CV Final Project

Dec. 29, 2024

Comparison of the Model on Single Image Reflection Removal

Group 29

Presenter: Guan-Wei Lai, Wei-Chien Cheng

Institute of Electronics, National Yang Ming Chiao Tung University

Outline

- **Introduction**
- **Model-I**
- **Model-II**
- **Model-III**
- **Experimental Result**
- **References**

Introduction

- **Domain**
  - Single Image Reflection Removal (SIRR)

- **Goal**
  - Separate reflection and transmission layers from a single image

- **Motivation**
  - Reflection artifacts often degrade the quality of photographs captured through glass surfaces. By exploring different approaches in SIRR, we aim to advance this field and improve visual quality for various applications such as photography, computer vision, and autonomous systems.

Introduction

- **Dataset**
  - Transmission layer
  - Reflection layer
  - Blended

(Blended)

[7] C. Lei *et al.*, *CVPR*, 2022.

Introduction

- **Focus**
  - Explore different network architectures and loss functions used in SIRR
    - **Model-I (*CVPR* 2018) [1]**
    - **Model-II (CVPR 2019) [2]**
    - **Model-III (CVPR 2020) [3]**

Outline

- Introduction
- **Model-I**
- Model-II
- Model-III
- Experimental Result
- References

Model-I

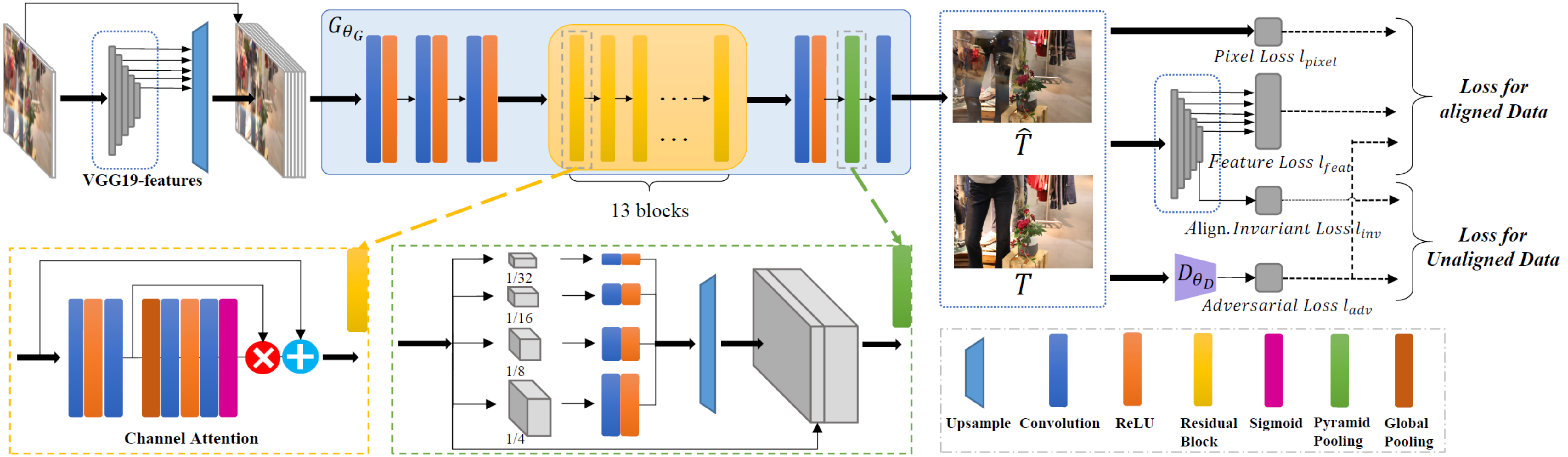
- **Model:** Single Image Reflection Separation with Perceptual Losses
- **Architecture**:
  - First layer: 1×11 convolution to reduce input dimensions.
  - Subsequent layers: 3×3 dilated convolutions (dilations vary from 1 to 128).
  - Final layer: Linear transformation for RGB color space.
- **Input**: Image (***I***) with reflections.
- **Output**: Two layers:
  - Transmission Layer: $f_T(I;\theta)$
  - Reflection Layer : $f_R(I;\theta)$
- The model is trained on a dataset $D=\{(I,\ T,\ R)\}$, where:
  - ***T*:** Ground truth transmission layer.
  - ***R*:** Ground truth reflection layer.
- **Loss Functions**
  - Feature Loss (L_{feat}): Matches features of predicted and ground-truth images.
  - Adversarial Loss (L_{adv}): Ensures realistic image refinement.
  - Separation Loss (L_{excel}): Promotes clear separation of transmission and reflection layers.
- **Combined Loss**: $L(\theta)=w_1L_{\text{feat}}+w_2L_{\text{adv}}+w_3L_{\text{excel}}$

