Colorful Image Colorization

School of Industrial and Management Engineering, Korea University

Jin Yong Jeong

KOREA UNIVERSITY

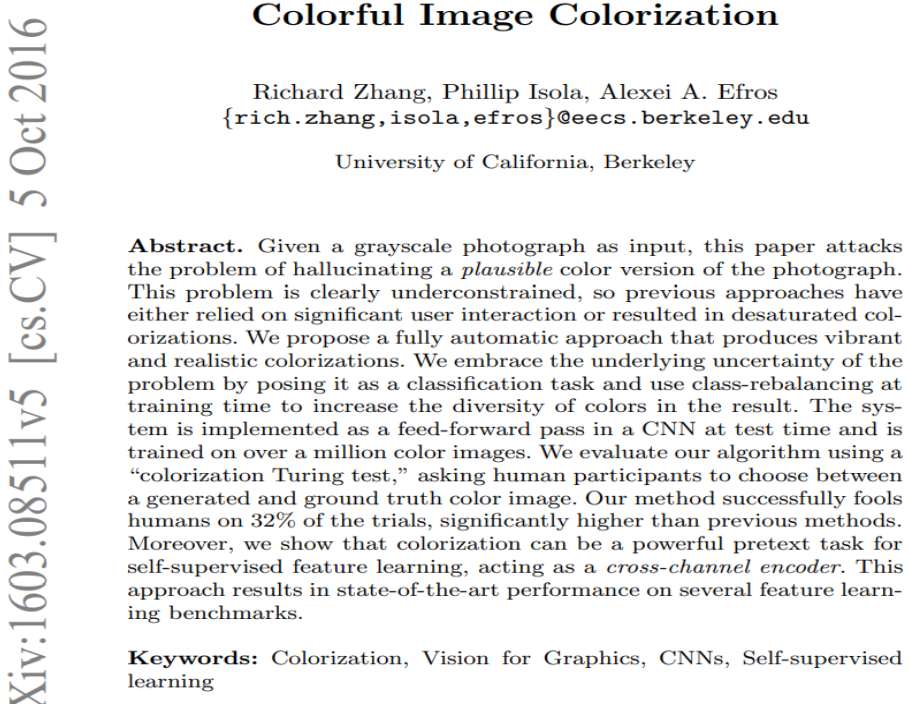
DMQA hcai Human-Centered Artificial Intelligence Center

Contents

Research Purpose

❖ Colorful Image Colorization (ECCV 2016)

- University of California, Berkeley에서 연구하였고 2022년 5월 5일 기준 2530회 인용
- 논문 내 제안 방법론은 Gray Scale 이미지를 **Colorization하는 방법론**
- **Colorization 방법론을** Self-supervised Learning 방법론 중 **Pretext task로써 추가 적용**



Colorful Image Colorization

Richard Zhang, Phillip Isola, Alexei A. Efros
{rich.zhang,isola,efros}@eecs.berkeley.edu

University of California, Berkeley

Abstract. Given a grayscale photograph as input, this paper attacks the problem of hallucinating a *plausible* color version of the photograph. This problem is clearly underconstrained, so previous approaches have either relied on significant user interaction or resulted in desaturated colorizations. We propose a fully automatic approach that produces vibrant and realistic colorizations. We embrace the underlying uncertainty of the problem by posing it as a classification task and use class-rebalancing at training time to increase the diversity of colors in the result. The system is implemented as a feed-forward pass in a CNN at test time and is trained on over a million color images. We evaluate our algorithm using a "colorization Turing test," asking human participants to choose between a generated and ground truth color image. Our method successfully fools humans on 32% of the trials, significantly higher than previous methods. Moreover, we show that colorization can be a powerful pretext task for self-supervised feature learning, acting as a *cross-channel encoder*. This approach results in state-of-the-art performance on several feature learning benchmarks.

