Smart Media
한국스마트미디어학회

Colorization of Natural Scene Image using U-net

Tram-Tran Nguyen-Quynh, Nhu-Tai Do, Soo-Hyung Kim

School of Electronics and Computer Engineering,

Chonnam National University, Korea

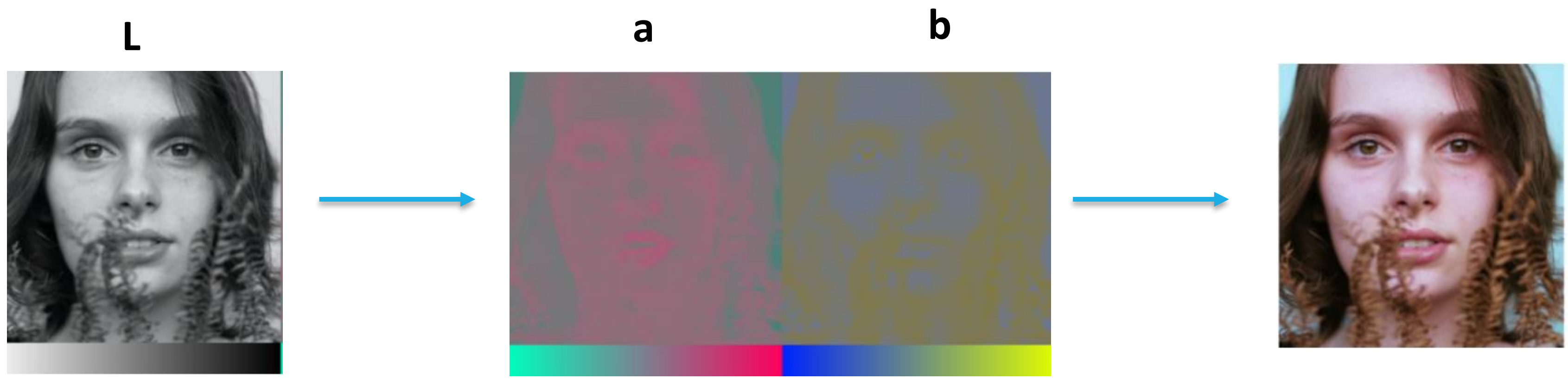
tramtran2@gmail.com, donhutai@gmail.com, shkim@chonnam.ac.kr

May 23rd, 2020

Agenda

1. INTRODUCTION

- **Our problem: Fully Automatic Colorization**
  - Given the **grayscale image**, produce ***a plausible colorization to fool a human observer***.
  - **Input**: Grayscale or L channel of image, output ab channel of image



1. INTRODUCTION

- **Challenges of Fully Automatic Colorization**:
  - **Averaging effect**: ***grayish***, ***desaturated*** results due to 94% of the cells in our eyes determine brightness, only 6% for colors. Grayscale image is a lot sharper than the color layers.

  - **Rare colors in images**: ***strongly biased due to the appearance of backgrounds*** such as clouds, pavement, dirt, and walls.

  - **Semantic information matters**: a system must ***interpret the semantic composition*** of the scene (what is in the image: tree, sky, ocean, . . .) as well as ***localize objects*** (where things are).

GT: lagoon
top-1: balcony interior (0.136)
top-2: beach house (0.134)
top-3: boardwalk (0.123)
top-4: roof garden (0.103)
top-5: restaurant patio (0.068)

1. INTRODUCTION

- **Our objectives**:
  - The model of human cognition proceed information about color and meaning of an image depending on their **previous experiences**
  - Image Colorization integrated ***exploiting the scene-context and the uncertainty in the scene classification***.