[1] X. Zhang *et al.*, "Single Image Reflection Separation with Perceptual Losses", *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

Outline

- Introduction
- Model-I
- **Model-II**
- Model-III
- Experimental Result
- References

Model-II



- **Simplify the basic residual block (Remove BN layer)**
- **Increase the capacity by widening the network feature maps**
- **Training for unaligned data**

[2] K. Wei *et al.*, "Single Image Reflection Removal Exploiting Misaligned Training Data and Network Enhancements", *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

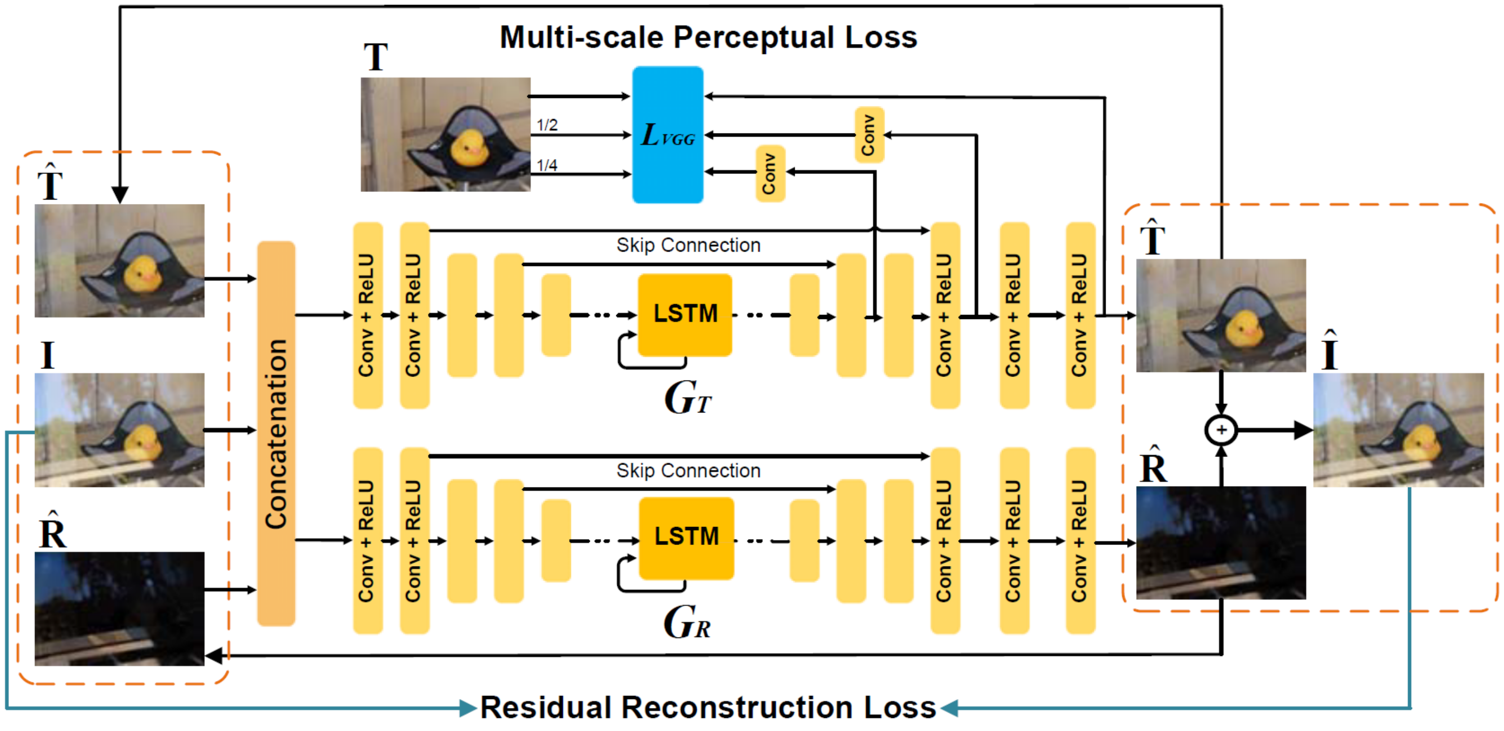
Outline

- Introduction
- Model-I
- Model-II
- **Model-III**
- Experimental Result
