Colorful Image Colorization

ECCV' 16

Jung Soo Lee
19.10.08

Colorful Image Colorization ECCV' 16

- Goal : grayscale photograph
  -> color version of photograph
- Color: Lab color space
- Grayscale photograph: photograph only with light (no a and b)
  - Not real black and white images in training

Colorful Image Colorization ECCV' 16

- Goal : Plausible colorization that could potentially fool human observers
  - Color prediction: multimodal
    - Ex) Apple could be red, green, yellow. But not blue or orange
    - To make colors that people think it is plausible

- Contribution:
  - 1) automatic image colorization
    - 1. Objective function for multimodal uncertainty
    - 2. novel framework for testing colorization algorithms (New evaluation method)
    - 3. High-water mark on the task (SOTA)
  - 2) self-supervised representation

Colorful Image Colorization ECCV’ 16

Prior work & Concurrent work

- Non-parametric
  - Reference image -> Image Analogies framework
    - Applying diverse filters such as texture synthesis, super-resolution, texture-transfer, etc.
- Parametric
  - Prediction function
  - Concurrent work / classification: this work, Larsson, Iizuka

Colorful Image Colorization ECCV' 16

Concurrent work

- loss functions
  - This work: loss – rebalanced rare classes
  - Larsson: loss – un-rebalanced
  - Iizuka: regression loss

- Architecture
  - This work: single-stream / Image-net
  - Larsson: hyper columns / Image-net
  - Iizuka: two-stream / Places

Colorful Image Colorization ECCV' 16

Architecture

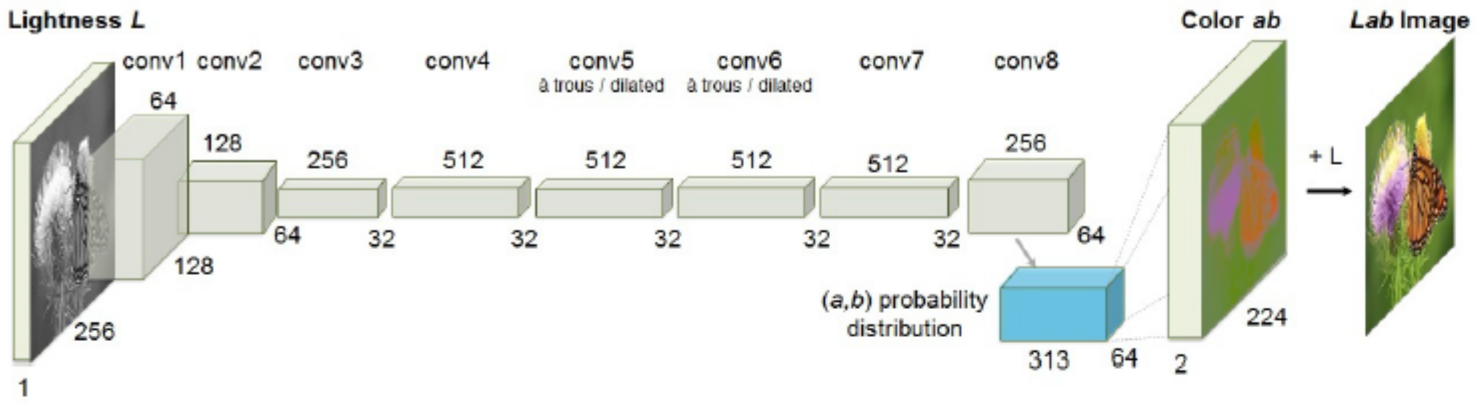


Fig. 2. Our network architecture. Each `conv` layer refers to a block of 2 or 3 repeated `conv` and `ReLU` layers, followed by a `BatchNorm` [30] layer. The net has no `pool` layers. All changes in resolution are achieved through spatial downsampling or upsampling between `conv` blocks.

