Supplemental Material for Evaluating 'Graphical Perception' with CNNs

Daniel Haehn, James Tompkin, and Hanspeter Pfister

We present several plots which contain complete results for the elementary perceptual task experiment, for which, given the number of experiments, the main paper presented only a selection. We report MLAE for all added parameters to the stimuli, rather than just the most complex parameterization as in the main paper (Figures 3 and 4). From this, we see that performance is largely equal across parameters, showing that most networks have sufficient capacity for the given parameterizations.

We also report complete results for the cross-network variability experiment on the elementary perceptual task experiment (Figure 5). As in the main paper, this shows that our networks are not able to generalize to additional translation or stroke width parameters without representative training data.

Further, we show how the errors for each network are distributed, across elementary perceptual tasks and across different cross-validation splits (Figure 6). Most errors are approximately normally distributed, though our CNN with less parameters (LeNet) and our MLP often have errors which are farther from a normal distribution, and show structure.

1 Experiment: Anti-aliasing

We also include an experimental result showing that, for our stimuli, adding anti-aliasing to the line generation was not important to the tested CNNs (Figure 1).

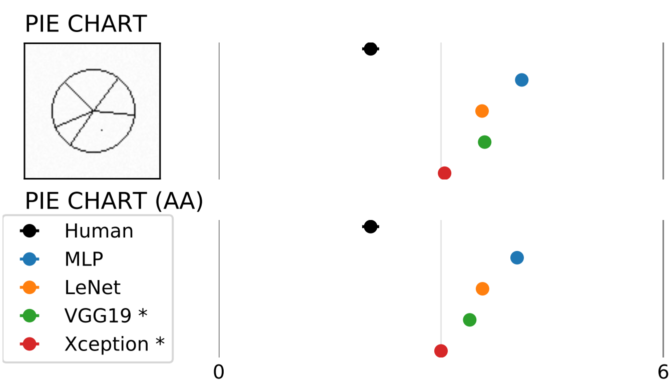


Fig. 1: **Anti-aliasing.** We test whether anti-aliasing effects the performance of our networks on pie charts by measuring MLAE. The difference is not statistically significant ($F(1,30) = 0.341, p > 0.5$).

2 Experiment: Noise

For all our experiments, we add subtle 5% noise to every pixel to enhance variability. We did not observe a significant effect on regression performance when comparing the weber-fechner's law experiment with and without noise averaged over 4 runs (Figure 2). However, the variability of our stimuli increases.

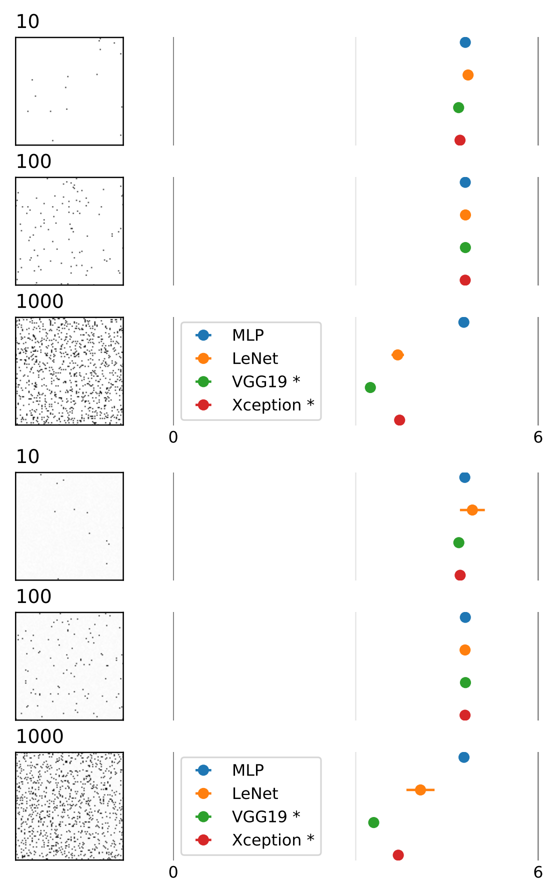


Fig. 2: **Noise.** We test whether noise (top: off, bottom: on) effects the performance of our networks on the weber-fechner's law experiment by measuring MLAE. With noise, $MLAE = 4.511$ ($SD = 0.512$) and without noise $MLAE = 4.491$ ($SD = 0.543$). The difference is not statistically significant ($F(1,22) = 0.008, p > 0.5$).

