# Two-Phase Progressive Multiple Exposure Fusion via Intermediate Exposure Generation

Yu-Fu Huang

Advisor: Yung-Yu Chuang Ph.D.

# Introduction

Problem Definition

- Real-world scenes: $10^{-3} \sim 10^{5}$
- Display limitation: 8 $\text{bits} \rightarrow 2^{8} = 256$

# Introduction

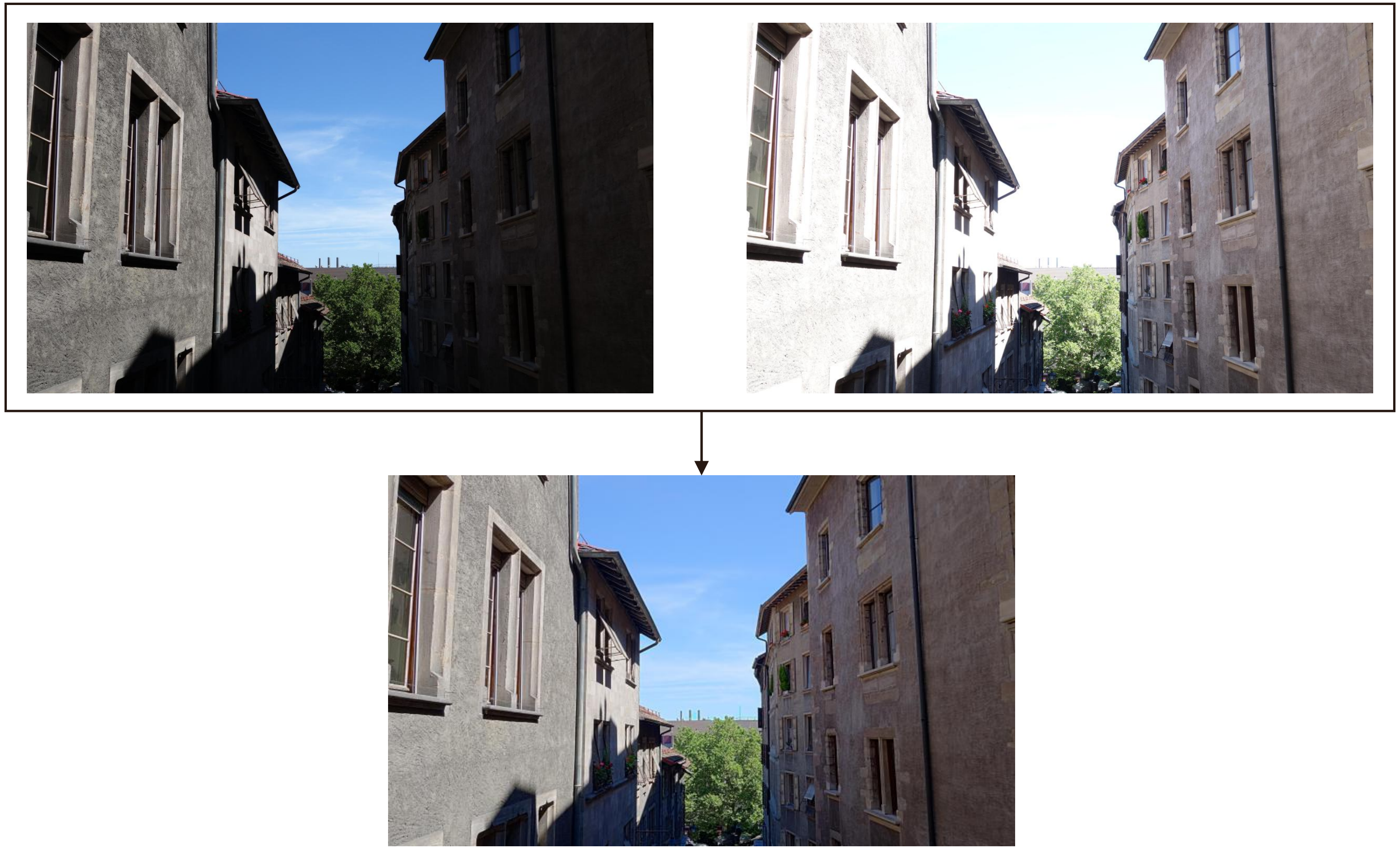
Problem Definition

- Real-world scenes: $10^{-3} \sim 10^{5}$
- Display limitation: $8 \text{ bits} \rightarrow 2^{8} = 256$
- Result: Detail loss in extremely dark/bright regions

# Introduction

## Problem Definition

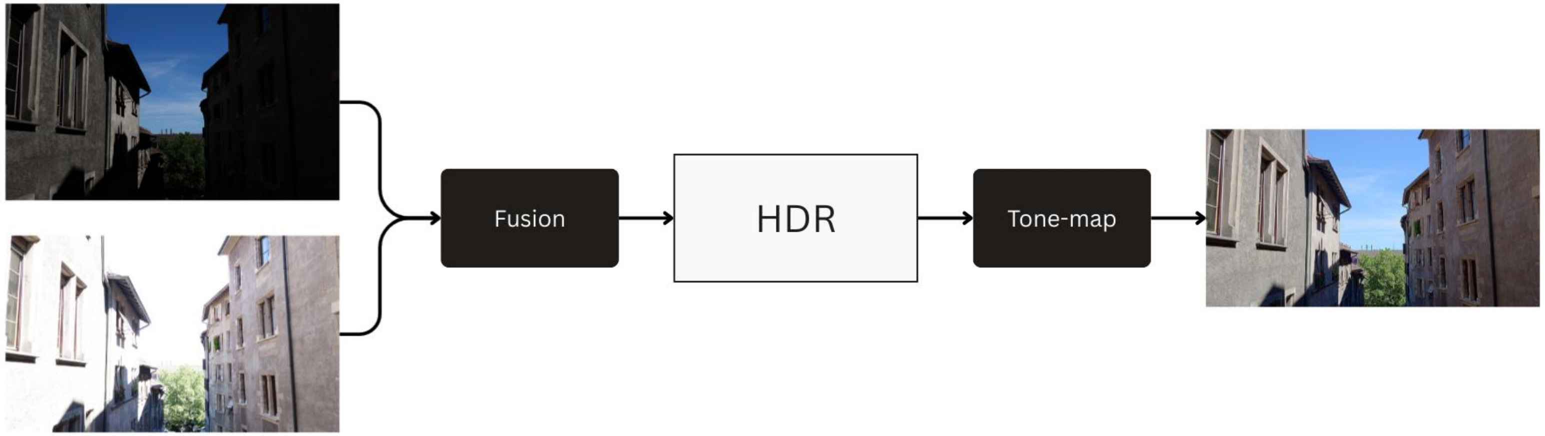
- Solution: Multiple exposure capture



# Introduction

HDR vs. MEF

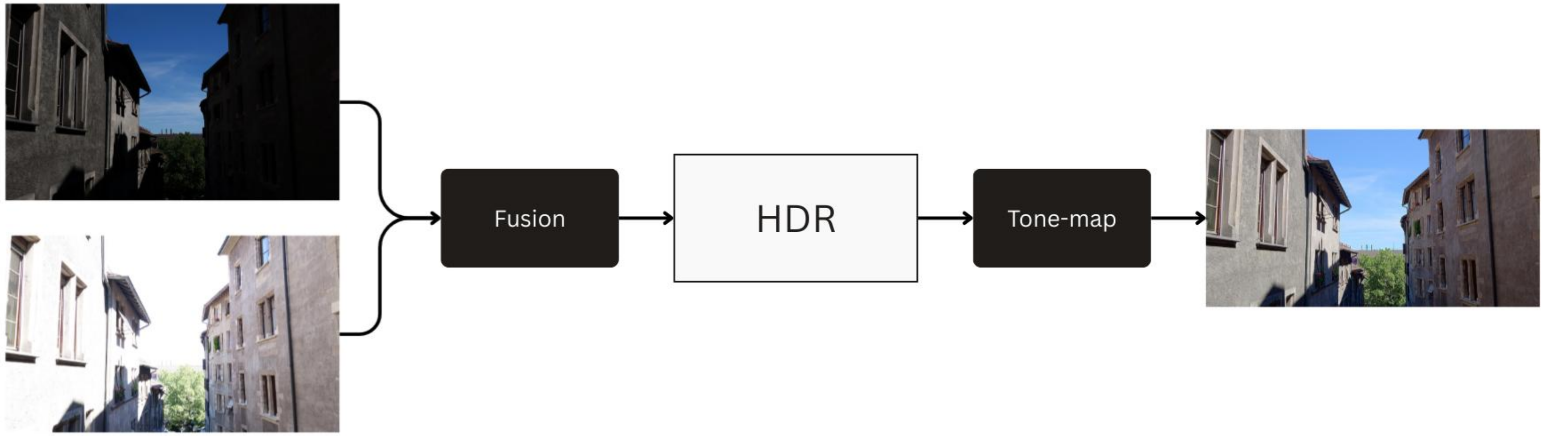
- HDR Reconstruction:



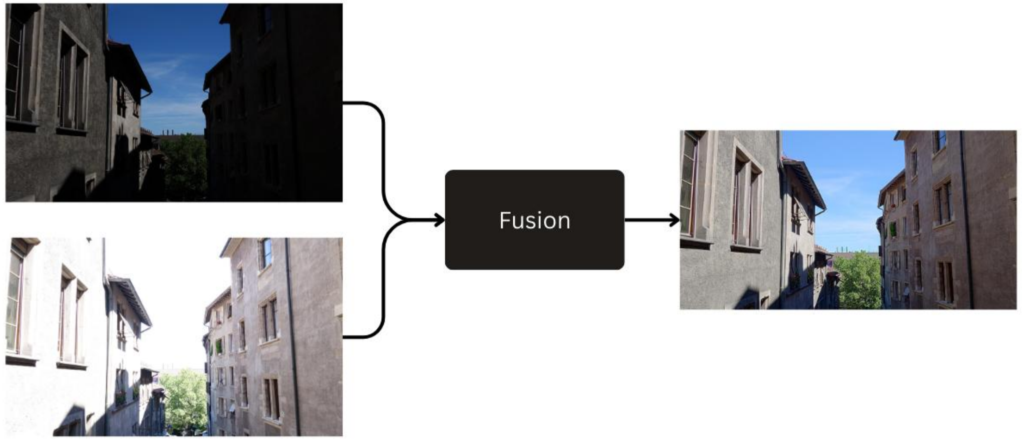
# Introduction

HDR vs. MEF

- HDR Reconstruction:



