Supplementary Material. Semantic Pre-Supplement for Exposure Correction

Zhen Zou
University of Science and Technology of China
He Fei, China
zouzhen@mail.ustc.edu.cn

Wei Yu
University of Science and Technology of China
He Fei, China
patrick914y@mail.ustc.edu.cn

Jie Huang
University of Science and Technology of China
He Fei, China
hj0117@mail.ustc.edu.cn

Feng Zhao
University of Science and Technology of China
He Fei, China
fzhao956@ustc.edu.cn

1. Inpainting model

Li et al. [2] find that the shadow removal network pre-trained on the image inpainting dataset can remove shadows in images very well. Considering the similarity between shadow removal and repair of severely exposed areas (especially underexposure), we believe that this strategy can also be well applied in the scenes of this paper.

A simple framework based on the encoder-decoder and feature fusion module as shown in Fig. 1 is proposed in [2]. This paper uses the framework as our inpainting model. The encoders ϕ and ψ are the same. They are composed of three blocks consisting of a convolution layer followed by a ReLU layer. The extracted features are predicted by a convolution layer to get the fusion weights. The fusion weights are subjected to sigmoid operation and then fused with the features to obtain the fused features. The fused features are stacked and convolved before being sent to the decoder. The decoder consists of 8 Resnet blocks and 3 transposed convolution layers. The decoder outputs the final recovery result.

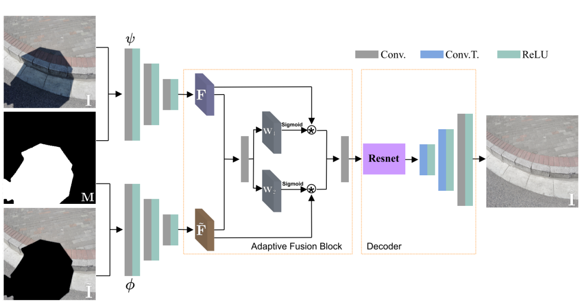


Figure 1. The framework of inpainting model.

2. FECNet

Huang et al. [1] proposed FECNet, which consists of an amplitude sub-network and a phase sub-network to gradually reconstruct the representation of brightness and structure components. This reconstruction order is consistent with our claim that the inpainting module provides sufficient semantic information to guide color transformation.

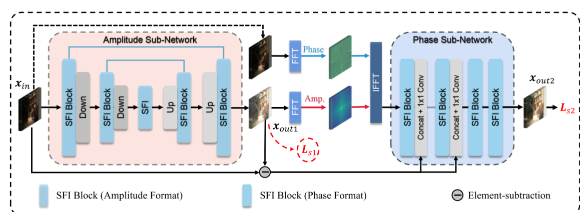


Figure 2. The framework of the complete FECNet. An amplitude sub-network and a phase sub-network progressively reconstruct the representation of lightness and structure components. Parts other than the amplitude sub-network are omitted in the paper.

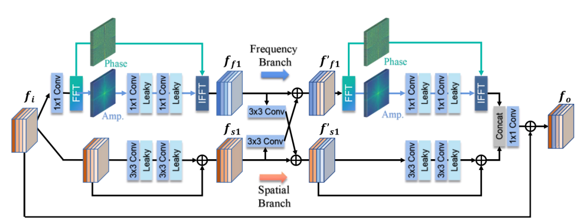


Figure 3. Amplitude format of the SFI block, which can be transformed into its phase form by simply switching the positions of the amplitude spectrum and the phase spectrum. The frequency branch interacts with the space branch in parallel, and the frequency branch processes the amplitude component and bypasses the phase component.

The framework of FECNet is shown in Fig. 2, and some details are simplified in the article to highlight the key points. The Spatial-Frequency Interaction (SFI) block is shown in Fig. 3.