- *Daniel Haehn, and Hanspeter Pfister are with Harvard University. E-mail: {haehn,pfister}@seas.harvard.edu.*
- *James Tompkin is with Brown University. E-mail: james_tompkin@brown.edu.*

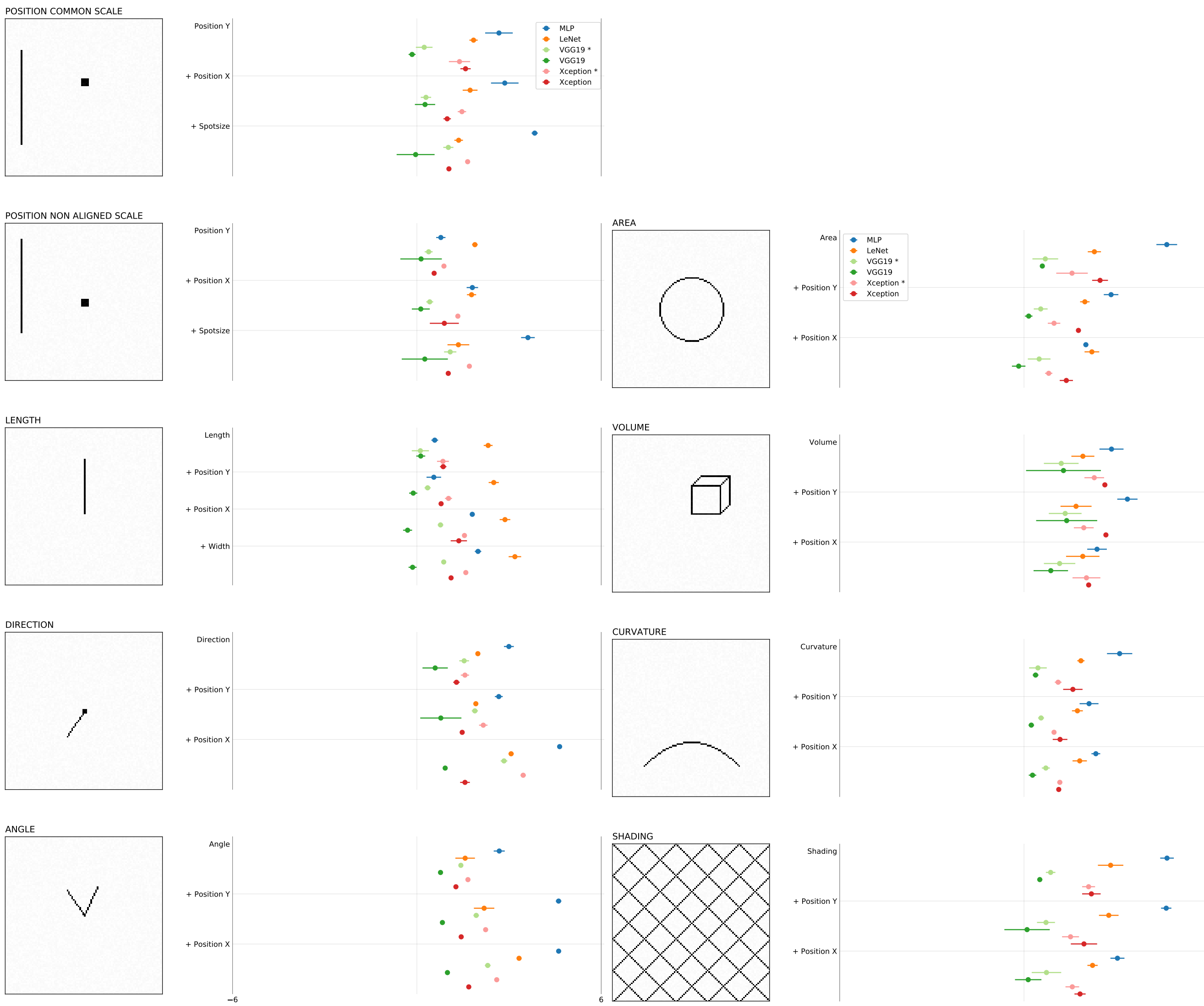


Fig. 3: **Elementary perceptual tasks.** Midmean logistic absolute errors (MLAE) for all generated stimuli and across all networks. The * indicates networks which use ImageNet weights instead of bein trained from scratch.