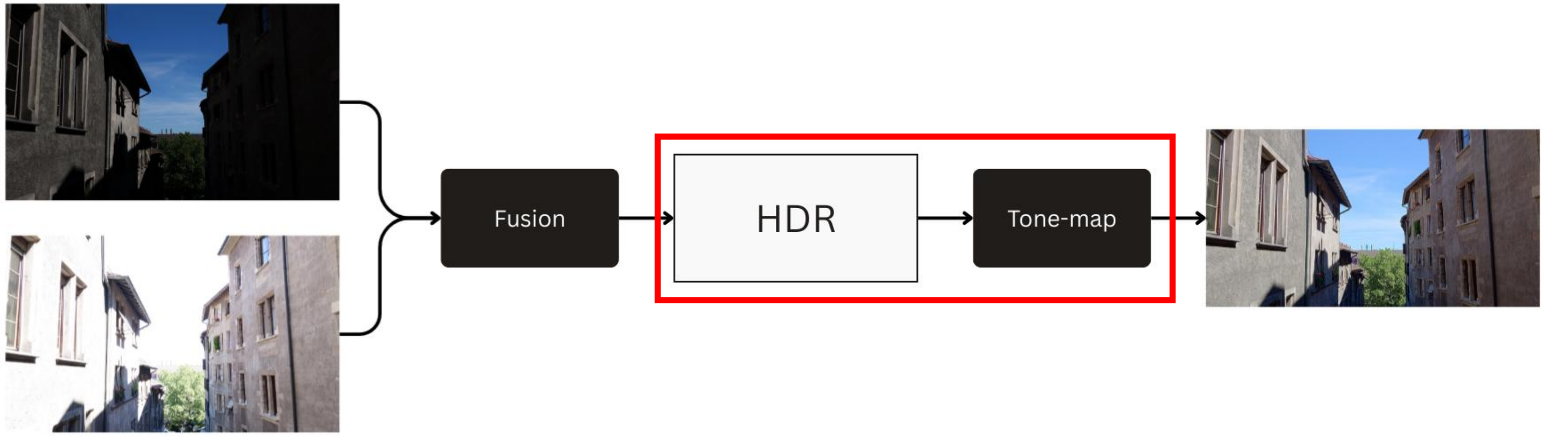

- Multiple Exposure Fusion (MEF)



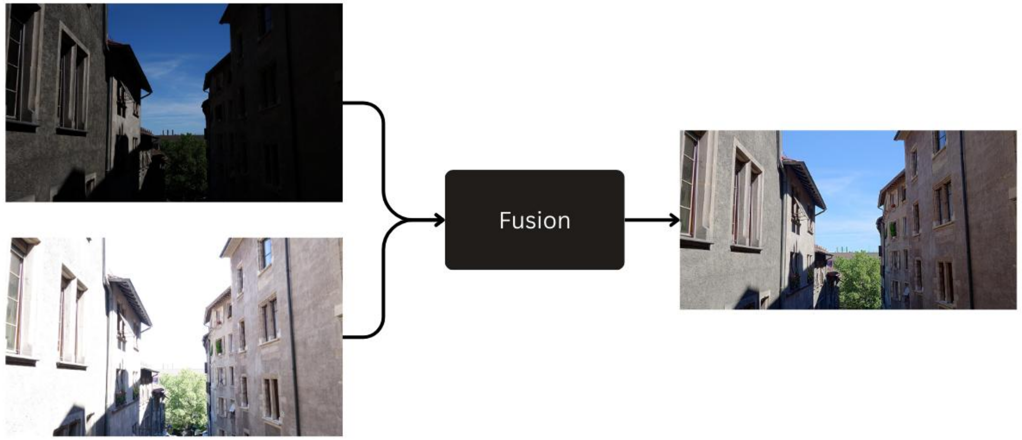
# Introduction

HDR vs. MEF

- HDR Reconstruction:




- Multiple Exposure Fusion (MEF)



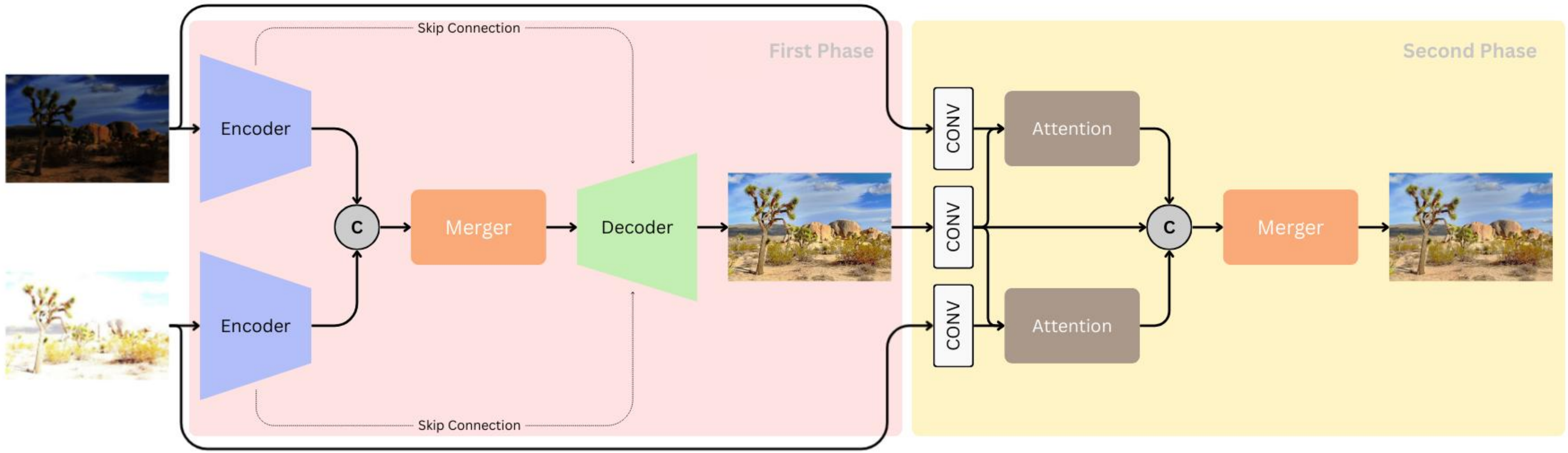
# Introduction

Contribution

Novel Two-Phase Progressive Architecture with Decoupled Training

First to progressively fuse first-stage output with original inputs

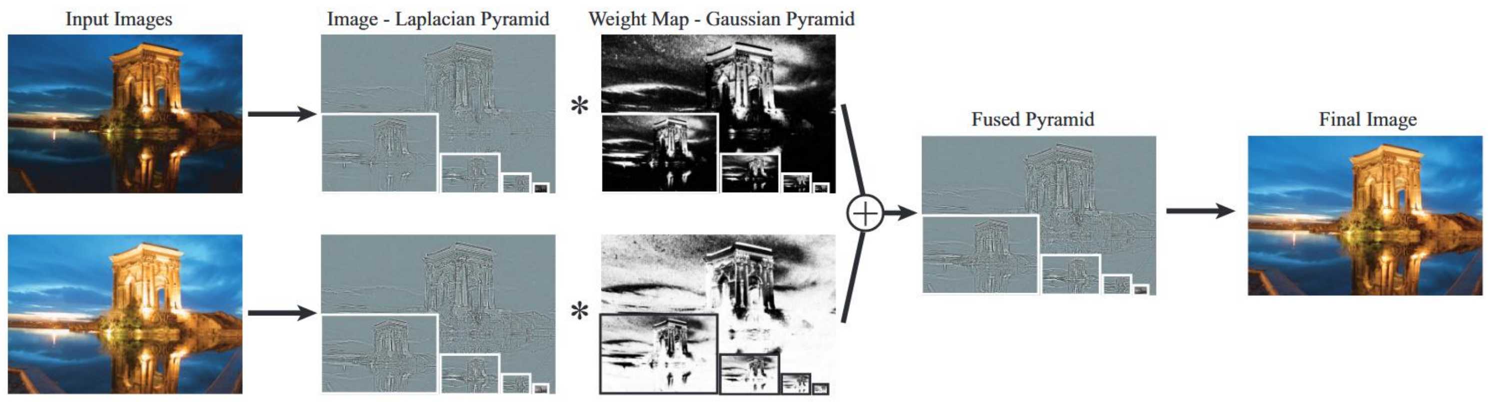
- Phase 1: Intermediate Exposure Generation with UNet for Global Information
- Phase 2: Final Resutl with Attention-Based Refinement for Local Information



# Related Works

Traditional Methods

- Mertens et al. (2007) – Exposure Fusion
  - Laplacian pyramid decomposition
  - Quality measures: Contrast, Saturation, Well-exposedness



# Related Works

Traditional Methods

- Mertens et al. (2007) – Exposure Fusion
    - Laplacian pyramid decomposition
    - Quality measures: Contrast, Saturation, Well-exposedness
- Ma et al. (2017) – Structural Patch Decomposition
    - Patch level instead of pixel level
    - Decomposes into 3 components: signal strength, signal structure, mean intensity

$$
\begin{aligned}
\mathbf{x}_k &= \left\|\mathbf{x}_k - \mu_{\mathbf{x}_k}\right\| \cdot \frac{\mathbf{x}_k - \mu_{\mathbf{x}_k}}{\left\|\mathbf{x}_k - \mu_{\mathbf{x}_k}\right\|} + \mu_{\mathbf{x}_k} \\
&= \left\|\tilde{\mathbf{x}}_k\right\| \cdot \frac{\tilde{\mathbf{x}}_k}{\left\|\tilde{\mathbf{x}}_k\right\|} + \mu_{\mathbf{x}_k} \\
&= c_k \cdot \mathbf{s}_k + l_k
\end{aligned}
$$

# Related Works

## Deep Learning Methods

- DPE-MEF: Multi-exposure image fusion via deep perceptual enhancement
    - Output in YCbCr space

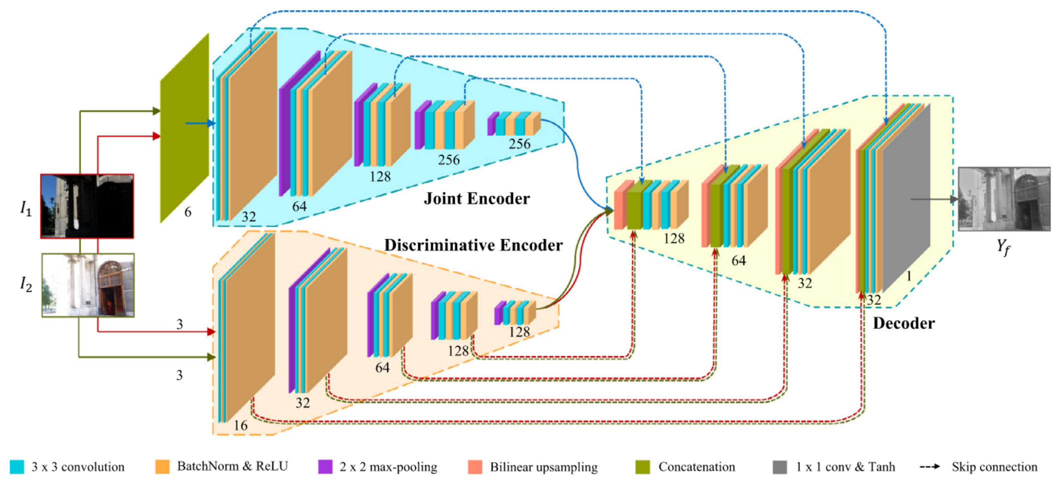
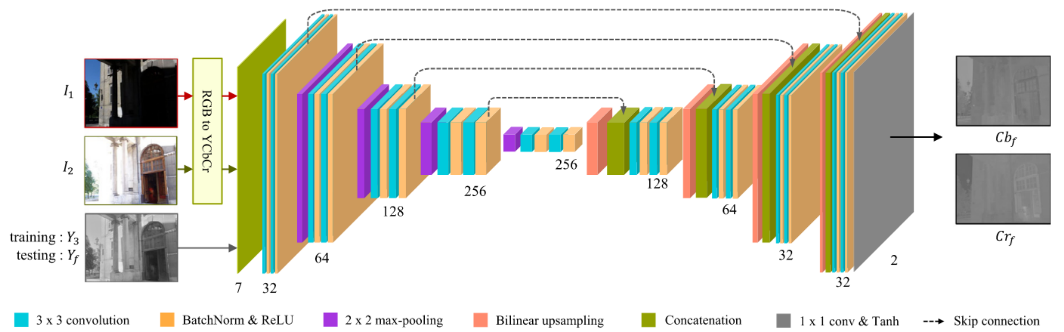