Keywords: Colorization, Vision for Graphics, CNNs, Self-supervised learning

Research Purpose

❖ Colorful Image Colorization (ECCV 2016)

- 해당 논문의 Contributions는 **Automatic image colorization**과 **Self-supervised pretext task** 두 가지로 나뉨
- 논문의 궁극적인 **목표는** Image Colorization을 통해 Ground-Truth와 일치하는 사진을 만드는게 아니라 **사람을 속일 수 있는 그럴듯한 사진을 만드는 것**

Fig. 1. Example input grayscale photos and output colorizations from our algorithm. These examples are cases where our model works especially well. Please visit `http://richzhang.github.io/colorization/` to see the full range of results and to try our model and code. Best viewed in color (obviously).

Background

❖ Lab color space

- L*a*b* 색 공간은 RGB, CMYK 등이 표현할 수 있는 모든 색역을 포함하는 색 공간
  - ✓ L : 밝기를 의미하는 0 ~ 100 사이 값, 커질수록 검은색
  - ✓ a : 음수이면 초록, 양수이면 빨강과 보라
  - ✓ b : 음수이면 파랑, 양수이면 노랑
  - ✓ a, b 색 조합은 총 313가지

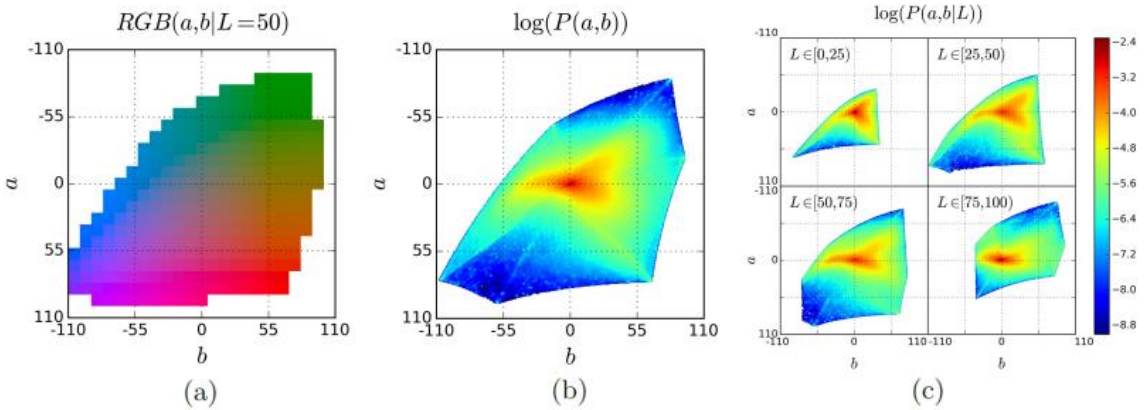


Fig. 3. (a) Quantized *ab* color space with a grid size of 10. A total of 313 *ab* pairs are in gamut. (b) Empirical probability distribution of *ab* values, shown in log scale. (c) Empirical probability distribution of *ab* values, conditioned on *L*, shown in log scale.

Proposed Method

❖ Network architecture

- Gray scale 이미지를 의미하는 Lightness channel L을 모델에 입력
- L*a*b* color space 상에서 L에 상응하는 실질적인 이미지 색인 a와 b channels를 예측
  - ✓ Lightness channel $X \in \mathbb{R}^{HXWX1}$이고 a, b 색상 $Y \in \mathbb{R}^{HXWX2}$일 때, $\hat{Y} = F(X)$ 를 매핑하는 함수 F(·)를 학습하는 것이 목표, **색 분포에서 가능한 색을 찾는다는 의미**
- 예측된 색상 a와 b에 L을 추가하여 Lab image를 출력