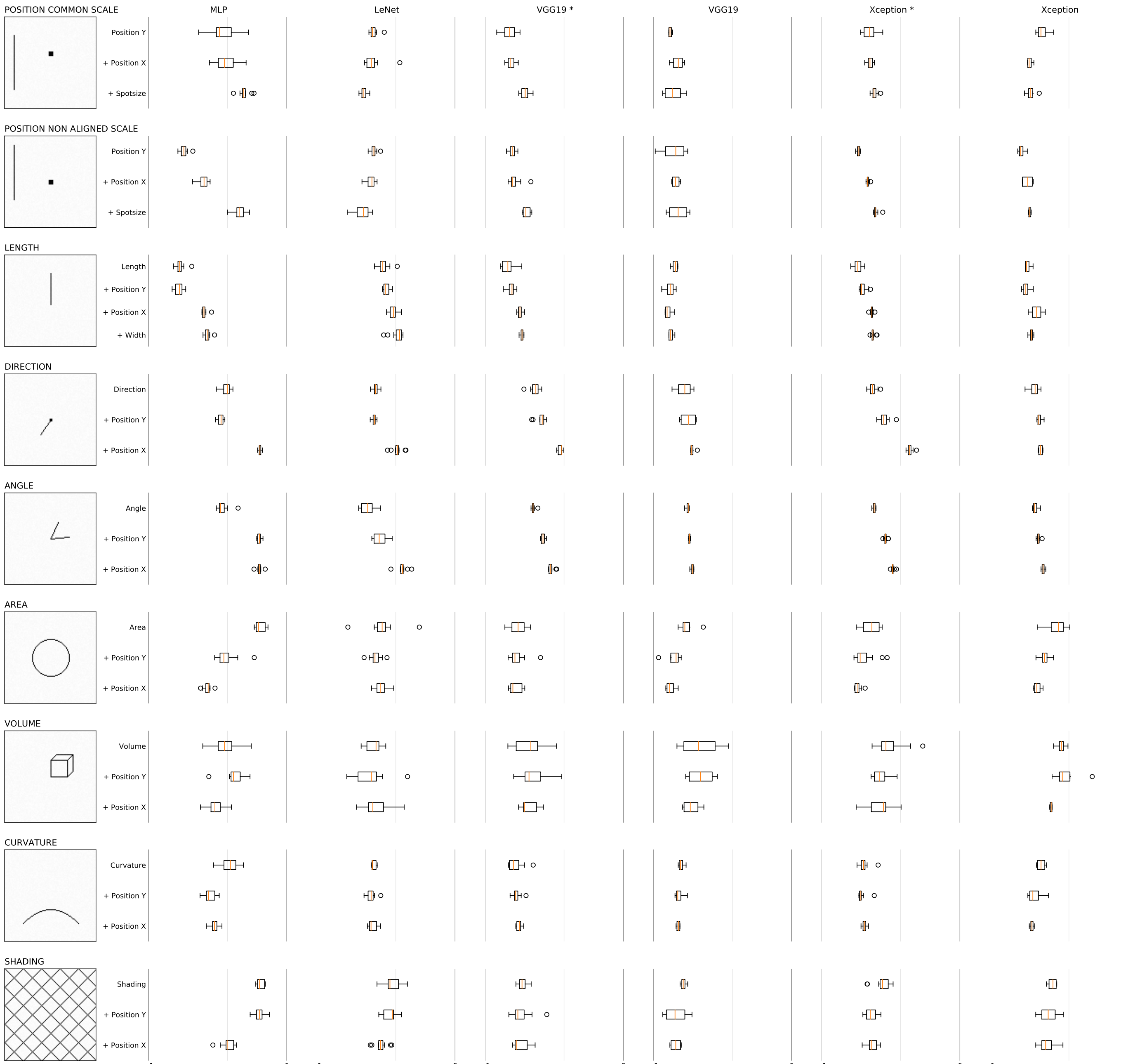


Fig. 4: **Elementary perceptual tasks.** Midmean logistic absolute errors (MLAE) visualized as box plots.

