6 Supplementary material

Figure 8: Random colorized images from the ILSVRC 2012 validation set *(part 1)*.

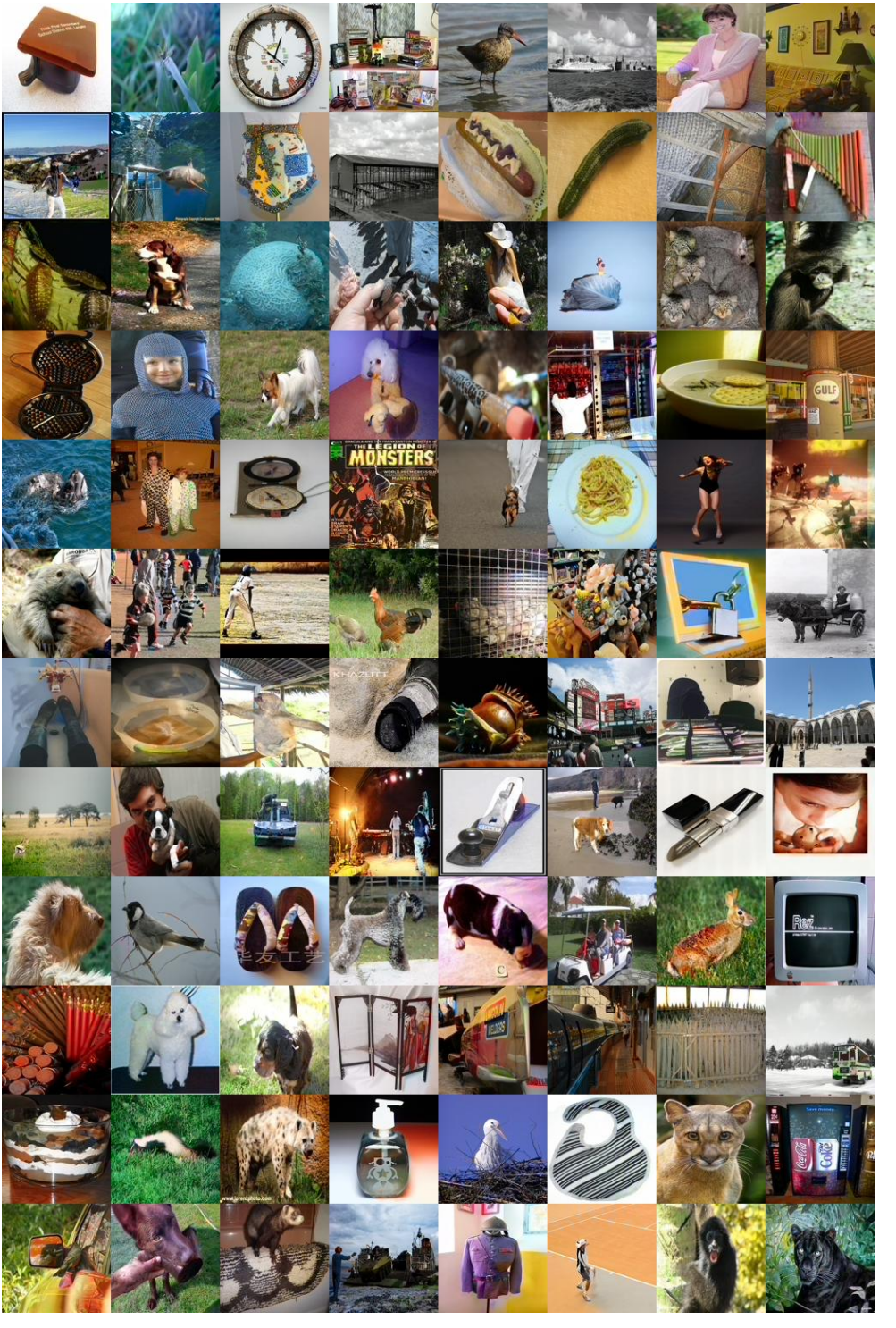

Figure 9: Random colorized images from the ILSVRC 2012 validation set *(part2)*.

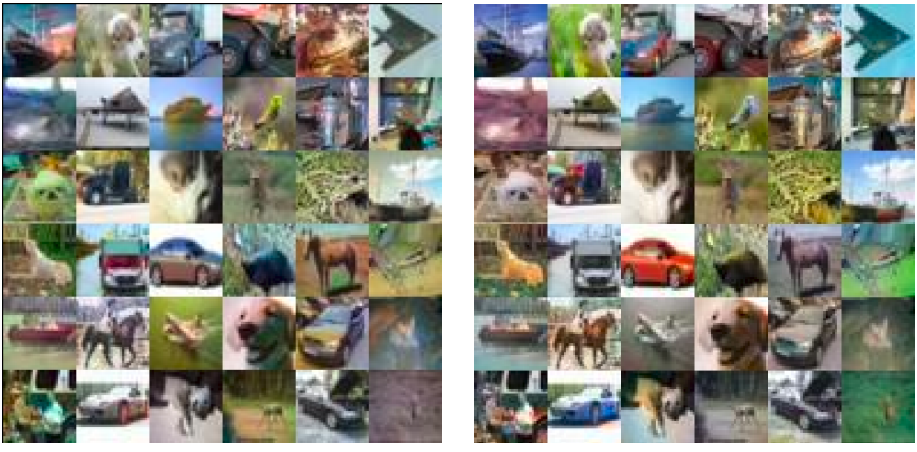

Figure 10: Comparison of CIFAR-10 colorization samples obtained without gating *(left)* and with gating *(right)*.

6.1 Model selection and gating non-linearity

Recently it was demonstrated that gatining non-linearity is useful for the task of natural image modeling [76]. Thus, we expect that gating should be also beneficial for colorization, which is closely related to image modeling.

One way to verify whether gating is useful is to perform qualitative sample analysis. We illustrate samples from the PIC model, as well as the samples from the identical model without gating non-linearity in Figure 10. We observe that even though the samples obtained with the gating mechanism appear to have higher visual quality, i.e. have slightly more saturated colors and better global consistency, it is hard to make definitive conclusion.

Thus, we also perform quantitative analysis using the likelihood measure. PIC with gating achieves the negative log-likelihood of 2.72, while its counterpart without gating achieves 2.78. The quantitative evaluation is consistent with our preliminary qualitative evaluation.

We argue negative log-likelihood on the hold-out image set may be used as a principled measure for model selection. Importantly, our metric measures how well *the joint distribution* of image colors is explained by the model. And unlike all previous metrics, which were used to evaluate image colorization performance, our metric accounts for intrinsic uncertainty of the task and, at the same time, for modeling complex interactions between pixels within one image.