Supplementary Material for "Learning to Generate Chairs with Convolutional Neural Networks"

Alexey Dosovitskiy Jost Tobias Springenberg Thomas Brox

Department of Computer Science, University of Freiburg

`{dosovits, springj, brox}@cs.uni-freiburg.de`

We show several additional experiments related to the analysis of the generative network, as well as some representative 'simple' and 'difficult' image pairs from the test set we used to evaluate the quality of correspondences found by different methods.

1. Analysis of the network

1.1. Activating single neurons and groups of neurons

In the main paper we show images generated from single neuron activations in various fully connected layers of the network. However, as pointed out in the main paper, the images generated from individual neurons in each layer look fairly similar, independent of which single neuron is activated. This suggests that 1) the amount of variation may be dependent on the activation strength of a neuron and 2) larger variations can be obtained by activating multiple neurons. Here we experimentally test both these hypotheses for the FC-1 layer.

Figure 1 shows the variation in the generated depending on the value of the activation of a single FC-1 neuron (one neuron per column, one value per row). These are the same neurons as in Figure 7 of the main paper. The activation values vary, top to bottom, from 2 to 25. In Figure 7 of the main paper the activation value was 10, so it coincides with the 4th row of Figure 1. Clearly, larger activations induce larger changes, but extreme activations lead to images which do not resemble chairs anymore.

In Figure 2 we choose two neurons from FC-1 and vary their activations simultaneously between 0 and 9, spanning a 2D grid of generated images. Clearly, activating both neurons simultaneously leads to a combination of the effects of these two neurons.

Finally, we activate random subsets of several FC-1 neurons. We randomly select a number of neurons and set their activations to the same constant (we selected the value manually depending on the number of activated neurons to obtain the most visually appealing results). The results are shown in Figure 3. Each row corresponds to a different
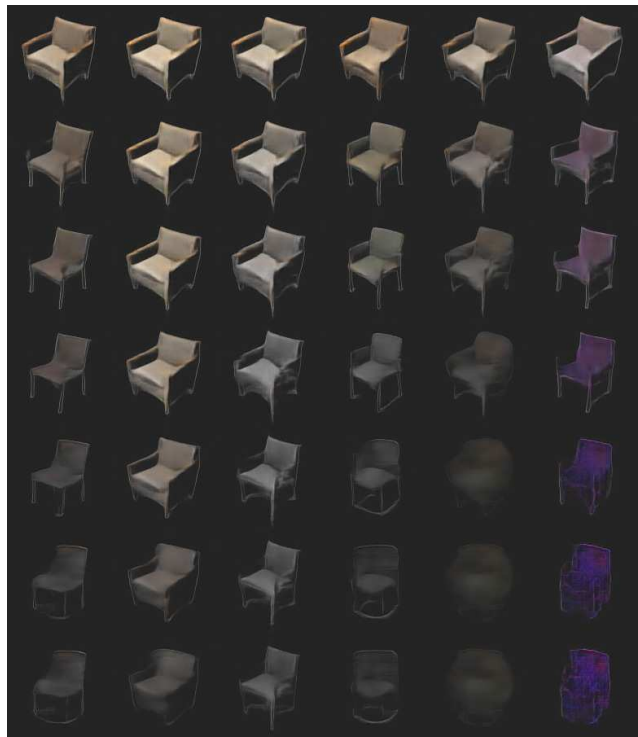

Figure 1. Generating from single FC-1 neurons with varying activation, one neuron per column, one activation value per row. The activation values vary, top to bottom, from 2 to 25.

number of randomly selected active neurons, increasing top to bottom. Clearly, more active neurons lead to more variance in chair appearance, and the number of active neurons affects the chair style: chairs with fewer neurons are more 'square' or armchair-like.

1.2. Analysis of the hidden layers

In Figure 4 we show some representative feature maps of different layers from the generating streams (FC-5 to uconv-3). For better viewing, the feature maps are modified by cutting 1 percent of the darkest and brightest values.
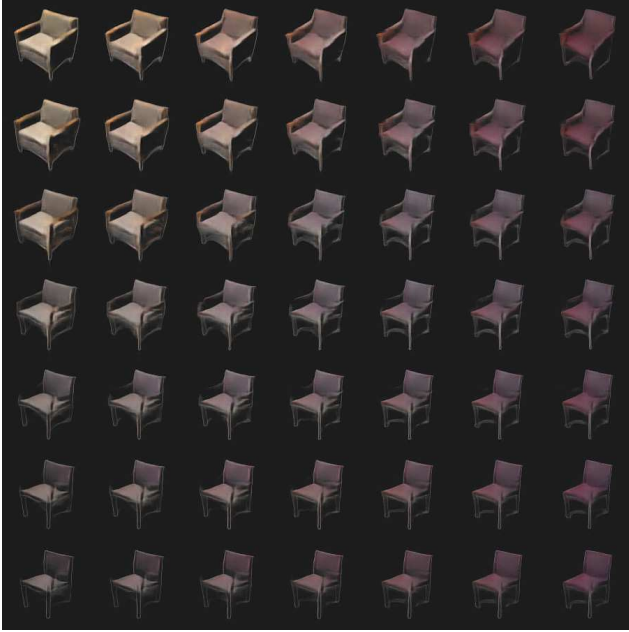
Figure 2. Generating from two FC-1 neuron activations. The value of eaach neuron's activation is varied between 0 and 9.
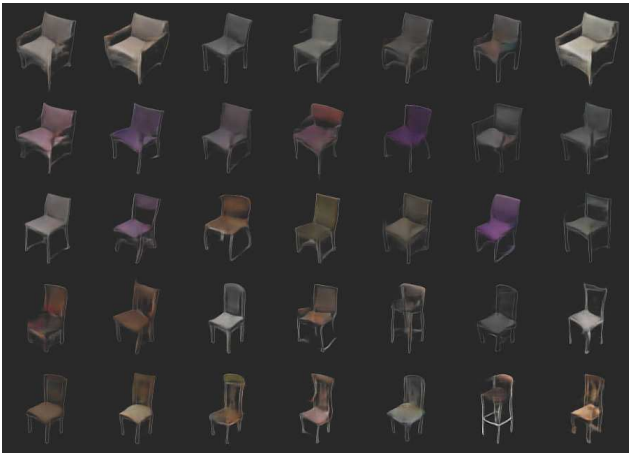

