Learning to Generate Chairs, Tables and Cars with Convolutional Networks

Generative
model

- Learning the distribution from which the image are to be generated
- Learning the generator which produces an image conditioned on a vector from this distribution

Given high-level
descriptions

- model number (style), viewpoint,color,brightness etc.

Model Description

Dataset: examples D = $\{(c^1, v^1, \theta^1), \ldots, (c^N, v^N, \theta^N)\}$ + target $O = \{(x^1, s^1), \ldots, (x^N, s^N)\}$

c: one-hot encoding of the model identity

v: azimuth and elevation of the camera position

θ: transformation parameters

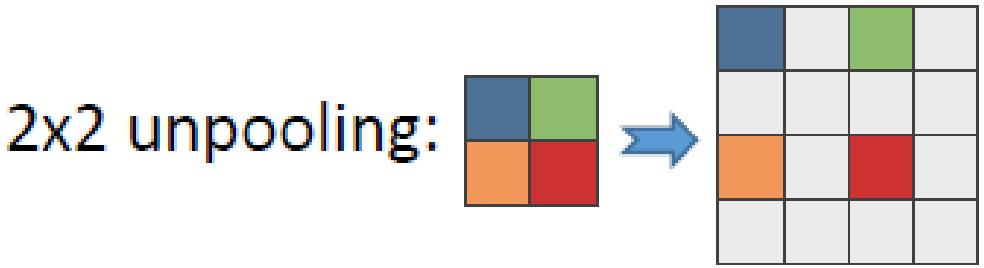
x: RGB output image

s: segmentation mask

up-
convolution

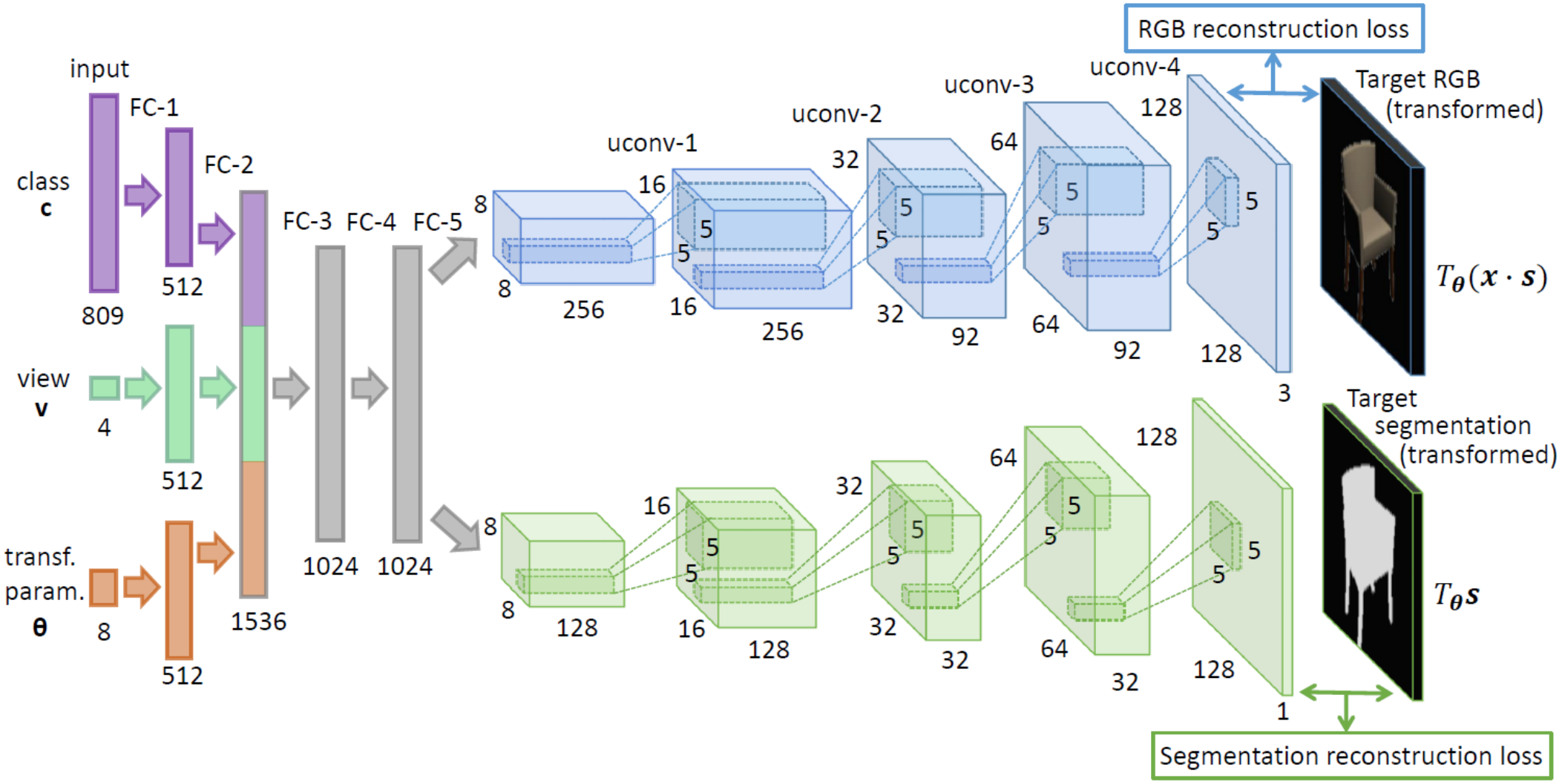
- inverse operation of the convolution+pooling performed in a standard CNN