Scene Type **Probability** **Scene Label**

Smart Media
한국스마트미디어학회

1. INTRODUCTION

- **Our objectives**:
  - Multi-Task Learning based on U-Net:

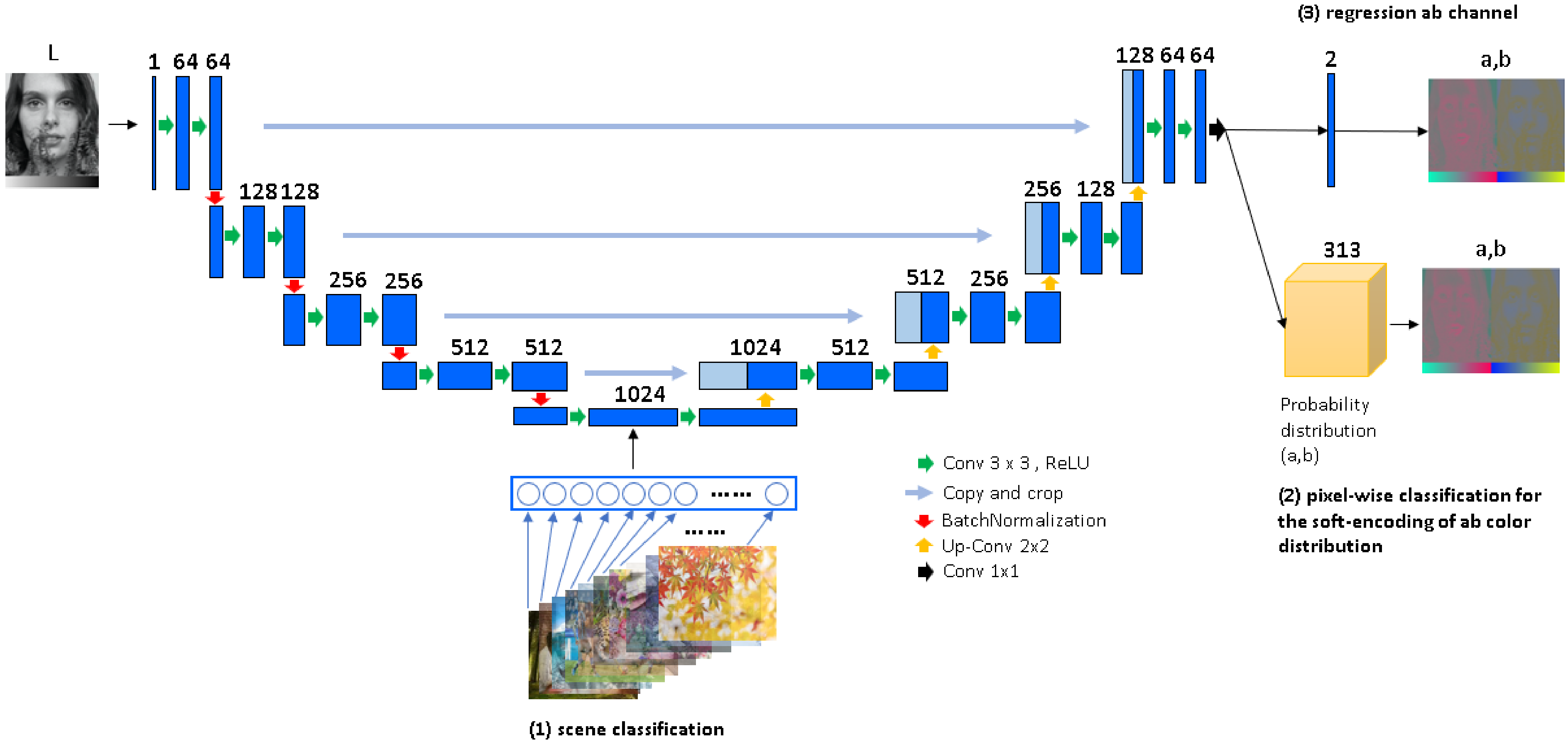
    (1) scene classification ***to exploit the global information***

    (2) pixel-wise classification for the soft-encoding of ab color probability vector ***to encourage the rare color and rebalance color***

    (3) ab channel regression ***to keep the accuracy from content***
  - Make experiments on ***Coco-Stuff for training***, ***DIV2K for testing*** and compare with the state-of-the-art methods.

2. PROPOSED METHOD

- ***Unet network:*** take advantage of skip connections between the contracting and expanding path at the same depth level.