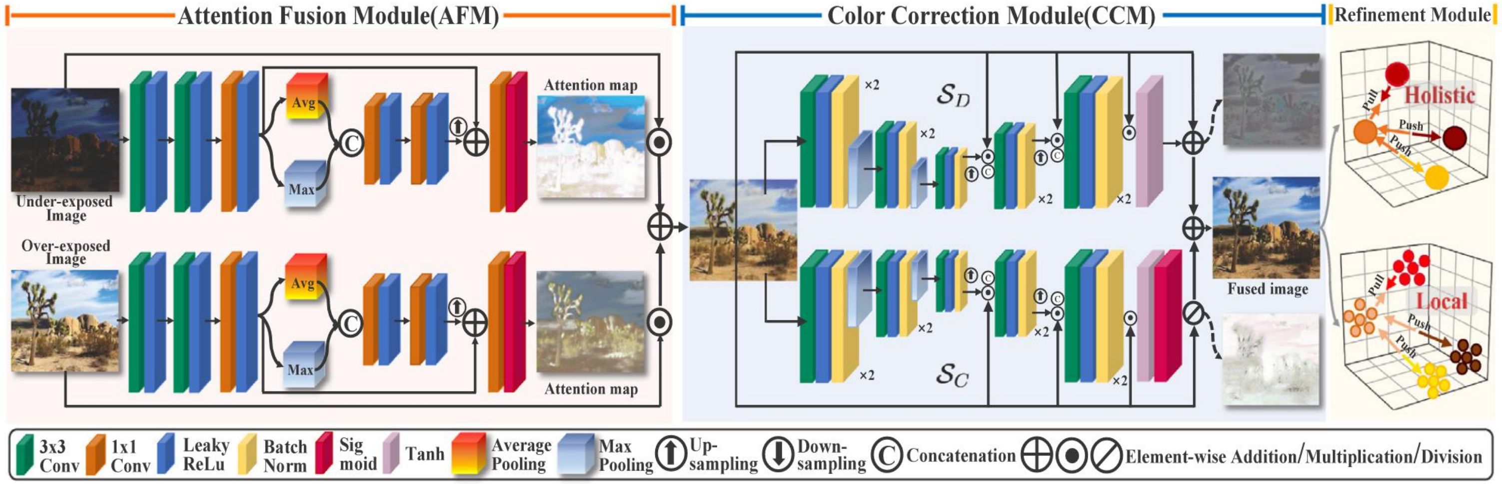

**Fig. 4.** The architecture of detail enhancement module. The numbers indicate the channel amounts.



# Related Works

Deep Learning Methods

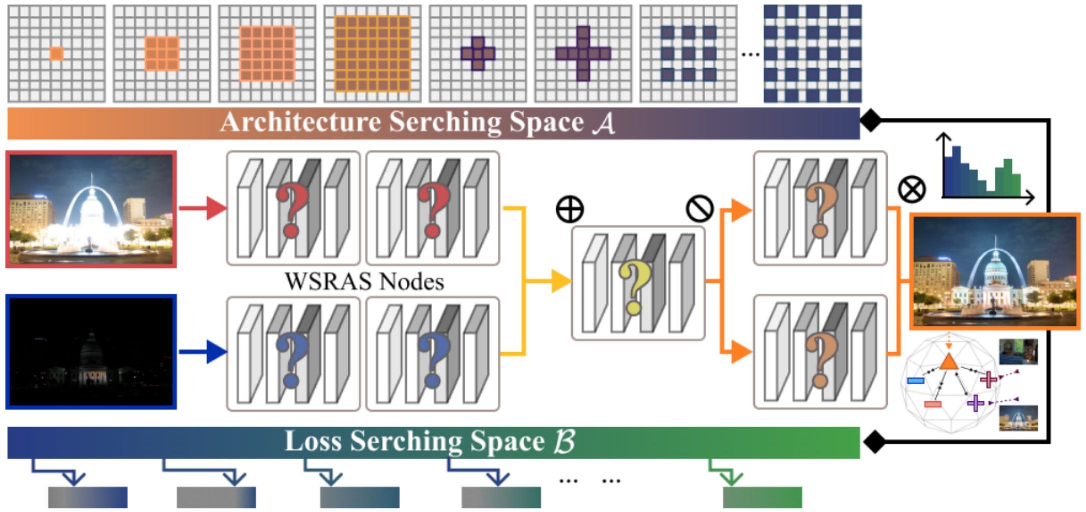
- DPE-MEF: Multi-exposure image fusion via deep perceptual enhancement
    - Output in YCbCr space
- HoLoCo: Holistic and local contrastive learning network for multi-exposure image fusion
    - Attention Map + UNet based on Retinex Theory
    - Contrastive Learning



# Related Works

Deep Learning Methods – Automated Design

- HSDS-MEF: Hybrid-Supervised Dual-Search: Leveraging Automatic Learning for Loss-Free Multi-Exposure Image Fusion
    - Search for optimized loss parameters and network layers




**Algorithm 1:** Dual Search for Structure and Loss Function

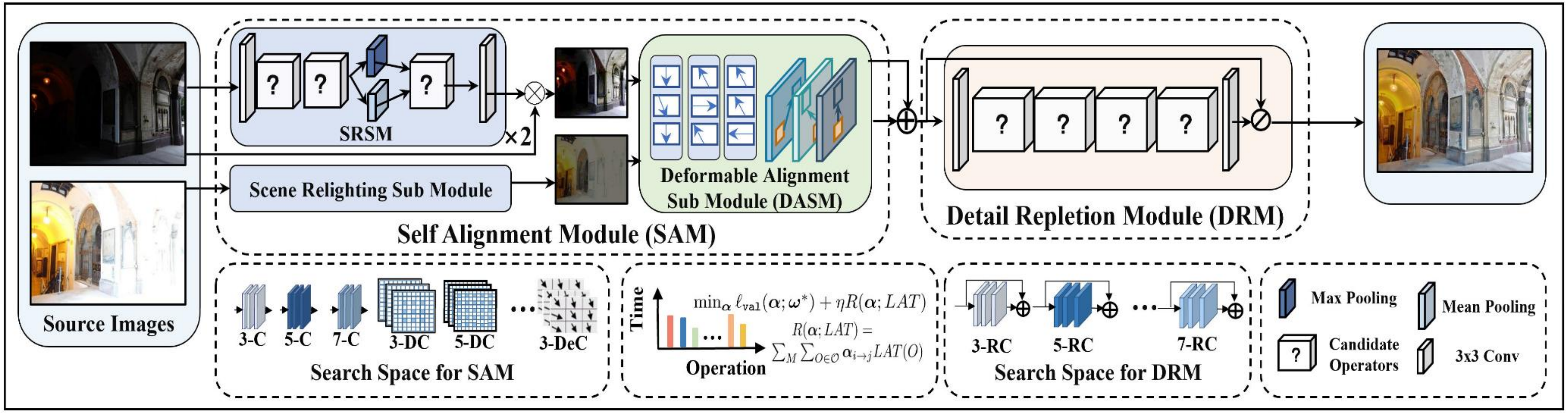
**Require:** Fusion loss $\mathcal{L}^{\mathrm{f}}$, hybrid-supervised contrast constraint $\mathbf{\Gamma}_{\mathrm{h}}$, search spaces $\mathcal{A}$ and $\mathcal{B}$, and other necessary hyper-parameters.
**Ensure:** Optimal parameters $\boldsymbol{\alpha}^*$ and $\boldsymbol{\beta}^*$.

```
while not converged do
    % Optimizing the image fusion network
    ω ← ω − ∇L^f_train ∼ (α; ω; β)
    % First-order approximation to optimize β
    β ← β − ∇_β Γ_H ∼ (α; ω)
    % First-order approximation to optimize α
    α ← α − ∇_α L^f_val ∼ (α; ω; β)
    while |A| > P do
        if min α_i < θ then
            % Prune the operation with the smallest weight in search space A
            A ← RefineSmallestWeight(A)
        end if
    end while
end while
return Top-P operations based on α*, β*.
```

# Related Works

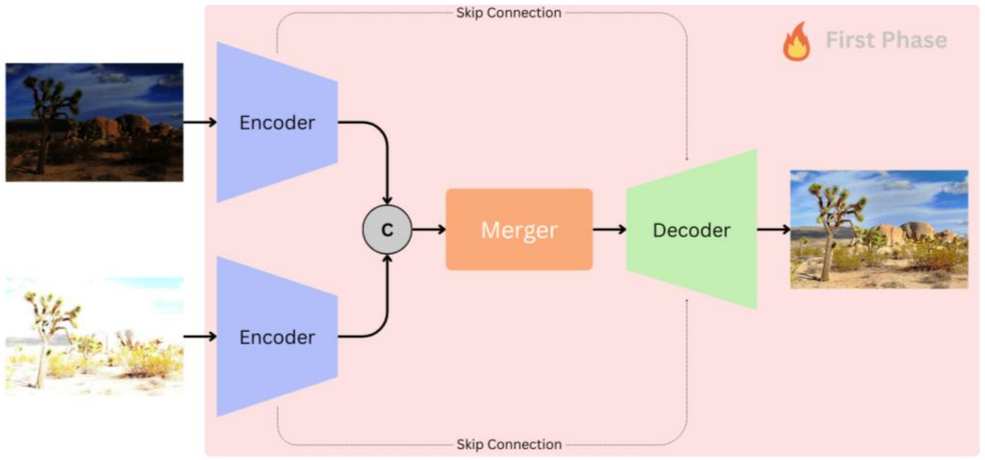
## Deep Learning Methods – Automated Design

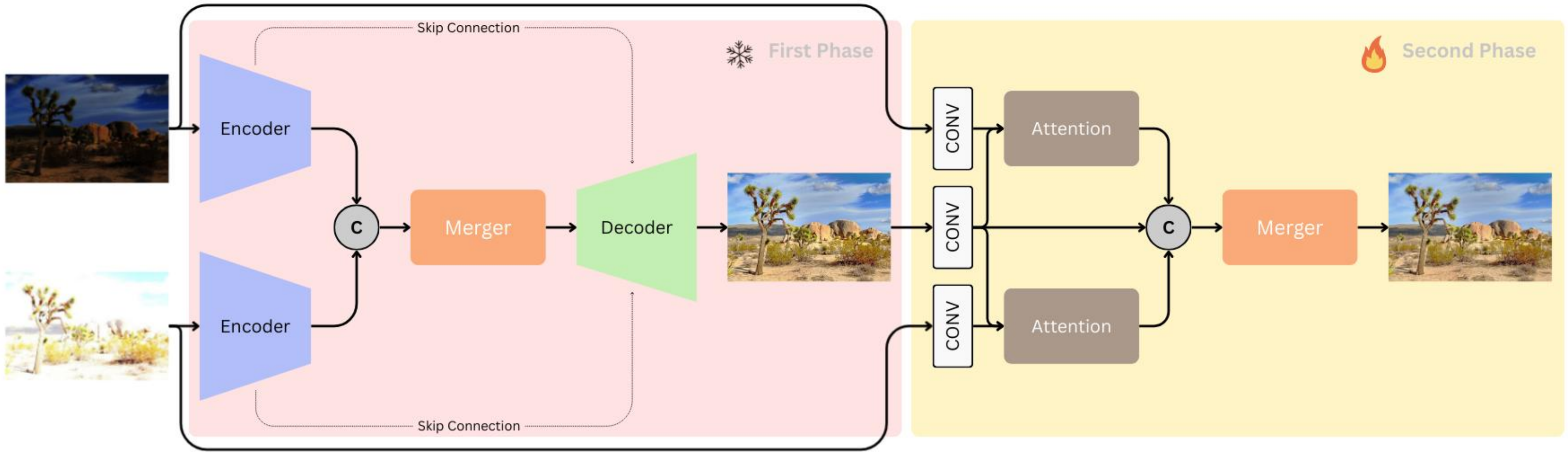
- HSDS-MEF: Hybrid-Supervised Dual-Search: Leveraging Automatic Learning for Loss-Free Multi-Exposure Image Fusion
  - Search for optimized loss parameters and network layers
- Searching a Compact Architecture for Robust Multi-Exposure Image Fusion
  - Hardware constraint



