A review and analysis of the existing literature on grayscale photography colorization using CNNs

Alexandru Marian Adăscăliței
Babeș-Bolyai University

WeADL 2021 Workshop

The workshop is organized under the umbrella of WeaMyL, project funded by the EEA and Norway Grants under the number RO-NO-2019-0133. Contract: No 26/2020.

Working together for a green, competitive and inclusive Europe

Outline

1. Problem statement, introduction, and motivation
2. Research Questions
3. Colorization Patterns
4. Colorization Models
5. Results Analysis
6. Conclusions and Future Work

What is colorization?

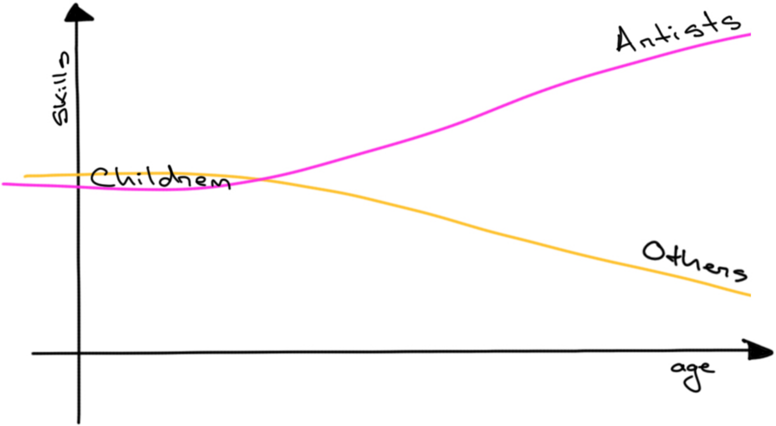


Figure: Colorization learning curve as seen from a human perspective.

Problem Statement

Photography colorization, in our context, is the task of artificially reconstructing color information in a picture that has never been captured on a storage medium capable of recording color.

Introduction

Figure: The Paper Time Machine, by Wolfgang Wild and Jordan J. Lloyd

Introduction

- deep learning algorithms are predicting the chromaticity through either a discriminative, or generative learning
- artists, such as those from Dynamichrome [3], are closing the gap through the manually constructed layers which often come from intuition
- fooling the human perception of truth is the main goal of any method, as monochromatic areas of a picture may have multiple plausible colorization

Introduction

Figure: Visual decomposition of the RGB and LAB layers.

Motivation

Why would someone invest in colorization?

- **Medicine**: improved user interfaces for diagnostic purposes
- **Communications**: improvements in compression algorithms, decreasing the waiting time
- **Games**: rendering photo-realistic scenes
- **Arts**: restoring old Hollywood movies, comics, and legacy photography
- **Computational Intelligence**: proxy for other learning tasks

Motivation

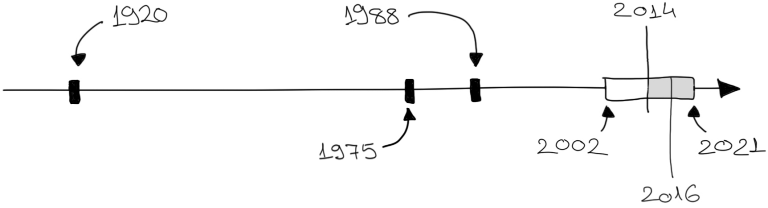


Figure: The role of timing in seizing research opportunities, starting with Wilson Markle and Brian Hunt, and ending up with research initiatives published a couple of months ago.

Research Questions

- What patterns and models are usually followed?
- What are the implications of Convolutional Neural Network?
- How well would these methods perform in professional applications?

Colorization Patterns

Data-Driven Colorization

- early iterations heavily relied on human interventions
- leveraging large-scale datasets and GPU performance, fully-automatic colorization became achievable

Human-in-the-Loop Colorization

With data-driven approaches, user preferences were not taken into consideration, hence the need for additional solutions:

- based on textual descriptions
- based on color hints
- based on reference color images

Based on Textual Descriptions

- notes were often placed on the back of legacy photography
- **social media platforms** are improving their indexing systems
  - words and sentences associated with the visual content
- building on the idea that particular colors are associated with **complex semantic concepts**
  - language specific colors: English has eleven basic color categories, Russian twelve
  - a language may have only three basic color categories
- imagine that a *cold evening* varies in nuances of blue, while the *golden hour* covers everything in warm colors

Based on Textual Descriptions

- models that **join textual** and **visual** feature maps, with expensive computational costs due to the number of parameters
- balancing image segmentation - Hu et al. [4], and fusion modules - Chen et al. [2]
- for parameters efficiency we may apply feature-wise linear modulation - Perez et al. 2018

Based on Color Hints

Figure: Capture from the application proposed in Zhang et al. [9].

