JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

L-CoIns: Language-based Colorization with Instance Awareness

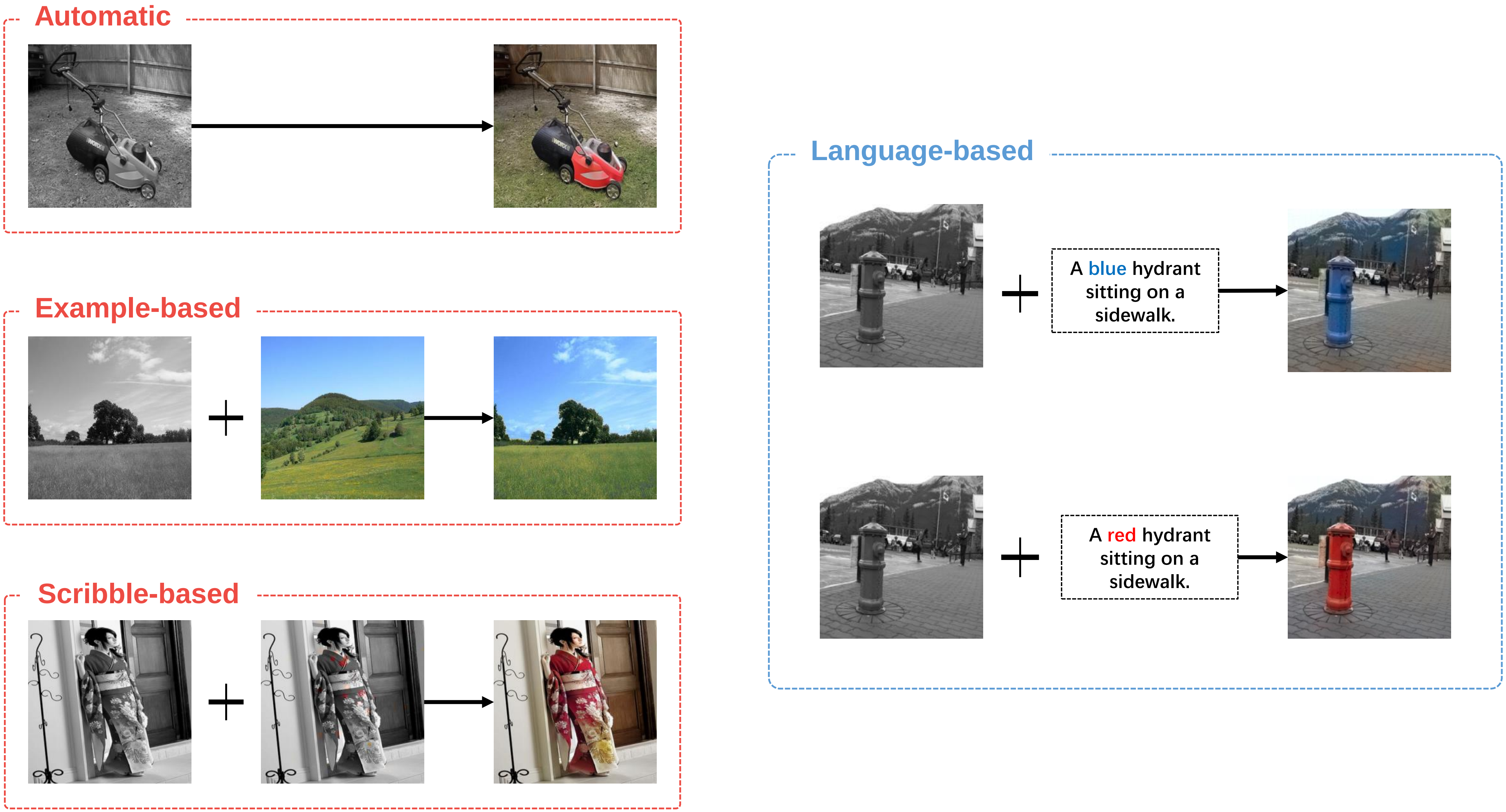
Zheng Chang[1#] Shuchen Weng[2,3#] Peixuan Zhang[1] Yu Li[4] Si Li[*1] Boxin Shi[2,3]

[1]School of Artificial Intelligence, Beijing University of Posts and Telecommunications
[2]National Key Laboratory for Multimedia Information Processing, School of Computer Science, Peking University
[3]National Engineering Research Center of Visual Technology, School of Computer Science, Peking University
[4]International Digital Economy Academy

