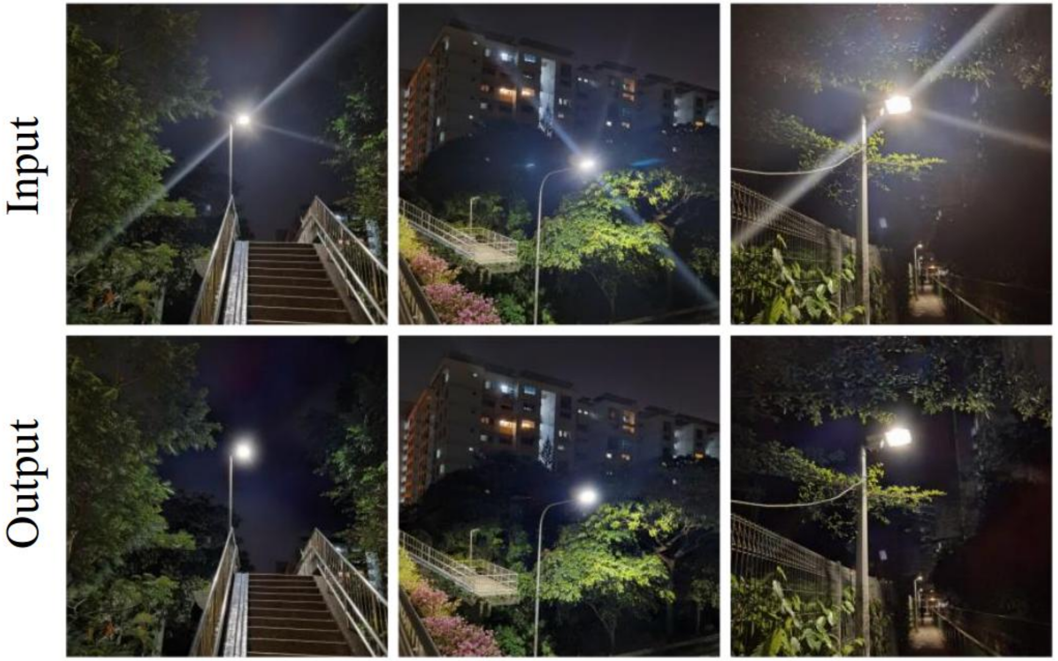
# LPFSformer: Location Prior Guided Frequency and Spatial Interactive Learning for Nighttime Flare Removal

Guang-Yong Chen, Wei Dong, Guodong Fan, Jian-Nan Su,
Min Gan and C. L. Philip Chen

TCSVT, 2024

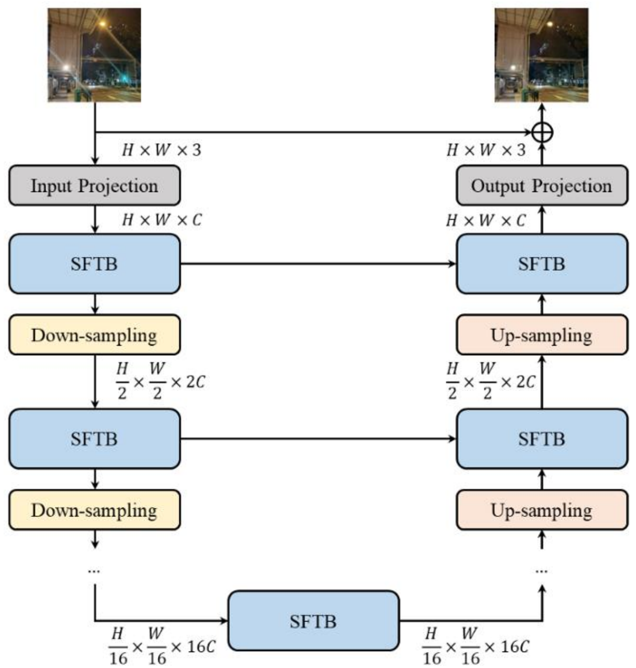
# Nighttime Flare Removal

Lens flare is an optical phenomenon in which intense light is scattered and/or reflected in an optical system. It leaves a radial-shaped bright area and light spots on the captured photo. The effects of flares are more severe in the nighttime environment due to the existence of multiple artificial lights. This phenomenon may lead to low contrast and suppressed details around the light sources, degrading the image's visual quality and the performance of vision algorithms.