VGG Network (modified)
- input: L
- conv – relu – batchnorm
- output: Lab
- X pool layers

Colorful Image Colorization ECCV' 16

Method & Previous Objective Function

Input : image with only light $\mathbf{X} \in \mathbb{R}^{H \times W \times 1}$
- not black/white images
- colored images -> only light channel

Learn mapping : $\widehat{\mathbf{Y}} = \mathcal{F}(\mathbf{X})$

Ground Truth : $\mathbf{Y} \in \mathbb{R}^{H \times W \times 2}$

Previous: minimize Euclidean distance
- not robust to inherent ambiguity and multimodal nature of colorization problem
- optimal solution:
  convex: mean -> grayish
  nonconvex -> implausible result

$$L_2(\widehat{\mathbf{Y}}, \mathbf{Y}) = \frac{1}{2} \sum_{h,w} \|\mathbf{Y}_{h,w} - \widehat{\mathbf{Y}}_{h,w}\|_2^2$$

Colorful Image Colorization ECCV' 16

Objective Function of this work

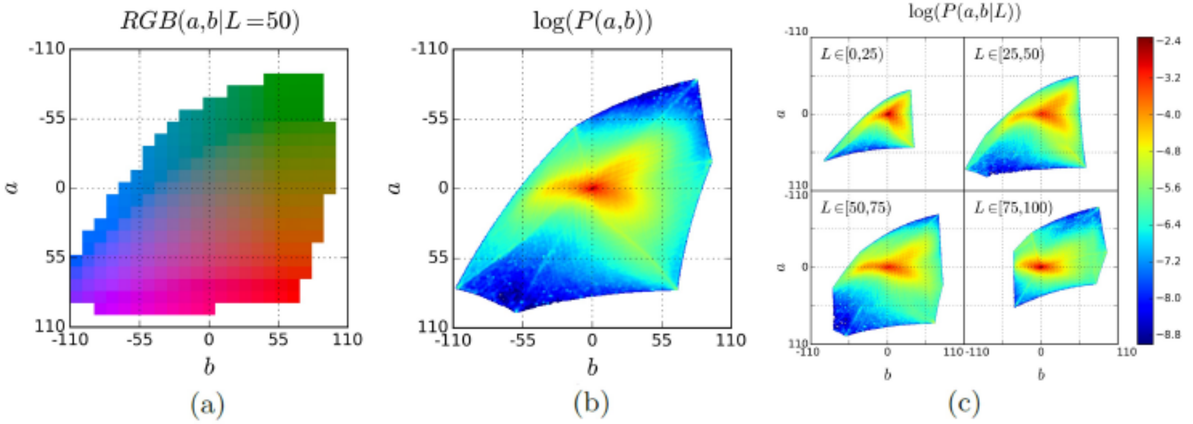


Fig. 3. (a) Quantized *ab* color space with a grid size of 10. A total of 313 *ab* pairs are in gamut. (b) Empirical probability distribution of *ab* values, shown in log scale. (c) Empirical probability distribution of *ab* values, conditioned on *L*, shown in log scale.

(a): divide ab color space into 10 grid sizes & select 313 values of pairs (you can see the grids if you look carefully)
L
(b): probability distribution of log scale
(c): probability distribution of log scale based on light

L: 0-black / 100-white
a: negative-red / positive-green
b: negative-blue / positive-yellow

Colorful Image Colorization ECCV' 16

Objective Function of this work

Input : $\mathbf{X} \in \mathbb{R}^{H \times W \times 1}$

Mapping : $\widehat{\mathbf{Z}} = \mathcal{G}(\mathbf{X})$, $\widehat{\mathbf{Z}} \in [0,1]^{H x W x Q}$, Q = number of quantized ab values (313)

Compare prediction with Ground Truth, $\mathbf{Z} = \mathcal{H}_{gt}^{-1}(\mathbf{Y})$
- find 5-nearest neighbors to T and weight them proportionally to their distance

$$\mathrm{L}_{cl}(\widehat{\mathbf{Z}}, \mathbf{Z}) = -\sum_{h,w} v(\mathbf{Z}_{h,w}) \sum_{q} \mathbf{Z}_{h,w,q} \log(\widehat{\mathbf{Z}}_{h,w,q})$$

This work: loss tailored to the colorization problem
- multi-modal cross-entropy based on re-balanced
- v : weighting term for rebalance