- References

Model-III

- **Model:** Iterative Boost Convolutional LSTM Network (IBCLN)

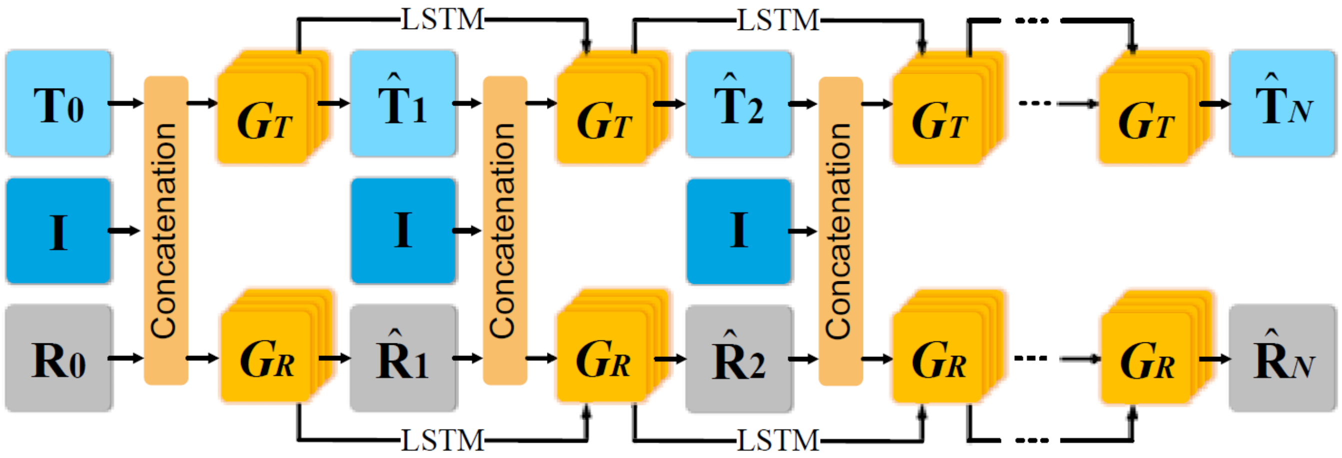
- **Convolutional LSTM:** Transfers information through gates (Input, Forget, Output, Cell).
- **A cascaded framework with two sub-networks:**
  - Transmission Prediction Network (G_T)
  - Reflection Prediction Network (G_R)
- **Each sub-network includes:**
  - Encoder-Decoder structure with **Conv + ReLU** layers.
  - Skip **connections** to prevent output blurring.
  - A **Convolutional LSTM** unit for information transfer between steps.



[3] C. Li *et al.*, "Single Image Reflection Removal Through Cascaded Refinement", *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