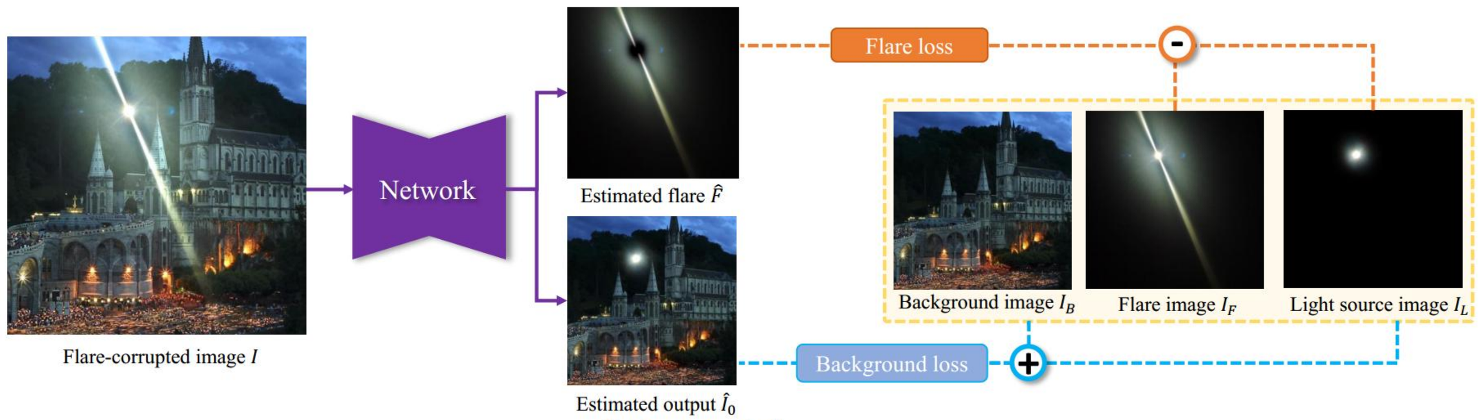


# Motivation

Existing methods often fail to adequately address the differences between flare-affected and non-flare-affected regions when processing images. This oversight can easily lead to incomplete removal or overprocessing of flare regions, as well as erroneous processing of non-flare-affected areas, resulting in discrepancies between the visual quality of the reconstructed images and the actual scenes.



(a) FF-Former



(b) Flare7K++ Pipeline


[1] Dai, Y., Li, C., Zhou, S., Feng, R., Luo, Y., Loy, C.C.: Flare7k++: Mixing synthetic and real datasets for nighttime flare removal and beyond. arXiv preprint arXiv:2306.04236 (2023)
[2] Zhang, D., Ouyang, J., Liu, G., Wang, X., Kong, X., Jin, Z.: Ff-former: Swin fourier transformer for nighttime flare removal. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2824–2832 (2023)

# Motivation


Fig. 1. (a) Real-world flare-corrupted images without ground truth provided by Dai et al.[1]. (b) Visualization of the proposed Flare Location Prior Guidance (LPG), which can identify and effectively localize flare signals. (c) The flare removal effect of the method Flare7K[1] in the field of flare removal. (d) The flare removal effect of the method Flare7K++[5] in the field of flare removal. (e) The flare removal effect of our proposed method. The red box delineates an enlarged area within the image, with arrows indicating the locations of the defects.

# Contributions

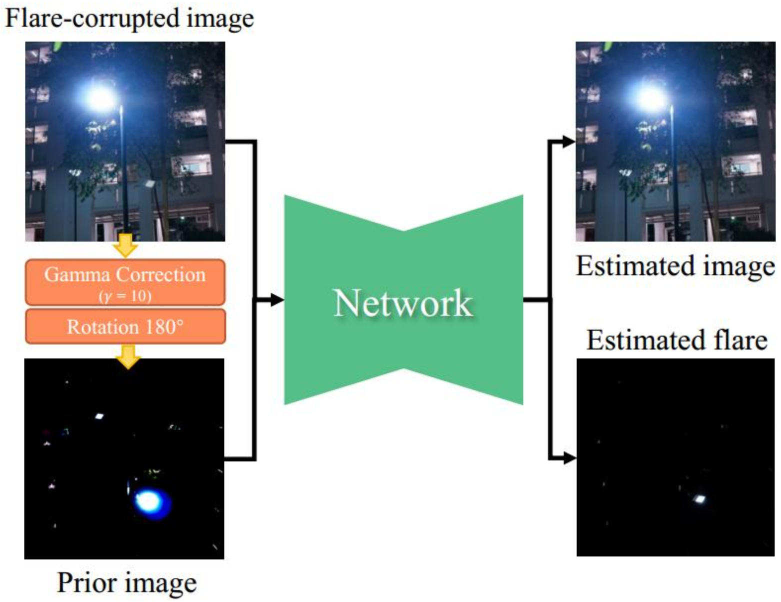
- **Flare Location Prior Guidance**: we considered the critical impact of differentiating between flare-affected and nonflare-affected regions on image reconstruction performance, and subsequently propose Location Prior Guidance (LPG). This method integrates learned flare location information into the network through Location Prior Injection (LPI), guiding the model to effectively focus on flare regions within the image and significantly enhancing flare removal performance.
- **Frequency and Spatial Domain Interaction Learning**: To integrate the flare location information learned by LPG, we introduce the LPFSformer. The designed FCB and STB modules enable the LPFSformer to effectively perceive and distinguish flare features within images, thereby allowing targeted processing of flare-affected regions, which is crucial for flare removal.
- **SOTA Performance**: Extensive benchmark experiments demonstrate that our method outperforms existing stateof-the-art methods in terms of performance and shows excellent generalization ability in real-world scenarios.