Colorful Image Colorization ECCV' 16

Class rebalancing

Fig. 3.b) Desaturated color dominant : need to take care of this issue
Solution: Class rebalance for rare classes -> similar with resampling training space

$$v(\mathbf{Z}_{h,w}) = \mathbf{w}_{q^*}, \text{ where } q^* = \arg\max_q \mathbf{Z}_{h,w,q}$$

$$\mathbf{w} \propto \left((1-\lambda)\widetilde{\mathbf{p}} + \frac{\lambda}{Q}\right)^{-1}, \quad \mathbb{E}[\mathbf{w}] = \sum_q \widetilde{\mathbf{p}}_q \mathbf{w}_q = 1$$

p: empirical probability of colors from ImageNet training set G_σ $\mathbf{p} \in \Delta^Q$
$\tilde{p}$: smoothed empirical distribution from p with Gaussian $\widetilde{\mathbf{p}} \in \Delta^Q$
Formula: mix $\tilde{p}$ distribution with a uniform distribution with weight $\lambda \in [0,1]$
q: quantize value (out of 313)
$\lambda = \frac{1}{2}, \sigma = 5$

Colorful Image Colorization ECCV' 16

Class Probabilities to Point Estimates

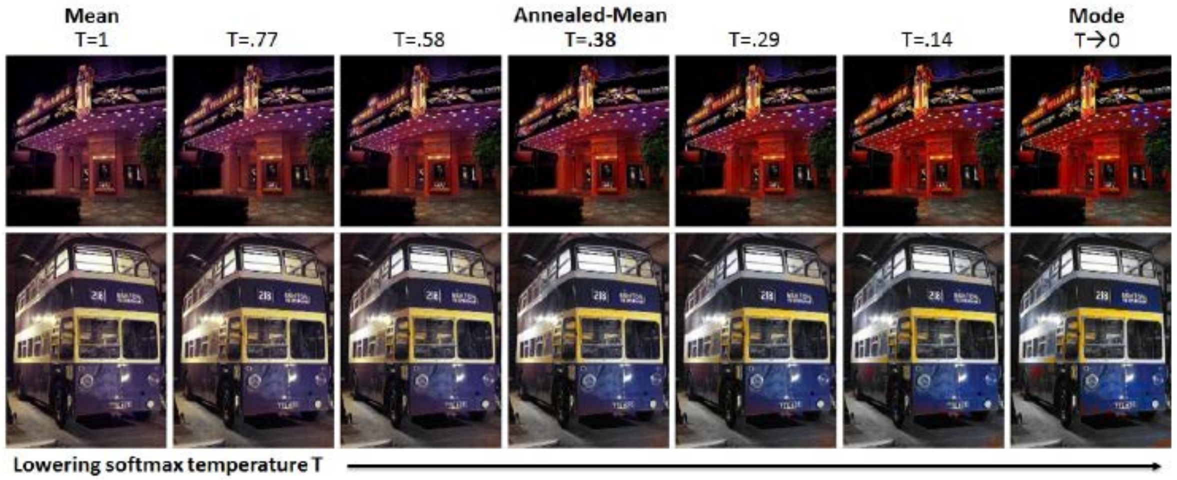


Fig. 4. The effect of temperature parameter T on the *annealed-mean* output (Equation 5). The left-most images show the means of the predicted color distributions and the right-most show the modes. We use $T = 0.38$ in our system.

Annealed Mean

$$\mathcal{H}(\mathbf{Z}_{h,w}) = \mathbb{E}\big[f_T(\mathbf{Z}_{h,w})\big], \quad f_T(\mathbf{z}) = \frac{\exp(\log(\mathbf{z})/T)}{\sum_q \exp(\log(\mathbf{z}_q)/T)}$$

T smaller -> peaked distribution (ex) T=0: one-hot encoding
T = 0.38 best

- Function H: (predicted distribution $\hat{Z} \rightarrow$ point estimate $\hat{Y}$)
- Consideration:
  - 1) prediction right away
    - spatially inconsistent (right-most 2 columns)
  - 2) mean prediction
    - Spatially consistent but desaturated
      - Left most
  - Solution: readjust softmax with temperature T

Output: $\widehat{\mathbf{Y}} = \mathcal{H}(\widehat{\mathbf{Z}})$

Colorful Image Colorization ECCV' 16

Evaluation – 1) Colorization quality (Quantitative Evaluation)

Colorization Results on ImageNet							
Method	Model Params (MB)	Model 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**

