MLEU: Multi-Level Embedding U-Net for Fully Automatic Image Colorization

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

Jan 18th, 2020

Agenda

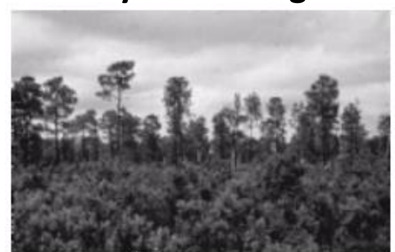
1. INTRODUCTION

- **Problem**: Image Colorization is the task of colorizing gray-scale images.
- **Practical applications**: coloring old black and white images, movies etc.
- **Main approaches:** Scribble-based, Example-based, and <u>**Fully Automatic**</u>.

User Stroke on Image

Reference image

Gray-scale image



Gray-scale image

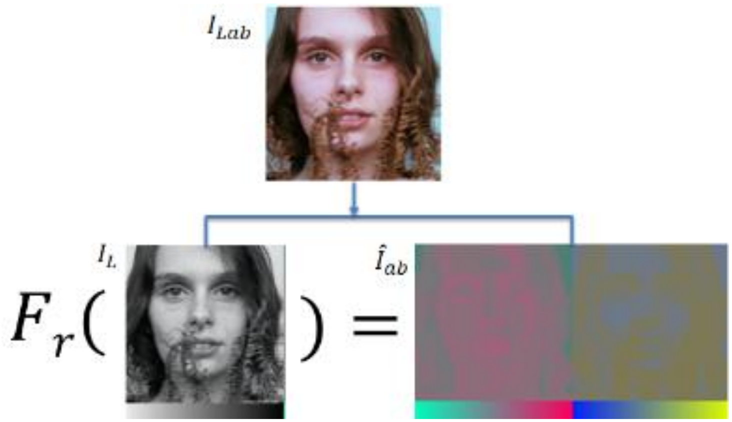
Scribble-based colorization

Example-based colorization

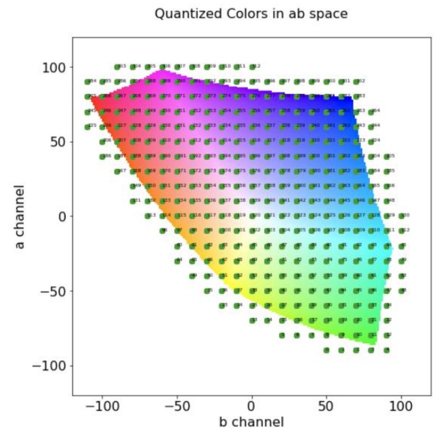
Fully Automatic colorization

1. INTRODUCTION

- **Our problem** focuses on **Fully Automatic Colorization**: Given the **grayscale image**, produce ***a plausible colorization to fool a human observer***.



Regression mapping function



Quantized colors in ab space

	Regression Approach	Classification Approach
Input	$I_L \in \mathcal{R}^{h \times w \times 1}$	$I_L \in \mathcal{R}^{h \times w \times 1}$
Output	$\hat{I}_{ab} \in \mathcal{R}^{h \times w \times 2}$ ab channels in CIE color space Lab	$\hat{Z}_{ab} \in \mathcal{R}^{h \times w \times n}$ one-hot encoding of the quantized colors in ab space

