What do CNNs gain by imitating the visual development of primate infants? (Supplement)

BMVC 2020 Submission # 196

1 Model implementation and training details

The initial LR for all experiments is 0.1. For ImageNetH200, the LR is reduced by 0.1 every 30 epochs for a total of 100 epochs. For CIFAR10 and CIFAR100, it is reduced by 0.1 first at 150 epochs, and then at 250 epochs for a total of 350 epochs. Additionally, all results are reported over 5 trials for CIFAR10 and CIFAR100, and 2 trials for ImageNetH200. Training for all models was performed over 2 GPUs with a batch size of 128 and implemented in PyTorch. Additionally, for color control experiments, we evaluated performance wherein saturation and contrast was applied randomly with an equivalent probability to the used stage specific training occurances (for example, for CIFAR10, given that saturation ratio of 0.9 was applied for 15 epochs in stage setup, correspondingly, for control, it was randomly applied with a probability of $15/350$ for CIFAR10 training in setups with a "Static input").

2 Filter visualizations for Resnet50 training

To visualize the difference in the learning process for Resnet50 when trained in stages with a refining input distribution in comparison to when trained with a static input distribution, we analyzed the filters induced at the end of each training stage. We use the standard filter visualization technique wherein the input is optimized to maximally activate a selected set of filters. Figure 1 shows the difference in filters induced in the second convolutional layer of the Resnet50 model when trained with a gradually refining distribution in comparison to the conventional training setup for ImageNetH200. The impact of color sensitivity in terms of saturation and contrast can be seen in the earlier few stages, with filters responding to sharper and more distinct color shades in the case of a gradually refining input distribution. The final best performing models have more similar filter visualizations compared to previous stages; however, in case of the refining input distribution, there are few filters that maximally respond to the green shade which is not present for the static input distribution.

3 Input refinement illustration

Figure 2 illustrates the gradual refinement of visual input in stages and the individual effects of saturation, contrast and resolution on sample images from ImageNetH200.