Figure 3. Generating from multiple FC-1 neuron activations. Each row shows several random combinations with the same number of neurons. Number of neurons, top to bottom: 2, 5, 10, 20, 50, 100, 200.

Feature maps uconv-2 and uconv-3 contain some chair-like contours, but FC-5 and uconv-1, which are further from the output image, are rather abstract and difficult to interpret.

Similarly to the 'zoom neuron' described in the main paper, there exists a separate single neuron in the layer FC-4 for more or less every artificial transformation we applied during training. The effect of increasing their values given a feature map of a real chair is shown in Figure 5, one neuron per row, activation increasing from left to right.
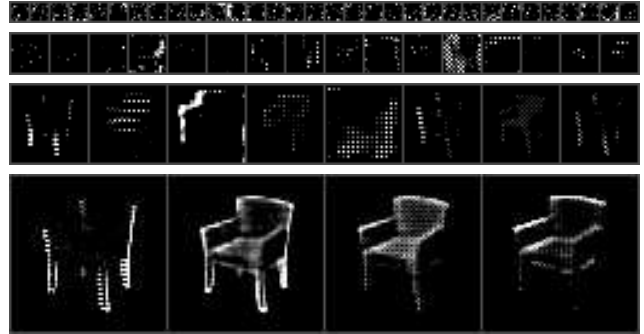

Figure 4. Representative feature maps of different convolutional layers. Top to bottom: FC-5, uconv-1, uconv-2, uconv-3. Relative scale of different maps is correct.

It is very surprising that such complex operation as applying various transformations, especially spatial ones, happens as late as in the last fully connected layer FC-5. This suggests that all potential transformed versions of a chair are already contained in the FC-4 feature maps, and the 'transformation neurons' only modify them to give more relevance to the activations corresponding to the required transformation. To gain some understanding of how this happens, we show weights connected to the 'transformation neurons' in Figure 6. The order of the rows is the same as in Figure 5. Each neuron of FC-4 is connected to a $8 \times 8 \times 256$ blob in FC-5. We do not show all 256 channels, but rather only the ones which exhibit most interesting and interpretable behavior (the same set of channels for all neurons). Spatial-related and color-related neurons influence different sets of feature maps. Color-related neurons have a roughly constant value over the whole image, while the spatial-related ones often affect two halves of the image differently.

Finally, to analyze the robustness of the hidden representation, we visualize the effect of setting some of the weakest neuron activations in the layer FC-1 to zero. This is shown in Figure 7. Each column corresponds to a different chair class, each row - to a different ratio of weakest non-zero FC-1 activations set to zero, top to bottom: 0, 0.2, 0.5, 0.75, 0.9, 0.95, 0.98. Up to the ratio 0.5 there are basically no changes, for 0.75 the generated images look slightly deformed, and starting from 0.9 the images look very distorted. This suggests that the hidden representation of layer FC-1 is distributed and reasonably robust.

2. Interpolation between classes

We show several more examples of 'morphings' generated by the network in Figure 8. More examples of morphings are shown in a supplementary video available at our website or https://www.youtube.com/watch?v=QCSW4isBDL0. The video shows consecutive morphing

Figure 5. Examples of the effect of specialized neurons in the layer FC-4. Each row shows the result of increasing the value of a single FC-4 neuron given the feature maps of a real chair. Row 2 shows the effect of the 'zoom neuron' from the main paper. Effects of all neurons, top to bottom: translation upwards, zoom, stretch horizontally, stretch vertically, rotate counter-clockwise, rotate clockwise, increase saturation, decrease saturation, make violet.
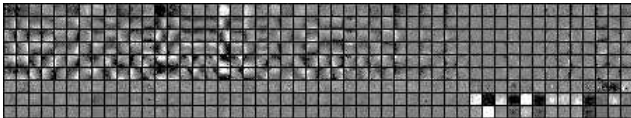

Figure 6. Some of the neural network weights corresponding to the transformation neurons shown in Figure 5. Each row shows weights connected to one neuron, in the same order as in Figure 5. Only a selected subset of most interesting channels is shown.

of 50 random chair styles one into another.