unpooling



Kernel: 2x2
stride: 2

2-stream network



Network training

The network parameters W are trained by minimizing the error of reconstructing the segmented-out chair image and the segmentation mask

$$\min_{\mathbf{W}} \sum_{i=1}^{N} L_{RGB}\left(T_{\theta^i}(\mathbf{x}^i \cdot \mathbf{s}^i),\, u_{RGB}(h(\mathbf{c}^i, \mathbf{v}^i, \theta^i))\right) + \lambda \cdot L_{segm}\left(T_{\theta^i}\mathbf{s}^i,\, u_{segm}(h(\mathbf{c}^i, \mathbf{v}^i, \theta^i))\right),$$

λ is a trade-off parameter.

Knowledge Transfer

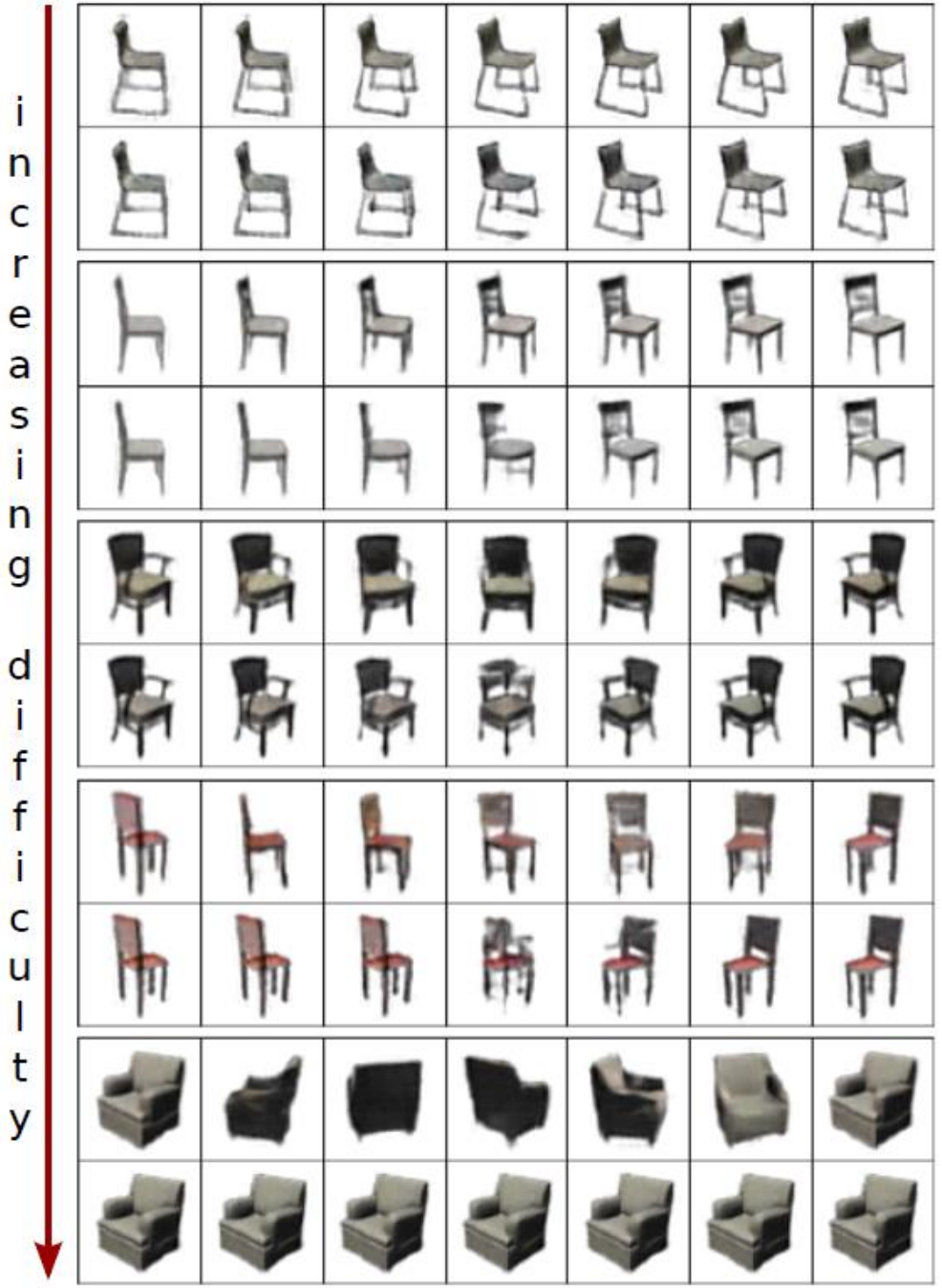
- Knowledge transfer within an object class
- Knowledge transfer between classes

