Leveraging Transfer Learning to Imitate Human Behaviour for Image Colorization

BEUTH HOCHSCHULE FÜR TECHNIK BERLIN
University of Applied Sciences

Christoph Geis-Schroer, Sebastian Jäger

Learning from Images
WS 2019/2020
Prof. Dr. Kristian Hildebrand

Abstract/Introduction

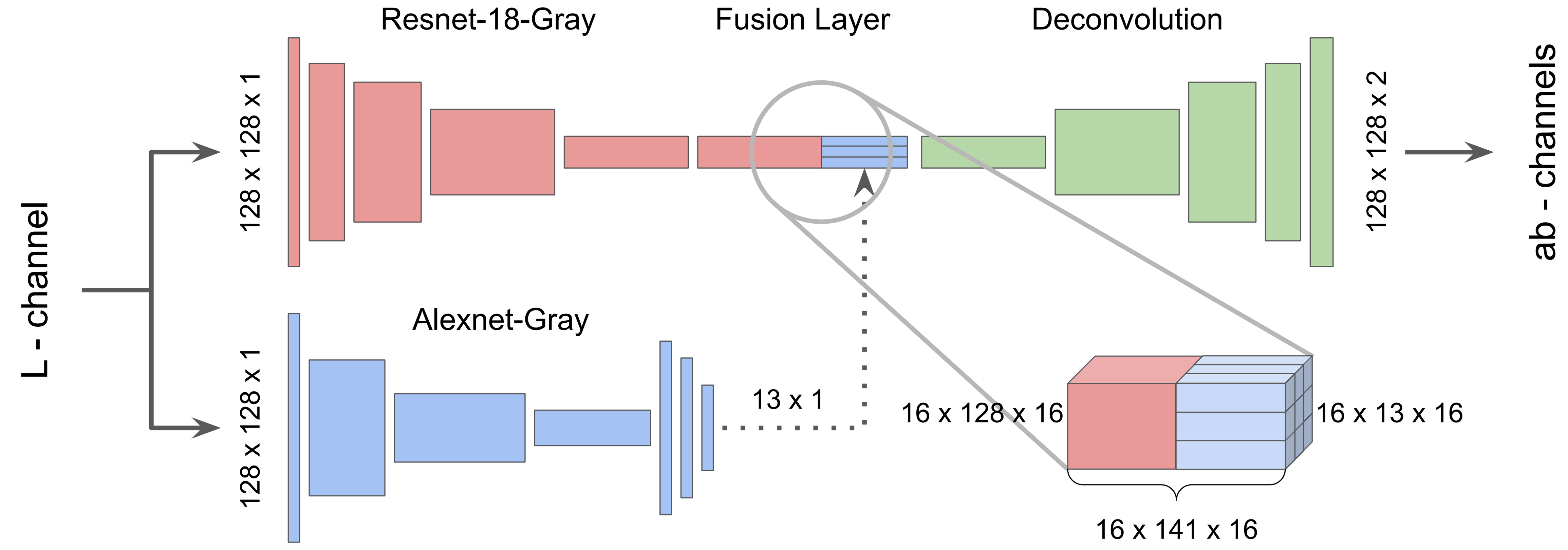
We implemented an Ensemble model for image colorization, the task of generating colored images from grayscale image inputs.

- We built upon the work by Melas-Kyriazi and Han [1]
- We leveraged transfer learning for the classification, Alexnet [2]

By treating image colorization as a classification task and using other advancements, the best results are possible [4]. Regression task settings are easier but perform worse [3]. We tried to imitate human behavior, first classify the scene and then choose sensible colors, with a regression-based (MSE loss) Ensemble model that adds separated classification results as mid-features (Fusion Layer).