# Method

## Overview

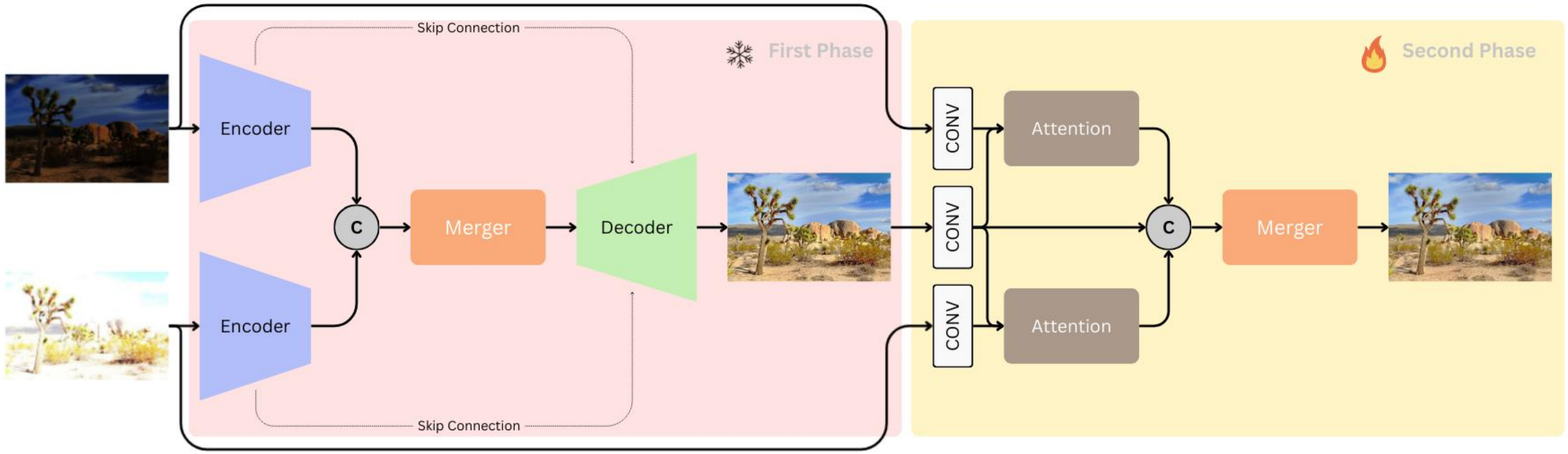


# Method

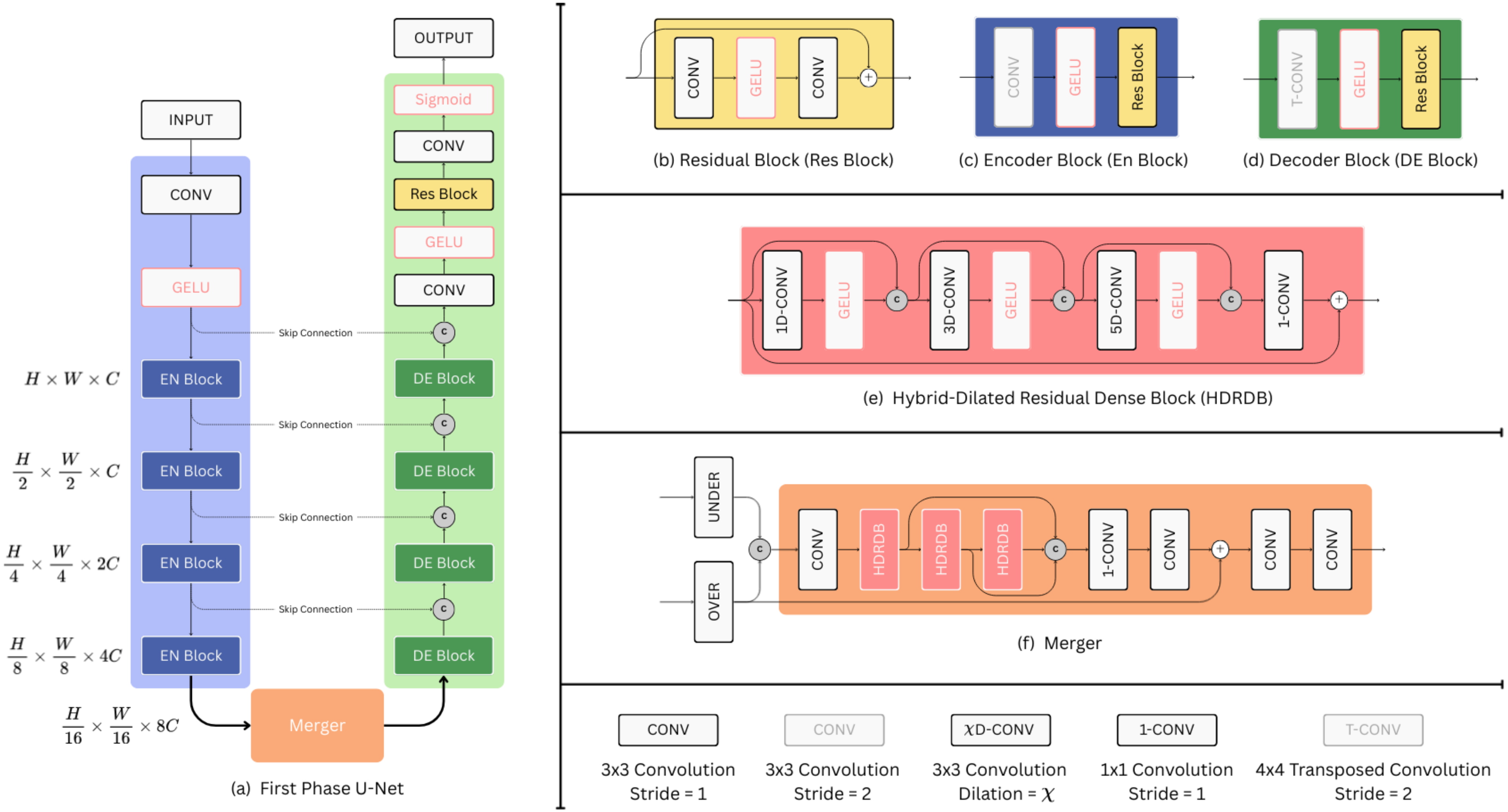
## Overview

- Propose an architecture that explicitly separates global and local fusion
- Develop a decoupled training strategy for specialized optimizations.
- Introduce intermediate exposure generation to bridge extreme inputs
- Demonstrate that reduced exposure differences improve attention mechanism performance in MEF



