# Multi-Task Image Colorization and Classification – Deep Machine Learning

Filip Anjou, Stanislaw Kolodziejczyk

Chalmers University of Technology
Electrical Engineering Department

CHALMERS
UNIVERSITY OF TECHNOLOGY

## Introduction

**Image colorization — does the model get any better if it knows what it's looking at?**

We trained and compared three models on a subset of ImageNet to evaluate whether there was any benefit in performing two tasks at once.

### Research questions

- Is colorization easier for a model that is also trained to classify the image?
- Is classification easier for a model that is also trained to colorize the image?

## Background

### Color spaces

- RGB
  - Intuitive
  - Sub-optimal (color information in all three channels)
- CIELAB
  - One channel for lightness ($L^*$)
  - Remaining two channels ($a^*$ and $b^*$) store color information
  - Easy to separate lightness from color information

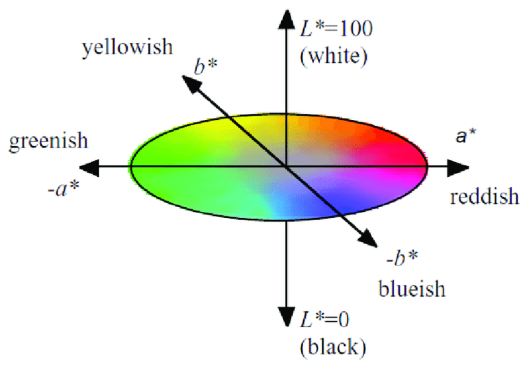


Figure 1: CIELAB color space

### Colorization head

- Transposed convolutions (deconvolution)
- Upsampling from feature space to ($a^*$ and $b^*$) color channels of image

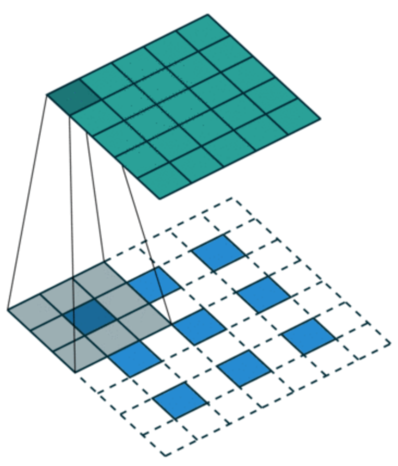

Figure 2: Transposed convolution

### Previous work on image colorization

- Plain feed-forward networks
- Generative Adversarial Networks (GAN)
  - pix2pix (cGAN)
- Diffusion models
  - Palette
  - Uses RGB color space

**Previous work on multi-task learning** has shown improvement in the main task when the model also performs other adjacent tasks [1].

### Performance metrics

- L1 loss (Mean Absolute Error)
  - How close to the correct color is each pixel?
- Adversarial loss (GAN loss)
  - How authentic does another model think this image looks?
- Fréchet Inception Distance (FID)
  - How similar is the distribution of the generated images to that of the ground-truth?

## Method

**Train three models on an ocean-themed subset of ImageNet (5 classes)**

- Single-task classification
- Single-task colorization
- Multi-task model

Each model has a pre-trained **ResNet-18 feature extractor base**, with either a classification or a colorization head (or both in the case of the multi-task model). We used **GAN** architecture for the colorization model.