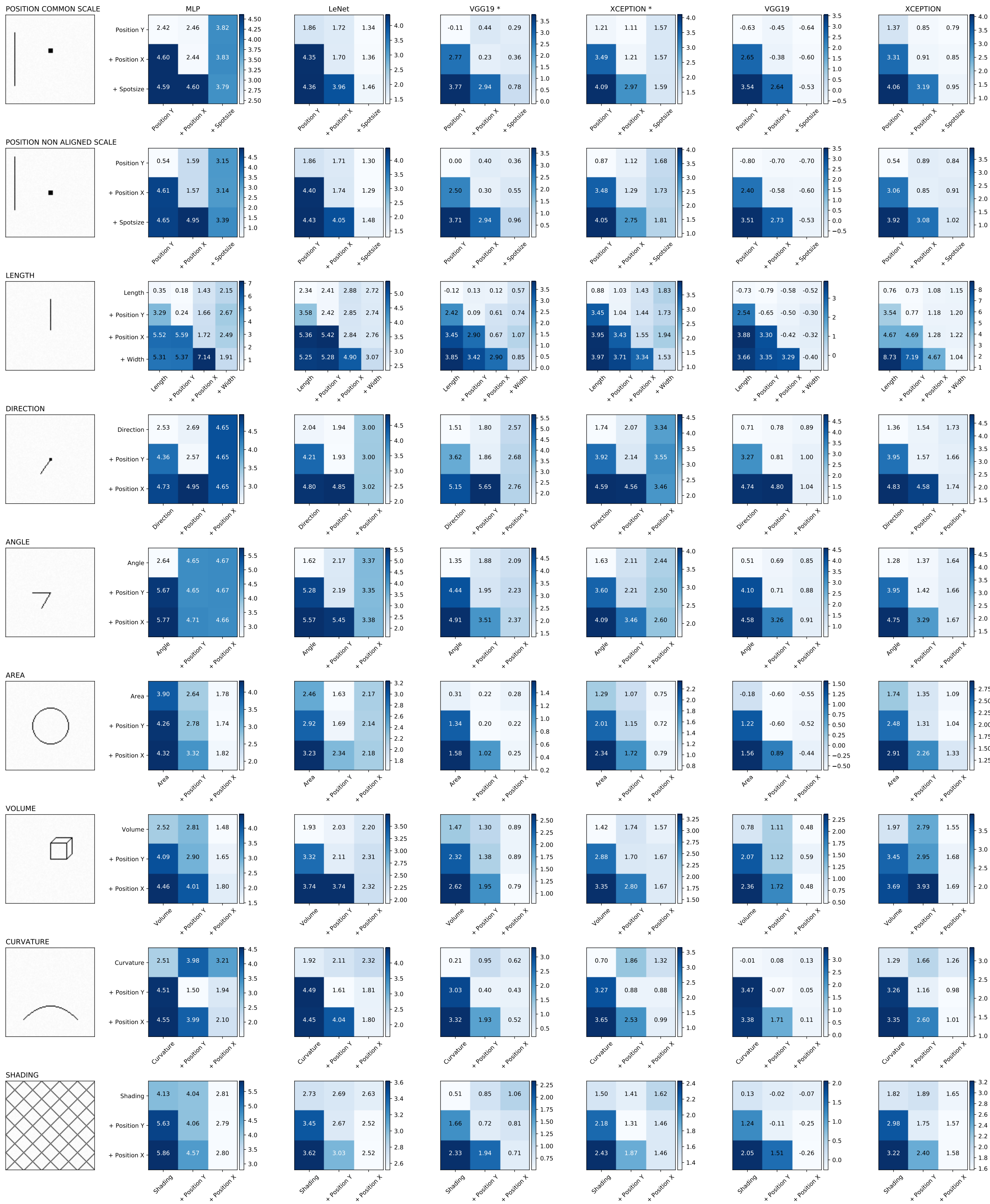


Fig. 5: **Cross-network variability.** Our networks fail when the stimuli changes through translation or stroke width. The x-labels indicate the training configuration while the y-labels indicate the stimuli variation. Numbers represent MLAE.