2.1. Correspondences

In Figure 9 we show examples of 'simple' and 'difficult' pairs from the test set. Examples of keypoint tracking using optical flow are shown in a supplementary video available at our website or https://www.youtube.com/watch?v=LAfmJQK4UW0. Optical flow does not always track the points perfectly, but in most cases the results look
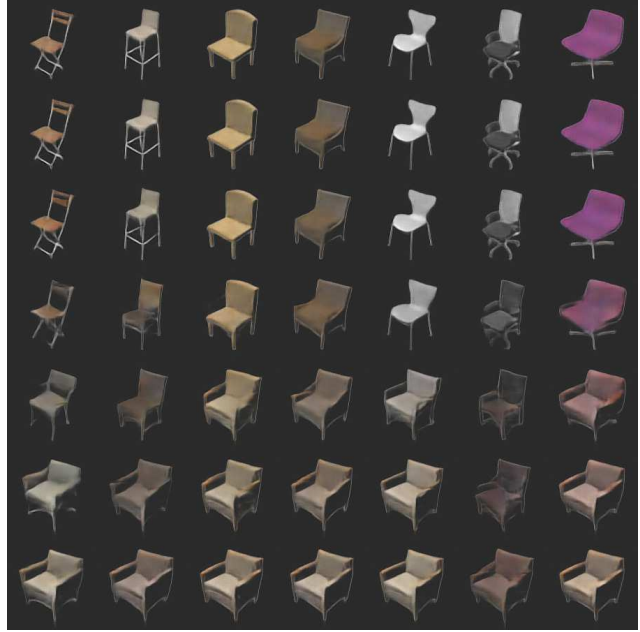

Figure 7. Setting weak activations to zero. The ratio of weak neurons set to zero increases top to bottom: 0, 0.2, 0.5, 0.75, 0.9, 0.95, 0.98.

Figure 8. Examples of morphing different chairs, one morphing per row.

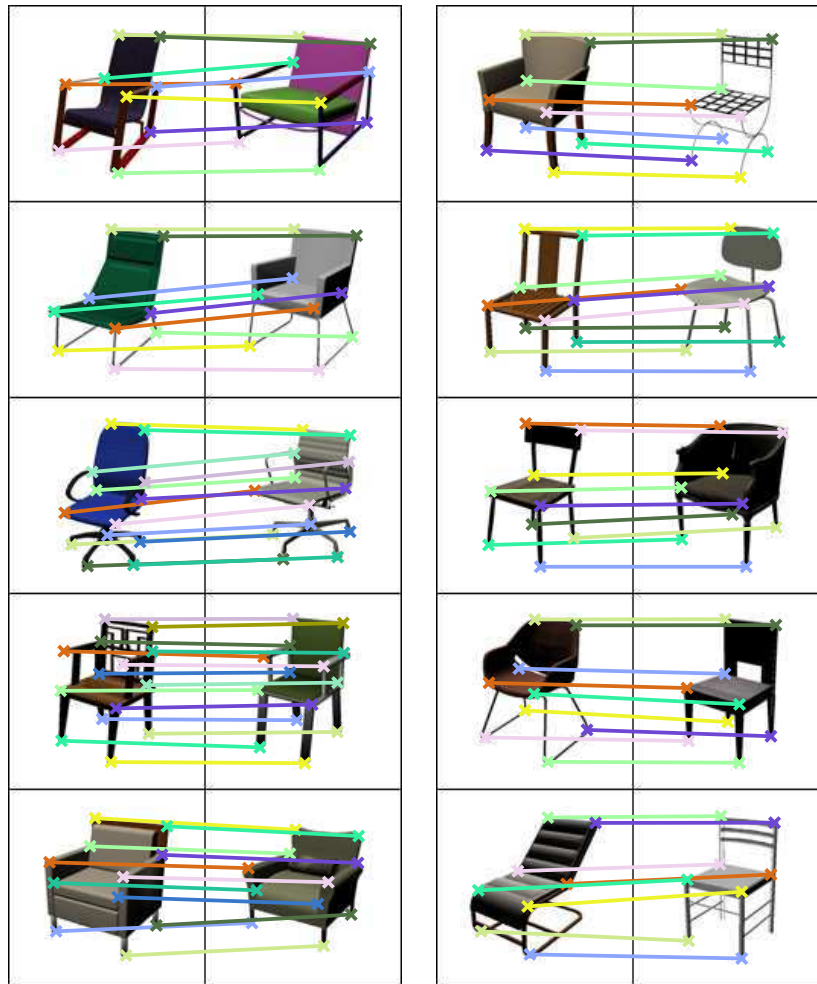qualitatively good, which is also supported by the numbers from the main paper.

Figure 9. Exemplar image pairs from the test set with ground truth correspondences. **Left:** 'simple' pairs, **right:** 'difficult' pairs