Methodology

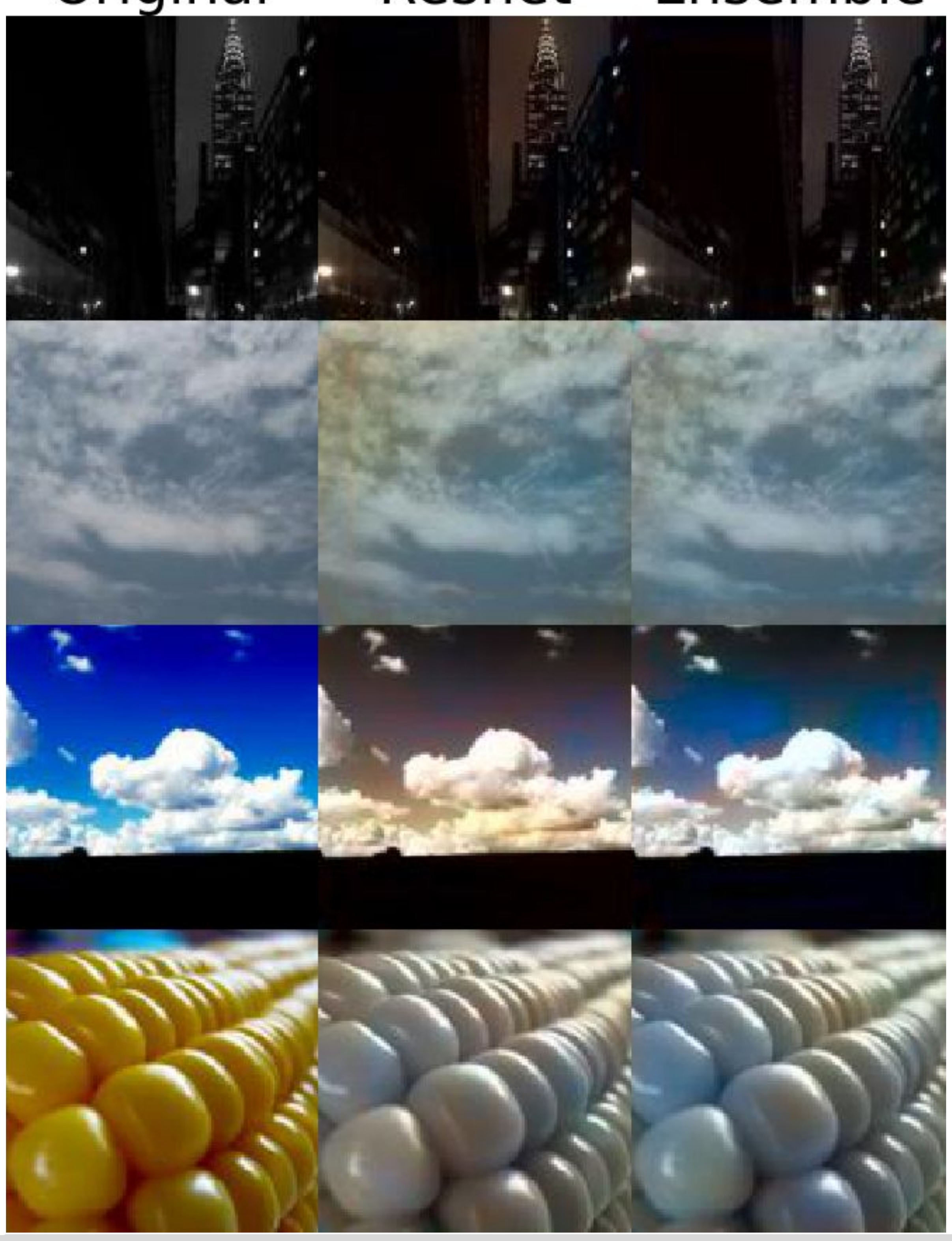
Our dataset is a subset of ImageNet [2] with only 13 classes (~10k train, ~2k test), resize to 128x128 pixels. Interpreting the images in the Lab color space allows using the lightness (L) channel as input and the color channels (ab) as output.
First, we fined tuned a pre-trained Alexnet on our dataset. Best top-3 classification accuracy: 75.73%.
Best hyperparameters after grid search are batch size: 512, L2 regularization: 0.00025, learning rate: 0.00005, Adam optimizer, and no one-hot-encoding of the classification output.