Outline

- Introduction
  - Background
  - Problem and Improvement
- Method
  - Pipeline
  - Luminance Augmentation
  - Aggregating Similar Patches
- Result
  - Comparison with Language-based Colorization
  - Comparison with Automatic Colorization
  - Ablation
  - Application

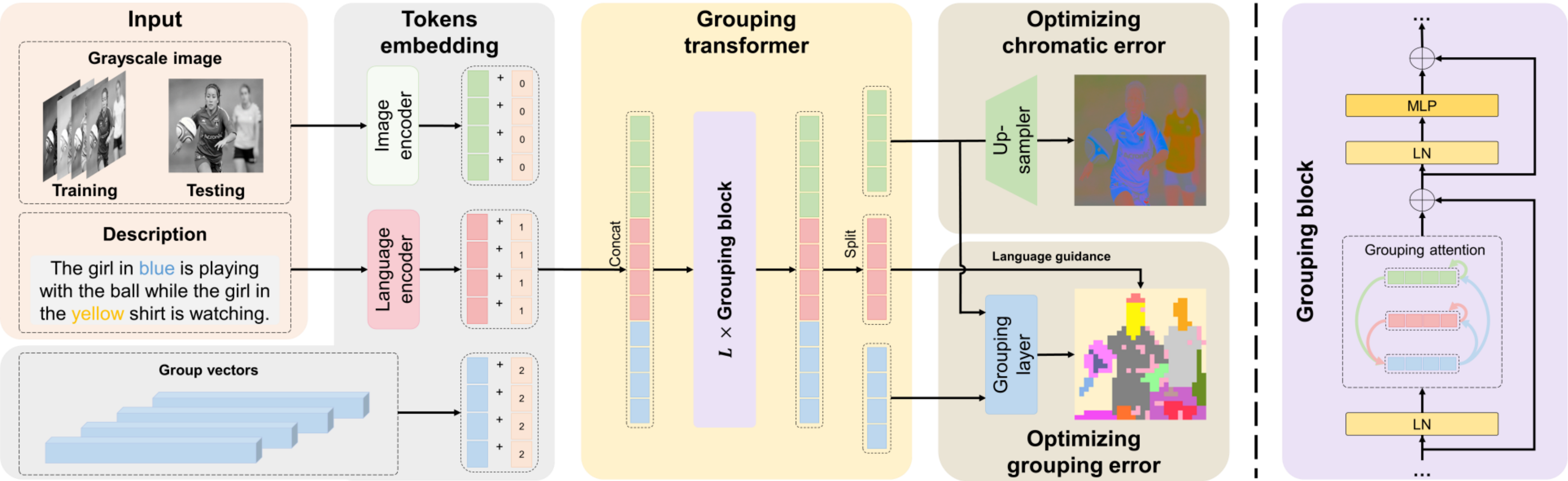
Background


Automatic
Example-based
Scribble-based
Language-based
A blue hydrant sitting on a sidewalk.
A red hydrant sitting on a sidewalk.

Problem and Improvement

Three woman are dress in pink.	The middle woman is in yellow, the left woman in blue, the right woman in red.	The woman in the right was dressed in shirt of violet color.	The middle woman is dressed in orange and the right woman in yellow.