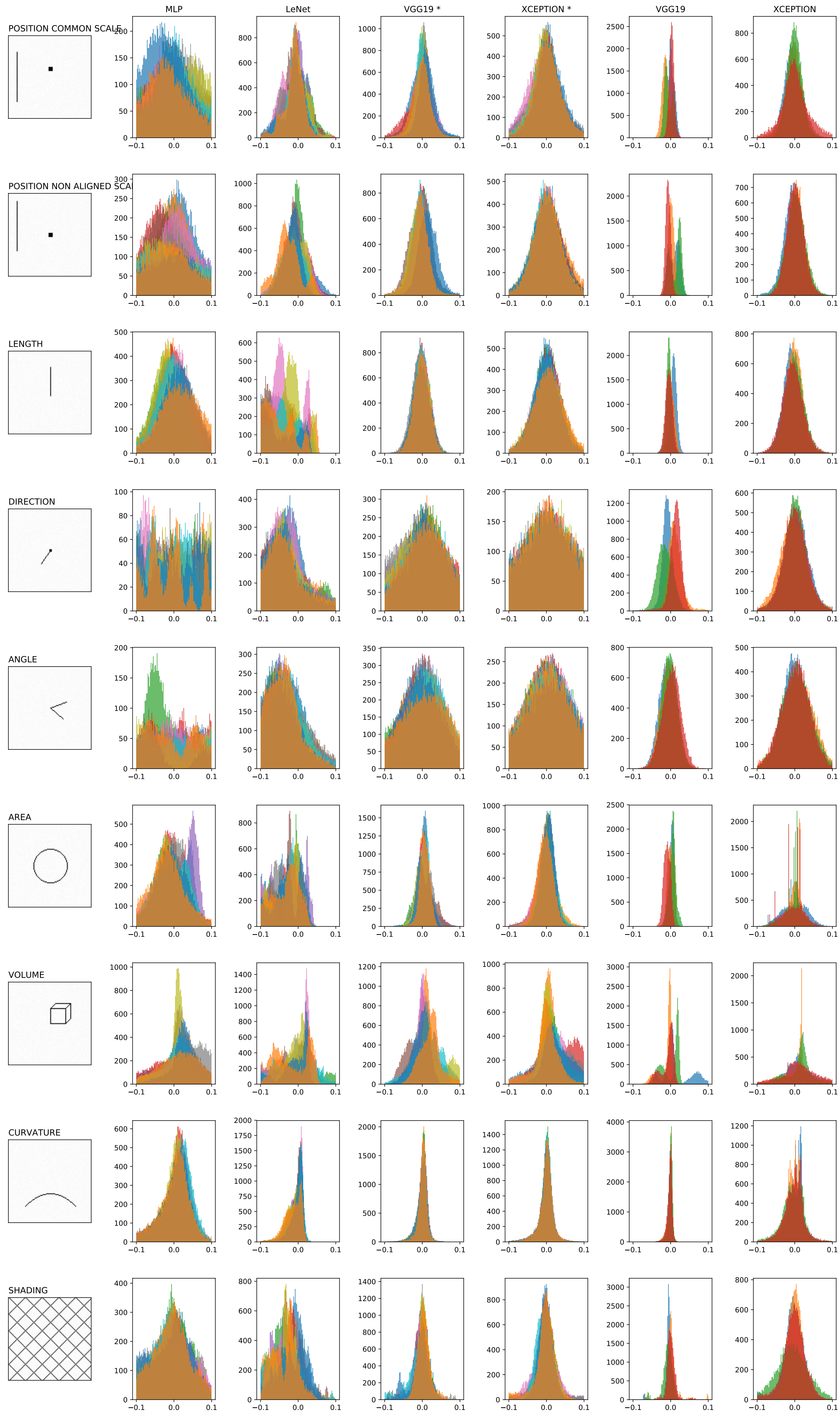


Fig. 6: **Error distributions.** Error distributions of our networks when decoding elementary perceptual tasks.