# Proposed Method

## Task-specific priors

Dai et al. introduced a novel optical center symmetric prior, specifically designed for reflective flare. However, it relies on the assumption that most smartphone cameras satisfy the optical center symmetry prior, a condition that does not always apply to certain professional cameras.

In contrast, our LPG is specifically designed to address scattered flare. It locates scattered flare by capturing flare features within the image, enabling effective localization without the constraints of the optical center symmetry prior. As a result, the LPG is more flexible and robust in real-world scenarios.





[1] Dai, Y., Luo, Y., Zhou, S., Li, C., Loy, C.C.: Nighttime smartphone reflective flare removal using optical center symmetry prior. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 20783–20791 (2023)

# Proposed Method

## Overall pipeline

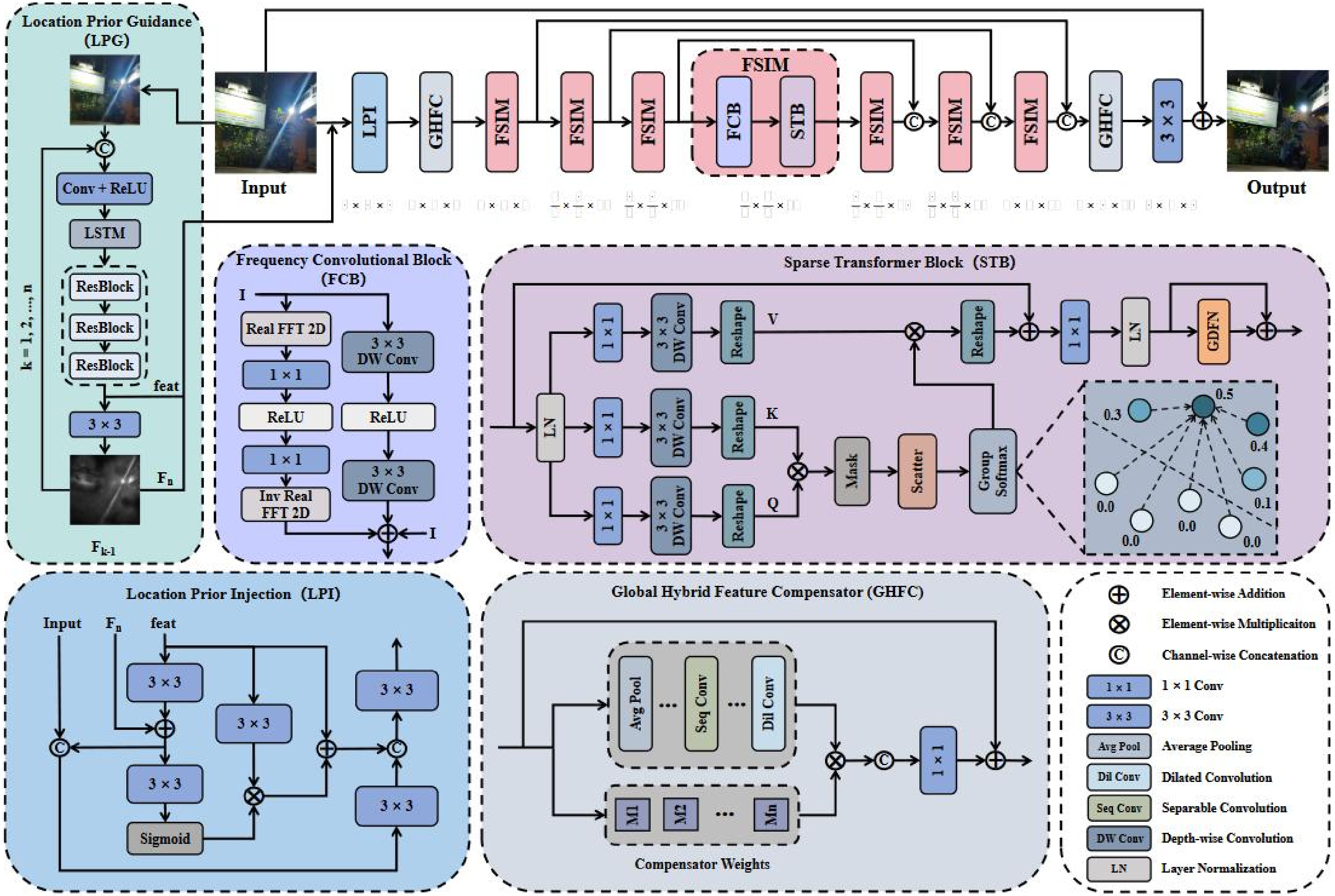
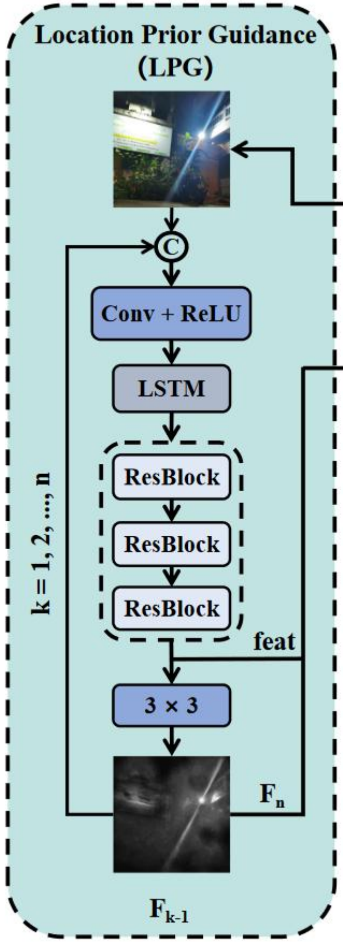