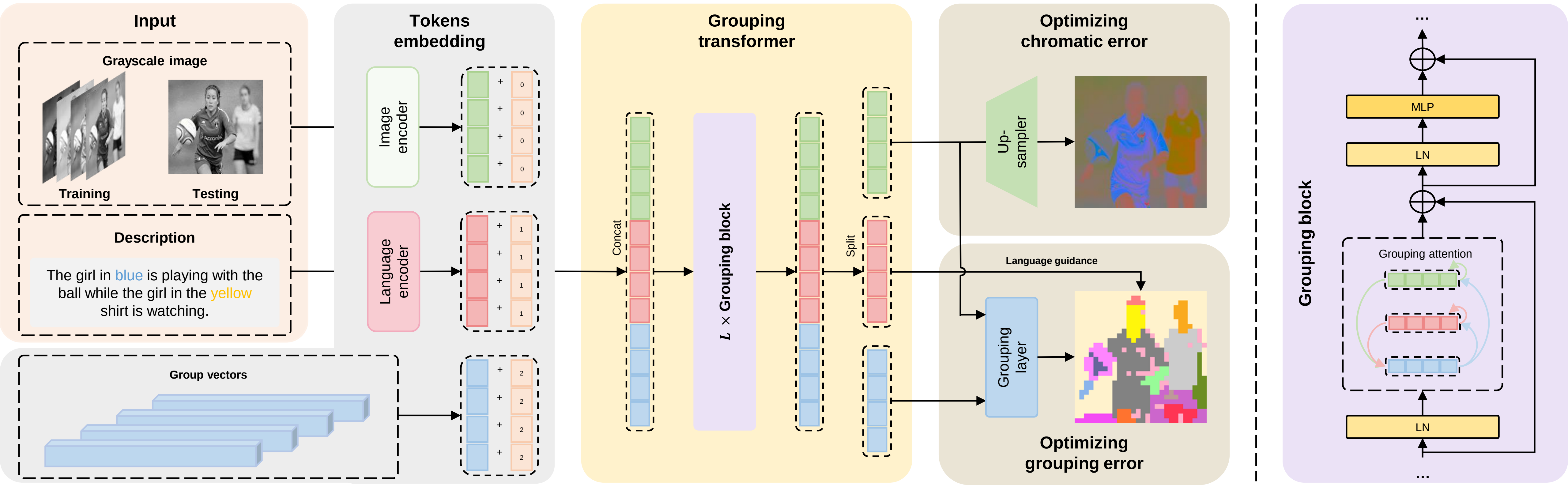
L-CoDer



Outline

- Introduction
  - Background
  - Problem and Improvement
- Method
  - Pipeline
  - Luminance Augmentation
  - Aggregating Similar Patches
- Result
  - Comparison with Language-based Colorization
  - Comparison with Automatic Colorization
  - Ablation
  - Application

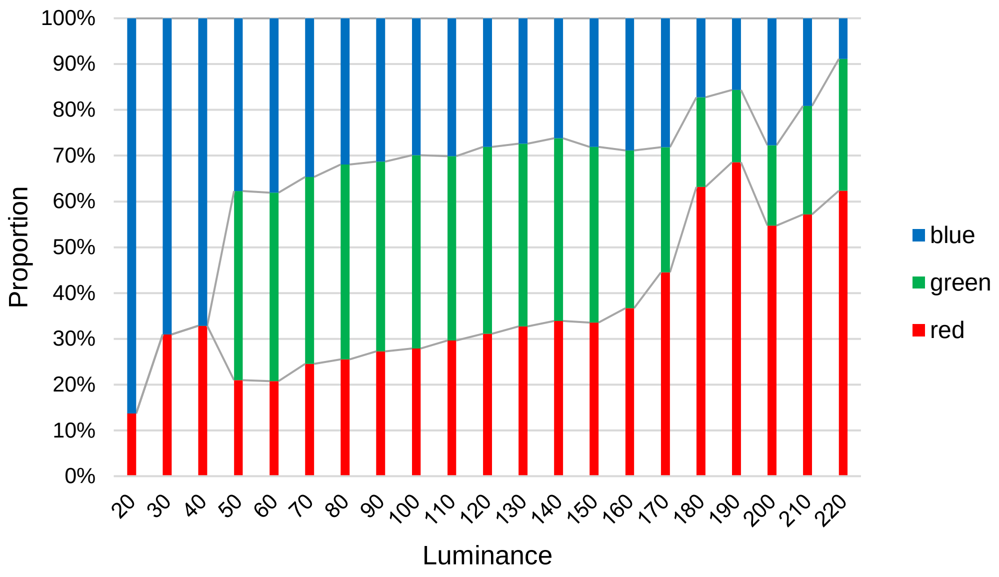
Pipeline



Training Loss：

$$L_{\text{total}} = \alpha L_{\delta} + \beta L_{\text{ctr}} \ \text{, where} \ L_{\delta} = \frac{1}{N_{\text{p}}} \sum \frac{1}{2} (\hat{I}_{\text{ab}} - I_{\text{ab}})^2 \mathbb{1}_{\{|\hat{I}_{\text{ab}} - I_{\text{ab}}| < \delta\}} + \frac{1}{N_{\text{p}}} \sum \delta (|\hat{I}_{\text{ab}} - I_{\text{ab}}| - \frac{1}{2}\delta) \mathbb{1}_{\{|\hat{I}_{\text{ab}} - I_{\text{ab}}| \geq \delta\}}$$