Based on Color Hints

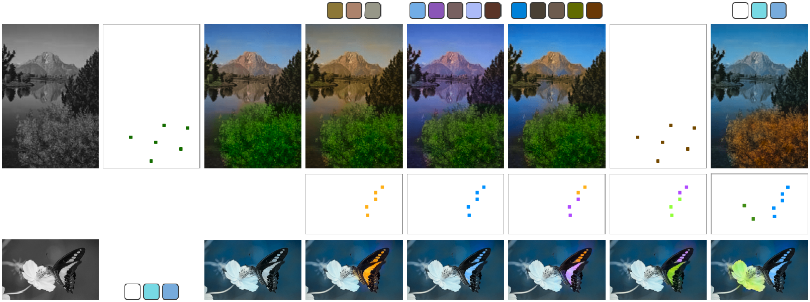
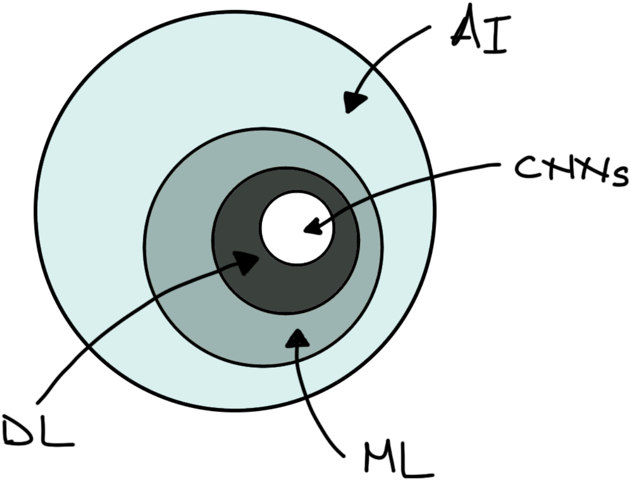


Figure: Capture from the model proposed in Xiao et al. [7].

Based on Reference Color Images

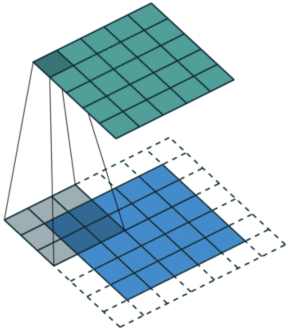
- **transferring** the **chromaticity** information from a semantically related color image to a target grayscale image
- allows for a multi-modal colorization
- the user may provide an image, or the system may retrieve the appropriate one
- imagine passing colors from a cherry blossom to a black and white Californian coast image, obtaining synthetic, but artistic pink waves

Deep Learning Models



CNN-based Models

- the network's most important aspect are the convolutional layers, made up of convolutional kernels (filters)
- when convolved with the input image, these filters are generating the feature maps



CNN-based Models

- these features are collected from various components and compressed, then later up-scaled to the original image size
- the image ratio must be preserved (using padding), and distortions must be prevented (using stride instead of pooling)

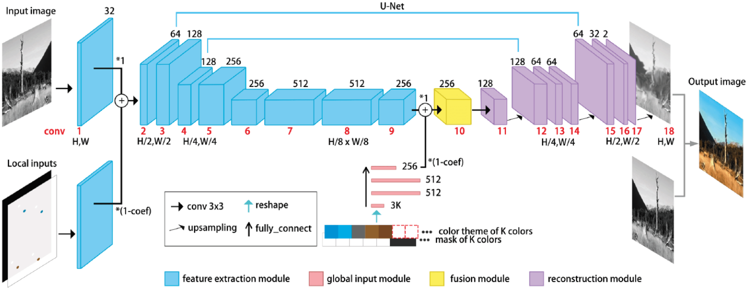


Figure: Network architecture from Xiao et al. [7].