Fig. 2. The overall architecture of the proposed Transformer network for flare removal in night-time images (LPFSformer). It primarily includes: (1) Location Prior Guidance (LPG), (2) Location Prior Injection Module (LPI), (3) Coupled Module for Joint Learning in Frequency and Spatial Domains (FSIM), which consists of Frequency Convolution Block (FCB) and Sparse Transformer Block (STB), and (4) Global Hybrid Feature Compensator (GHFC).

# Proposed Method

## The Location Prior Guidance (LPG)




We designed a ResBlock combination for flare feature extraction. Additionally, we introduced recursive expansion in the ResBlock stages, enabling the progressive learning of flare information while reducing network parameters. The LPG network consists of four components: (1) a convolutional layer $f_{in}$, (2) a convolutional Long Short-Term Memory (LSTM) layer as a recurrent layer $f_{recurrent}$, (3) multiple residual blocks $f_{res}$, and (4) a convolutional block $f_{out}$.

$$\overline{\mathbf{F}}_{ks} = f_{in}(I, \overline{\mathbf{F}}_{k-1}), \quad k = 1, \ldots, N,$$

$$\overline{\mathbf{S}}_{k} = f_{recurrent}(\overline{\mathbf{S}}_{k-1}, \overline{\mathbf{F}}_{ks}),$$