1. INTRODUCTION

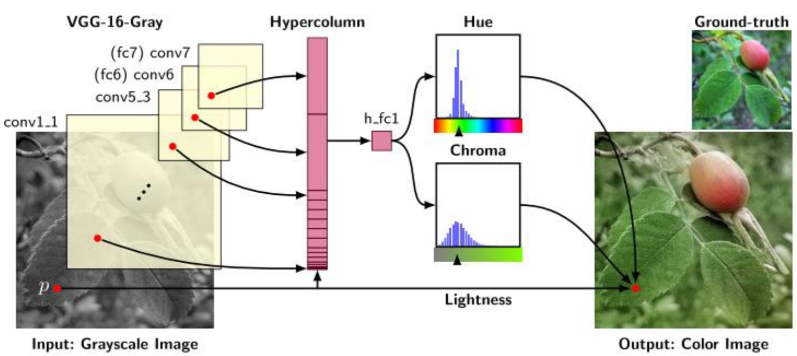
- **Challenges of Fully Automatic Colorization:**
  - ***Averaging effect:*** grayish, desaturated results due to 94% of the cells in our eyes determine brightness, only 6% for colors → grayscale image is a lot sharper than the color layers.
  - ***Rare colors in images***: strongly biased due to the appearance of backgrounds such as clouds, pavement, dirt, and walls.
  - ***Usage of a huge database to train***: to encode meaningful features that help to colorize the image.
- **Our objectives**:
  - ***Construct Multi-Level Embedding UNet*** (MLEU) for transfer learning from the pre-trained ImageNet weight, and gradient flow enhancing by skip connections.
  - ***Improve the distribution over quantized color*** by interpolation and smoothness for reducing the unbalance colors and focusing on rare colors.
  - ***Make experiments on DIV2K*** and compare with the state-of-the-art method from Zhang et. al. work [1].

[1] R. Zhang, P. Isola, and A. A. Efros,**Colorful Image Colorization**, ECCV, pp. 649–666, 2016

2. RELATED WORKS

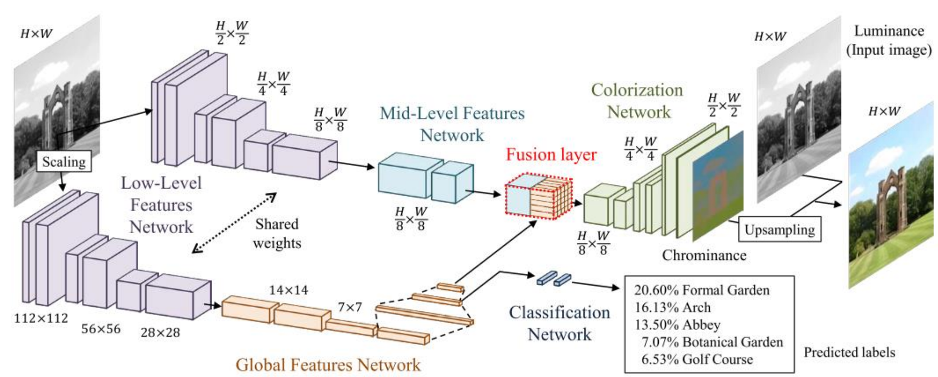
- **Deep Learning Approach**:
  - **Larsson et al.**[1]: use un-rebalanced classification **loss**, build on hyper-columns on a VGG **network**, train on **ImageNet**, evaluate on **PSNR, RMSE**.
  - **Iizuka et al.**[2]: use a regression **loss**, build a ***two-stream architecture*** fusing global and local features, train on ***Places scene dataset***, evaluate on ***naturalness*** of the colorizations by ***user asking***

Regression



Larsson et al.

Classification



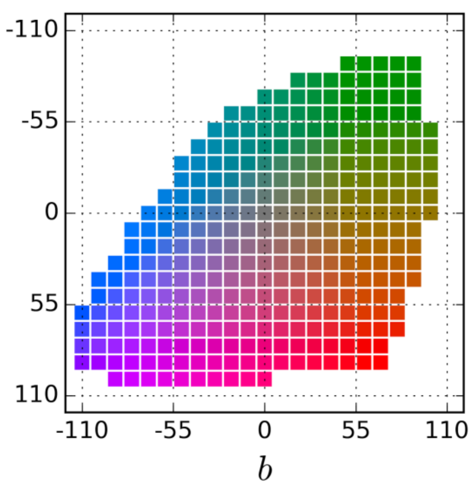
Iizuka et al.

