Supplementary Material for "Learning Context Enhancement Prior with deep unfolding network for Snapshot Compressive Imaging"[⋆]

Mengying Jin$^{1[0000-0001-5582-1015]}$, Zhihui Wei$^{1[0000-0002-4841-6051]}$, and Liang Xiao$^{1[0000-0003-0178-9384]}$

[1] Nanjing University of Science and Technology, Nanjing 210094, USA
[2] lncs@springer.com
http://www.springer.com/gp/computer-science/lncs
[3] ABC Institute, Rupert-Karls-University Heidelberg, Heidelberg, Germany
{abc,lncs}@uni-heidelberg.de

1 Mathematical model of CASSI system

This section will briefly describe the CASSI system's working principle and mathematically model light propagation in CASSI. According to Fig.1., the CASSI system contains an objective lens, a coded aperture, a relay lens, a dispersion prism, and a 2D detector planr. The spectral information captured by the objective lens first, modulated by the coded aperture, passes through the relay lens, then dispersed by the dispersion element, and falls overlappingly on the detector plane to be collected. At this point, the spatial and spectral scales of the image are no longer consistent with the original data.

1.1 Light Propagation Model.

To mathematically model this physical imaging process, we assume the coded aperture is set to obey $T(x, y)$. The spatio-spectral power spectral density before and after the coded aperture should be:

$$f_1(x, y, \lambda) = T(x, y) f_0(x, y, \lambda) \tag{1}$$

The coded aperture is spatially modulated in each spectral slice of the original spectral cube. Also, in this equation, to simplify the model, the effect of the point diffusion equation arising from the light as it propagates is ignored. Then, after the data pass through the relay lens and the linear dispersion element, the power spectral density is:

$$\begin{aligned} f_2(x, y, \lambda) &= \sigma(x' - (x + S(\lambda)))\sigma(y' - y) f_1(x', y'; \lambda) dx' dy' \\ &= f_0(x + S(\lambda), y; \lambda) T(x + S(\lambda), y) \end{aligned} \tag{2}$$

[⋆] This research was funded by the National Natural Science Foundation of China under Grant 61871226;supported by the Fundamental Research Funds for the Central Universities under Grant NO. JSGP202204 in part by the Jiangsu Provincial Social Developing Project under Grant BE2018727.

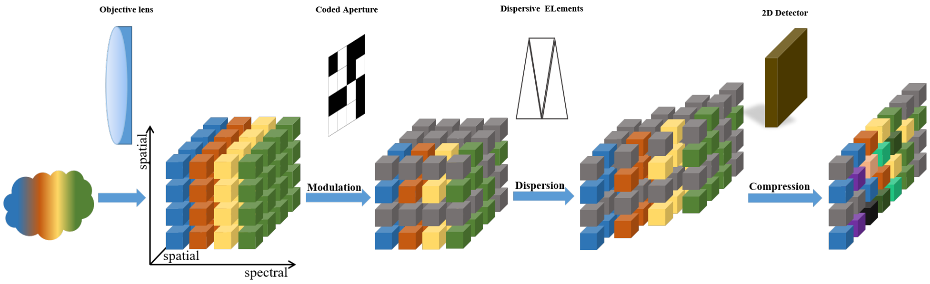

Fig. 1. The visualization of CASSI system.

with:

$$S(\lambda) = a(\lambda)(\lambda - \lambda_c) \tag{3}$$

describes the dispersion produced by the dispersive element whose central wavelength is λ_c, where $a(\lambda)$ is the dispersion coefficient generated by the dispersion element. $a(\lambda)$ is a constant when and only when the dispersion element is linear. Finally, the image at the 2D detector plane can be modeled as the integral of the 3D data along the spectrum Λ:

$$S(\lambda) = \int_{\Lambda} f_0(x + S(\lambda), y; \lambda) T(x + S(\lambda), y) d\lambda \tag{4}$$