Luminance Augmentation



Luminance augmentation consists of two steps:

I. Randomly rotate the hue as:

$$I_{\mathrm{r}} = \left[\mathrm{F}_{\mathrm{rotate}}\left(I_{\mathrm{o}}^{\mathrm{h}}, \lambda\right), I_{\mathrm{o}}^{\mathrm{s}}, I_{\mathrm{o}}^{\mathrm{v}}\right]$$

II. Adjust the global luminance as:

$$\hat{I}_{\mathrm{g}} = A I_{\mathrm{g}}^{\mathrm{F}_{\mathrm{inv}}(\gamma)}$$

Original grayscale

Augmented grayscale 1

Augmented grayscale 2

Augmented grayscale 3

Augmented grayscale 4

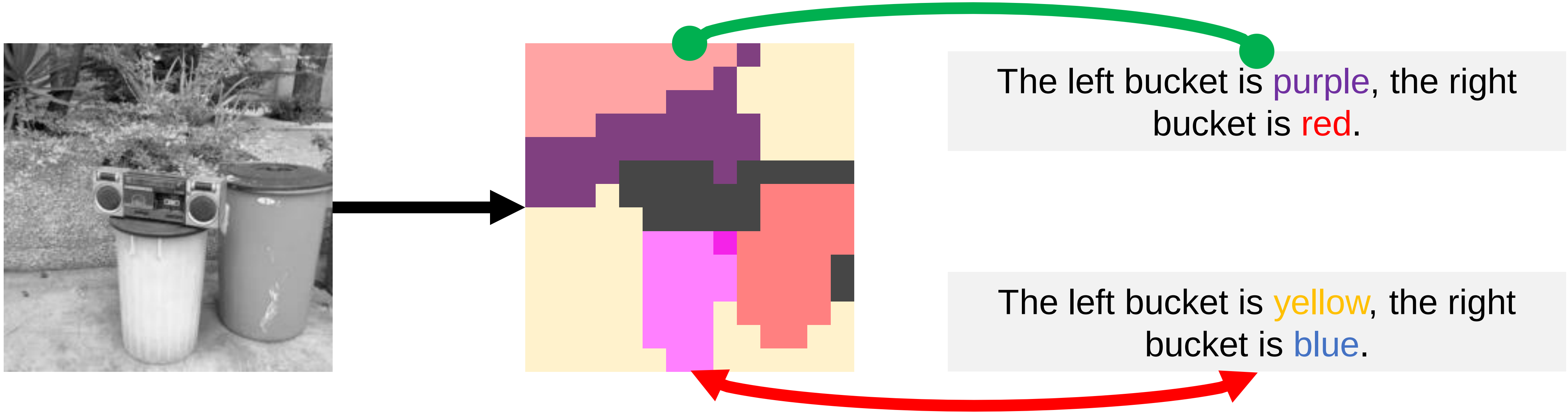
Aggregating Similar Patches

I. Merging image tokens corresponding to the same group to concrete the instance representation:

$$\bar{Z}^{L+1}_{\mathrm{grp},i} = Z^{L+1}_{\mathrm{grp},i} + \frac{\sum_{j=1}^{N_{\mathrm{I}}} \hat{A}_{i,j} \, W_{\mathrm{v}} Z^{L+1}_{\mathrm{img},j}}{\sum_{j=1}^{N_{\mathrm{I}}} \hat{A}_{i,j}}$$

II. Using the counter-color loss to optimize the grouping error:

$$\begin{aligned} L_{\mathrm{ctr}} \quad = \quad & -\log\left(\mathrm{F_{sim}}\left(R_{\mathrm{lag}}, R_{\mathrm{grp}}\right)\left(1 - \mathrm{F_{sim}}\left(R'_{\mathrm{lag}}, R_{\mathrm{grp}}\right)\right)\right) \\ & -\log\left(\mathrm{F_{sim}}\left(R'_{\mathrm{lag}}, R'_{\mathrm{grp}}\right)\left(1 - \mathrm{F_{sim}}\left(R_{\mathrm{lag}}, R'_{\mathrm{grp}}\right)\right)\right) \end{aligned}$$