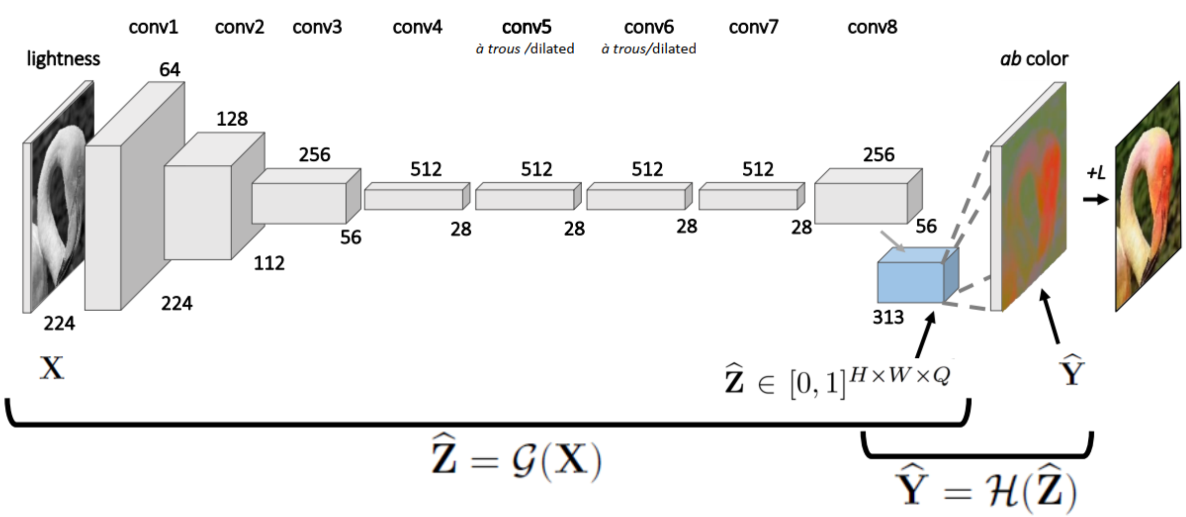
[1] G. Larsson, M. Maire, and G. Shakhnarovich, "***Learning Representations for Automatic Colorization***," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9908 LNCS, 2016, pp. 577–593.
[2] S. Iizuka, E. 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 Classificatio***," *ACM Transactions on Graphics*, vol. 35, no. 4, pp. 1–11, Jul. 2016.

2. RELATED WORKS

- **Zhang et. at[1]: Main idea**
  - ***Multinomial classification*** problem by ***quantize ab space*** into grid size 10, keep 313 bins in gamut.
  - Cross entropy loss with ***class rebalancing*** to encourage learning of rare colors.
  - **Post-processing**: per-pixel color distribution to single point estimate **by interpolating between mean and mode with annealed-mean.**

Colors in *ab* space
(discrete)