3. Reviews and Rebuttals

3.1. Reviewer1

Review: This paper proposes inpainting image guidance to train an exposure correction model. Authors claim that the input image of overexposed or underexposed regions loses the texture or semantic structure. It loses the information to enhance the detail or contrast of the exposure-corrected image. So they propose an inpainted image as the other input for training the other twin model and distilled to the original model from the twin model to guide the encoder-decoder feature of the model. It is an interesting idea and a reasonable framework to realize their concept. The experimental evaluation shows the method's advantage and realizes the SOTA performance supported by their concept of the idea. One thing, I would like to see is how much improvement just comes from the masked area and whether there are the side effects such as additional artifacts due to the inpainting guidance. That kind of detailed quantitative evaluation and qualitative evaluation would be also interesting to see.
Rebuttal: Thank you for your comments and suggestions. We provide the responses below.

1. Yes, there are some unexpected artifacts. Examples in Fig. 5 and 9 in the main body show that it's inevitable. We address this issue by leveraging mask area mixing and the refining capability of subsequent networks.
2. Compared with FECNet, the effectiveness of the inpainting strategy is demonstrated by the great quantitative and qualitative improvement, which comes from the good reasoning ability displayed in the mask area.

3.2. Reviewer3

Review:
Strengths:

1. This paper integrates pre-structural repair using an inpainting model to help the exposure correction.
2. The method effectively leverages contextual semantic information for image restoration, particularly in severely exposed areas.
3. The authors conduct extensive experiments across multiple datasets to demonstrate the effectiveness and superiority of their method in the exposure correction task.

Weaknesses:

1. Although the inpainting model assists in reconstructing structure, it may also produce artifacts, particularly in regions with complex textures or patterns that the training dataset does not adequately represent.
2. I have doubts about the necessity of the Inpainting Model branch. If it is for feature alignment, why not use the ground truth image as the input for the lower FECNet in Figure 4? Because no matter how good the Inpainting Model is, the inpainting result won't be better than the ground truth image.
3. In the real world, some photos may have bright backgrounds that are not actually overexposed areas. I am concerned that this method might have limitations when applied to real-world photos, as the inpainting model focuses on areas of exposure, which could simply be bright areas rather than overexposed ones.

Rebuttal: Thank you for your comments and suggestions. We provide the response below.

1. We address this issue by leveraging mask area mixing and the refining capability of subsequent networks.
2. Given that GT also has a better structure than input, we do not introduce GT because it would compromise the network's original ability to correct. After all, when the input and target are both GT, the network tends to be equivalent to self-mapping.
3. We acknowledge this weakness, real-world images aren't always underexposed in dark areas. Therefore, the current mask region selection method is not optimal. We will explore better mask strategies in the real world in future works.

3.3. Reviewer4

Review:
The paper "Semantic Pre-Supplement for Exposure Correction" presents a novel exposure correction strategy named "Inpainting Assisted Exposure Correction" (IAEC) to guide exposure correction by pre-performing image structure repair on severely exposed areas. The authors argue that existing methods treat the entire image equally without considering the varying difficulty of recovery between differently exposed areas. IAEC is based on the principle that contextual semantic information in the image structure can help overall image recovery, especially where there is a significant loss in severely exposed areas. The weaknesses of the paper are that (1) the novelty of the paper should be addressed more and (2)More experiments are needed to verify the generalization of the method.
Rebuttal: Thank you for your comments and suggestions. We provide the response below.

1. For the first time, we applied image inpainting to assist in exposure correction and designed an effective pipeline.
2. We added experimental results on the LCDP dataset. Our method has achieved good performance on current mainstream datasets.