$$\overline{\mathbf{F}}_{k} = f_{out}(f_{res}(\overline{\mathbf{S}}_{k})),$$

# Proposed Method

## Frequency and Spatial Domain Interaction Learning

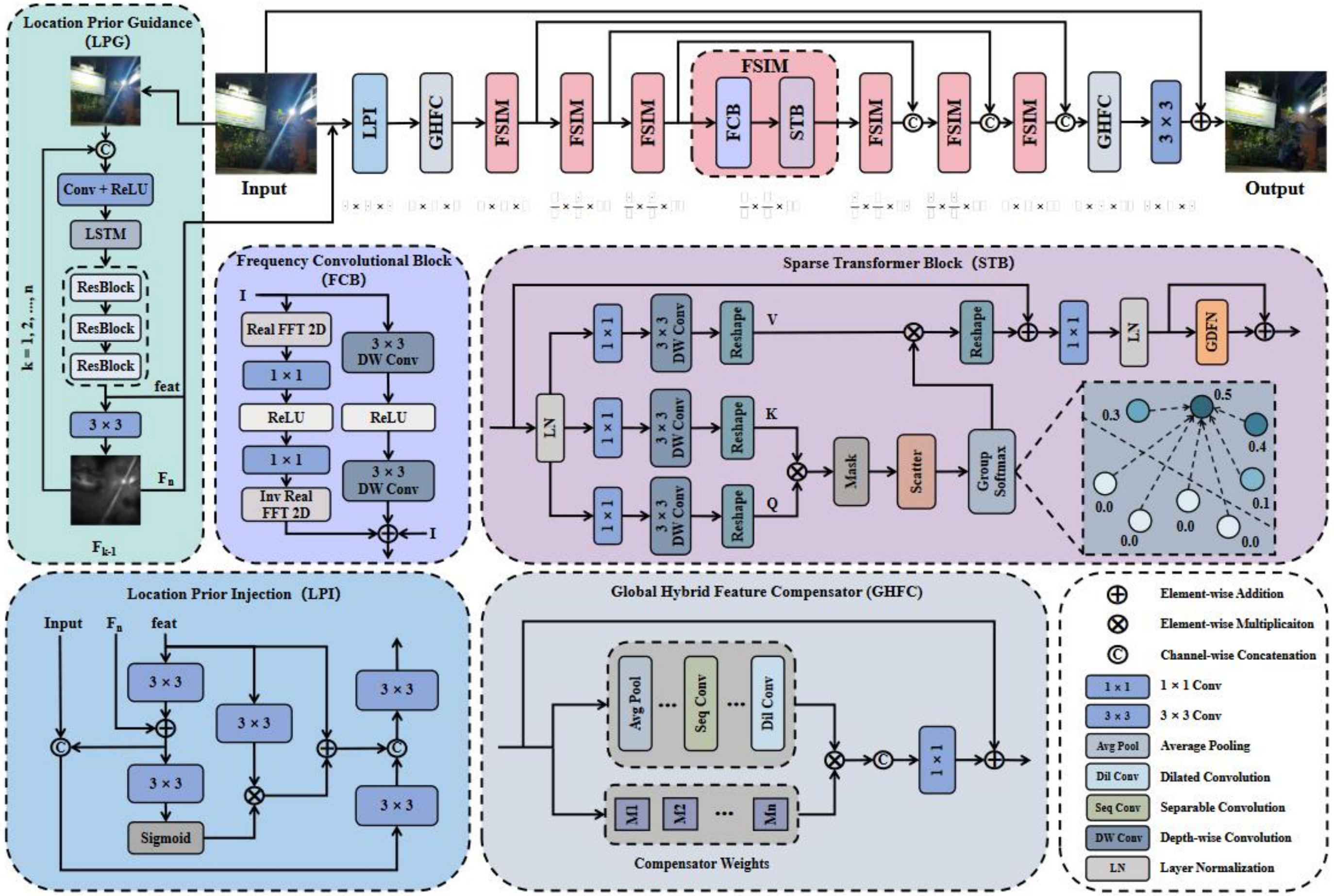



Fig. 2. The overall architecture of the proposed Transformer network for flare removal in night-time images (LPFSformer). It primarily includes: (1) Location Prior Guidance (LPG), (2) Location Prior Injection Module (LPI), (3) Coupled Module for Joint Learning in Frequency and Spatial Domains (FSIM), which consists of Frequency Convolution Block (FCB) and Sparse Transformer Block (STB), and (4) Global Hybrid Feature Compensator (GHFC).

### The Location Prior Injection (LPI)

When the flare color is similar to the background tone, directly integrating LPG priors often results in suboptimal removal of flare artifacts. This is because, the contrast between flare-affected and non-flare-affected regions is insufficiently pronounced. Inspired by this observation, we specifically designed LPI. LPI further updates the prior knowledge in a residual manner with attention weights then generated based on the updated prior knowledge.