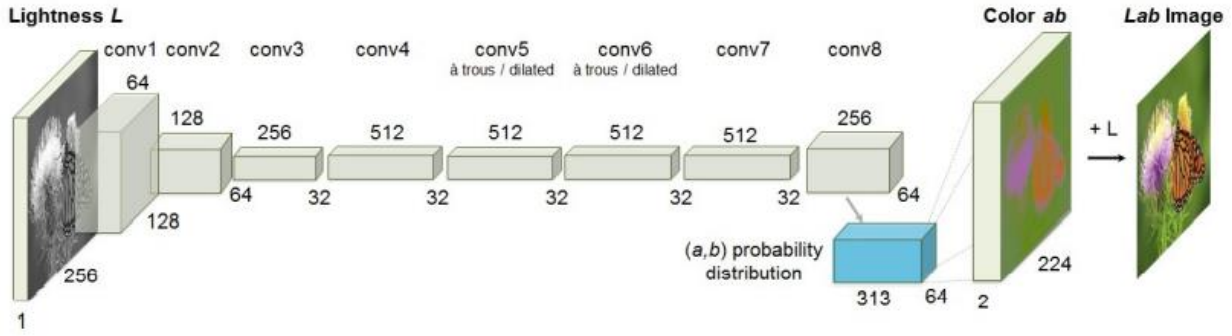


Fig. 2. Our network architecture. Each `conv` layer refers to a block of 2 or 3 repeated `conv` and `ReLU` layers, followed by a `BatchNorm` [30] layer. The net has no `pool` layers. All changes in resolution are achieved through spatial downsampling or upsampling between `conv` blocks.

Proposed Method

❖ Approach – Objective Function

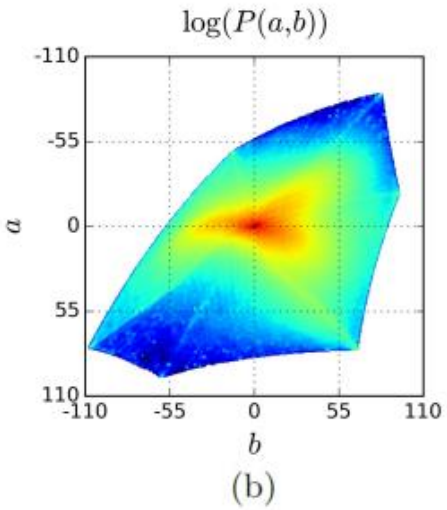
- Euclidean loss $L2(\hat{Y}, Y) = \frac{1}{2}\sum_{h,w}\left\|Y_{h,w} - \hat{Y}_{h,w}\right\|_2^2$
- L2 loss는 ab의 평균이 최적해가 되므로 색상 예측에서 회색 빛이 도는 결과를 보임
- a와 b의 313가지 전체 색 조합 출력 공간에 매핑 시킬 수 있는 **새로운 목적 함수 필요**

- **Multinomial cross entropy loss** $\boldsymbol{L_{cl}(\hat{Z}, Z) = -\sum_{h,w} v(Z_{h,w}) \sum_q Z_{h,w,q} \log(\hat{Z}_{h,w,q})}$
- 입력 X를 가능한 색상들의 확률 분포 $\hat{Z} \in [0,1]^{H X W X Q}$에 매핑시킴
  - ✓ Q = 313, a와 b의 전체 색 조합 가지 수
- Ground truth color Y는 soft-encoding scheme을 사용하여 vector Z로 변환
- 수식에서 v (·)는 weighting term
  - ✓ Approach – Class rebalancing에 자세하게 설명

Proposed Method

❖ Approach – Class rebalancing (1/2)

- 학습 이미지로 사용된 ImageNet은 구름, 포장도로, 흙과 같은 자연 이미지들이 다수 존재
  - ✓ **ab값 분포가 낮은 값으로 편향**되어 있음
  - ✓ Colorization을 할 때 a와 b가 낮은 값으로 채색될 확률이 높다는 의미
- **색의 불균형을 해결하기 위해** 학습 시 각 픽셀이 가지는 색의 rarity를 기반으로 **loss**를 **reweighting**하는 방법을 사용



(b)

Proposed Method

❖ Approach – Class rebalancing (2/2)