CNN-based Models

- low, middle, and global features extraction
- predictions are not always deterministic, but often probabilistic
- discriminative models: VGG variants and U-Net based architectures
- generative model: Pixel Convolutional Neural Network
- end-to-end learning is often used
  - alleviate the bias encapsulated with various decisions
  - reduce artifacts
  - no need for hand-designed components

CNN-based Models

We often noticed the following objective function strategies being applied to the networks:

- **Huber Loss**, **L2**, *Kullback–Leibler divergence*, *Perception Loss*, *cross-entropy*, *Color Embedding*, *Color Generation*, and *Semantic Loss*

Open problems

- conservative guess (everything can be brown)
- lack of color normalization
- color bleeding
- small objects are ignored

Results Analysis

- since early 80's, the number of solutions proposed in literature remained small (aprox. 85 papers)
- the human eye may be fooled by only a dozen of these algorithms
- we wondered if we can reproduce the results on a manually curated dataset

Results Analysis

Paper	Colorization Metrics						Recommended
	↓ LPIPS	σ	↑ PSNR	σ	↑ SSIM	σ	types of images
Antic et al. [1]	0.18389	0.08614	13.36557	3.55204	0.73828	0.12560	all
Iizuka et al. [5]	0.18068	0.06863	15.80264	3.94617	0.77813	0.12155	events, portraits, landscapes
Zhang et al. [8]	0.22174	0.08790	13.60779	4.01649	0.77388	0.11998	landscapes
Kumar et al. [6]	0.30766	0.07357	11.22693	3.14602	0.53996	0.15731	close-up portraits, landscapes

Table: Performance evaluation made on **urban landscapes** and **events**, **objects**, and **portraits**.

Results Analysis

Figure: A visual validation of the results obtained with Antic et al. [1].

Metrics

- **Most used metrics**: Peak Signal-to-Noise Ratio, Structural Similarity Index Measure, Learned Perceptual Image Patch Similarity
- **Alternative metrics**: Patch-based Contrast Quality Index and the Underwater Image Quality Measure
- **Turing Test** - having a person assessing the colorization results is the golden standard at the moment

Metrics

- **LPIPS** uses deep network activations as a perceptual similarity metric, which works surprisingly well, and comes closer to the human preference in raking
- in general, metrics account for the mean **luminosity**, change in **contrast**, **structural distortion**, **sharpness**, and **colorfulness**

Colorization Software Reliability

- only a few colorization algorithms are available online
- the setup and hardware requirements are a challenge
- GitHub repositories are not often well maintained

How well would these methods perform in professional applications?

- integrated into products targeting the general public
- Zhang et al. [9] was included in Photoshop Elements 2020

Conclusions and Future Work

Our work sets the grounds for further colorization initiatives.

Future Work

- extend the experimental evaluation
- contribute on making these models more accessible to the general public
- improve on the existing CNN-based approaches

Thank you!
Questions?

ANTIC, J.
jantic/deoldify: A deep learning based project for colorizing and restoring old images (and video!).
github.com/jantic/DeOldify [Online; accessed Dec 4, 2020].

CHEN, J., SHEN, Y., GAO, J., LIU, J., AND LIU, X.
Language-based image editing with recurrent attentive models.

github.com/Jianbo-Lab/LBIE, 2018.

DYNAMICHROME.
Showcase.
dynamichrome.com [Online; accessed Dec 4, 2020].

HU, R., ROHRBACH, M., AND DARRELL, T.
Segmentation from natural language expressions, 2016.

IIZUKA, S., SIMO-SERRA, E., AND ISHIKAWA, H.

Let there be color!: joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification.
ACM Transactions on Graphics 35 (07 2016), 1–11.

KUMAR, M., WEISSENBORN, D., AND KALCHBRENNER, N.
Colorization transformer.
github.com/google-research/google-research/tree/master/coltran, 2021.

XIAO, Y., ZHOU, P., AND ZHENG, Y.
Interactive deep colorization with simultaneous global and local inputs, 2018.

ZHANG, R., ISOLA, P., AND EFROS, A. A.
Colorful image colorization, 2016.

ZHANG, R., ZHU, J.-Y., ISOLA, P., GENG, X., LIN, A. S., YU, T., AND EFROS, A. A.

Real-time user-guided image colorization with learned deep priors.
github.com/junyanz/interactive-deep-colorization, 2017.