# Gravitated Latent Space Loss Generated by Metric Tensor for High-Dynamic Range Imaging


Heunseung Lim [1], Jungkyoo Shin [2], Hyoungki Choi [1], Dohoon Kim [1], Eunwoo Kim [2], Joonki Paik [1, 2, *]

Department of Image [1], Department of Artificial intelligence [2], Chung-Ang University, Seoul, Korea

hslim@ipis.cau.ac.kr, neo293@cau.ac.kr , Hyunggi120@ipis.cau.ac.kr ,dhkim@ipis.cau.ac.kr, eunwoo@cau.ac.kr, hslim@ipis.cau.ac.kr


## Introduction

Low Dynamic Range (LDR) High Dynamic Range (HDR)

### High-Dynamic Range Imaging

- High Dynamic Range (HDR) imaging seeks to enhance image quality by combining multiple Low Dynamic Range (LDR) images captured at varying exposure levels.
- Traditional deep learning approaches often employ reconstruction loss, but this method can lead to ambiguities in feature space during training.

### Traditional Loss Function

- Mean square error (MSE, $\ell_2$)

$$d^2 = (x \quad y)\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix}$$

- $x$ and $y$ respectively represent the ground truth and predicted images, $d^2$ the distance between these two in the latent space, and unit matrix is a metric tensor.

## Proposed Method

**Goal: Gravitated latent space loss generated by metric tensor for high-dynamic range imaging**

**Contribution**

1. **Incorporating spatial attention in HDR reconstruction from LDR inputs**
2. **Architecture and components of transformer-based U-shape network (TUnet)**
3. **Incorporating curvature in latent space via gravitated latent space loss**

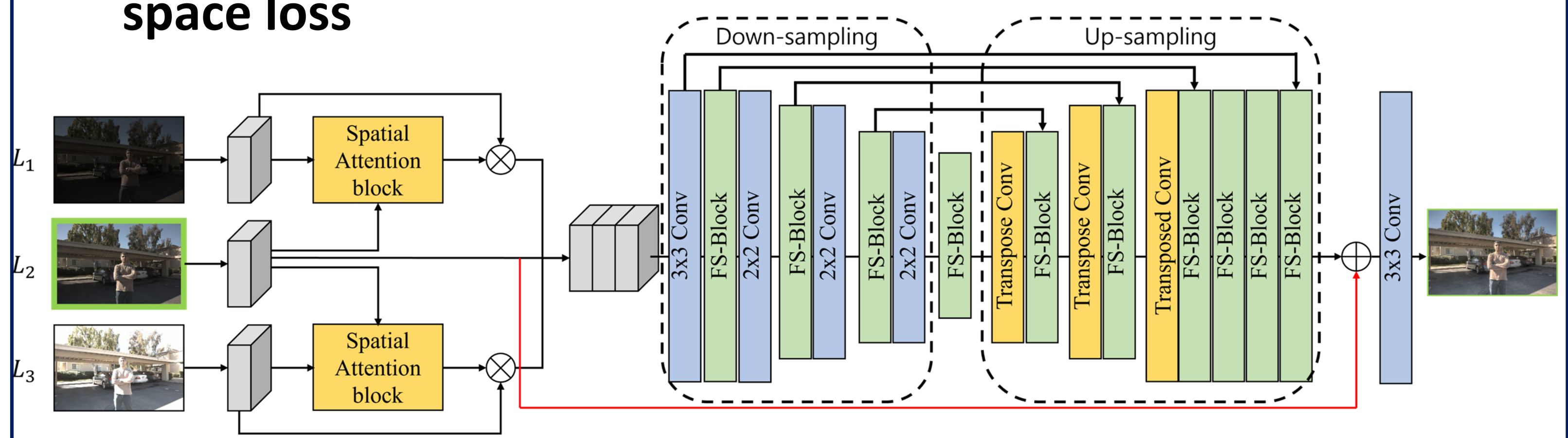



**Fig. 1. Schematic representation of the transformer-based u-shape network (TUnet) used in our proposed method**

### 1. Incorporating spatial attention in HDR reconstruction from LDR inputs

- The proposed TUnet is a feature map estimated in the spatial attention mechanism, and features maps that can minimize the ghosting artifacts occurring in multiple-LDR inputs are input in Fig. 1.