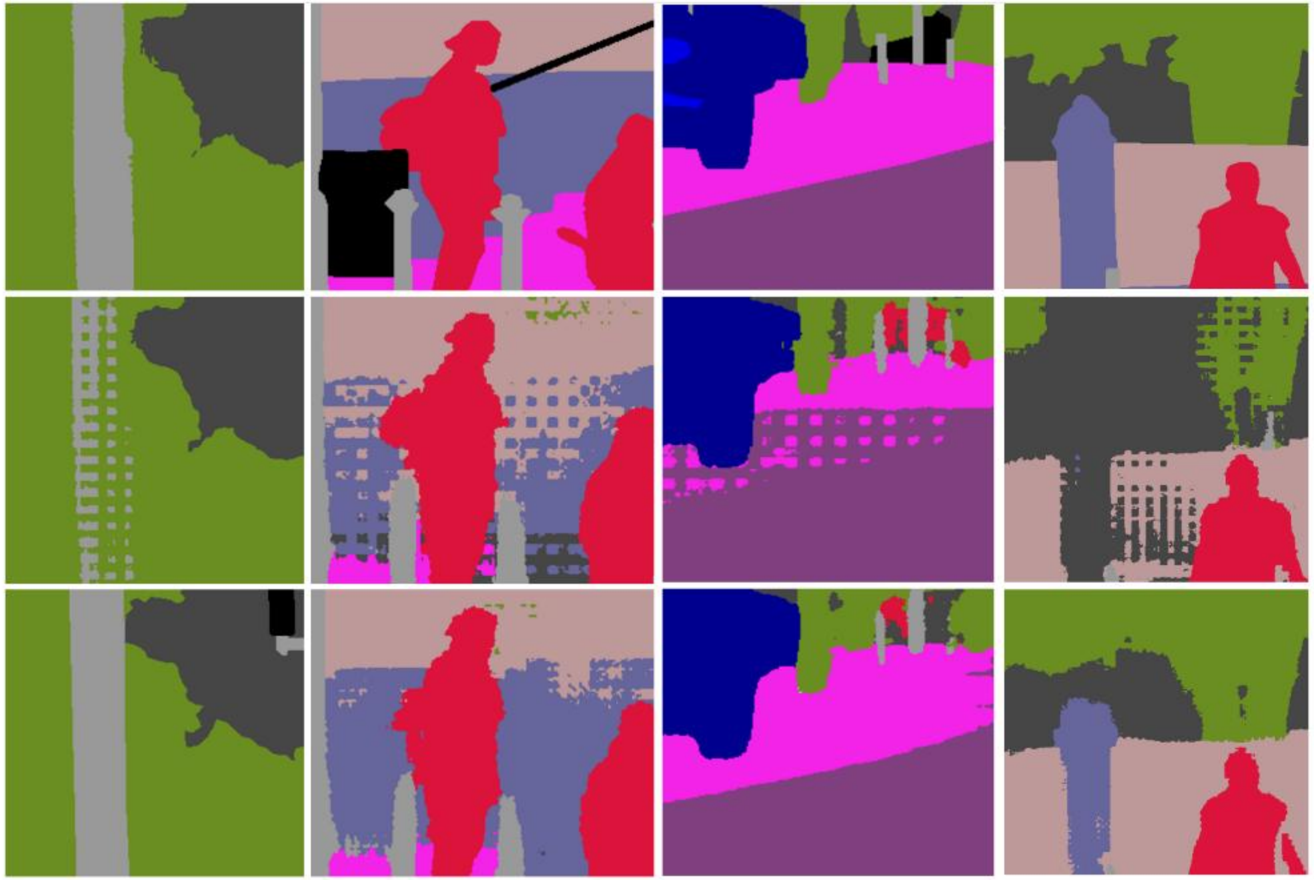
# Method

## First Phase



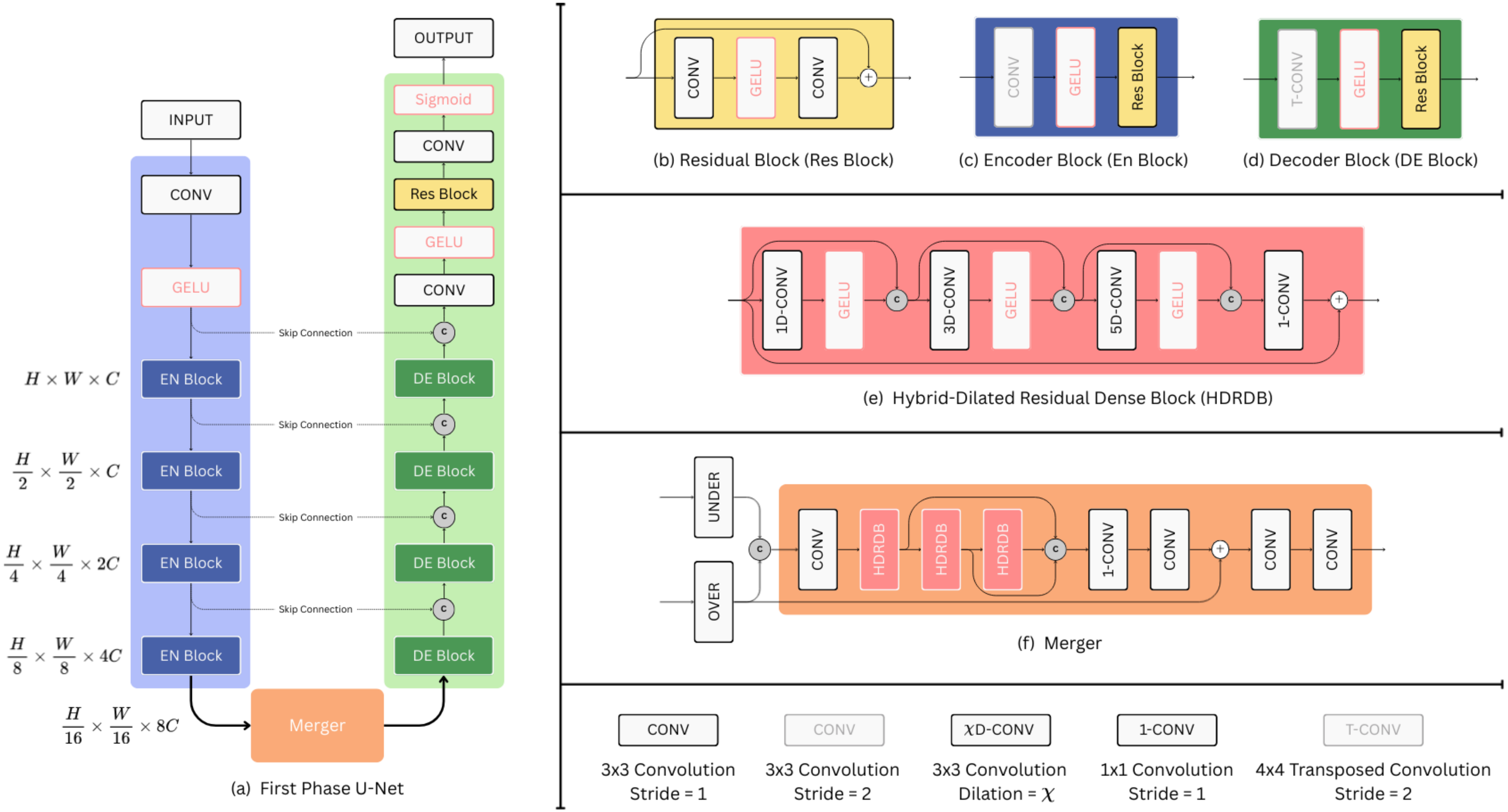
# Method

First Phase



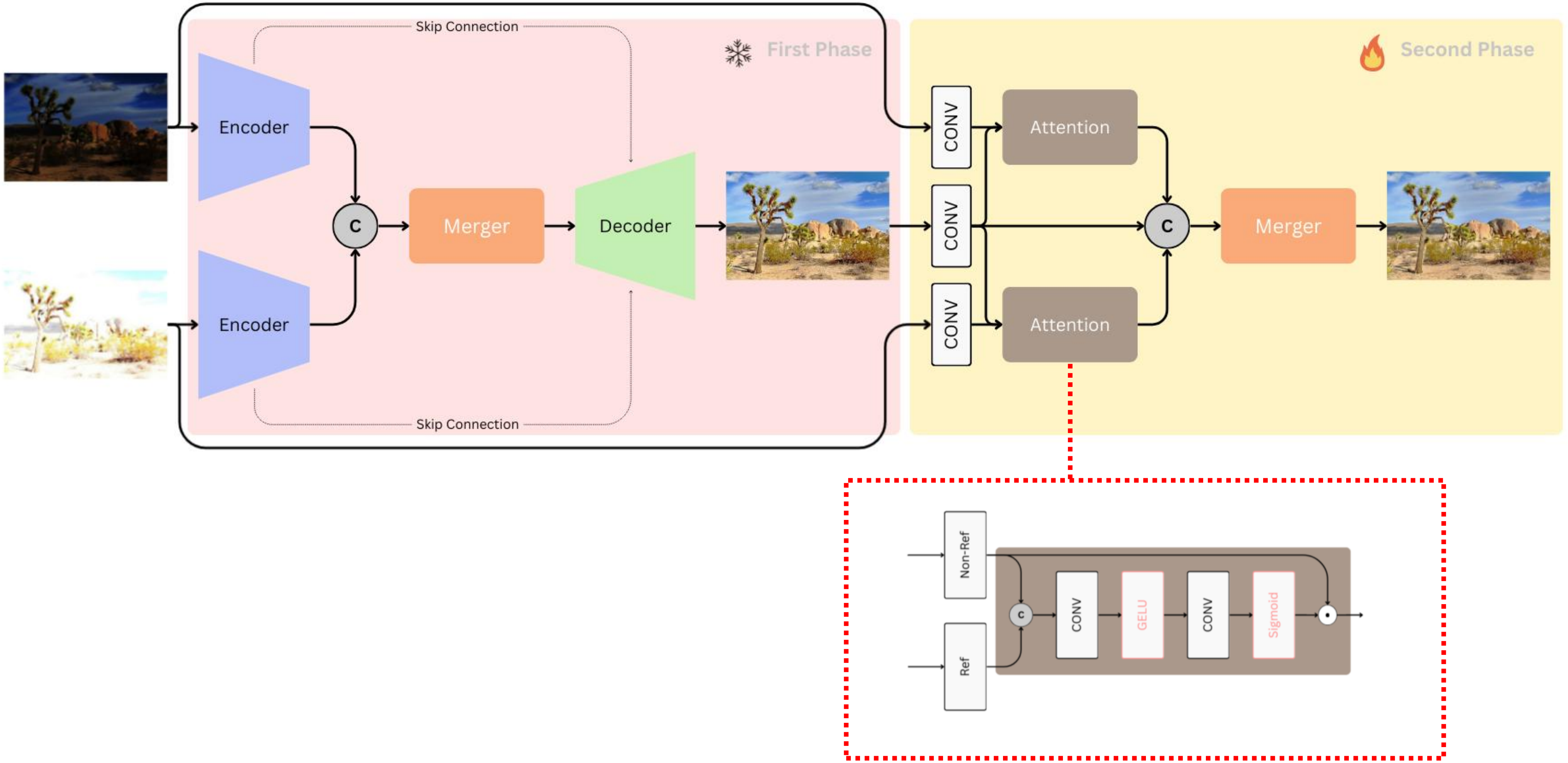
# Method

First Phase



# Method

Second Phase



# Method

Loss Function

- First Phase

$$I_{inter} = \mathrm{UNet}(I_u, I_o)$$

$$\mathcal{L} = \mathcal{L}_{\mathrm{L1}}(I_{inter}, I_{gt}) + \lambda_{vgg}\mathcal{L}_{\mathrm{VGG}}(I_{inter}, I_{gt}) + \lambda_{ssim}\mathcal{L}_{\mathrm{SSIM}}(I_{inter}, I_{gt})$$

# Method

Loss Function

- First Phase

$$I_{inter} = \text{UNet}(I_u, I_o)$$

$$\mathcal{L} = \mathcal{L}_{\text{L1}}(I_{inter}, I_{gt}) + \lambda_{vgg}\mathcal{L}_{\text{VGG}}(I_{inter}, I_{gt}) + \lambda_{ssim}\mathcal{L}_{\text{SSIM}}(I_{inter}, I_{gt})$$

- Second Phase

$$I_{inter} = \text{UNet}_{frozen}(I_u, I_o)$$

$$I_{final} = \text{AMNet}(I_u, I_{inter}, I_o)$$

$$\mathcal{L} = \mathcal{L}_{\text{L1}}(I_{final}, I_{gt}) + \lambda_{vgg}\mathcal{L}_{\text{VGG}}(I_{final}, I_{gt}) + \lambda_{ssim}\mathcal{L}_{\text{SSIM}}(I_{final}, I_{gt})$$

# Experiments

Dataset

- SICE Dataset:
  - Original: 589 sequences
  - After filtering: 453 high-quality sequences
- Ground Truth:
  - 13 fusion algorithms → user votingTraining Set: 360 sequences
  - 13 amateur photographers + 5 volunteers