Model-III

- **Characterizing IBCLN with increasing number of time steps:**



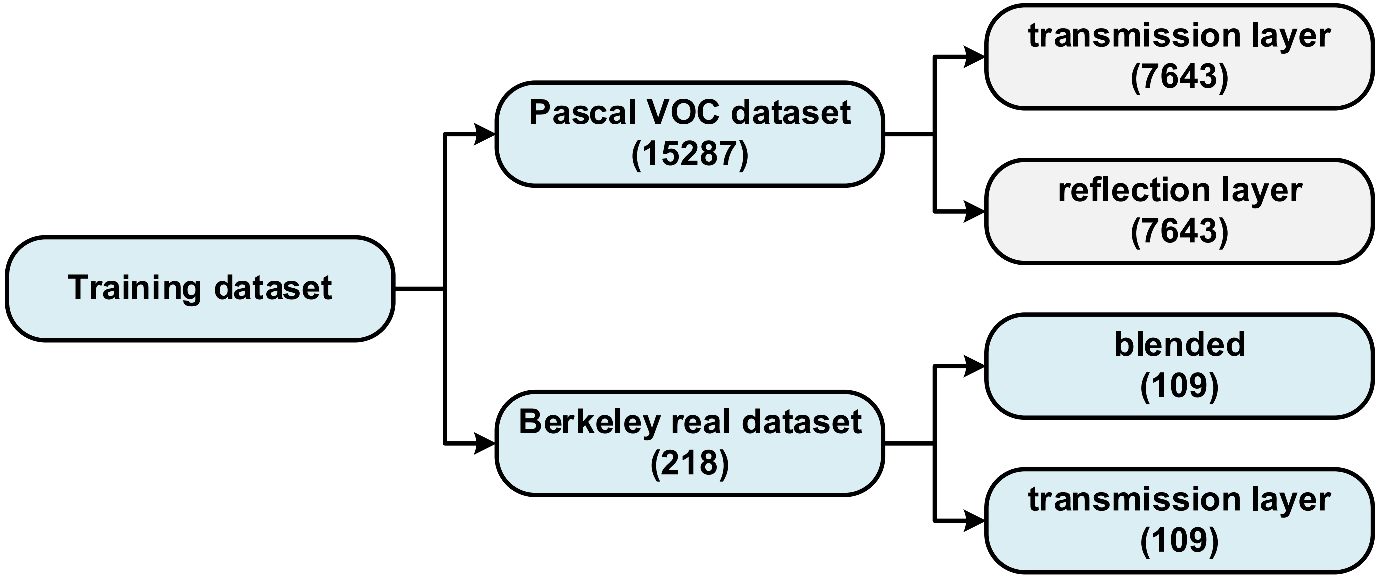
- **Key Features:**
  - **Convolutional LSTM:** Transfers information through gates (Input, Forget, Output, Cell).
  - Iterative process enhances both transmission and reflection predictions.
- **Advantages:**
  - Accurate reflection removal through **collaborative refinement**.
  - Prevents blurring with **skip connections**.

Outline

- Introduction
- Model-I
- Model-II
- Model-III
- **Experimental Result**
- References