### 2. Architecture and components of transformer-based U-shape network (TUnet)

- Transformer-based u-shape network (TUnet), whose architecture is elaborated in Fig. 1.
- TUnet architecture incorporates various components such as the swin-transformer and residual block. Spectral transform (SpT) component comprises a sequence of operation: Fast Fourier Transform (FFT), convolution, and inverse FFT.

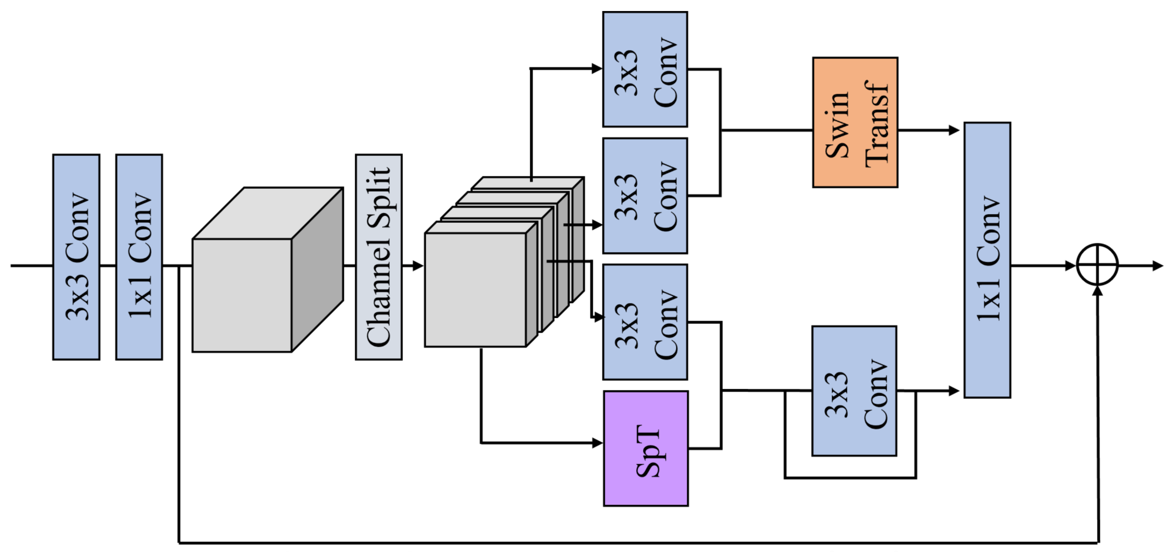



**Fig. 2. Structure of the frequency-spatial (FS) block in Tunet.**

### 3. Incorporating curvature in latent space via gravitated latent space loss

- In the proposed method, we modify the unit matrix as follows, similar to applying the form of "**virtual gravity**" to introduce a curved surface into the latent space by replacing the metric sensor with a learnable parameter.

$$\ell_{GLS} = (T(x) \quad T(y))\begin{pmatrix}\sigma(g_1) & \sigma g_2)\\ \sigma(g_3) & \sigma(g_4)\end{pmatrix}\begin{pmatrix}T(x)\\ T(y)\end{pmatrix}$$

- $\ell_{GLS}$ represents the proposed GLS loss function, $\sigma(\cdot)$ the sigmoid function, $T(\cdot)$ the tone-mapping operator.