# Proposed Method

## Frequency and Spatial Domain Interaction Learning

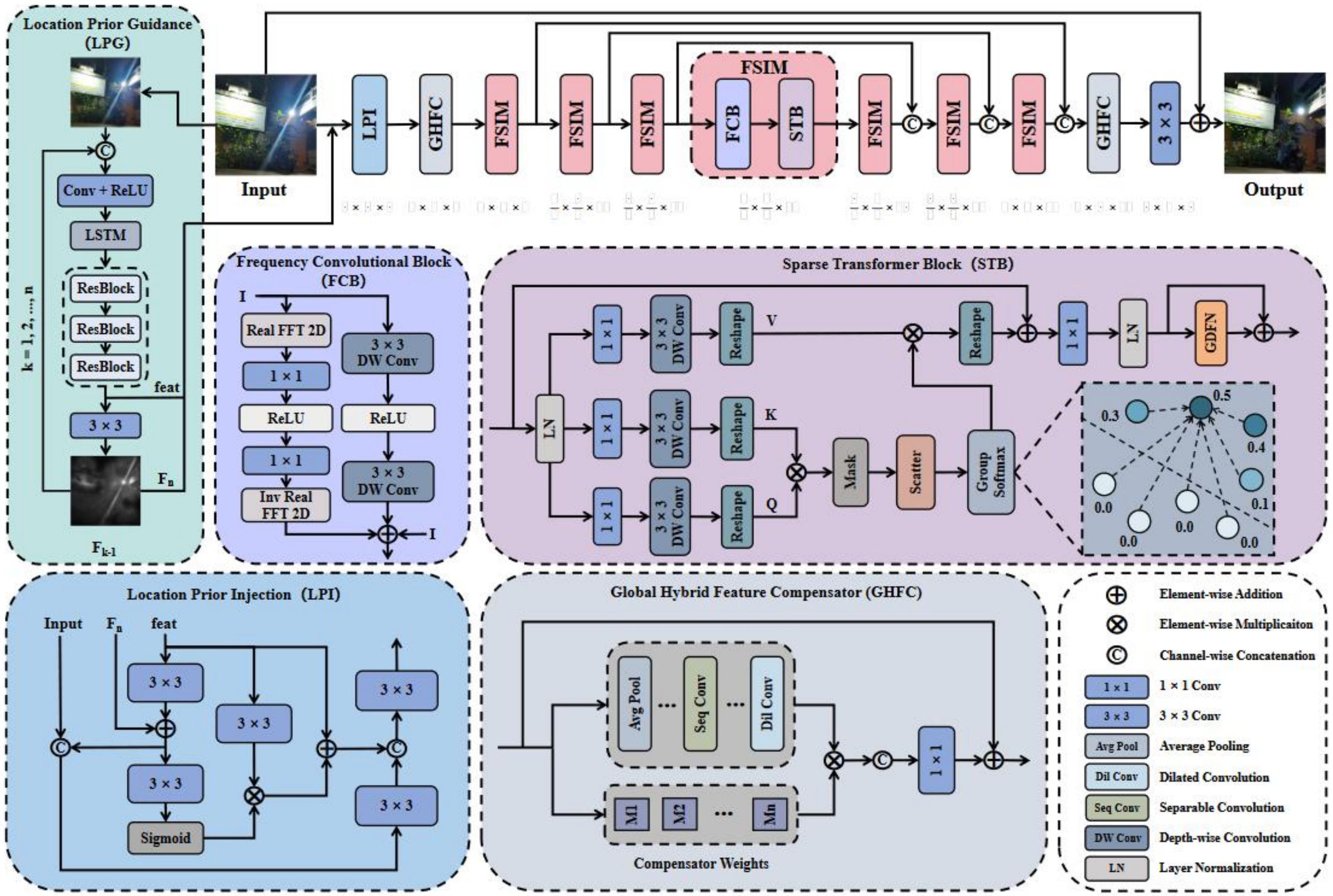



Fig. 2. The overall architecture of the proposed Transformer network for flare removal in night-time images (LPFSformer). It primarily includes: (1) Location Prior Guidance (LPG), (2) Location Prior Injection Module (LPI), (3) Coupled Module for Joint Learning in Frequency and Spatial Domains (FSIM), which consists of Frequency Convolution Block (FCB) and Sparse Transformer Block (STB), and (4) Global Hybrid Feature Compensator (GHFC).

## The Frequency-Space Interaction Module (FSIM)

**Frequency Convolution Block (FCB):**

The introduction of the FCB can effectively enhance the frequency representation of weak flare-affected regions in the image, thereby improving the network's locality in the frequency domain.