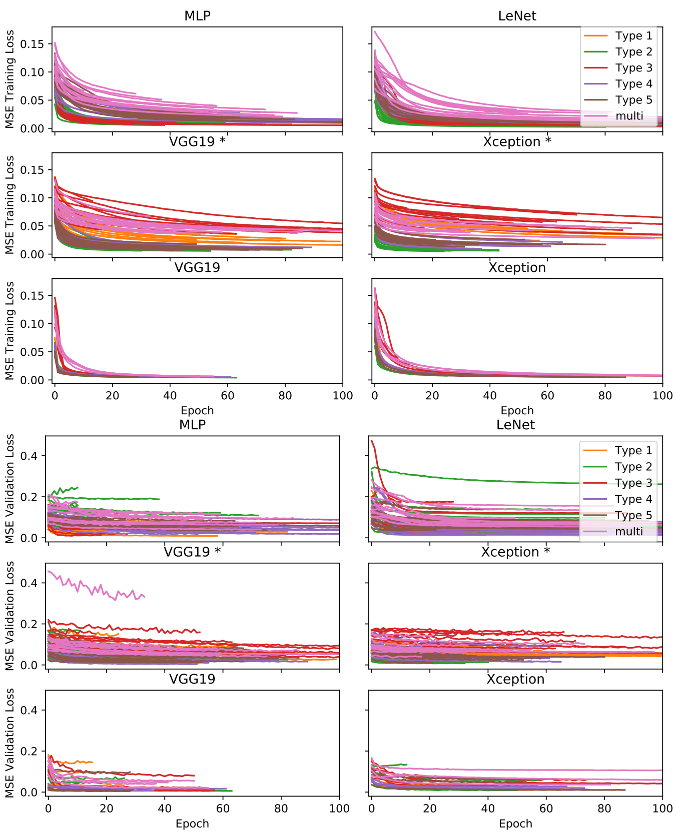


Fig. 7: **Loss plots for the position-length experiment.** We visualize the MSE loss on training data and for unseen validation data after each epoch. There is no significant difference