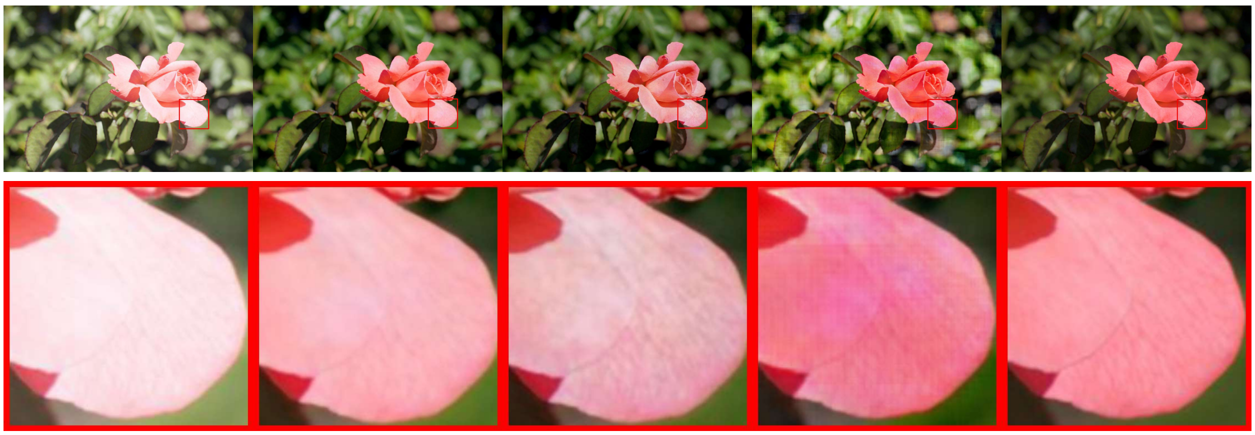
4. Changes according to the reviews

1. The output of the inpainting network is pointed out in Fig. 5 and 9 in the main body and artifact effects are emphasized.

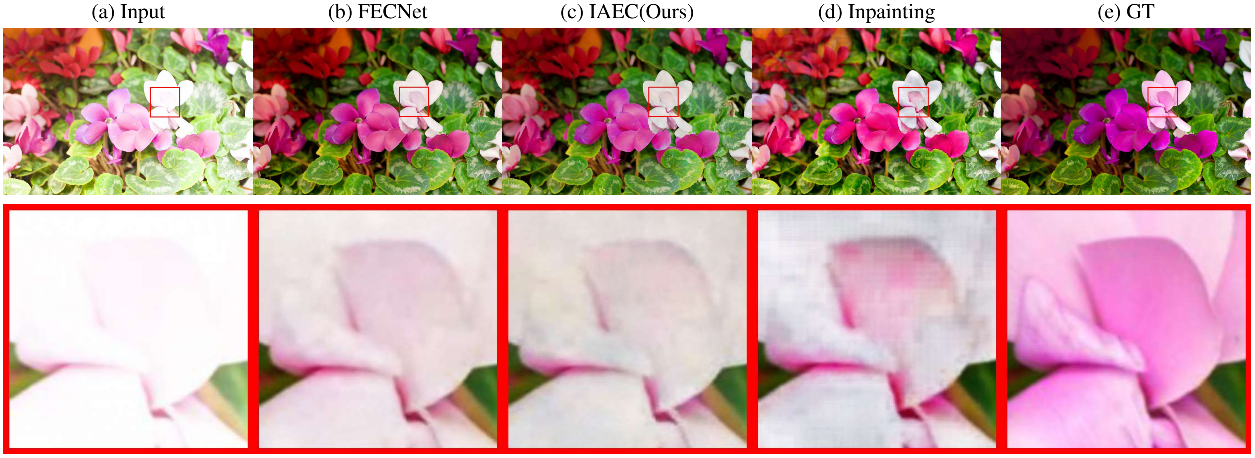
2. We emphasized how artifacts can be avoided and handled properly.
3. We present experimental results on the LCDP dataset.
4. We add the limitations of the paper in Section 4.4.

5. More Visualization Results

Due to the length limit of the main body, we provide more visual results here to demonstrate the excellent repair capabilities of the inpainting module.