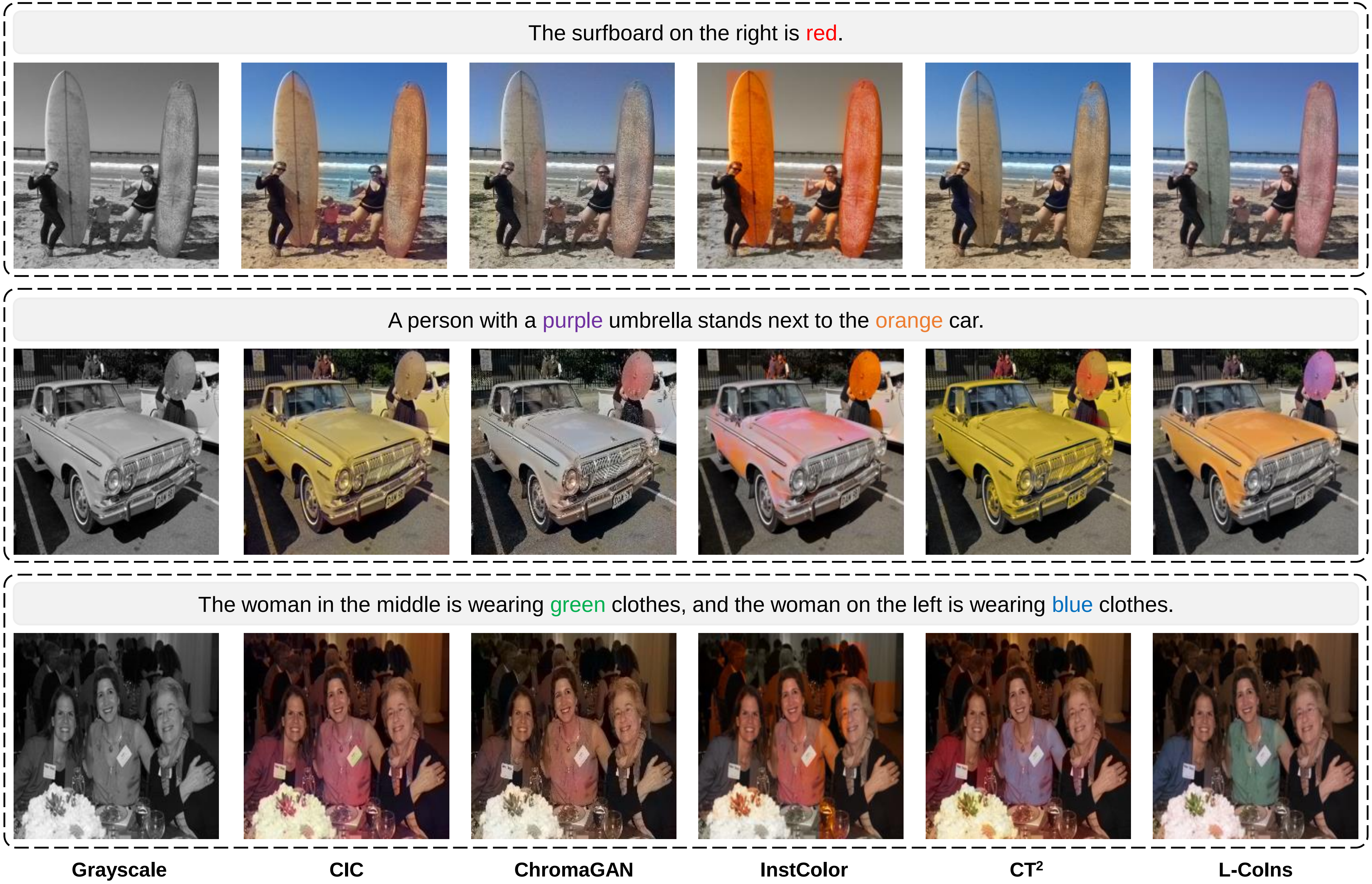


Outline

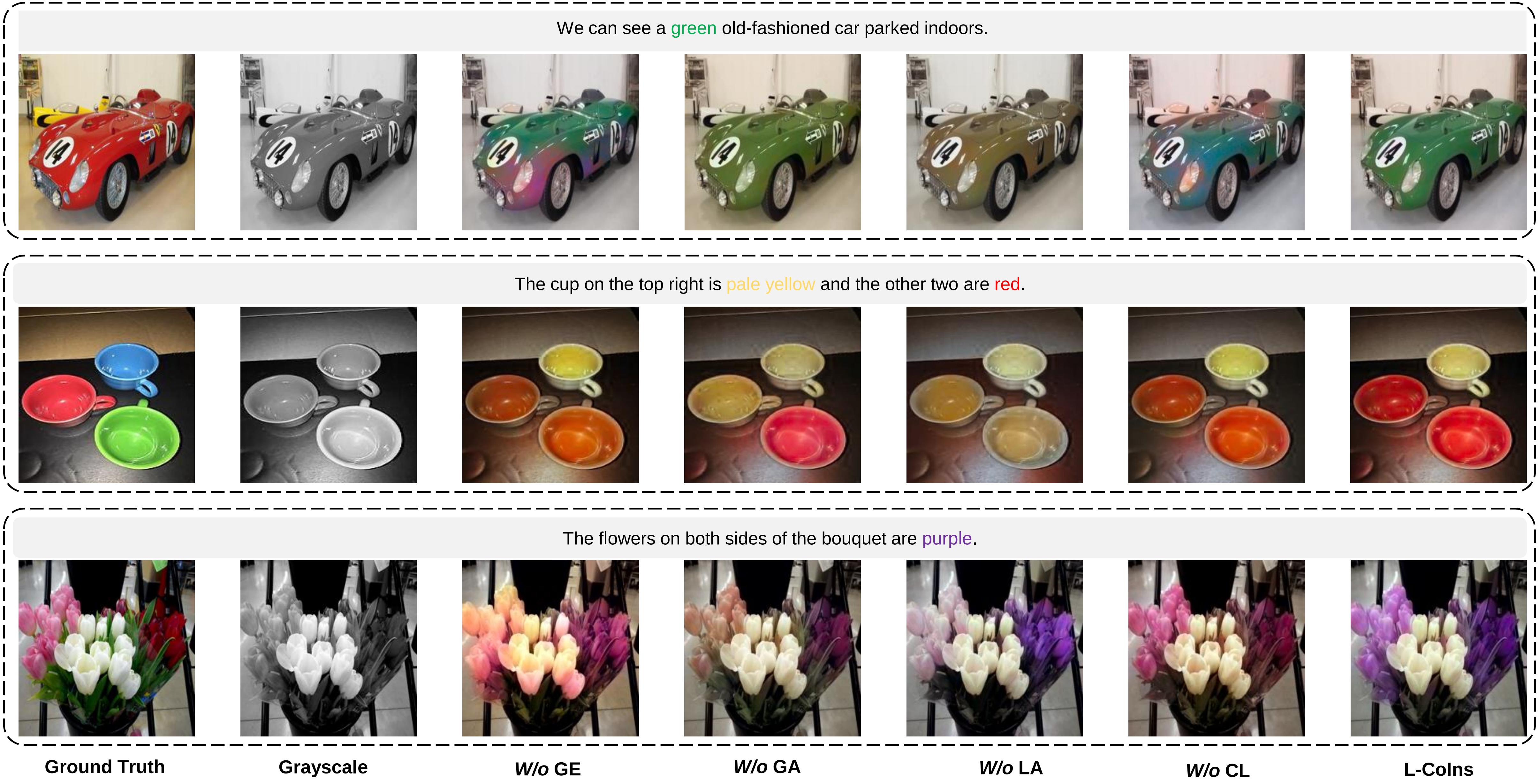
- Introduction
  - Background
  - Problem and Improvement
- Method
  - Pipeline
  - Luminance Augmentation
  - Aggregating Similar Patches
- Result
  - Comparison with Language-based Colorization
  - Comparison with Automatic Colorization
  - Ablation
  - Application

Comparison with Language-based Colorization

Comparison with Automatic Colorization



Ablation


We can see a green old-fashioned car parked indoors.
The cup on the top right is pale yellow and the other two are red.
The flowers on both sides of the bouquet are purple.
Ground Truth
Grayscale
W/o GE
W/o GA
W/o LA
W/o CL
L-CoIns

Application

1909. "Central station."

There is a red house under the blue sky surrounded by green grasses.

1923. " Jewett touring car on mountain road ."

A red car stopped on the yellow dirt road.

1940. "Three boys from Los Angeles who are looking for work in an airplane factory."

The man in the middle is wearing a gold coat.

The two boys on the left are wearing green coats.

The boy in the middle wears a yellow coat, and the men on both sides wear blue coats.