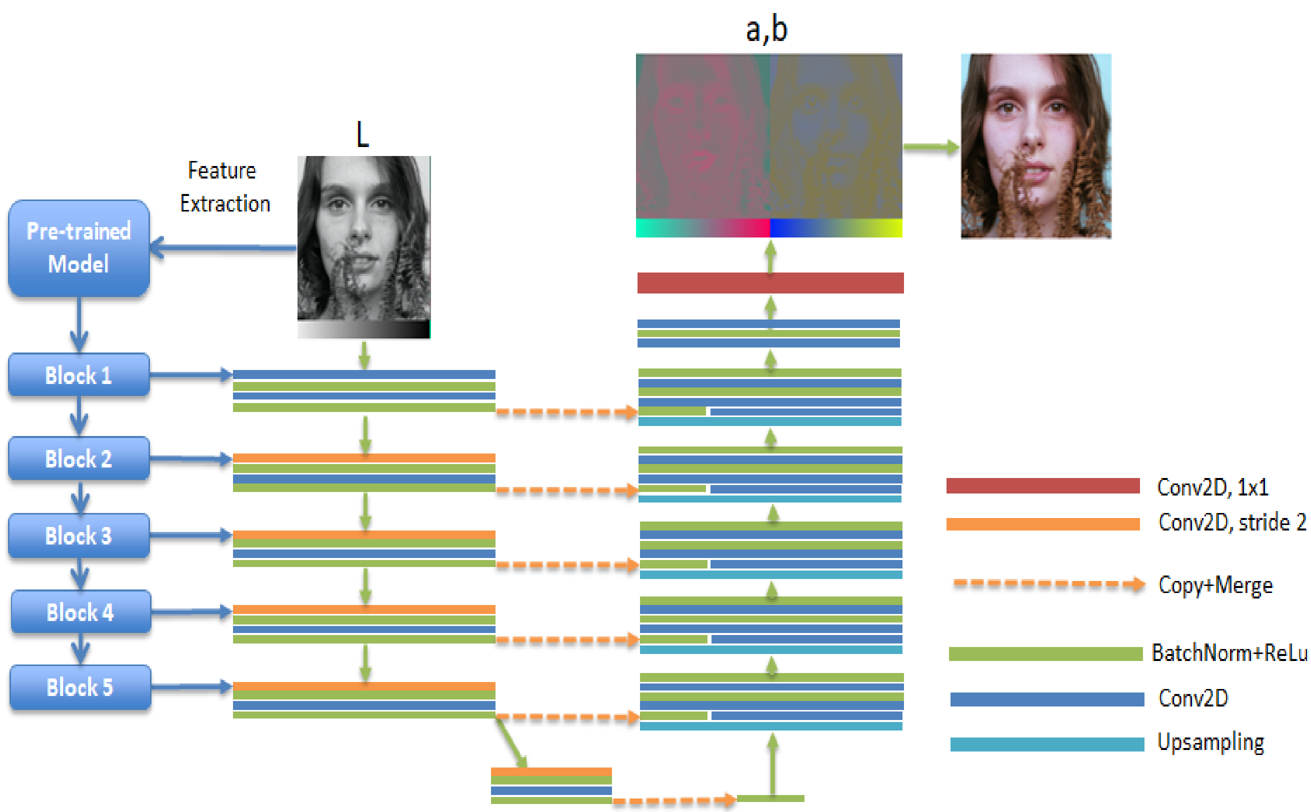


Deep network model

[1] R. Zhang, P. Isola, and A. A. Efros,**Colorful Image Colorization**, ECCV, pp. 649–666, 2016

3. PROPOSED METHOD

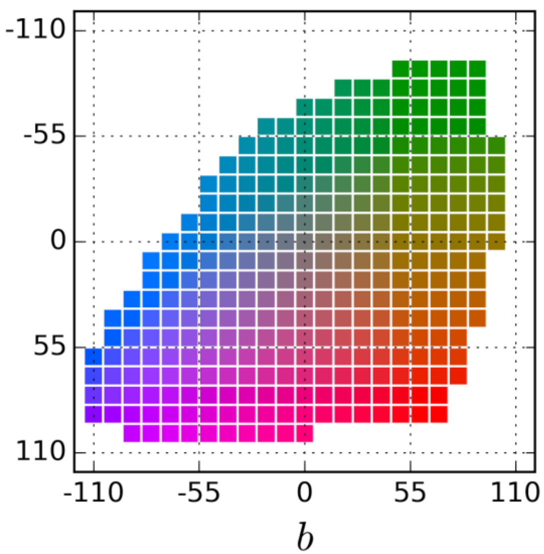
- **Network Architecture**:
  - ***Apply Unet network:*** take advantage of skip connections between the contracting and expanding path at the same depth level.
  - ***Multi-Level Embedding***: integrate ImageNet features at every encoding layers for transferring learning.



3. PROPOSED METHOD

- **Quantization process** in classification approach from Richard Zhang et. al.:
  - ***Quantization*** Lab Color Space ***into 313 bins***
  - Using ***soft-encoding scheme*** instead of nearest searching
- Benefits from this quantization process to classify:
  - Prevent the averaging effect of regression loss: easy to favor grayish, desaturated results
  - Increase the correlation between nearest color pixels by soft-encoding.