Our Split: Training 360 sequences + Testing 93 sequences

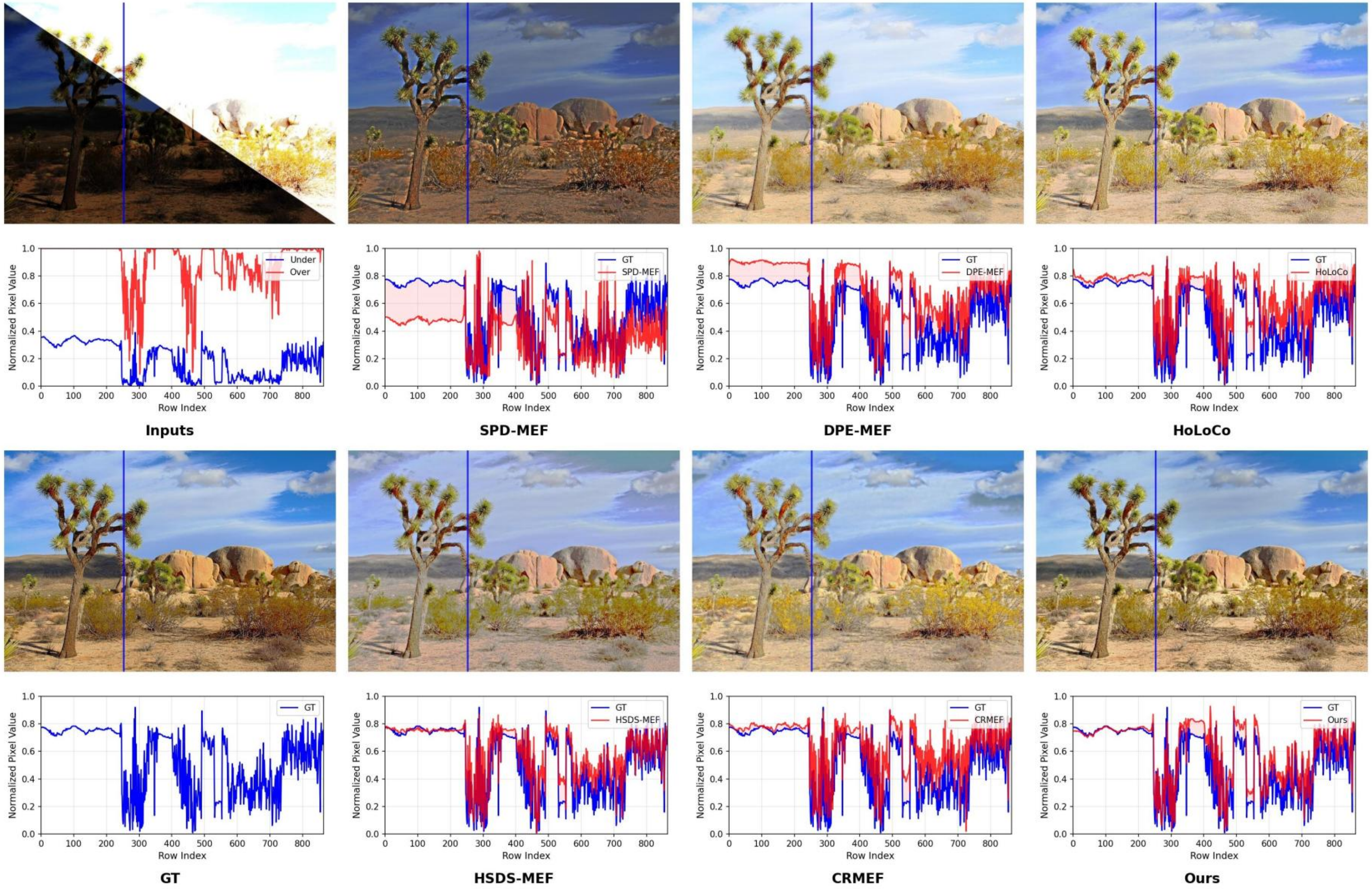
# Experiments

Quantitative

| Methods | PSNR | SSIM | MS-SSIM | VIF |
|---|---|---|---|---|
| SPD-MEF | 17.5891 | 0.7643 | 0.8599 | 0.6294 |
| DPE-MEF | 18.4047 | 0.8065 | 0.8963 | 0.6716 |
| HoLoCo | 20.0829 | 0.8236 | 0.8822 | 0.4678 |
| HSDS-MEF | 20.0444 | 0.7771 | 0.9058 | 0.5661 |
| CRMEF | 20.0893 | 0.8273 | 0.8740 | 0.6570 |
| Ours | 22.9934 | 0.8768 | 0.9359 | 0.7761 |