**Sparse Transformer Block (STB):**

Due to the standard Transformer's use of all tokens to globally compute self-attention, it may involve noisy interactions between unrelated features, which is not ideal for nighttime flare removal tasks. The STB can focus on the most relevant non-local information at each scale, enabling more accurate representations.

# Proposed Method

## Training details

In the first phase, we train the proposed the LPG by cropping the input images to 384 × 384. The location priors learned by LPG are compared with grayscale images of pure flare provided by the Flare7K++ synthetic pipeline, and optimization is performed using the $L_1$ loss function. After training, the obtained prior is used as the image preprocessing module for the second-phase flare removal model.

In the second phase, we train the flare removal model, LPFSformer, with images cropped to 384 × 384 × 3, performing 600K iterations on the 24K Flickr dataset. We combine reconstruction loss $L_1$ and perceptual loss $L_{per}$ for joint supervision during training. The final loss function is expressed as:

$$\mathcal{L} = \omega_1 * L_1 + \omega_2 * L_{per}.$$

setting the weights $\omega_i$ of both the reconstruction loss $L_1$ and the perceptual loss $L_{per}$ to 0.5.

# Experiments

TABLE I
QUANTITATIVE RESULTS ON REAL-WORLD NIGHTTIME FLARE IMAGES. "*" DENOTES MODELS WITH REDUCED PARAMETERS DUE TO THE LIMITED GPUMEMORY. THE BEST RESULT IS HIGHLIGHTED IN RED AND THE SECOND BEST RESULT IS HIGHLIGHTED IN BLUE.

| Dataset | | Flare7K: Real_test | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Metrics | | PSNR↑ | SSIM↑ | LPIPS↓ | G-PSNR↑ | S-PSNR↑ | Params (M) | MACs (G) |
| Input | | 22.561 | 0.857 | 0.0777 | 19.556 | 13.105 | - | - |
| Previous Data Synthesis Pipelines | Sharma [71] | 20.492 | 0.826 | 0.1115 | 17.790 | 12.648 | 22.365 | 285.12 |
| | Wu [29] | 24.613 | 0.871 | 0.0598 | 21.772 | 16.728 | 34.526 | 261.901 |
| | Flare7K [1] | 26.978 | 0.890 | 0.0466 | 23.507 | 21.563 | 20.429 | 159.643 |
| Flare7K++ Pipeline[5] | U-net [16] | 27.189 | 0.894 | 0.0452 | 23.527 | 22.647 | 34.527 | 261.953 |
| | HINet [17] | 27.548 | 0.892 | 0.0464 | 24.081 | 22.907 | 88.674 | 685.127 |
| | MPRNet* [18] | 27.036 | 0.893 | 0.0481 | 23.490 | 22.267 | 3.642 | 567.187 |
| | Restormer* [19] | 27.597 | 0.897 | 0.0447 | 23.828 | 22.452 | 2.981 | 57.975 |
| | Uformer [20] | 27.633 | 0.894 | 0.0428 | 23.949 | 22.603 | 20.429 | 159.643 |
| | InternImage [72] | 27.432 | 0.892 | 0.0488 | 23.623 | 22.420 | 59.015 | 944.246 |
| | GRL [73] | 27.642 | 0.895 | 0.0475 | 23.853 | 22.523 | 19.775 | 1108.47 |
| | Kotp and Torki [33] | 27.662 | 0.897 | 0.0422 | 23.987 | 22.847 | 129.306 | 271.419 |
| | LPFSformer(Ours) | 28.238 | 0.905 | 0.0422 | 24.793 | 23.876 | 13.733 | 525.442 |