Scene-context Image Colorization U-Net

2. PROPOSED METHOD

- **For uncertainty scene classification**: make scene ground-truth for training dataset
  - Extract the scene probabilities based on the pre-trained model of Places365[1]

Image

Pre-trained Model

0	1	2	3	...	n-1

Scene probability with n=365

  - **Label Smoothing[2] with top-5 prediction**: keep 5 highest probabilities, set all remain values to 0, and normalize the probabilities with sum 1.

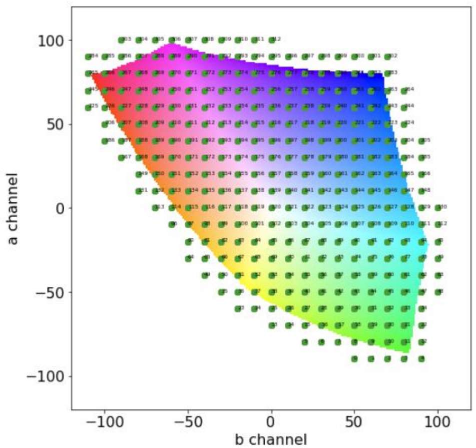
Top-1: Cafeteria (0.179)
Top-2: Restaurant (0.167)
Top-3: dining_hall (0.091)
Top-4: coffe_shop (0.086)
Top-5: restaurant_patio (0.080)

Top-1: Cafeteria (0.297)
Top-2: Restaurant (0.277)
Top-3: dining_hall (0.151)
Top-4: coffe_shop (0.143)
Top-5: restaurant_patio (0.132)

[1] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, “Places: A 10 Million Image Database for Scene Recognition,” IEEE transactions on pattern analysis and machine intelligence (TPAMI), vol. 40, no. 6, pp. 1452–1464, 2018
[2] R. Müller, S. Kornblith, and G. Hinton, “***When Does Label Smoothing Help?***,” In Advances in Neural Information Processing Systems (NeurIPS), pp.4696-4705, 2019.

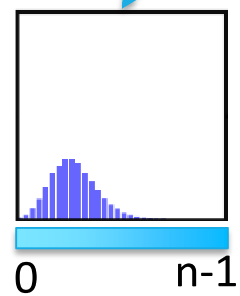
2. PROPOSED METHOD

- **For pixel-wise classification for the soft-encoding**:



Quantized ab color space with grid n=313

ab color probability vector per pixel



0	1	2	3	...	n-1

Vector representation of ab color space with n=313

2. PROPOSED METHOD

- **Multi-Task Losses:**
  - **Scene-context classification:** Category Cross-Entropy (CCE) loss:

$$CCE(y,\hat{y}) = -\sum_{i=1}^{C} y_i \log \hat{y}_i$$

Where C is the number of scene, y_i/ $\hat{y}_i$ is the ground-truth/predicted scene probability.

  - **Pixel Classification of ab color distribution**: Weighted Category Cross-Entropy Loss:

$$CCE(y,\hat{y}) = -\sum_{h,w} v(y_{h,w}) \sum_{i=0}^{N-1} y_{h,w,i} \log \hat{y}_{h,w,i}$$

Where h, w is the height and width of image, N is the number of quantized colors of ab color distribution, $\boldsymbol{v(y_{h,w})}$ **is the weighted of color-class at pixel (h,w) to encourage the rare-color**, $y_{h,w,i}$/ $\hat{y}_{h,w,i}$ is the ground-truth/prediction probability of the soft-encoding color i at pixel (h,w).