# Experiments

## Qualitative

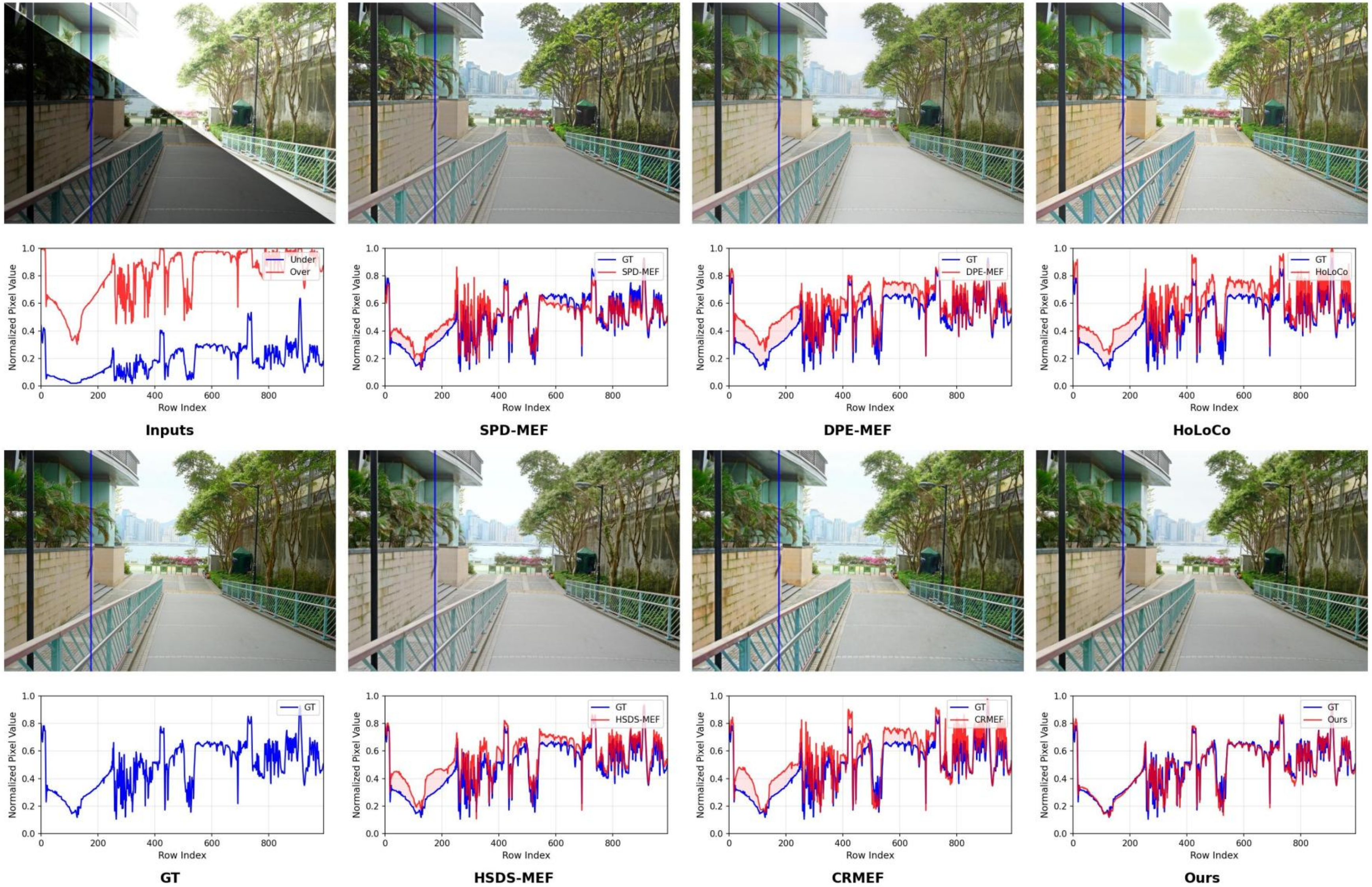


# Experiments

Qualitative

# Experiments

Qualitative

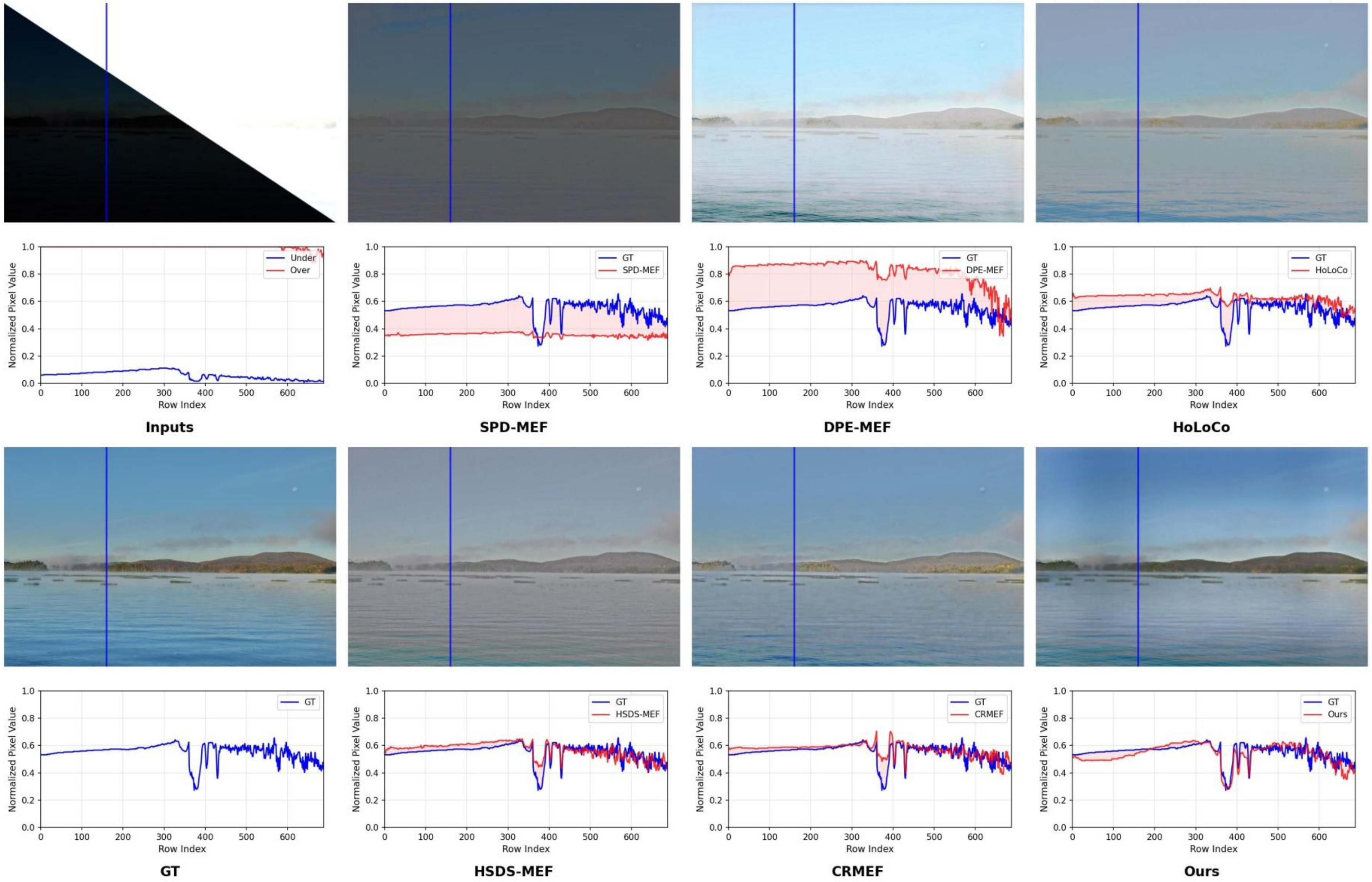


# Experiments

Qualitative

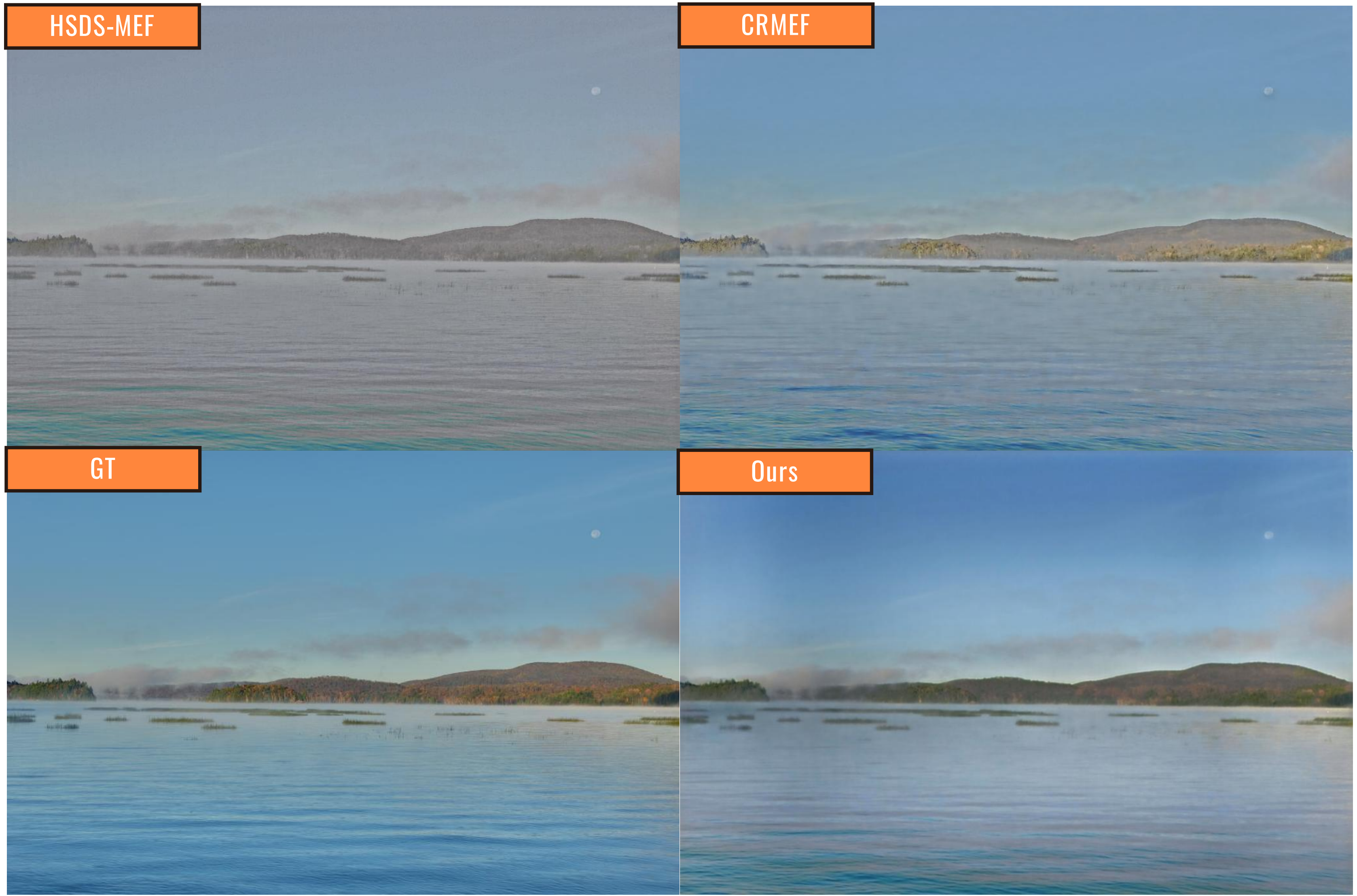
# Experiments

Qualitative



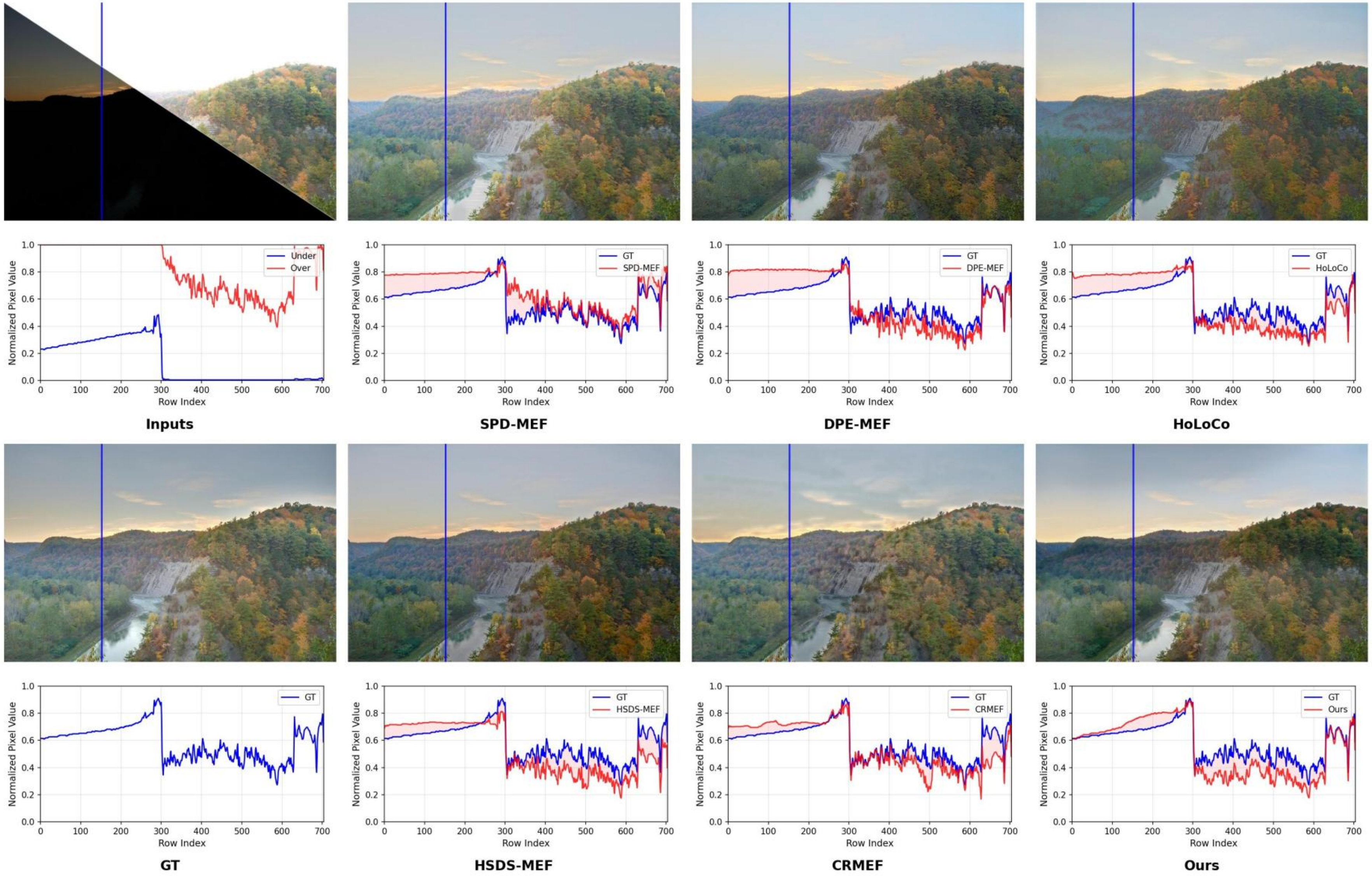
# Experiments

Qualitative



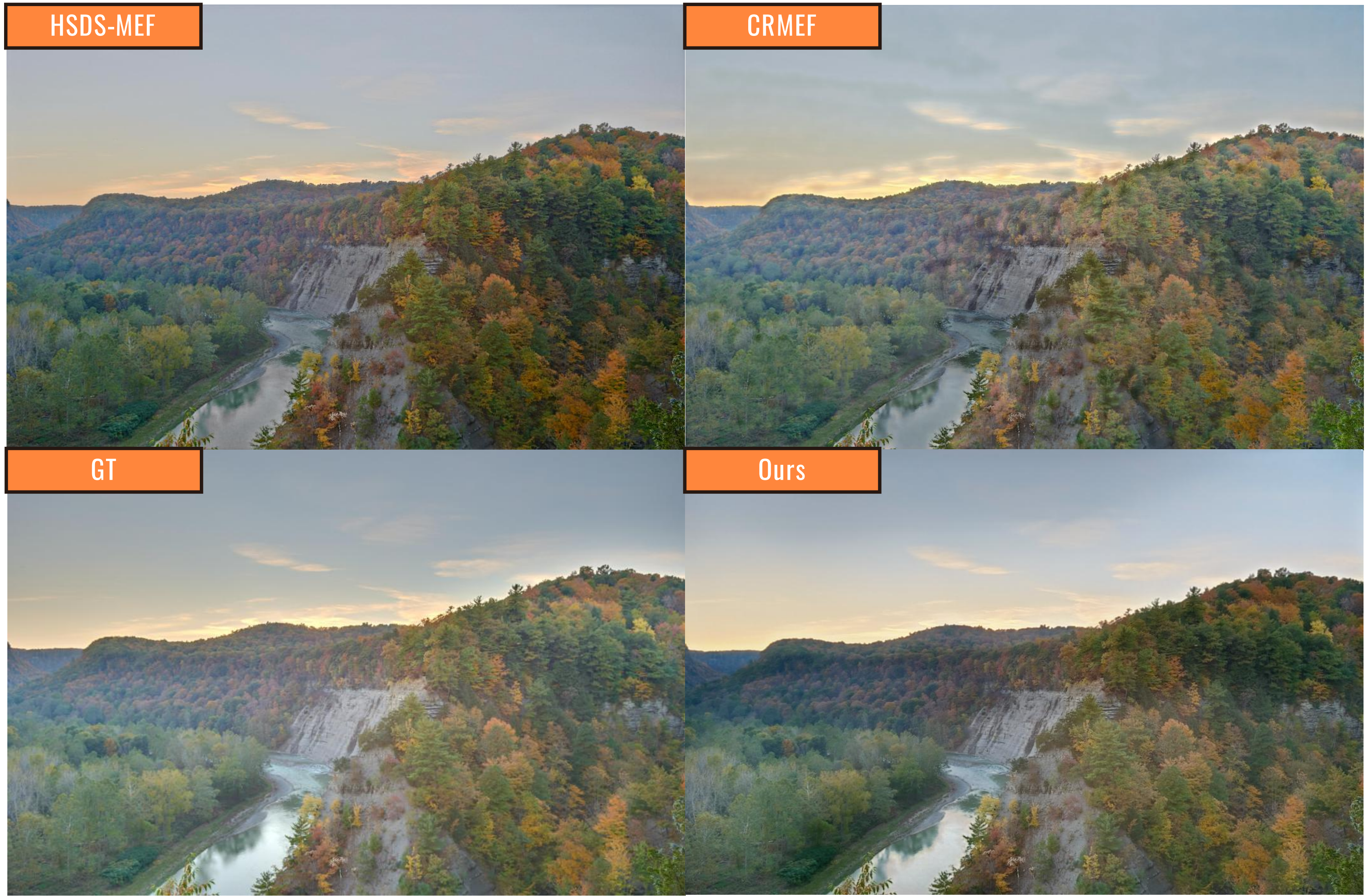
# Experiments

Qualitative



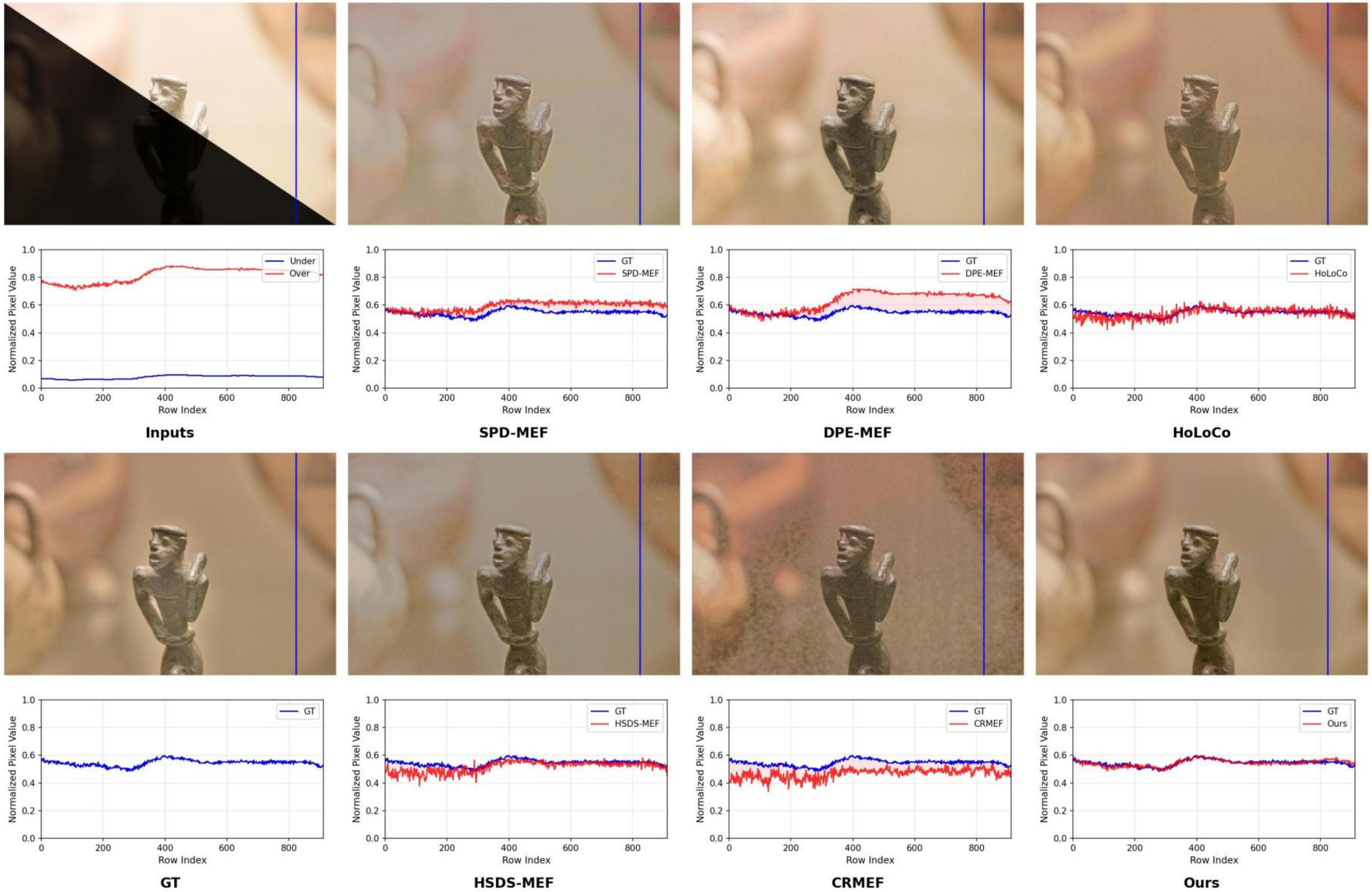
# Experiments

Qualitative