Colors in *ab* space
(discrete)



$$L(\hat{Z}, Z) = -\frac{1}{HW}\sum_{h,w} v(Z_{h,w}) \boxed{\sum_{q} Z_{h,w,q}} log(\hat{Z}_{h,w,q})$$

Rarity weighting   Target distribution   Predicted distribution

Category Cross entropy loss

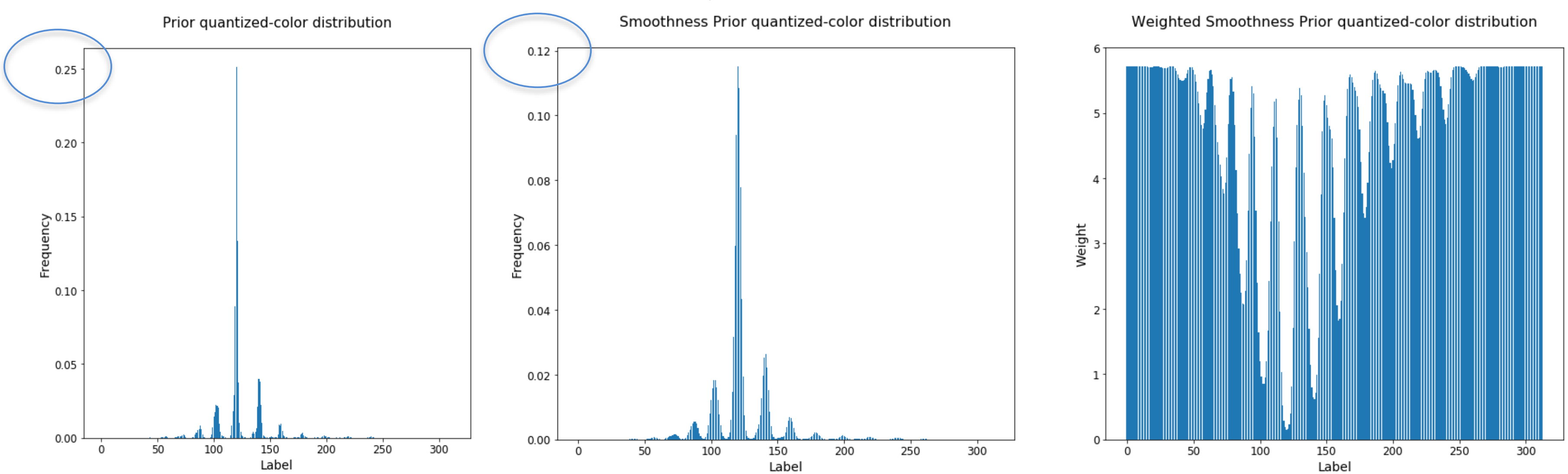
3. PROPOSED METHOD

- **Smoothness prior distribution**
  - Tackle the unbalance among the quantized colors.
  - Make the interpolation on the prior probability distribution **P**, apply the 1D-Gaussian kernel with σ =5 and normalize to unit-length to output $\mathbf{P_s}$:

$$\boldsymbol{P}_s = Interp(\boldsymbol{P}) * \mathbf{N}_\sigma, |\boldsymbol{P}_s| = 1$$

  - Use the equation of Zhang et. al to calculate the color-label weight based on smoothness prior distribution:

$$W \propto \left((1-\boldsymbol{\lambda})\boldsymbol{P}_s + \frac{\boldsymbol{\lambda}}{\boldsymbol{n}}\right)^{-\alpha}, \mathrm{E}[\boldsymbol{W}] = \sum_{i=1}^{n} \boldsymbol{P}_{si}\boldsymbol{W}_i = 1$$



3. PROPOSED METHOD

- **Loss function**:
  - Use the mean square error loss for the regress approach:

$$\boldsymbol{L_{MSE}}(I,\hat{I}) = \frac{1}{hw}\sum_{h,w}\left\|I-\hat{I}\right\|_2^2$$

  - Use the weighted category cross-entropy for the classification approach:

$$\boldsymbol{L_{WCCE}}(Z,\hat{Z}) = -\sum_{h,w} W_{h,w}\sum_{n} Z_{h,w,n}log\left(\hat{Z}_{h,w,n}\right)$$