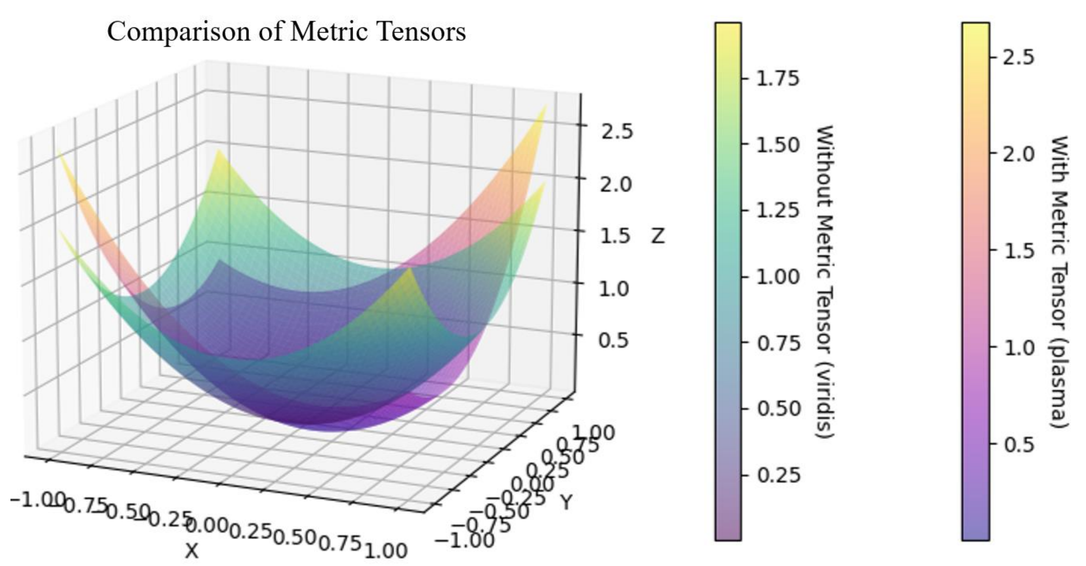
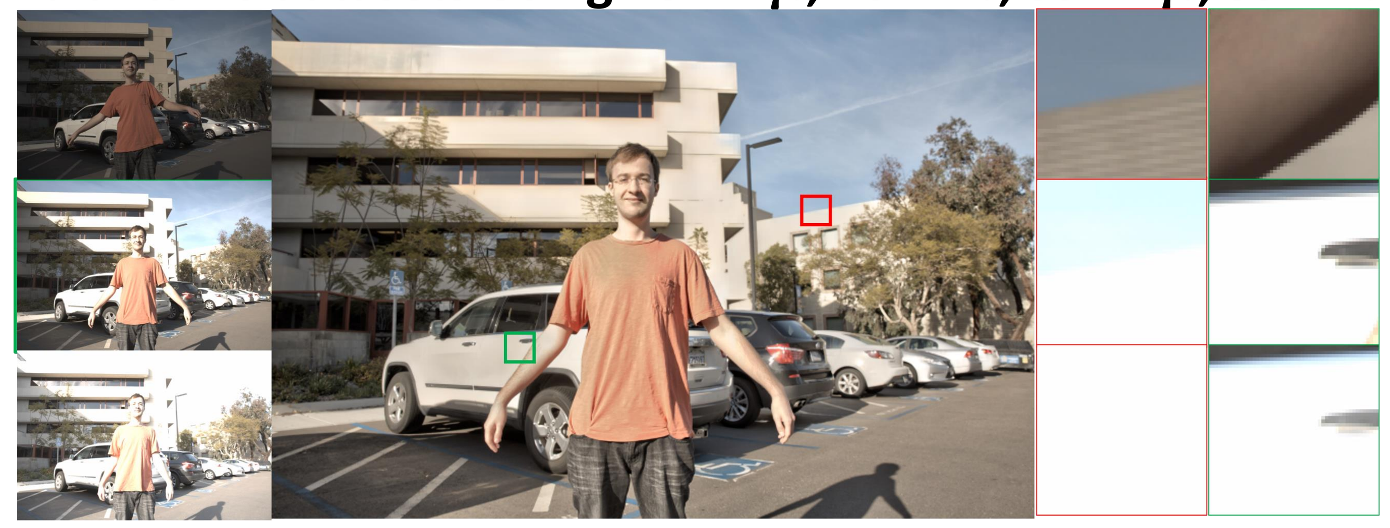