L2: L2 regression loss
L2,ft : fine-tuned from rebalancing network
class : no rebalance
full: method of this work

1. Perceptual Realism (AMT)
   - Given real & fake colors, choose fake colors
2. Semantic Interpretability (VGG classification)
   - Feed fake colorizes images
3. Raw Accuracy (AuC)
   - Plausibility

Colorful Image Colorization ECCV' 16

Evaluation – 1) Colorization quality (Qualitative Evaluation)

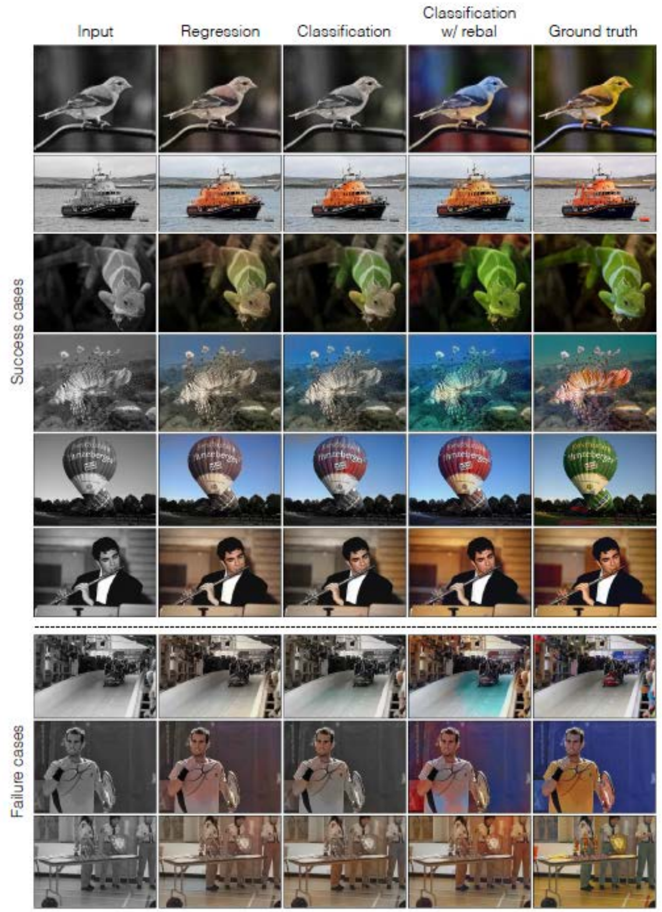
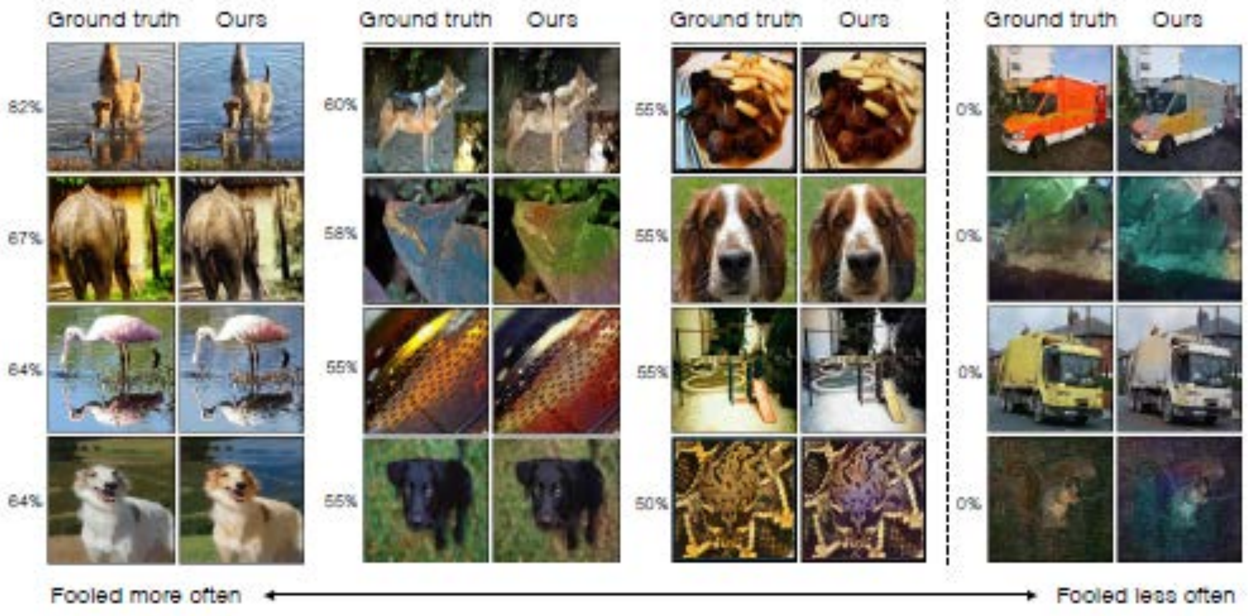