- Multinomial cross entropy loss $L_{cl}(\hat{Z}, Z) = -\sum_{h,w} \boldsymbol{v}(Z_{h,w}) \sum_q Z_{h,w,q} \log(\hat{Z}_{h,w,q})$
- $\mathbf{v}(\cdot)$는 weighting term으로 **색의 불균형을 해결하기 위해 사용**
- $v(Z_{h,w}) = w_{q^*}$, where $q^* = argmax_q Z_{h,w,q}$
- $w \propto ((1-\lambda)\tilde{p} + \frac{\lambda}{Q})^{-1}$, $E[w] = \sum_q \tilde{p}_q w_q = 1$
  - ✓ $\tilde{p}$는 empirical distribution with Gaussian kernel
  - ✓ Q = 313, a와 b의 전체 색 조합 수
- 각 픽셀은 w에 의해 reweighting
- 본 논문에서는 λ=1/2, σ=5일 때 가장 좋은 성능을 보인다고 함

Proposed Method

❖ Approach – Class probabilities to point estimates

- 예측된 분포 $\hat{Z}$를 실질적 색상 값으로 표현하기 위해 ab공간 상의 점 추정 값 $\hat{Y}$로 매핑함
- 점 추정은 각 픽셀에 독립적으로 작용함

- Anneled-mean of the distribution : $H(Z_{h,w}) = E[f_T(Z_{h,w})],\ f_T(Z) = \frac{\exp(\frac{\log(z)}{T})}{\sum_q \exp(\frac{\log(z_q)}{T})}$
  - ✓ 논문에서 T = 0.38일 때 최적의 색 분포가 나온다고 주장함

Fig. 4. The effect of temperature parameter T on the *annealed-mean* output (Equation 5). The left-most images show the means of the predicted color distributions and the right-most show the modes. We use $T = 0.38$ in our system.

Experiment

❖ 다른 Colorization 방법론들과 비교

- Class rebancing과 **AMT 지표에서 가장 높은 성능**을 보임
- **AMT는 사람에게 직접 평가하는 방법**이므로, 본 논문에서 '그럴듯한 ' 이미지를 만드는 것이 목적이기에 **AMT가 결과에 대한 더 정확한 평가라고 주장**함

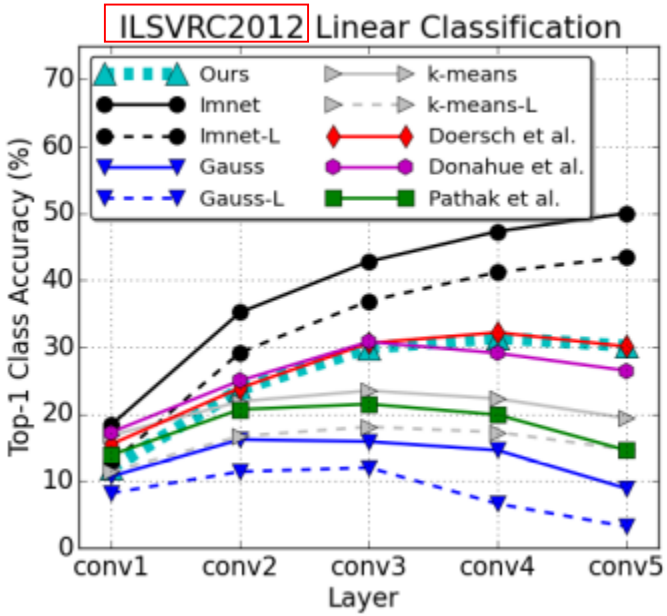
Colorization Results on ImageNet							
Method	Model Params (MB)	Feats (MB)	Runtime (ms)	AuC non-rebal (%)	AuC rebal (%)	VGG Top-1 Class Acc (%)	AMT Labeled Real (%)
Ground Truth	–	–	–	100	100	68.3	50
Gray	–	–	–	89.1	58.0	52.7	–
Random	–	–	–	84.2	57.3	41.0	13.0±4.4
Dahl [2]	–	–	–	90.4	58.9	48.7	18.3±2.8
Larsson et al. [23]	588	495	122.1	**91.7**	65.9	**59.4**	**27.2±2.7**
Ours (L2)	129	127	17.8	91.2	64.4	54.9	21.2±2.5
Ours (L2, ft)	129	127	17.8	91.5	66.2	56.5	23.9±2.8
Ours (class)	129	142	22.1	91.6	65.1	56.6	25.2±2.7
Ours (full)	129	142	22.1	89.5	**67.3**	56.0	**32.3±2.2**

