Deep Learning
Project Proposal

Group 12
Aydan Yagublu (s3980804), Eric Brouwer (s3934640),
Lauren Kersten (s3950905), Thijs van der Laan (s3986721)

March 12, 2023

For our project, we wanted to compare two well-known CNN architectures in the context of a U-Net model Ronneberger et al. (2015) for the task of recolouring black and white images. The two architectures we are considering investigating are ResNet-50 He et al. (2016) and MobileNetV2 Sandler et al. (2018), as we want to compare a lightweight model (MobileNetV2 $\sim$ 3.4 million parameters) to a more heavy-weight model (ResNet-50 $\sim$ 25 million parameters) to see their performance on the task of black and white recolouring. For the dataset, we were thinking of utilising the CIFAR-10 dataset and transforming the images to black and white.

Our motivation behind picking such a problem is based on determining the performance of two different CNN architectures in an effort to evaluate and determine which architecture is more suited for the task at hand.

For the short description of what we will do, our thinking goes as follows:

We implement two different CNN architectures integrated into a bigger U-Net architecture and train these in order to colour black and white images. For training, we intend to use the structural similarity index measure (SSIM) (Wang et al., 2004). We were also thinking of utilising L1 (mean absolute error) as the loss function and comparing the two loss functions on top of comparing the two architectures.

For this, we would need a dataset that contains both black-white and coloured versions of images. We could find one that already has this or one that takes a popular benchmark dataset and convert these to black and white ourselves. After training and testing, we could also make an attempt on historical pictures to see how the models perform on those.

Regarding the evaluation of the models, we intend to first separate the dataset into an 80/20 split, such that we perform 5-fold cross-validation on only 80% of the data to perform model selection. We then perform a final test on the remaining 20% to obtain a final metric. For the final test, we intend to use both the SSIM score and L1 score between the ground truth image and the predicted image.

References

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. volume 13, pages 600–612. IEEE.