Fig. 5. Example results from our ImageNet test set. Our classification loss with rebalancing produces more accurate and vibrant results than a regression loss or a classification loss without rebalancing. Successful colorizations are above the dotted line. Common failures are below. These include failure to capture long-range consistency, frequent confusions between red and blue, and a default sepia tone on complex indoor scenes. Please visit http://richzhang.github.io/colorization/ to see the full range of results.

The examples of AMT



Colorful Image Colorization ECCV’ 16

Evaluation – 2) Self-supervised representation learning

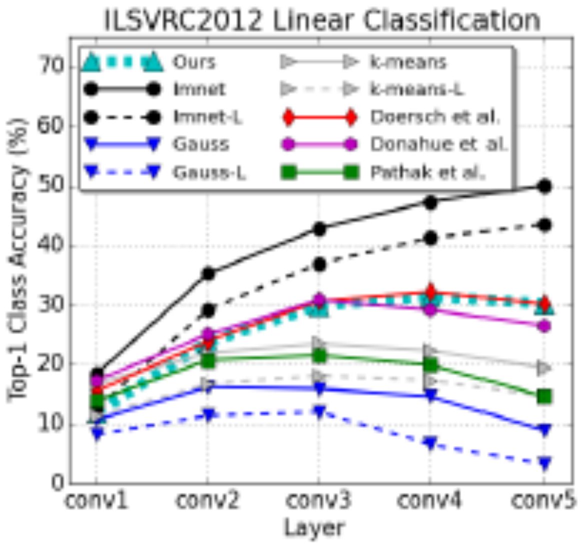


Fig. 7. ImageNet Linear Classification

Dataset and Task Generalization on PASCAL [37]

	Classification (% mAP)			Detection (% mAP)	Segmentation (% mIU)
fine-tuned layers	**fc8**	**fc6-fc8**	**all**	**all**	**all**
ImageNet [38]	76.8	78.9	79.9	56.8	48.0
Gaussian	–	–	53.3	43.4	19.8
Autoencoder	24.8	16.0	53.8	41.9	25.2
k-means [38]	32.0	39.2	56.6	45.6	32.6
Agrawal et al. [8]	31.2	31.0	54.2	43.9	–
Wang & Gupta [15]	28.1	52.2	58.7	44.0	–
*Doersch et al. [14]	44.7	55.1	**65.3**	**51.1**	–
*Pathak et al. [10]	–	–	56.5	44.5	29.7
*Donahue et al. [16]	38.2	50.2	58.6	45.1	34.9
Ours (gray)	**52.4**	**61.5**	**65.9**	46.9	35.0
Ours (color)	**52.4**	**61.5**	**65.6**	47.9	**35.6**

Table 2. PASCAL Tests

- Raw data: own source of supervision
- **Cross-channel encoder**
  Similar with autoencoder
  (except input & output different)

1. Task generalization
   ImageNet classification
   (linear classification of each layer)
2. PASCAL Tests
   1. Classification,
   2. Detection (Object)
   3. Semantic Segmentation

Colorful Image Colorization ECCV' 16

Application on Real Images

Fig. 8. Applying our method to legacy black and white photos. Left to right: photo by David Fleay of a Thylacine, now extinct, 1936; photo by Ansel Adams of Yosemite; amateur family photo from 1956; *Migrant Mother* by Dorothea Lange, 1936.

- Method of this work
    - Light channel of colorful image -> predict colors
- This application
    - Real black/white photos -> colorful images