Jan Čech
Center for Machine Perception
Department of Cybernetics, Faculty of Electrical Engineering
Czech Technical University in Prague

# Master's Thesis Review

Prague, January 18, 2018

**Title:** Colorization of black-and-white images using deep neural networks
**Author:** David Futschik
**Supervisor:** doc. Ing. Daniel Sýkora, Ph.D.
**Date received:** 09/01/2018

The thesis describes a method for gray-scale cartoon image colourization. The input of the algorithm is a gray-scale (single-channel) image, the output is the colour (three channel) image. The problem is intrinsically ambiguous, since many colours appears exactly the same intensity in the gray scale. The world of cartoon images is particularly challenging for automatic colourization as the limited amount of training data is available and the artistic drawing itself is a great simplification compared to real photographs. The cartoons has almost no texture which is a strong cue of semantics and thus colour, and the semantics is carried mostly by a shape of the objects.

The proposed solution is based on training deep convolutional neural network and eventually followed by a heuristic post-processing. The thesis uses a Rumcajs dataset of several episodes of a famous Czech cartoon series drawn by Radek Pilař. Training and test sets were disjoint, some of the episodes were used for training, the others for testing. Two network architectures were tested (plain CNN, resnet), and several loss functions ($L_2$-regression, and several classification losses).

The thesis is well structured. The introduction, a review of elementary building blogs of CNNs, and a review of related literature are summarized prior to the technical part. The method is presented in a reasonable level of details, together with the dataset and the training process. The experiments are carefully performed showing both qualitatively and quantitatively results of the considered algorithms. A simple, but apparently quite efficient, heuristic post-processing is proposed to remove the most striking artifacts of inhomogenous colourization. This is done by segmenting the images into homogeneous regions and flood-filling these regions with a single colour, an average of the predicted colour over the region and over the ensemble of colour prediction networks.

To name a few weaknesses of the thesis:

1. The proposed method is inspired by Zhang et al. [1], however it is unclear the extent of this inspiration. The architecture of the plain CNN is almost identical, the loss function is very similar. On the other hand, there are novel contributions over [1] which should have been made explicit. I would propose to show results of colourization by [1] (trained on natural photographs) without retraining to demonstrate the challenges of the cartoon world. The thesis only claims that the comparison would not be fair, which is true in one hand but a pity in the other.

2. Some of the technical explanations are not easily comprehensible, e.g. first two paragraphs on page 25 on channel estimation. More formal presentation and a forward reference to Chapter 6 of the CNN architecture would be worthy here. Loss function (5.2) is not exactly the Kullback-Leibler divergence as stated above.

3. I would suggest to show training plots, i.e. the training and validation error/loss as a function of the epoch. This would prove the training is healthy without over-fitting.

4. Ironically, the simplest model ($L_2$-regression) works the best, despite all the arguments for the classification-loss networks. This should have been commented more thoroughly.

5. I wonder, why the author did not experiment with Generative Adversial Networks (GANs), e.g Isola et al. [2]. The GANs are only shortly noted in Sec 9.1 as a possible future work. The advantage of GANs is that the source-target discrepancy loss is combined with the adversarial loss, which should preserve the output images stayed on the manifold of target training images. No classifier should distinguish the synthesised images from the real training images, that is desired in computer graphics.

On the other hand, the above weaknesses are minor and easily rectifiable. I especially liked the effort to justify the choices made in the algorithm design (e.g. dilated convolutions to enlarge the receptive field to capture the shape, or various loss functions), the elaborate experimental validation and fair results achieved considering challenges of the problem.

The candidate certainly proved that he is able to do a high quality-research and to deliver a solid report. Considering all above, I suggest evaluating the thesis as

A – excellent.

Ing. Jan Čech, Ph.D.

# References


[1] R. Zhang, P. Isola, and A. A. Efros, "Colorful image colorization," in *ECCV*, 2016.

[2] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial nets," in *CVPR*, 2017.