  - **Regression ab channel**: Using Mean Square Error (MSE) Loss:

$$MSE(y,\hat{y}) = \frac{1}{2hw} \sum_{h,w} \left\| y_{h,w,ab} - \hat{y}_{h,w,ab} \right\|_2^2$$

Where $y_{h,w,ab}$/ $\hat{y}_{h,w,ab}$ is the ground-truth/prediction of ab values at pixel (h,w)

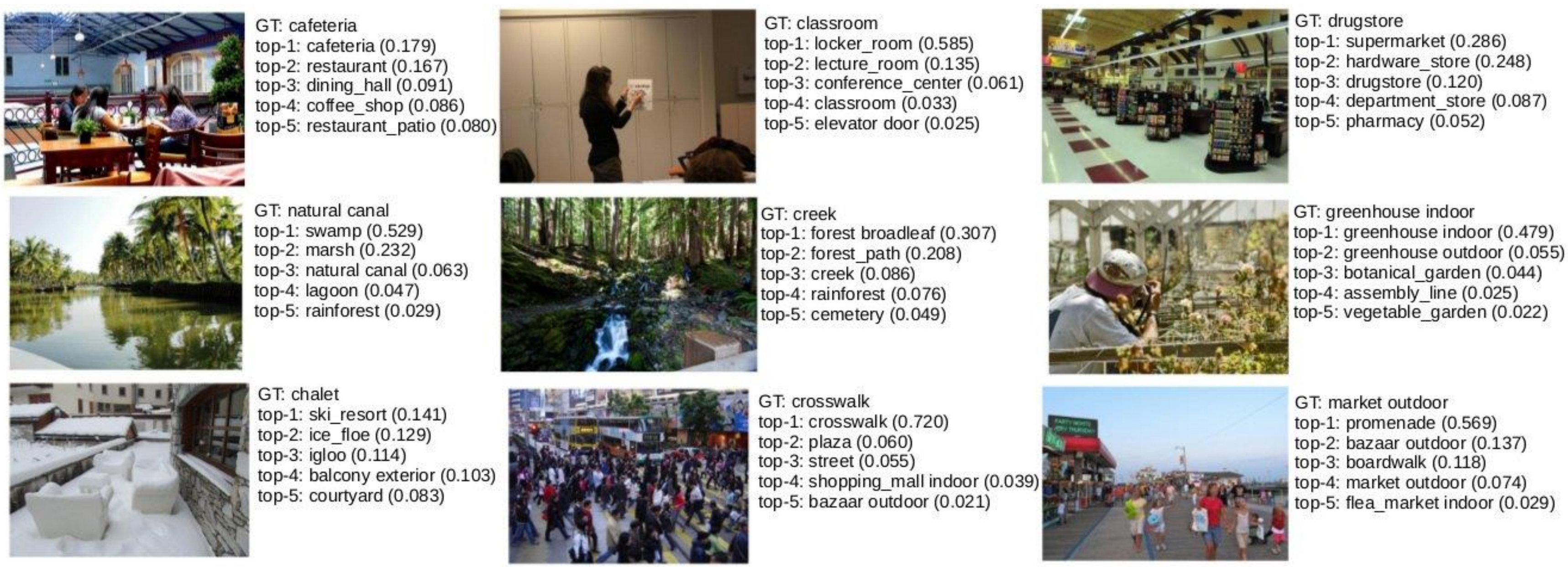
3. EXPERIMENTS AND DISCUSSION

- **Coco-Stuff Dataset (for training and validating)**
  - A large-scale object detection, segmentation, and captioning dataset
  - It involves **118.000 images for training** and **5.000 images for validation** set in which includes 172 classes containing 80 thing classes, 91 stuff classes and 1 class unlabeled.

http://cocodataset.org/

3. EXPERIMENTS AND DISCUSSION

- **Pre-trained Model on Places365 (for extracting scene-context probability)**
  – Places365-Standard is the latest subset of Places dataset with about 1.8 million images for training 365 different categories of scene/location, 5000 images per category