Table 1. Colorization results on 10k images in the ImageNet validation set [28], as used in [23]. AuC refers to the area under the curve of the cumulative error distribution over *ab* space [22]. Results column 2 shows the class-balanced variant of this metric. Column 3 is the classification accuracy after colorization using the VGG-16 [5] network. Column 4 shows results from our AMT *real vs. fake* test (with mean and standard error reported, estimated by bootstrap [34]). Note that an algorithm that produces ground truth images would achieve 50% performance in expectation. Higher is better for all metrics. Rows refer to different algorithms; see text for a description of each. Parameter and feature memory, and runtime, were measured on a Titan X GPU using *Caffe* [35].

Experiment

❖ Self-supervised learning의 pretext task로써 Colorization 성능 평가 (1/2)

- Colorization task로 학습된 network 가중치를 freeze시킨 후, 각 convolution layer 이후 linear classifier 를 학습함
- Imnet은 ImageNet으로 사전학습 된 AlexNet
- 제안 방법론은 Conv1에서 Gray scale image를 input으로 받았기 때문에 낮은 성능을 보임
- Conv1 이후 layer들에서는 다른 Self-supervised learning 방법론들 이상의 성능을 보임



Experiment

❖ Self-supervised learning의 pretext task로써 Colorization 성능 평가 (2/2)

- Computer vision 중 객체 인식 관련 대표적인 벤치마크 Pascal VOC 20**로 성능 평가
- 제안 방법론을 Gray image와 Color image로 각각 fine-tunning하여 실험
- 다른 Self-supervised learning 방법론들과 비교했을 때, Detection 성능을 제외하고 나머지 Classification과 Segmentation에서 가장 좋은 성능을 보임

Dataset and Task Generalization on PASCAL [37]								
		Class. (%mAP)			Det. (%mAP)		Seg. (%mIU)	
fine-tune layers	[Ref]	fc8	fc6-8	all	[Ref]	all	[Ref]	all
ImageNet [38]	-	76.8	78.9	79.9	[36]	56.8	[42]	48.0
Gaussian	[10]	–	–	53.3	[10]	43.4	[10]	19.8
Autoencoder	[16]	24.8	16.0	53.8	[10]	41.9	[10]	25.2
k-means [36]	[16]	32.0	39.2	56.6	[36]	45.6	[16]	32.6
Agrawal et al. [8]	[16]	31.2	31.0	54.2	[36]	43.9	–	–
Wang & Gupta [15]	–	28.1	52.2	58.7	[36]	47.4	–	–
*Doersch et al. [14]	[16]	44.7	55.1	**65.3**	[36]	**51.1**	–	–
*Pathak et al. [10]	[10]	–	–	56.5	[10]	44.5	[10]	29.7
*Donahue et al. [16]	–	38.2	50.2	58.6	[16]	46.2	[16]	34.9
Ours (gray)	–	**52.4**	**61.5**	**65.9**	–	46.1	–	35.0
Ours (color)	–	**52.4**	**61.5**	**65.6**	–	46.9	–	**35.6**

Table 2. PASCAL Tests

Conclusion

❖ Conclusion

- 목적 함수를 Colorization task에 알맞게 설정하여 실제 사진과 더 가깝게 생성함
- 본 논문의 제안 방법론은 Colorization을 통해 실용적인 Graphics output을 만들면서 동시에 Self-supervised representation learning에서 pretext task로 활용할 수 있음
- Pretext task로써 Pascal VOC 2007, 2012에서 State-of-the-art 달성


Thank You