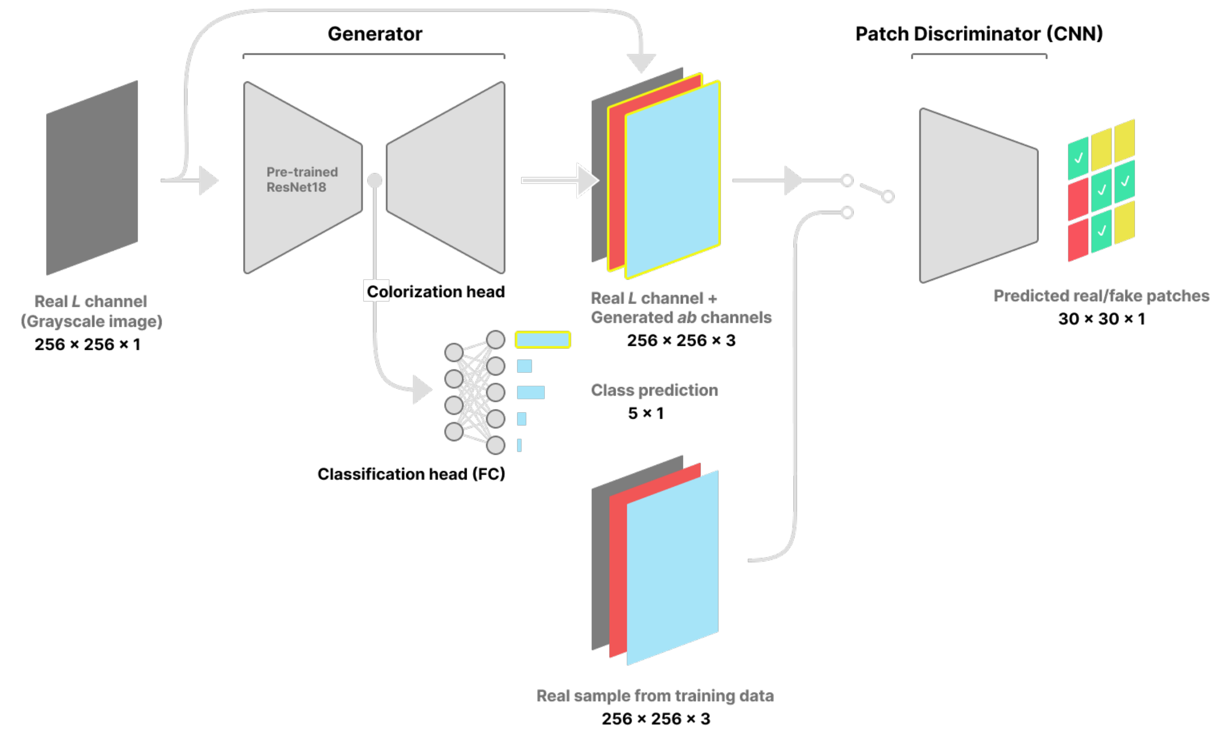


Figure 3: Multi-task learning setup

### Training schedule

1. Train only head(s) for 30 epochs, keeping base model's weights frozen
2. Unfreeze base model's weights
3. Fine-tune whole network for 15 more epochs

## Results

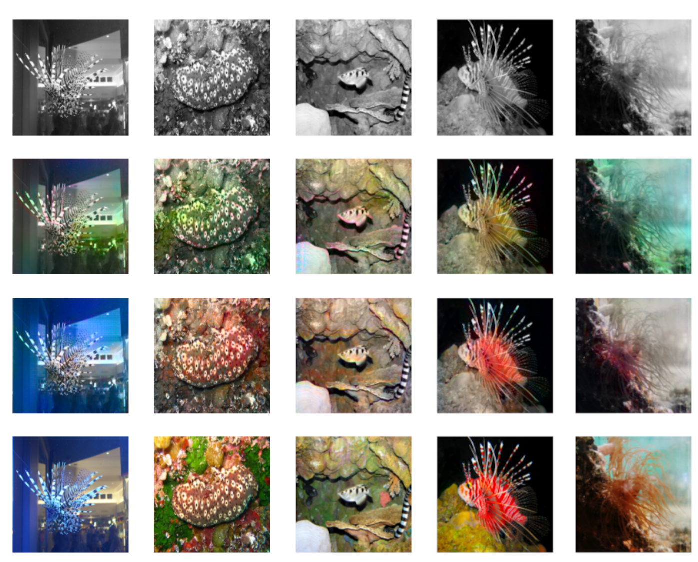

Figure 4: Single-task (2nd row) vs multi-task (3rd row) colorization on batch of samples from validation set. Top row is grayscale input, and bottom row is color ground-truth.

**Comparison!**

- Single-task model produces cooler blue-green colors
- Multi-task model produces warmer red colors
- Qualitatively, we think the multi-task model produces more natural-looking images
- Neither model is able to exactly reproduce the colors of the original image

Classification accuracy (validation set)

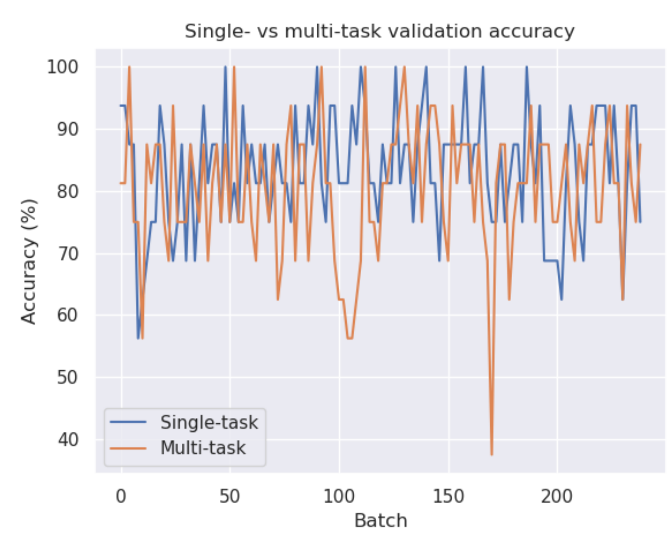


Figure 5: Classification accuracy of single- and multi-task models on the validation set

Colorization loss (L1) (validation set)

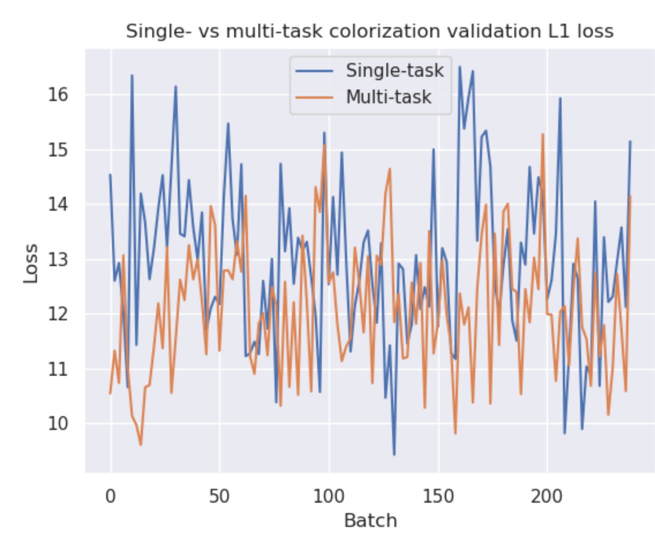


Figure 6: Colorization loss (L1) of the single- and multi-task models on the validation set

## References

[1] Zhanpeng Zhang, Ping Luo, Chen Change Loy, and Xiaoou Tang. Facial landmark detection by deep multi-task learning. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision – ECCV 2014*, pages 94–108, Cham, 2014. Springer International Publishing.


anjouf@chalmers.se