Where

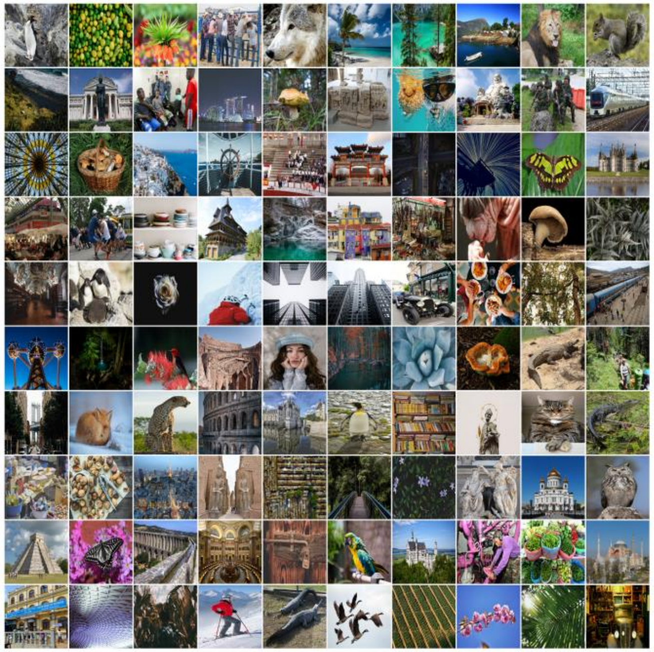
+ $W_{h,w}$ is the weighted smoothness prior distribution of quantized color

+ I ($\hat{I}$) is the ground-truth (predicted result) of the ab channel normalized in [-1,1]

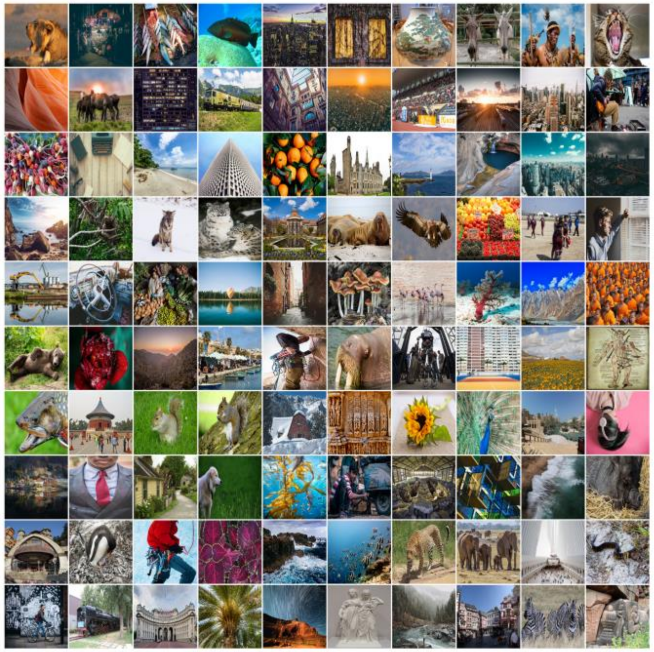
+ Z ($\hat{Z}$) is the ground-truth (predicted result) of the distribution over quantized color

4. EXPERIMENTS AND DISCUSSION

- **DIV2K Dataset**
  - 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/

4. EXPERIMENTS AND DISCUSSION

- Quantitative measures in ***Colorful Image Colorization Challenge NTIRE 2019***:
  - Peak Signal-to-Noise Ratio (**PSNR**)
  - The Structural Similarity index (**SSIM**)
  - **A nonreference manner**, three people who do not know the ground truth color image compare the colorization result

Team	Track 1 No guidance			Track 2 With guidance		
	PSNR	SSIM	Perceptual rank	PSNR	SSIM	Perceptual rank
IPCV_IITM	22.1232	0.9406	1	22.8605	0.9443	1
Athi	20.8710	0.9229	2	20.6470	0.7997	2
VIDAR	22.1949	0.9419	4	23.2707	0.9461	3
Team_India	17.9643	0.8472	3			
pksvision_mm	21.2248	0.9279	5			
ITU-GO	21.0773	0.8526	6			

Table 1. NTIRE 2019 Colorization Challenge results and final rankings. IPCV_IITM team is the winner of the challenge followed by Athi and VIDAR teams.