# Experiments

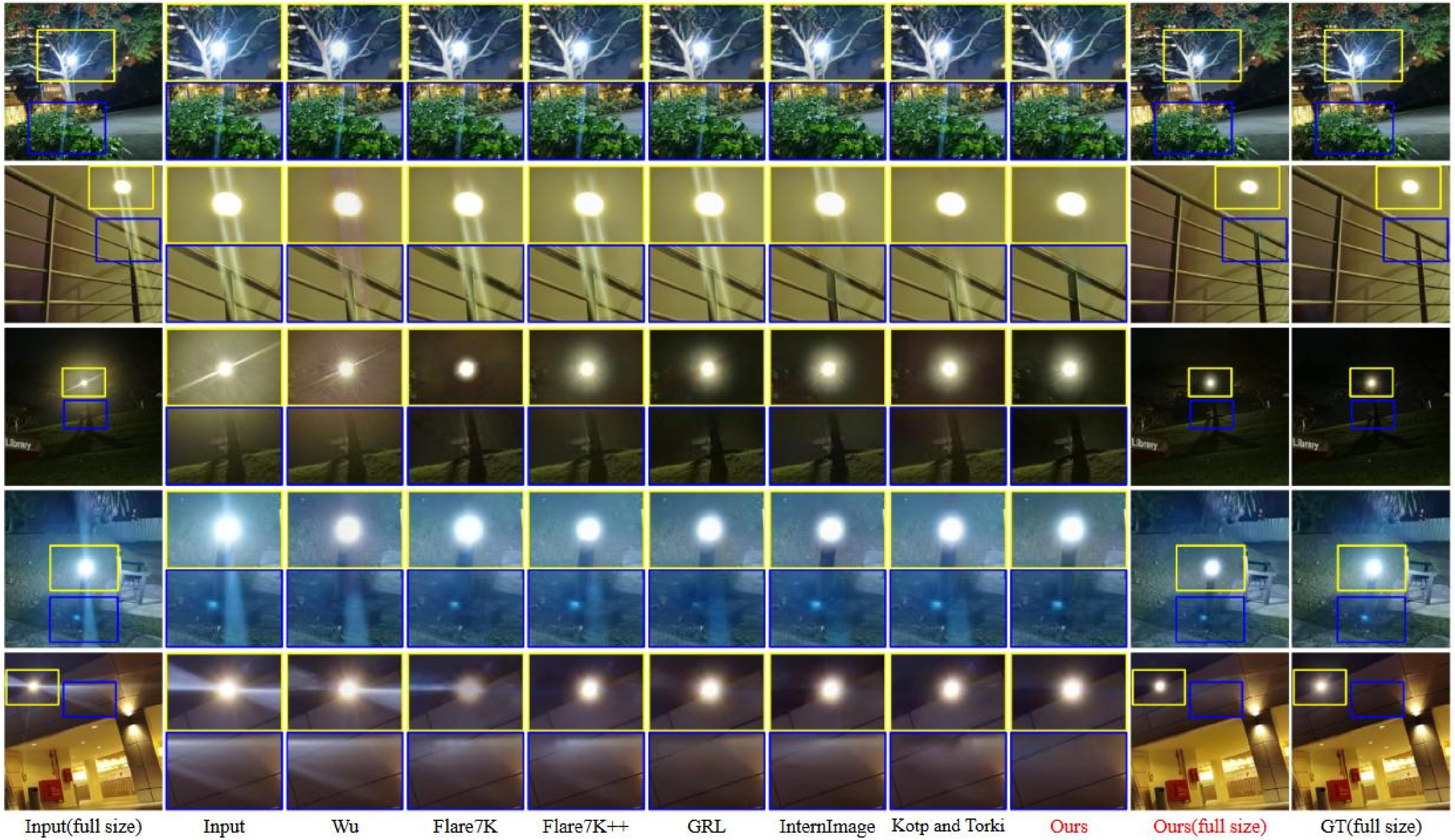



Fig. 3. Visual Quality Comparison on the Flare7K++ Real Dataset. We selected the high-performing Uformer[20] as the baseline network for both Flare7K[1] and Flare7K++[5] datasets. Additionally, we introduced InternImage[72], GRL[73] and Kotp and Torki[33], trained using the Flare7K++ pipeline, for a comprehensive comparison. The colored box delineates the enlarged area corresponding to the region within the image.

# Experiments

## Ablation Study

TABLE II
PERFORMANCE ANALYSIS OF ABLATION EXPERIMENTS FOR LPG AND THE EFFECTIVE FLARE REMOVAL MODULES. THE BEST RESULT IS HIGHLIGHTED.

| Model | PSNR↑ | SSIM↑ | LPIPS↓ | G-PSNR↑ | S-PSNR↑ |
|---|---|---|---|---|---|
| W/o LPG | 28.131 | 0.900 | 0.0433 | 24.492 | 23.492 |
| W/o LPI | 28.183 | 0.904 | 0.0423 | 24.744 | 23.837 |
| W/o FCB | 27.950 | 0.903 | 0.0423 | 24.315 | 23.541 |
| W/o GHFC | 28.103 | 0.903 | 0.0424 | 24.512 | 23.316 |
| LPFSformer | **28.238** | **0.905** | **0.0422** | **24.793** | **23.876** |

Input LPG W/o LPG W/ LPG

Fig. 6. Visual results of ablation for LPG. The LPG localization results presented in grayscale.