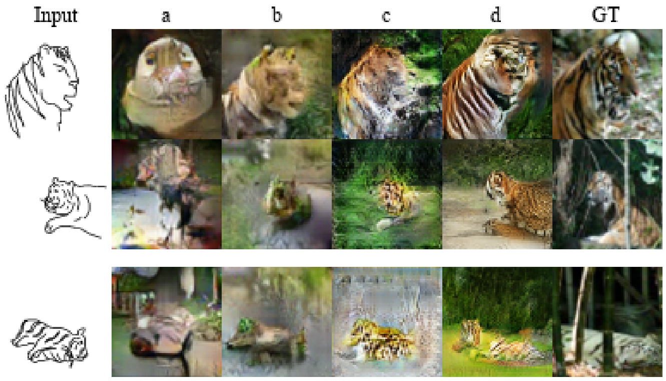
starfish



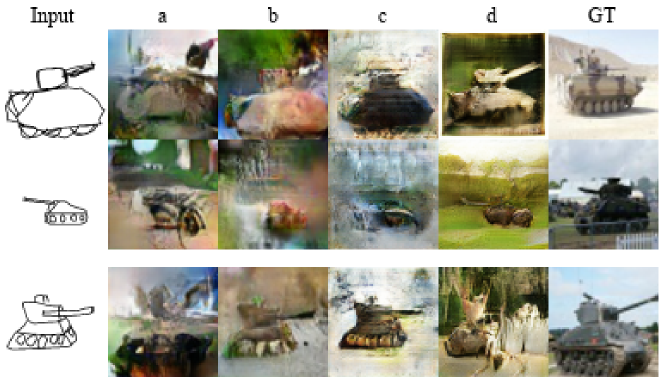
strawberry



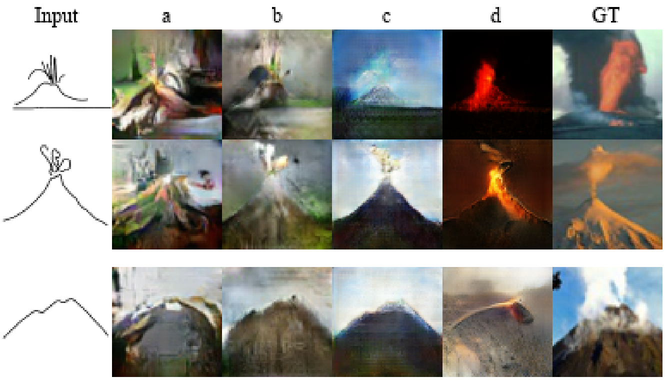
tiger



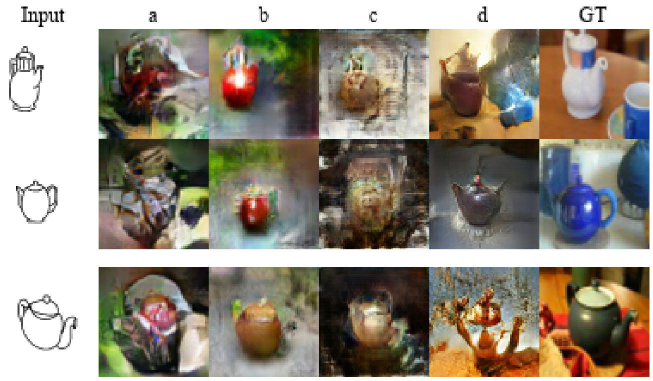
tank



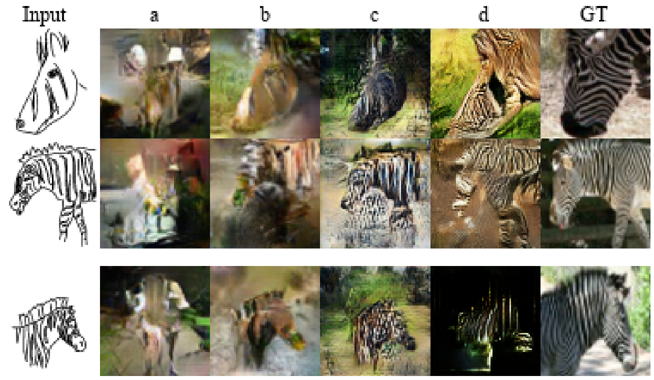
volcano



teapot



zebra



References

[1] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[2] K. He, X. Zhang, S. Ren, and J. Sun. Identity mappings in deep residual networks. In *European Conference on Computer Vision*, pages 630–645, 2016.

[3] M. Lin, Q. Chen, and S. Yan. Network in network. *arXiv preprint arXiv:1312.4400*, 2013.

[4] A. L. Maas, A. Y. Hannun, and A. Y. Ng. Rectifier nonlinearities improve neural network acoustic models. In *in ICML Workshop on Deep Learning for Audio, Speech and Language Processing*, 2013.

[5] R. K. Srivastava, K. Greff, and J. Schmidhuber. Training very deep networks. In *Advances in neural information processing systems*, pages 2377–2385, 2015.