in convergence for either encoding but spiky outliers due to monte-carlo cross validation are visible.

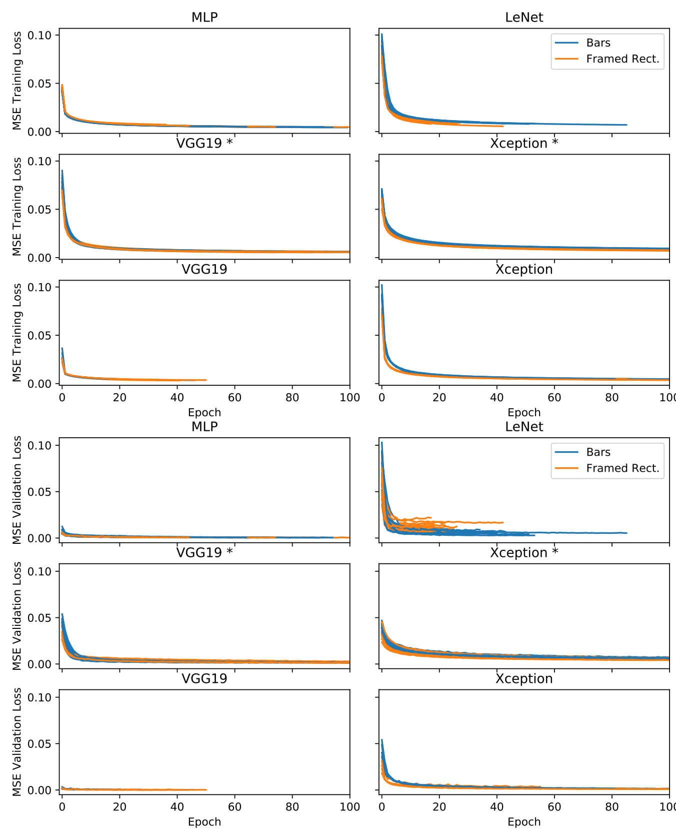
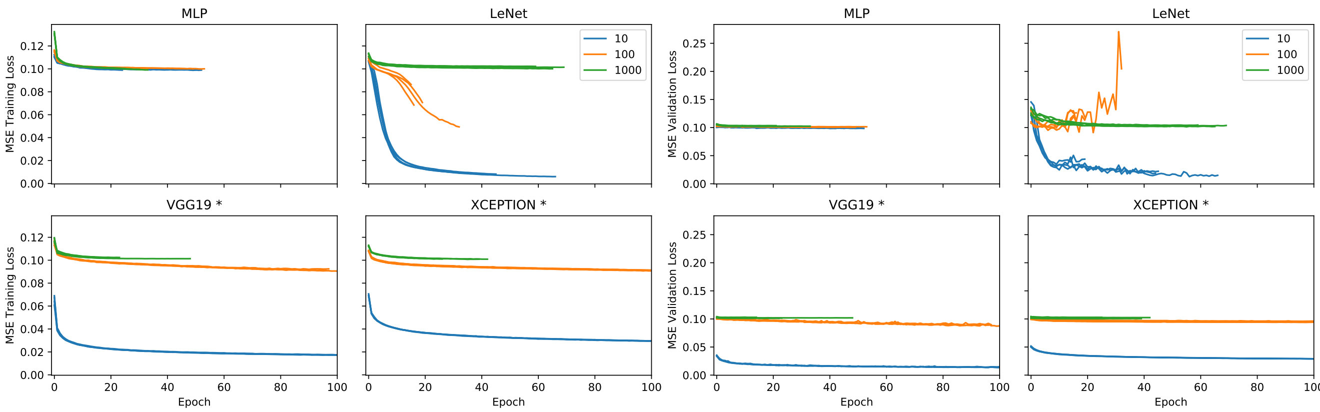
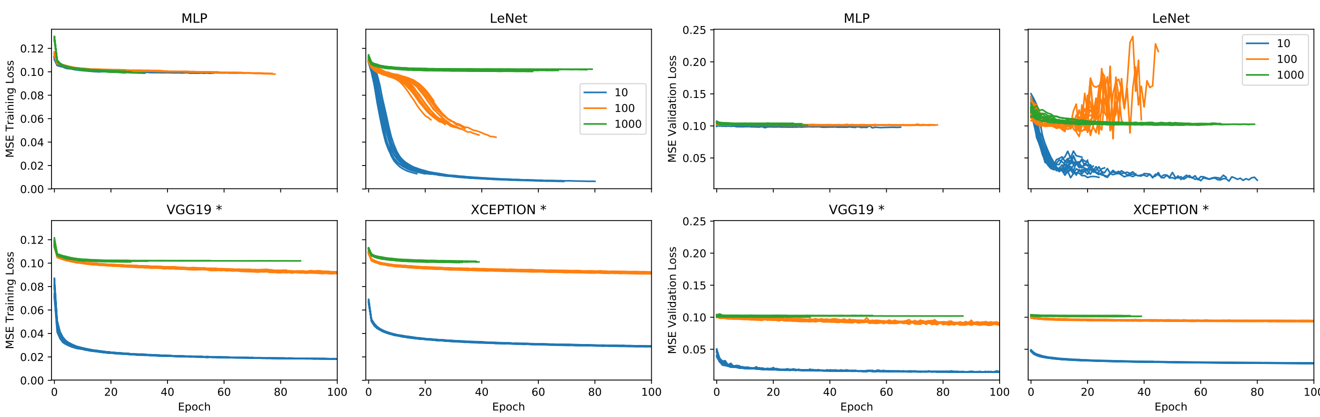