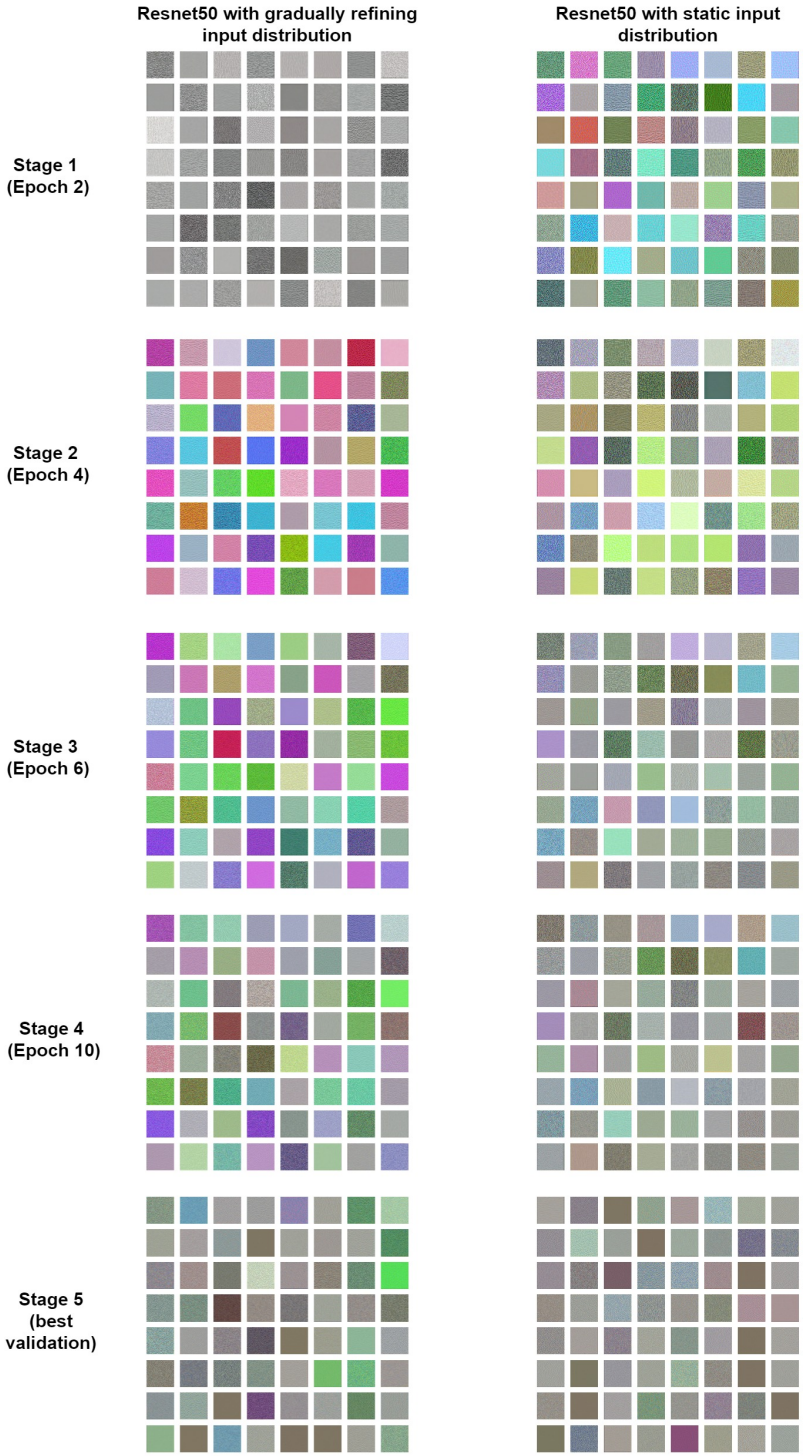

Figure 1: Comparison of maximally activating inputs for filters of the second convolution layer of Resnet50 when trained with a refining input distribution (left) and when trained with a conventional static input distribution (right) for ImageNetH200.

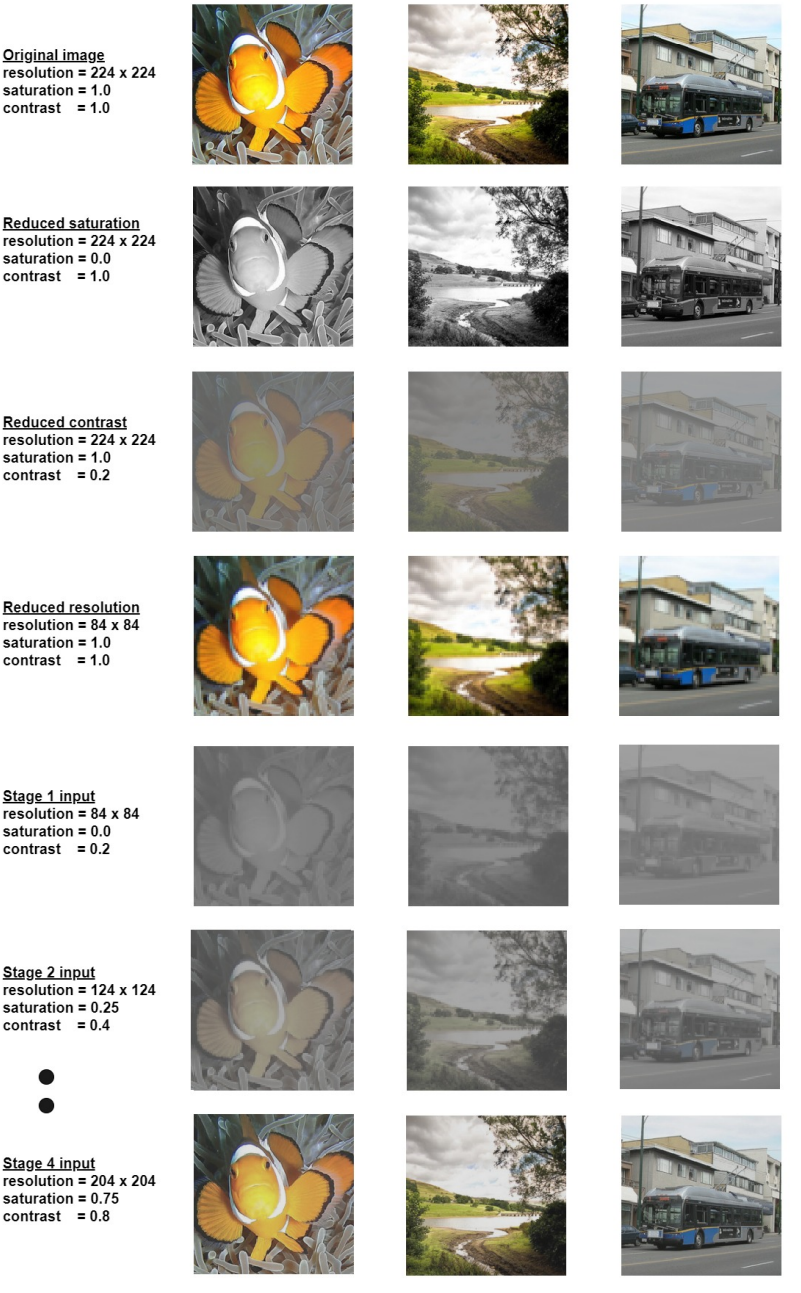

Figure 2: Illustration of the visual input refinement factors- contrast, saturation and resolution- and how stage wise training setup is performed. Stage 5 corresponds to the final stage wherein the input is fully formed (original image).