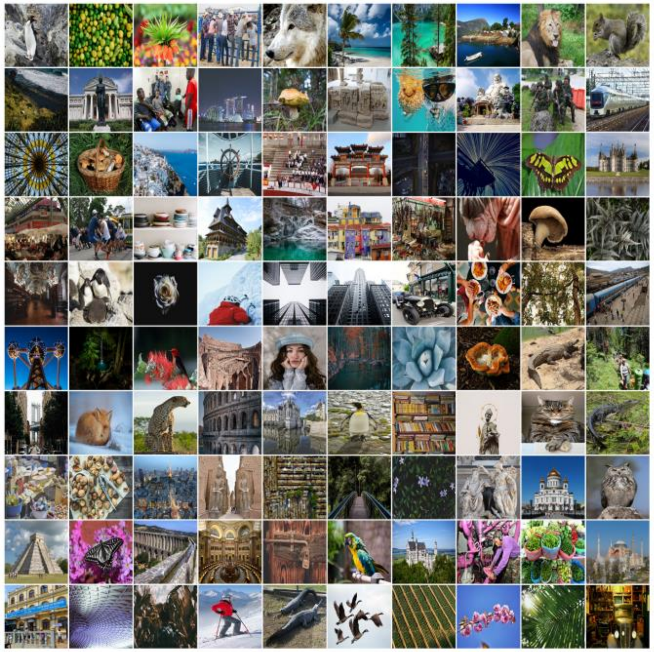


http://places2.csail.mit.edu/

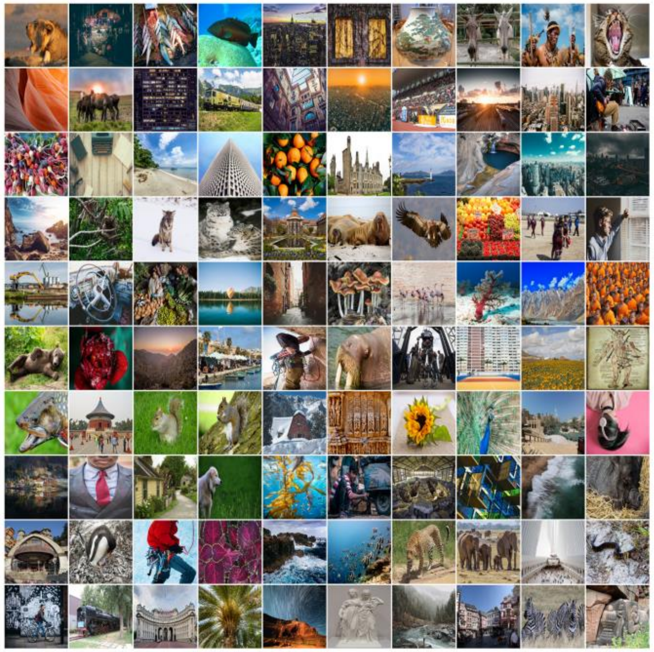
Smart Media 한국스마트미디어학회

3. EXPERIMENTS AND DISCUSSION

- **DIV2K Dataset (for testing)**
  - NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study
  - Using in Colorful Image Colorization Challenge NTIRE 2019



DIV2K 800 train images



DIV2K 100 test images

https://data.vision.ee.ethz.ch/cvl/DIV2K/

3. EXPERIMENTS AND DISCUSSION

- **Comparison methods**:

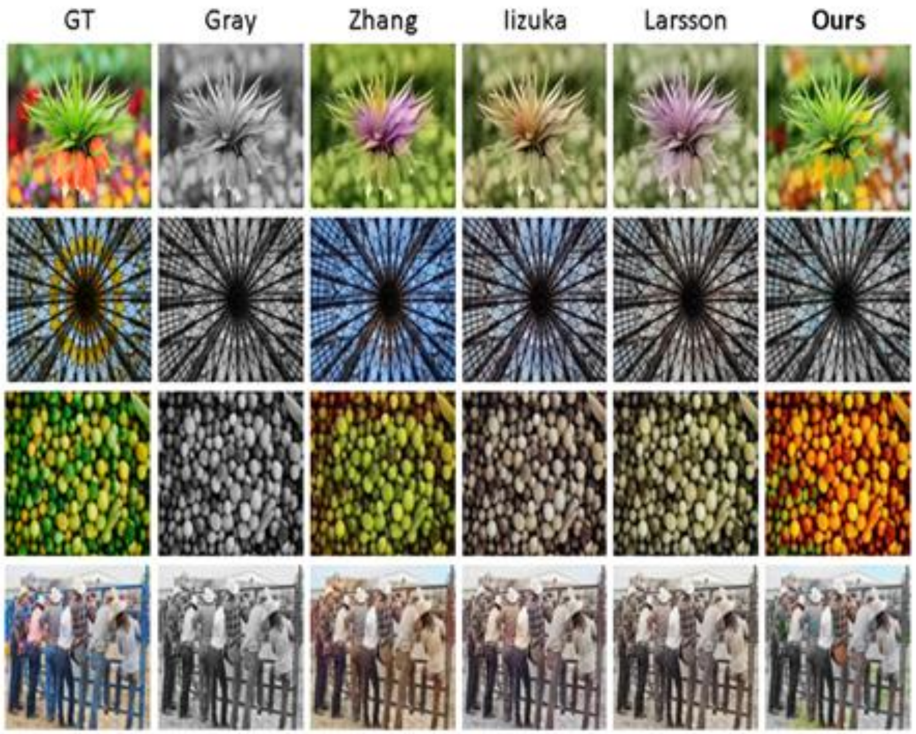
Method	Name	Train data	Test data
1	**Our method**	Coco-Stuff	DIV2K
2	**Iizuka et al.[1]**	Places	DIV2K
3	**Larsson et al.[2]**	ImageNet	DIV2K
4	**Zhang et al.[3]**	ImageNet	DIV2K

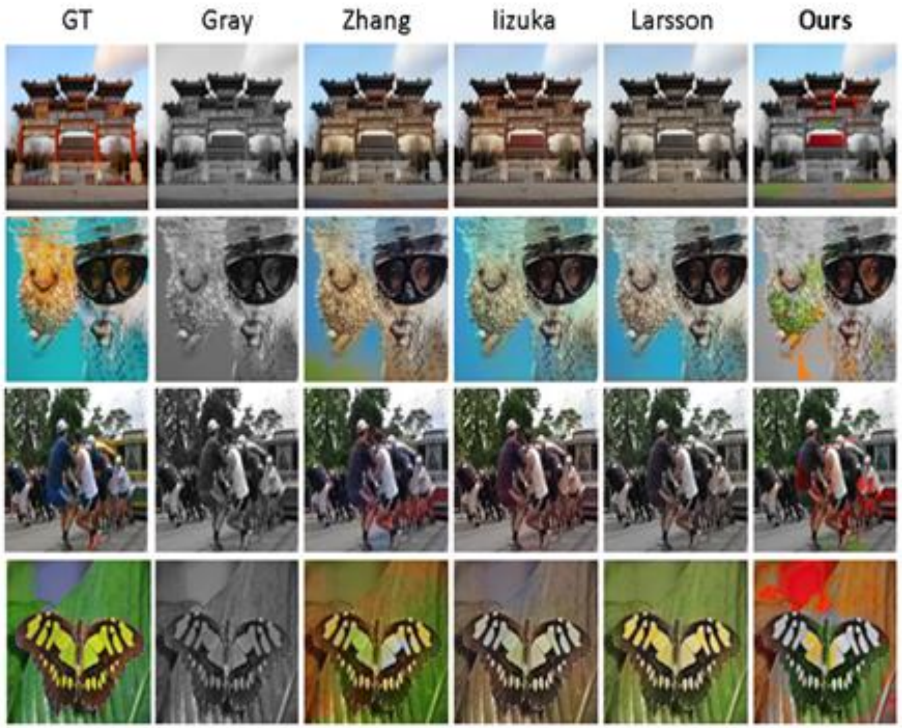
[1] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "***Let there be Color: Joint End-to-end Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification***," *ACM Transactions on Graphics*, vol. 35, no. 4, pp. 1–11, Jul. 2016.
[2] G. Larsson, M. Maire, and G. Shakhnarovich, "***Learning Representations for Automatic Colorization***," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9908 LNCS, 2016, pp. 577–593.
[3] R. Zhang, P. Isola, and A. A. Efros,"**Colorful Image Colorization**," ECCV, pp. 649–666, 2016