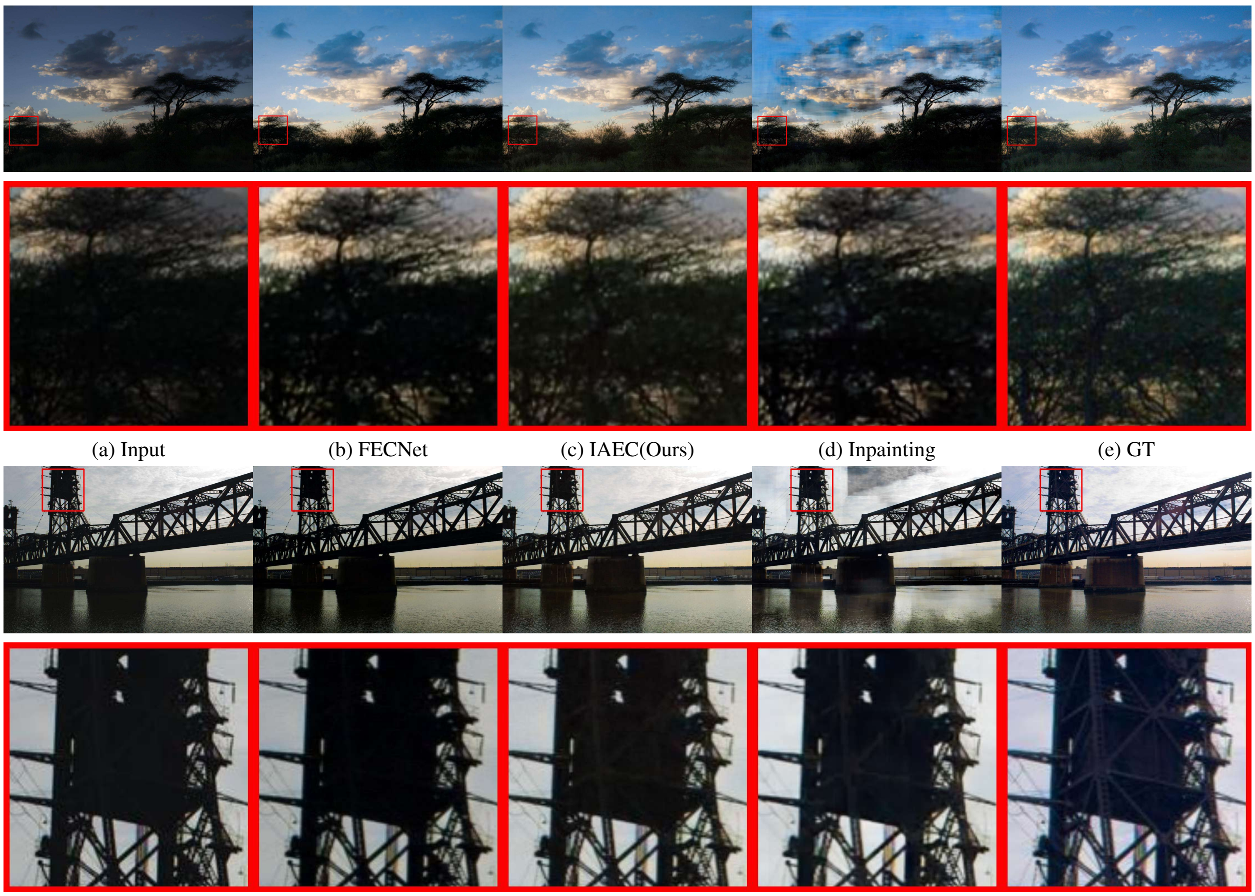


(a) Input (b) FECNet (c) IAEC(Ours) (d) Inpainting (e) GT



(f) Input (g) FECNet (h) IAEC(Ours) (i) Inpainting (j) GT

Figure 4. The visualization results on overexposed images of the ME dataset.



(a) Input (b) FECNet (c) IAEC(Ours) (d) Inpainting (e) GT

(f) Input (g) FECNet (h) IAEC(Ours) (i) Inpainting (j) GT

Figure 5. The visualization results on underexposed images of the ME dataset.

References

[1] Jie Huang, Yajing Liu, Feng Zhao, Keyu Yan, Jinghao Zhang, Yukun Huang, Man Zhou, and Zhiwei Xiong. Deep fourier-based exposure correction network with spatial-frequency interaction. In *Eur. Conf. Comput. Vis.*, pages 163–180. Springer, 2022. 1

[2] Xiaoguang Li, Qing Guo, Rabab Abdelfattah, Di Lin, Wei Feng, Ivor Tsang, and Song Wang. Leveraging inpainting for single-image shadow removal. *arXiv preprint arXiv:2302.05361*, 2023. 1