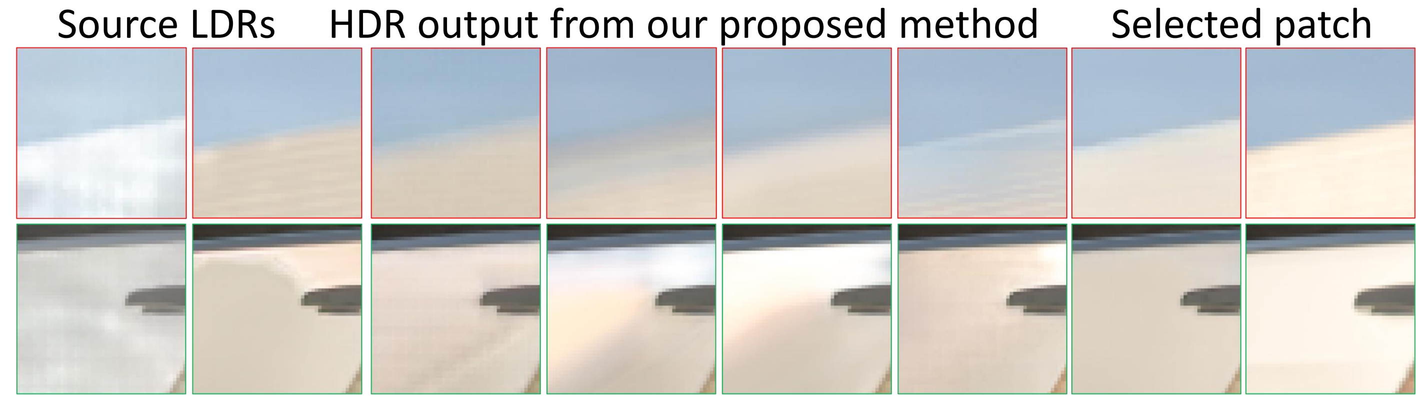

**Fig. 2. Comparative visualization of loss function outcomes with and without Metric tensor GLS. The minimum values observed are 9.2e-5 for the GLS loss and 2.0e-4 for the unit matrix approach.**

## Experimental Results

- **Quantitative and Qualitative Assessment of HDR Imaging Performance on Kalantari's datasets Using PSNR-$\mu$, PSNR-$l$, SSIM-$\mu$, SSIM-$l$**



Source LDRs HDR output from our proposed method Selected patch



TMO Kalantari DeepHDR AHDRNet HDR-GAN CV-ViT Our GT

| Methods | PSNR-$\mu$ | SSIM-$\mu$ | PSNR-$l$ | SSIM-$l$ |
|---|---|---|---|---|
| Kalantari | 42.83 | 0.9877 | 41.49 | 0.9858 |
| Prabhakar | 41.95 | 0.9873 | 41.82 | 0.9876 |
| HDR-GAN | 43.92 | 0.9905 | 41.57 | 0.9865 |
| CA-ViT | 44.32 | 0.9916 | 42.18 | 0.9884 |
| AHDRNet | 43.63 | 0.9900 | 41.14 | 0.9702 |
| AHDRNet+$\ell_{1GLS}$ | 44.15 | 0.9916 | 42.18 | 0.9702 |
| **Ours** | **44.46** | **0.9919** | **43.20** | **0.9904** |

- **Ablation study across various GLS loss function**

| | PSNR-$\mu$ | SSIM-$\mu$ | PSNR-$l$ | SSIM-$l$ |
|---|---|---|---|---|
| $\ell_2$ | 44.39 | 0.9916 | 43.00 | 0.9902 |
| $\ell_{2GSL}$ | 44.40 | 0.9919 | 43.02 | 0.9903 |
| $\ell_1$ | 44.42 | 0.9917 | 43.14 | 0.9903 |
| $\ell_{1GLS}$ | 44.46 | 0.9919 | 43.20 | 0.9904 |

$\ell_2$ $\ell_{2GLS}$ $\ell_1$ $\ell_{1GLS}$

**Fig. 3. Comparative Assessment of Metric Tensor**

## Reference


[1] Yan, Qingsen, et al. "Attention-guided network for ghost-free high dynamic range imaging." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.
[2] Niu, Yuzhen, et al. "HDR-GAN: HDR image reconstruction from multi-exposed LDR images with large motions." IEEE Transactions on Image Processing 30 (2021): 3885-3896.
[3] Liu, Zhen, et al. "Ghost-free high dynamic range imaging with context-aware transformer." European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022.