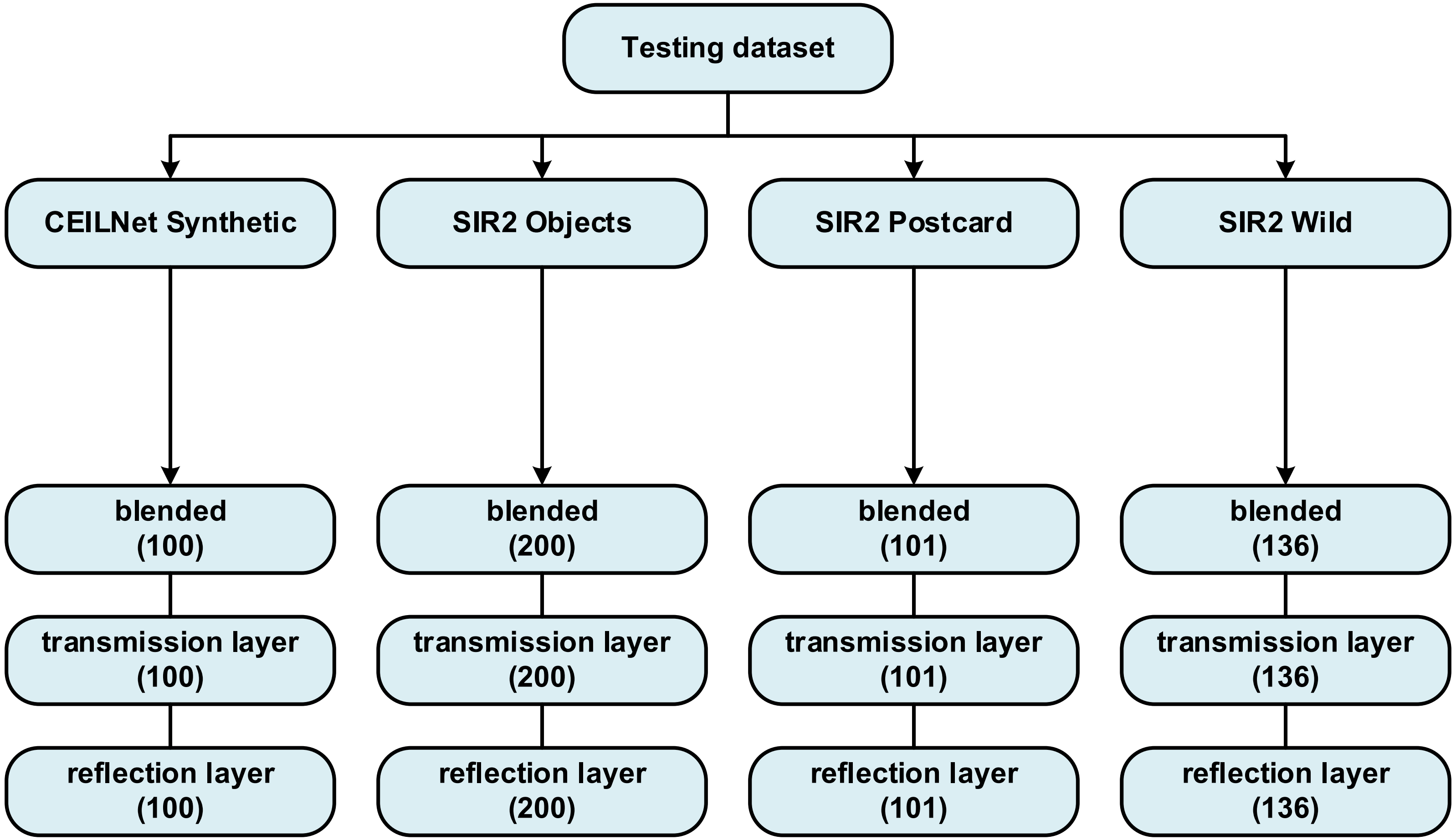
Experimental Result

- **Training Dataset**
  - **Real Data**
    - Extracted from **real-world** photographs
    - Includes **Blended** and **Transmission layers** (Reflection layer data is often missing)
  - **Synthetic Data**
    - Generated from both **Transmission layers** and **Reflection layers**



Experimental Result

- **Testing Dataset**



Experimental Result

- **Evaluation Metrix**
  - **PNSR (Peak Signal-to-Noise Ratio)**
    - Measures **mean squared error**
    - Quantifies **color fidelity**
    - Indicates the **maximum possible range** of the original image's pixel values
  - **SSIM (Structural Similarity Index)**
    - **Luminance**
    - **Contrast**
    - **Structure**

$$l(x,y)=\frac{2\mu_x\mu_y+C_1}{\mu_x^2+\mu_y^2+C_1},c(x,y)=\frac{2\sigma_x\sigma_y+C_2}{\sigma_x^2+\sigma_y^2+C_2},s(x,y)=\frac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3}$$

  - **LPIPS (Learned Perceptual Image Patch Similarity)**
    - A metric evaluated using **deep learning** models
    - It leverages **perceptual loss** from SRGAN to quantify differences in perceptual features between two images