- As the result table above, there are the problem in evaluation using PSNR and SSIM (ITU_GO Team). So, the challenge adds the perceptual rank in benchmark.
- ***In our experiments, we use PSNR and perceptual rank by comparison the prediction images.***

4. EXPERIMENTS AND DISCUSSION

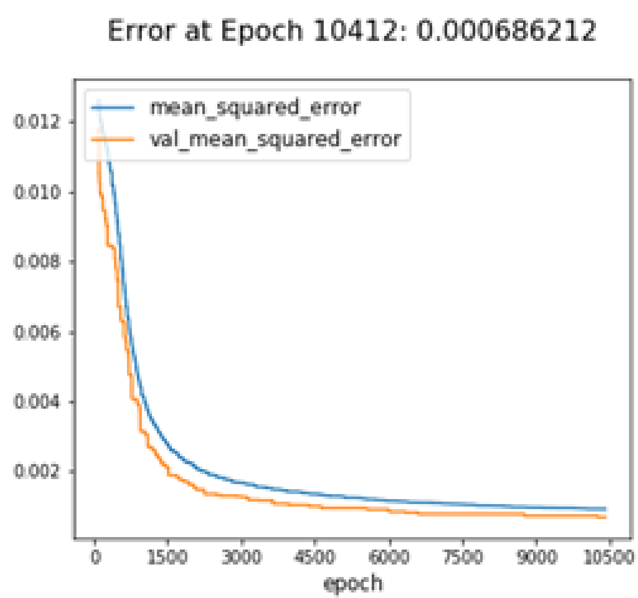
- **Comparison methods**:

Method	Name	Type	Train Data
1	MLEU Regression	Regression	DIV2K
2	MLEU Classification	Classification	DIV2K
3	Richard Zhang et. al. . [1]	Classification	ImageNet

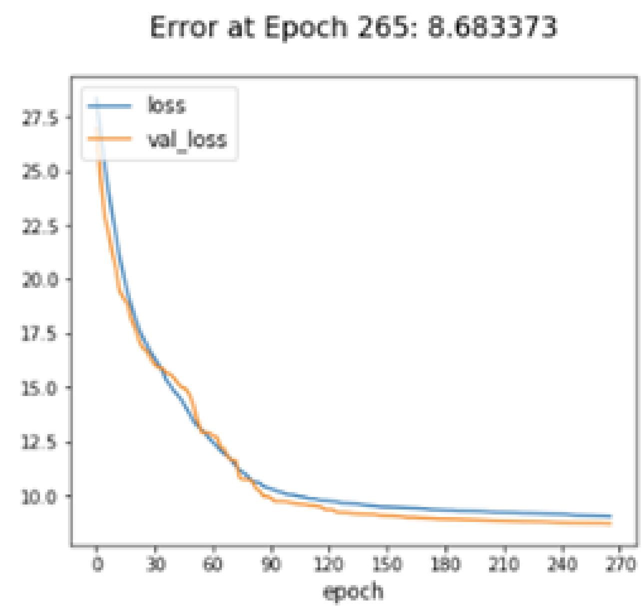
[1] R. Zhang, P. Isola, and A. A. Efros,**Colorful Image Colorization**, ECCV, pp. 649–666, 2016

4. EXPERIMENTS AND DISCUSSION

- **Training Process**:
  - Data augmentation: flipping, rotating and random cropping of the images
  - Optimization algorithm: train on Adam, SGD, and RMSProp with learning rate 0.004 and selected the best model.
  - Pre-trained model to combine features: pre-trained VGG-16 on ImageNet