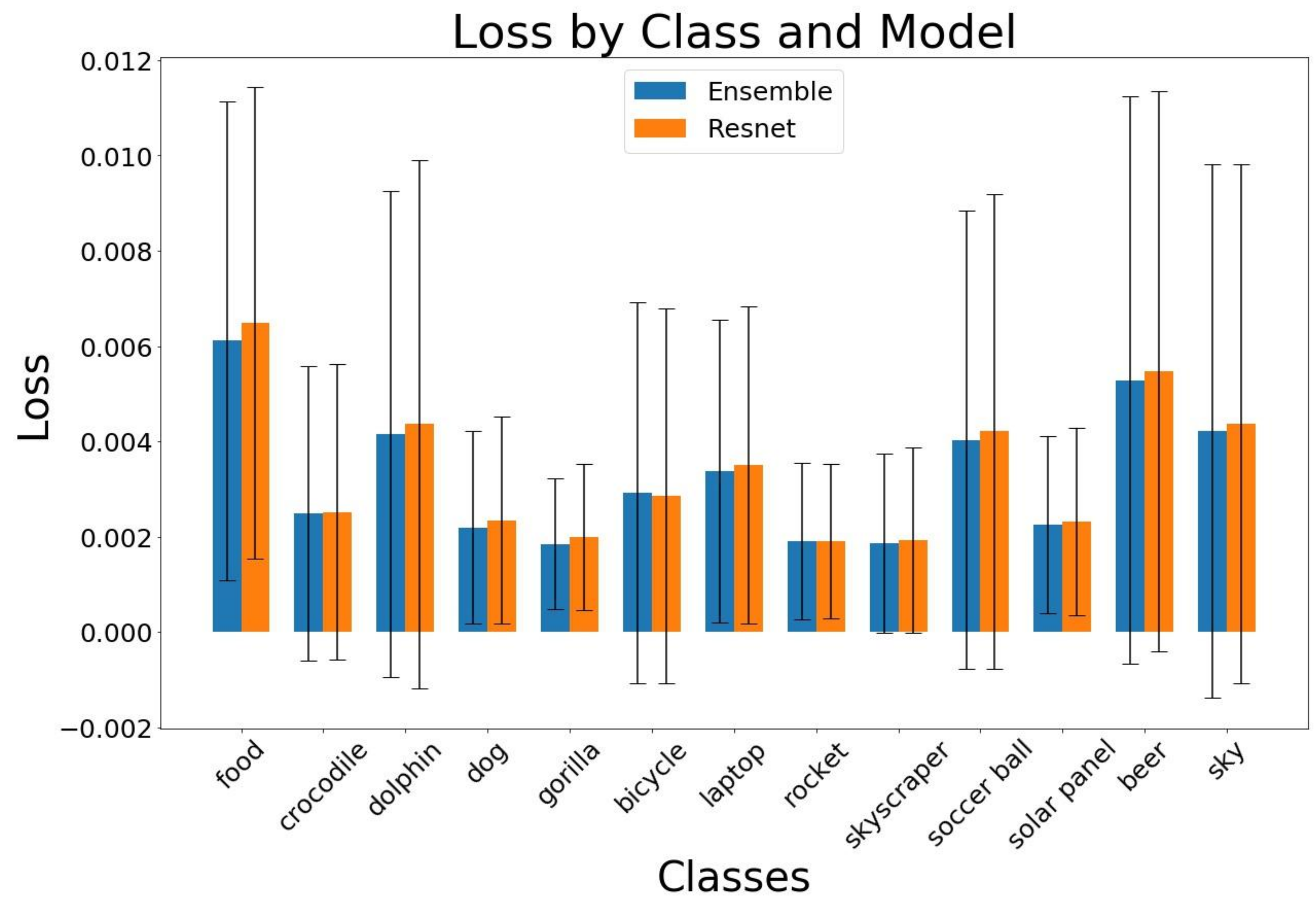


Results

The overall loss decreases with the Ensemble model by 3.82%. The small reduction leads to similar colorization results for both models. However, the ensemble model can omit most of the yellow areas the Resnet model predicts.

Taking human perception into account, the pre-classification of the images does not lead to recognizable improvements





For 11 of the 13 classes, the loss decreases with the use of the Ensemble model. Further, classes containing mostly less colorful images (i.e., rockets, skyscrapers, gorillas) yield lower losses. In most of the classes where the Ensemble model predicts better, the standard deviation decreased as well.

Conclusion/Future Work

Considering the small differences, we cannot determine the overall influence of our transfer learning approach. We suggest that more comprehensive results can be given by increasing the variance of the training dataset.
An additional issue is the image classification: It does only classify into larger groups, e.g., food or football. Here, a more narrow object detection could be implemented to recognize a strawberry (red) or the KNVB/Nederland sign (orange) and improve the colorization.

Literature

[1]: Melas-Kyriazi, L., Han G. (2018). Combining Deep Convolutional Neural Networks with Markov Random Fields for Image Colorization. https://lukemelas.github.io/image-colorization.html
[2]: Krizhevsky, A., Sutskever, I., & Hinton, G.E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. NIPS.
[3]: Charpiat, G., Hofmann, M., & Schölkopf, B. (2008). Automatic Image Colorization Via Multimodal Predictions. ECCV.
[4]: Zhang, R., Isola, P., & Efros, A.A. (2016). Colorful Image Colorization. ECCV.