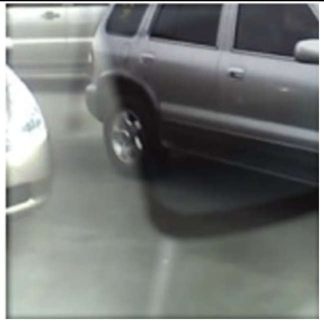
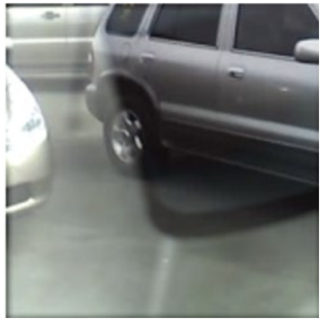
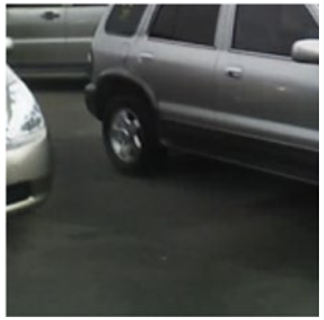
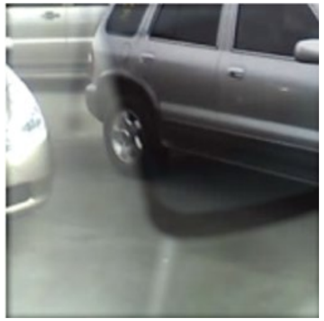
Experimental Result

***Table :** Evaluation dataset*

Dataset	Index	Model-I	Model-II	Model-III
CEIL_Net Synthetic (100)	PSNR(↑)	28.785	32.051	28.710
	SSIM(↑)	0.870	0.972	0.849
	LPIPS(↓)	0.1829	0.041	0.271
SIR2 Objects (200)	PSNR(↑)	29.608	29.929	29.840
	SSIM(↑)	0.937	0.940	0.943
	LPIPS(↓)	0.109	0.098	0.109
SIR2 Postcard (199)	PSNR(↑)	29.108	29.129	29.441
	SSIM(↑)	0.949	0.957	0.957
	LPIPS(↓)	0.164	0.159	0.168
SIR2 Wild (55)	PSNR(↑)	29.400	30.382	30.191
	SSIM(↑)	0.939	0.931	0.953
	LPIPS(↓)	0.116	0.106	0.117

Evaluation device: Nvidia A100 40GB, AMD EPYC 7302 16-Core, 256GB RAM

Experimental Result – CEIL_Net

	Input Image	Output (Transmission Layer)	Output (Reflection Layer)	Ground Truth (Transmission Layer)
Model-I				
Model-II				
Model-III				

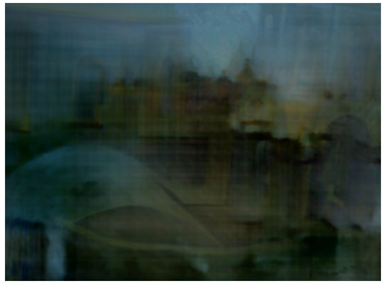
Experimental Result – Wild

	Input Image	Output (Transmission Layer)	Output (Reflection Layer)	Ground Truth (Transmission Layer)
Model-I				
Model-II				
Model-III				

Experimental Result – Object

	Input Image	Output (Transmission Layer)	Output (Reflection Layer)	Ground Truth (Transmission Layer)
Model-I				
Model-II				
Model-III				

Experimental Result – Postcard

	Input Image	Output (Transmission Layer)	Output (Reflection Layer)	Ground Truth (Transmission Layer)
Model-I				
Model-II				
Model-III				

Outline

- Introduction
- Model-I
- Model-II
- Model-III
- Experimental Result
- **References**

References

[1] X. Zhang *et al.*, “Single Image Reflection Separation with Perceptual Losses”, *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[2] K. Wei *et al.*, “Single Image Reflection Removal Exploiting Misaligned Training Data and Network Enhancements”, *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[3] C. Li *et al.*, “Single Image Reflection Removal Through Cascaded Refinement”, *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[4] M. Everingham *et al.*, “The Pascal Visual Object Classes (VOC) challenge,” [Data set] 2010. [Online]. Available: http://host.robots.ox.ac.uk/pascal/VOC/

[5] Q. Fan *et al.*, “CEILNet reflection removal dataset,” [Data set] 2017. [Online]. Available: https://github.com/YonghaoHe/CEILNet

[6] Wanet *et al.*, “Benchmarking single-image reflection removal algorithms,” [Data set] 2023. [Online]. Available: https://sir2data.github.io/

[7] C. Lei *et al.*, “A Categorized Reflection Removal Dataset with Diverse Real-world Scenes”, *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.