Interpolation between viewpoints

- 64x64 network
- Randomly separate the chair styles into subsets:
  90% source set + 10% target set
- Vary the number of viewpoints per style

  in both two subsets (no transfer)
  just in the source set (with transfer)

train a generative network



Elevation transfer and extrapolation

Transfer information about elevations from one class to another

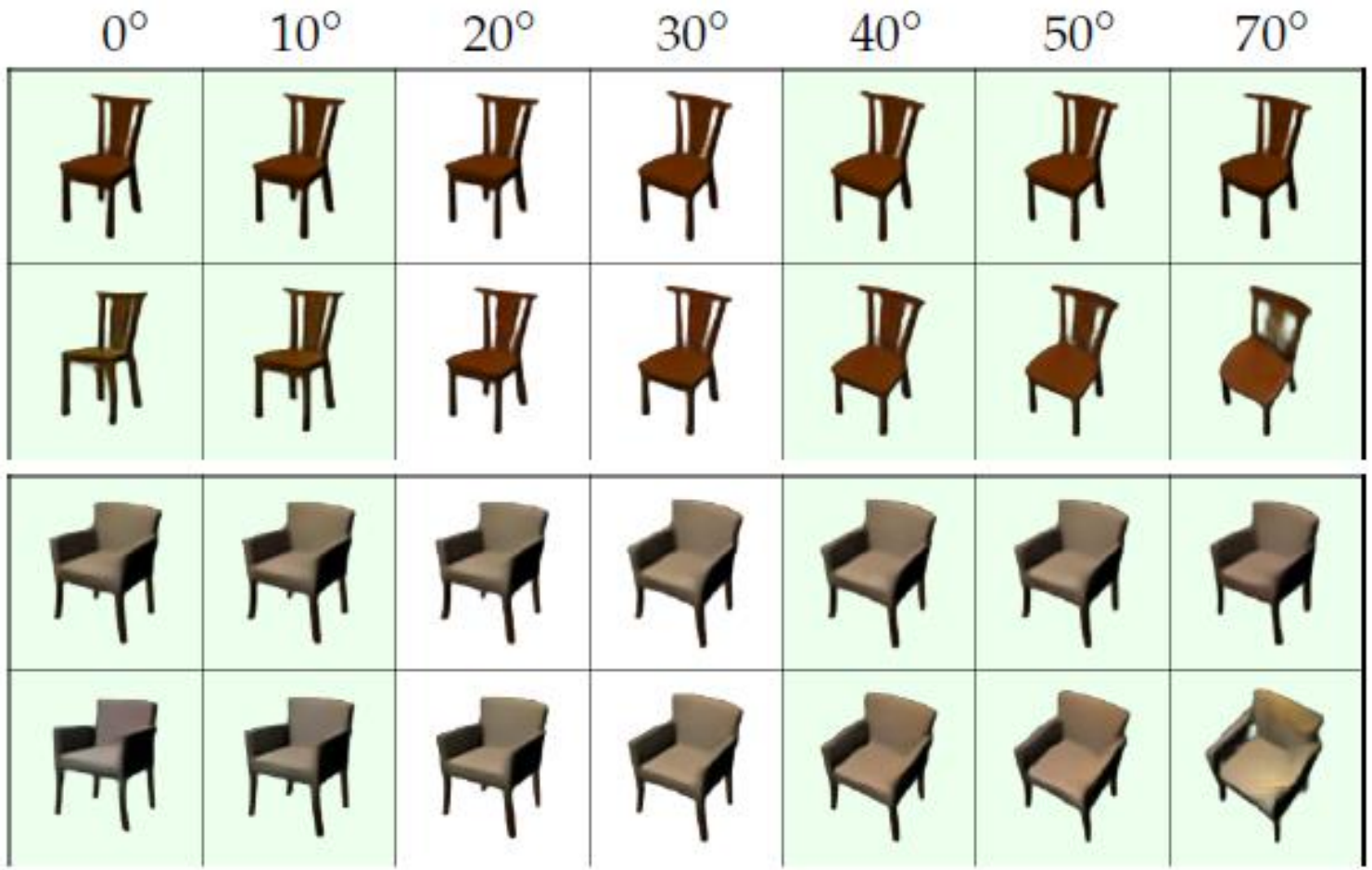
- Elevation angles:

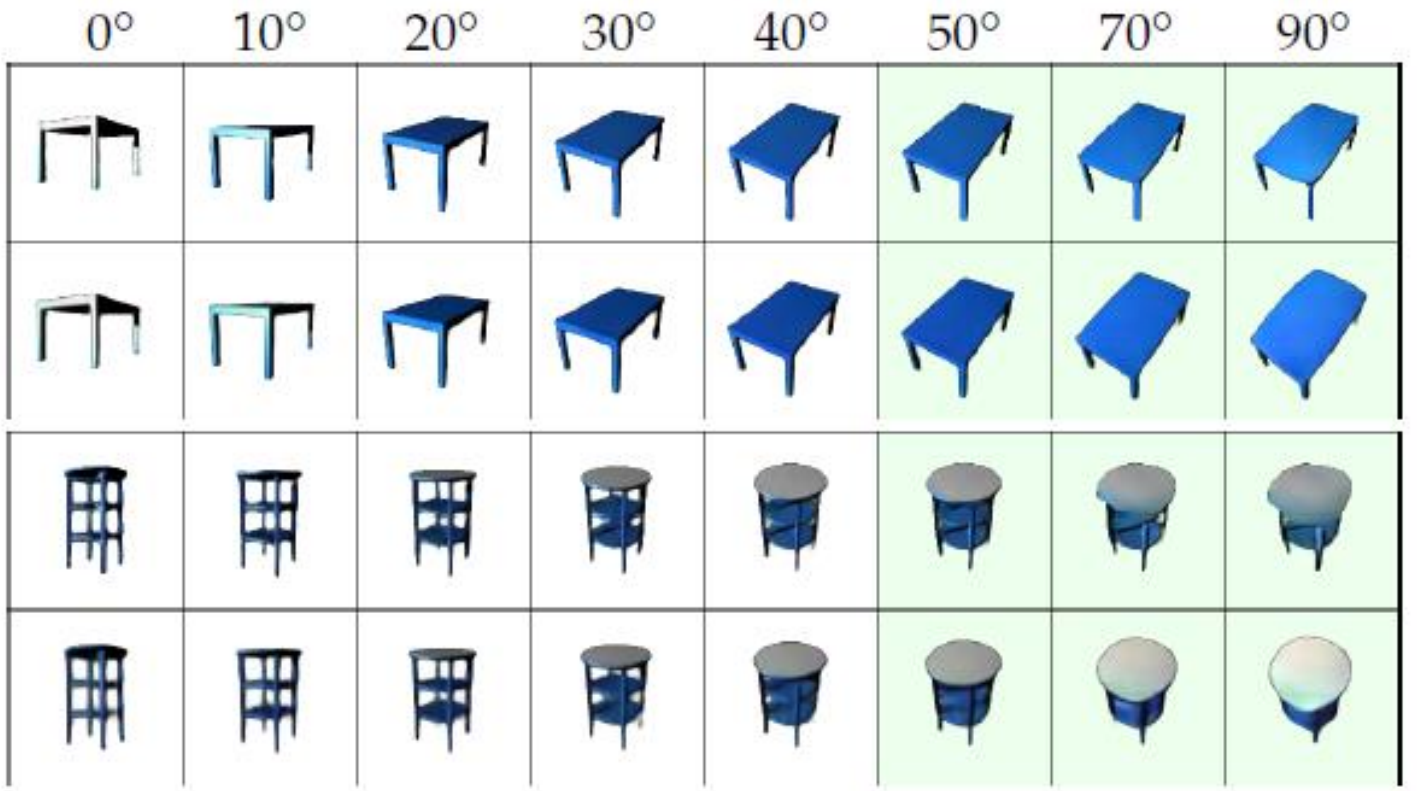
  in chairs dataset: 20 and 30

  in tables dataset: between 0 and 40

Train a network on both chairs and tables, and then generate images of chairs with elevations not present during training.

Baseline: a network trained solely on chairs.





The presence of chairs in the training set helps better extrapolate views of tables.

References

Dosovitskiy, Alexey, et al. "Learning to Generate Chairs, Tables and Cars with Convolutional Networks." *IEEE Transactions on Pattern Analysis & Machine Intelligence* 39.4(2017):692-705.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*. 2012.

Zeiler, Matthew D., and Rob Fergus. "Visualizing and understanding convolutional networks." *arXiv preprint arXiv:1311.2901* (2013).