Fig. 8: **Loss plots for the bars-and-framed-rectangles experiment.** We visualize the MSE loss on training data and for unseen validation data after each epoch. There is no significant difference in convergence for either encoding.



(a) without noise



(b) with noise

Fig. 9: **Loss plots for the weber-fechner's law experiment.** We visualize the MSE loss on training data (left) and for unseen validation data (right) after each epoch (a) without noise and (b) with subtle 5% noise per pixel. There is no significant difference when noise is added. The LeNet network seems to overfit with Weber Base 100 in both cases even with dropout regularization.

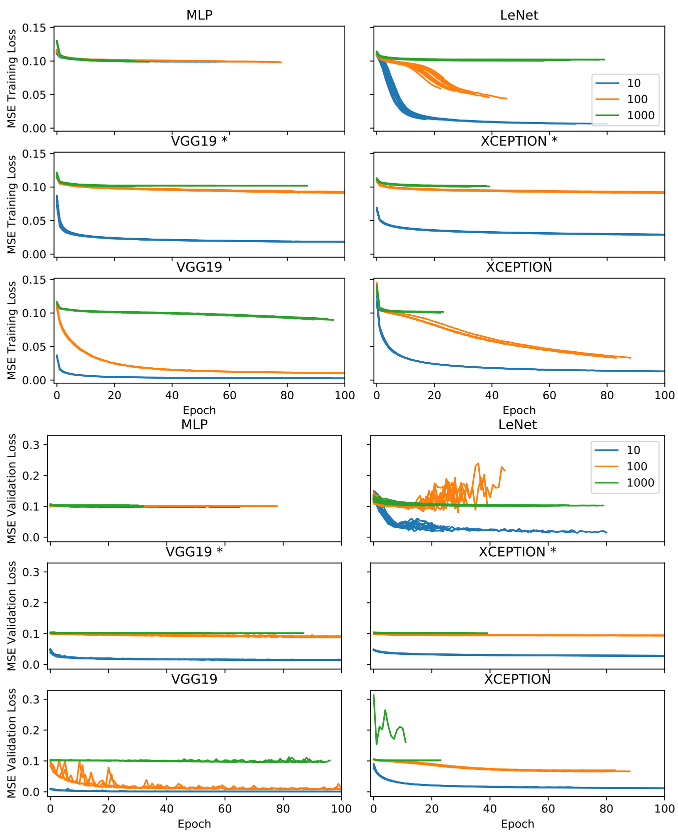


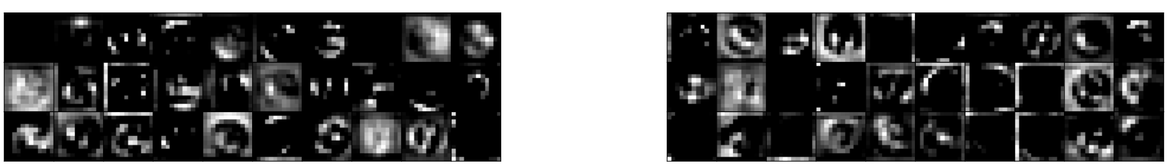
Fig. 10: **Loss plots for the weber-fechner's law experiment including VGG19 and Xception.** We visualize the MSE loss on training data and for unseen validation data after each epoch. This plot includes the VGG19 and Xception networks trained from scratch.

(a) VGG19 *, Block 3 (Conv. Layers 2+3)

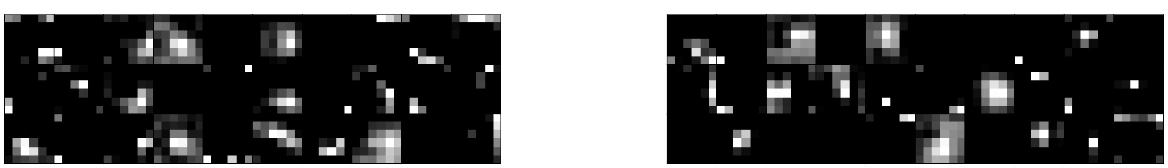
(b) VGG19 *, Block 4 (Conv. Layers 2+3)

(c) VGG19 *, Block 5 (Conv. Layers 2+3)

(d) VGG19, Block 3 (Conv. Layers 2+3)



(e) VGG19, Block 4 (Conv. Layers 2+3)



(f) VGG19, Block 5 (Conv. Layers 2+3)

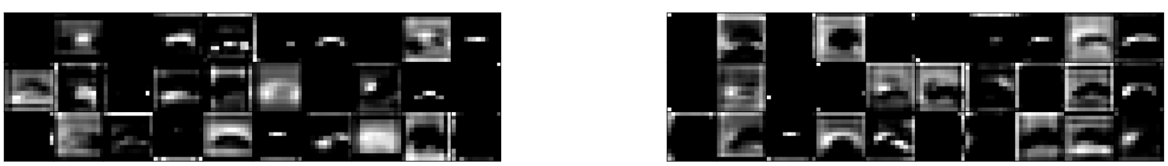
Fig. 11: **Convolutional Activation Maps for a Pie Chart.** (a)-(c) is VGG19 *, trained on ImageNet. The activation maps do not differ much which is surprising since VGG19 trained from scratch performs so much better in our experiments.

(a) VGG19 *, Block 3 (Conv. Layers 2+3)

(b) VGG19 *, Block 4 (Conv. Layers 2+3)

(c) VGG19 *, Block 5 (Conv. Layers 2+3)

(d) VGG19, Block 3 (Conv. Layers 2+3)



(e) VGG19, Block 4 (Conv. Layers 2+3)

(f) VGG19, Block 5 (Conv. Layers 2+3)

Fig. 12: **Convolutional Activation Maps for a Curvature stimuli.** (a)-(c) is VGG19 *, trained on ImageNet. The activation maps do not differ much which is surprising since VGG19 trained from scratch performs so much better in our experiments.