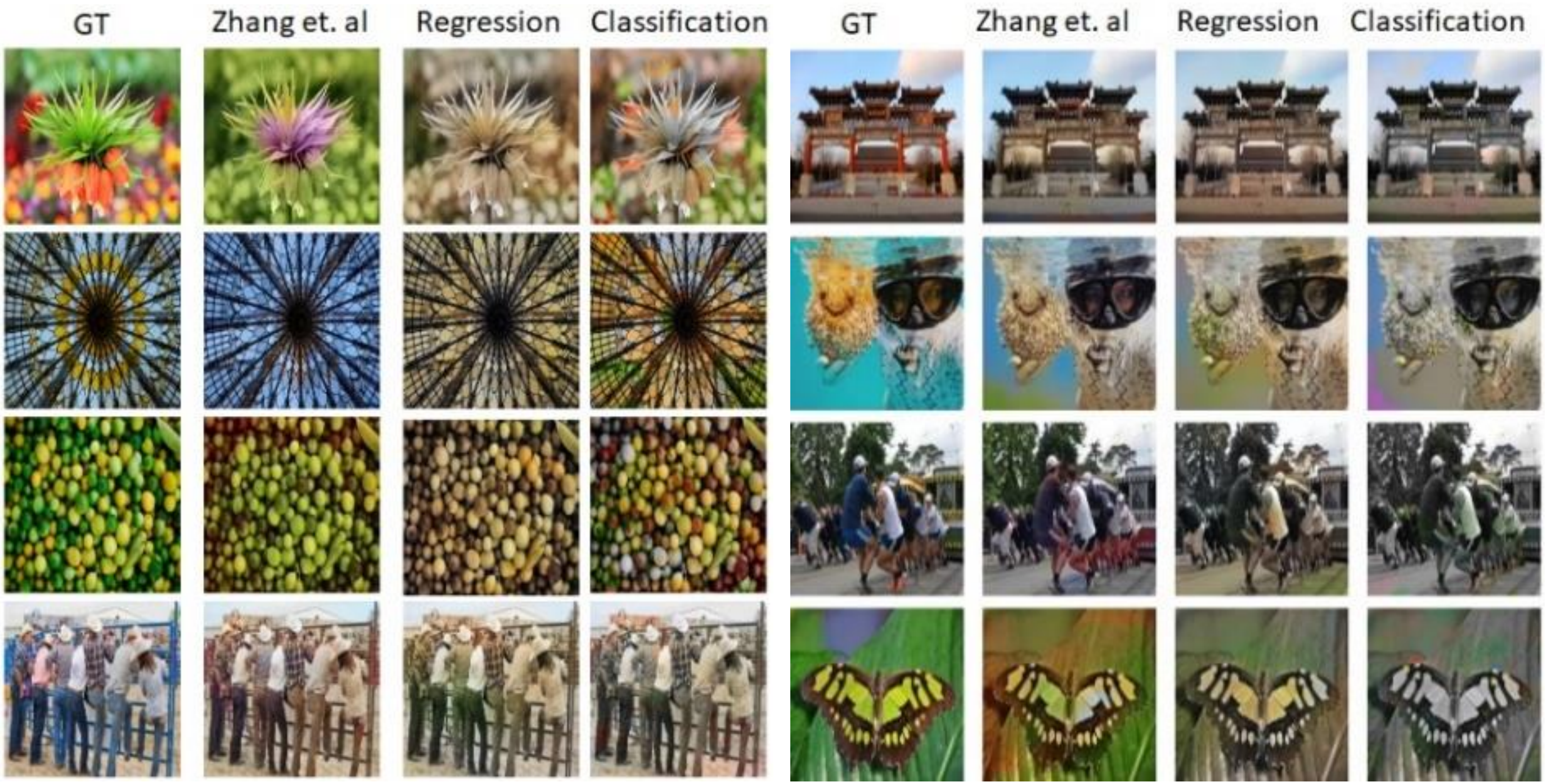
Regression



Classification

4. EXPERIMENTS AND DISCUSSION

- **Results**:



Best cases **Worst cases**

4. EXPERIMENTS AND DISCUSSION

- **Results**:

Method	Name	PSNR
1	MLEU Regression	20.9
2	MLEU Classification	**23.1**
3	Richard Zhang et. al. . [8]	22.8

About PSNR, we have the better results than Richard Zhang et. al.

5. CONCLUSIONS

- The colorization method was considered as a classification based on color quantization or regression method for comparison.
- We improved the original U-Net architecture to achieve the performance of transferring information well on a small dataset DIV2K.
- Our results remain some limitations in the quality of coloring images.
- To overcome it, we need:
  - cluster the color quantization space to avoid ambiguity in the quantizing process.
  - combine the regression and classification approaches to enhance quality.

THANK YOU FOR LISTENING