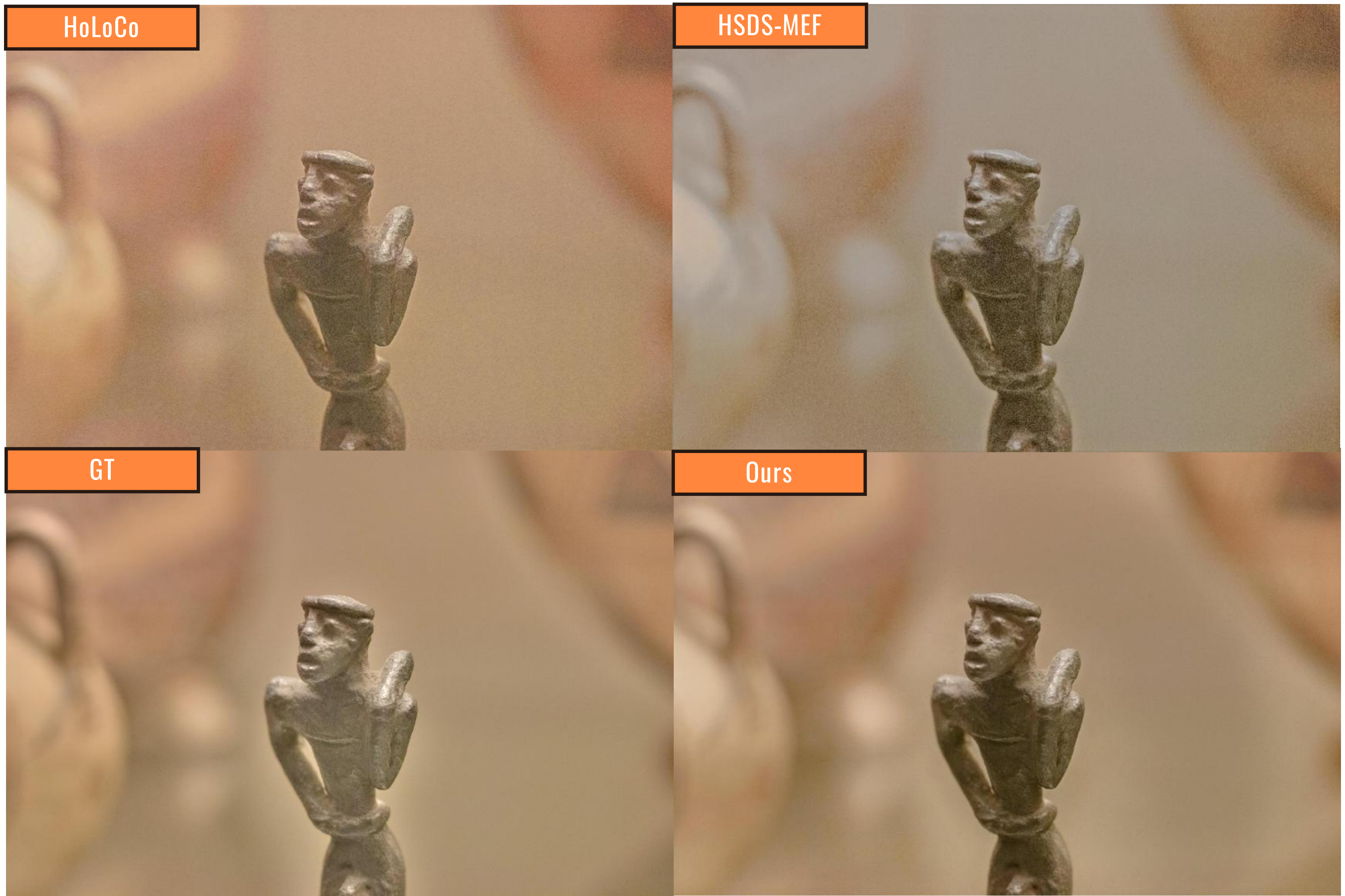
# Experiments

Qualitative



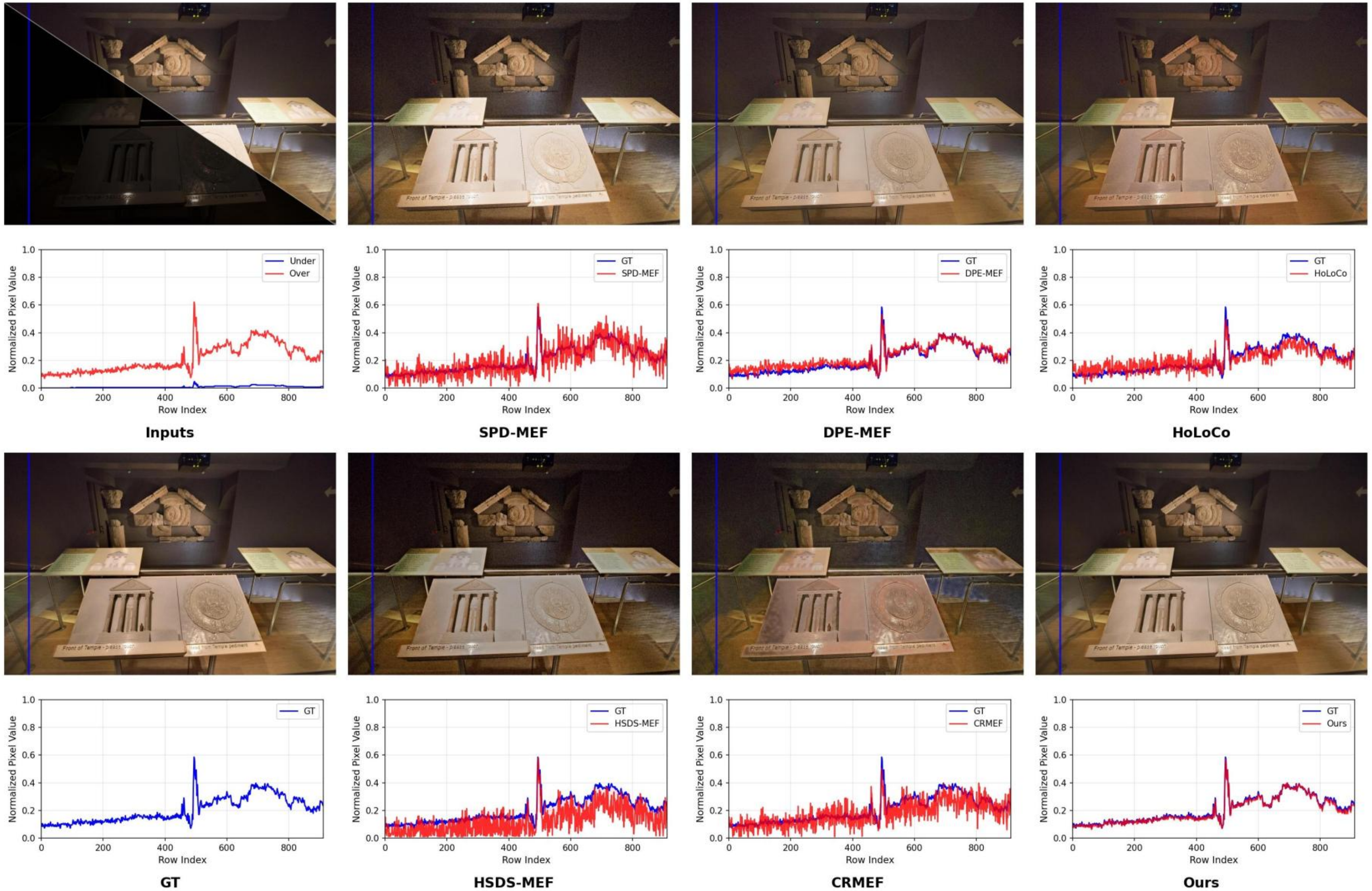
# Experiments

Qualitative



# Experiments

Qualitative



# Experiments

Qualitative

# Conclusion

- Insights
  - Reducing exposure differences improves attention weight computation
  - Decoupled training for specialized optimization
- Current Limitations
  - Static scenes only
- Future Directions:
  - Multi-Phase Progressive Architecture
  - Include motion detection in one of phases