Smart Media
한국스마트미디어학회

3. EXPERIMENTS AND DISCUSSION

- **Quality Results**:





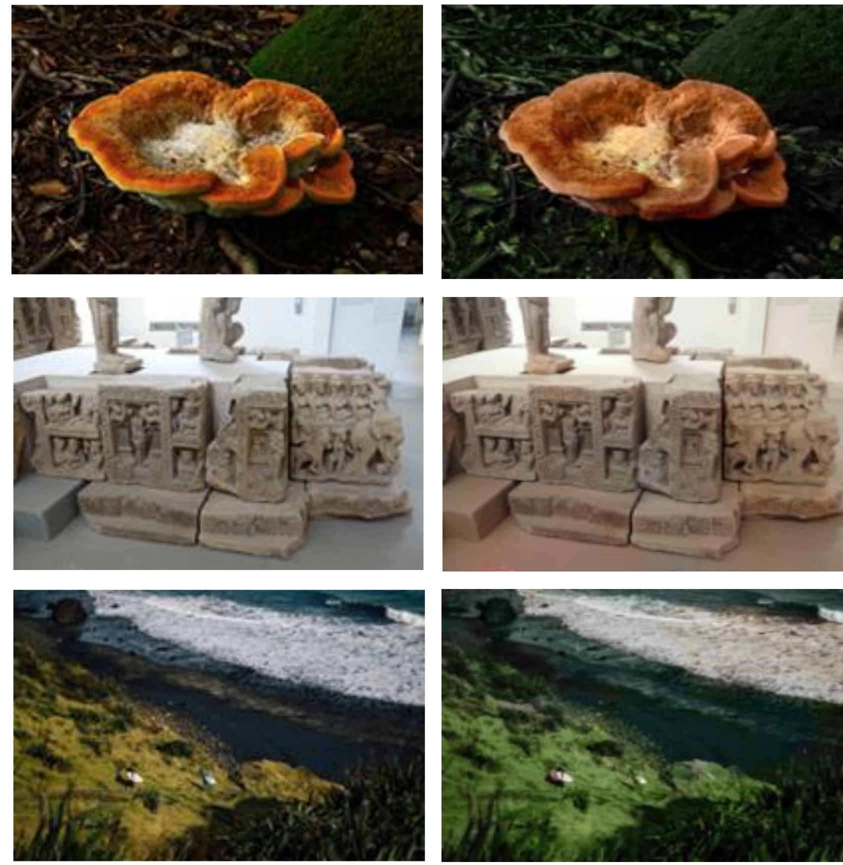
Best cases **Worst cases**

4. EXPERIMENTS AND DISCUSSION

- **Quality Results**:

GT Ours

GT Ours



Some results may fool a human observer

3. EXPERIMENTS AND DISCUSSION

- Quantitative results:
  - Peak Signal-to-Noise Ratio (**PSNR**)
  - The Structural Similarity index (**SSIM**)
  - Mean-Square Error (**MSE**)

In our experiments, we use ***PSNR*** and ***SSIM*** on Result Image, and ***MSE*** on ab channel.

Method	Name	PSNR	SSIM	MSE_{ab}
1	Our method	Soft:19.961 Reg: 22.263	Soft: 0.785 Reg: 0.867	Soft: **0.584** Reg: 0.605
2	Iizuka et al.	23.492	0.912	0.620
3	Larsson et al.	**23.809**	**0.914**	0.585
4	Zhang et al.	21.173	0.885	0.630

4. CONCLUSIONS

- In this paper, we ***exploit the uncertainty scene probability*** for image colorization problem by transfer learning from Places365 pre-trained model to Coco-Stuff dataset.

- We ***apply Multi-Task Learning*** with (1) uncertainty scene classification for global information (2) pixel-wise classification on ab color distribution (3) regression on ab channel.

- Our results ***remain some limitations*** in the quality of coloring images.

- To overcome it, we need:
  - **Building a tool for the perceptual rank (evaluating results by human)**
  - **Combining the segmentation approaches to improve quality.**

THANK YOU FOR LISTENING