1.2 Discrete Form.

Furthermore, in the discreste form, we assume the 2D measurement captured by the CASSI system is the $\mathcal{Y} \in R^{M \times N}$, the expected 3D spectral information is $\{\mathcal{X}_\lambda\}_{\lambda=1}^{\Lambda}$, then we have:

$$\mathcal{Y} = \sum_{\lambda=1}^{\Lambda} T_\lambda \odot \mathcal{X}_\lambda + N, \tag{5}$$

where $N \in R^{M(N+k(\lambda-1))}$ is the additive noise caused by the environment or the electronic element, etc., $\odot$ is the corresponding element product, and $\{\mathcal{T}_\lambda\}_{\lambda=1}^{\Lambda} \in R^{M(N+k(\lambda-1))}$ is defined as:

$$T_\lambda = T(x - k(\lambda - 1), y), \tag{6}$$

where $k(\lambda - 1)$ is the offset in the x-axis of the coding matrix corresponding to the λth band due to the dispersion element and the constant k is the dispersion coefficient due to the linear dispersion element. It is further written in matrix-vector form as:

$$y = \Phi \mathcal{X} + n \tag{7}$$

where $y \in R^{M(N+k(\lambda-1))}$ is the 2D measurment captured by the detector. $\mathcal{X} \in R^{MN\lambda}$ is the 3D data to be recovered and $n \in R^{M(N+k(\lambda-1))}$ is the additive noise. $\Phi \in R^{M(N+k(\lambda-1))\times MN\lambda}$ is the system forward response:

$$\Phi = [D_1, D_2, \cdots, D_\lambda] \tag{8}$$

where $\{\mathcal{X}_\lambda\}_{\lambda=1}^{\Lambda} \in R^{M(N+k(\lambda-1))}$ is the diagonal matrix, and obviously, $\Phi^T\Phi$ is also the diagonal matrix.

2 Training Details

2.1 Dataset

CAVE[8] The 32 images within CAVE are captured by a Cooled CCD camera (Apogee Alta U260), with the resolution of 512 × 512. They have 31 bands corresponding to the range of wavelength from 400nm to 700nm.

KAIST[1] The 36 images within KAIST are captured by Pointgrey Grasshopper 9.1MP Monochromatic (GS3-U3-91S6M-C) camera with a 2704 × 3376.

Following TSAnet[3] and DGSMP[2], we used 28 of the 31 bands and calibrated the wavelength range to between 453.3 nm and 648.1 nm.

Simulation. In our experiments, we select 30 of the 32 images in *CAVE* to be the training set and use 5000 randomly cropped 96 × 96 × 28 patches in each epoch. For each patch, we flip it randomly horizontally or vertically or rotate it by 90 degrees.

We use the real mask released in TSAnet[3] and also cut out a random mask patch of 96 × 96 at a time to simulate the CASSI images with a shift of 2 pixels.

Ten scenes from KAIST were selected for the test, and these images would not appear in the training set to ensure the reliability of the experiments. Each test sample has a resolution of 256 × 256.

2.2 Quantitative Metrics

In the paper, we offer numerical results based on six indicators.Here we will offer the definitions of the six indicators. We assume the ground truth images $\mathcal{X} \in R^{M \times N \times \lambda}$, the reconstructed images $\mathcal{Y} \in R^{M \times N \times \lambda}$.

Mean Squared Error (MSE) and Root Mean Squared Error (RMSE).

$$MSE(\mathcal{X}, \mathcal{Y}) = \frac{1}{MN\lambda} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \sum_{k=0}^{\lambda-1} (\mathcal{X}(i,j,k) - \mathcal{Y}(i,j,k))^2, \tag{9}$$

$$RMSE(\mathcal{X}, \mathcal{Y}) = \sqrt{\frac{1}{MN\lambda} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \sum_{k=0}^{\lambda-1} (\mathcal{X}(i,j,k) - \mathcal{Y}(i,j,k))^2}. \tag{10}$$

where the smaller RMSE, the better reconstruction quality.

Peak Signal-to-Noise Ratio (PSNR).

$$PSNR(\mathcal{X}, \mathcal{Y}) = 10 \cdot log_{10}(\frac{MAX(\mathcal{X})^2}{MSE}). \tag{11}$$

For the hyperspectral images in this paper, we calculate the PSNR for each band and take its average.

Structural Similarity Index (SSIM). SSIM is based on three comparative measures between sample $\mathcal{X}$ and $\mathcal{Y}$: luminance$l(x, y)$, contrast$c(x, y)$ and structures$s(x, y)$.

$$\begin{aligned} SSIM(\mathcal{X}, \mathcal{Y}) &= [l(\mathcal{X}, \mathcal{Y})]^{\alpha} [c(\mathcal{X}, \mathcal{Y})]^{\beta} [s(\mathcal{X}, \mathcal{Y})]^{\gamma} \\ &= [\tfrac{2\mu_{\mathcal{X}}\mu_{y}+c_1}{\mu_{\mathcal{X}}^2+\mu_{\mathcal{Y}}^2+c_1}]^{\alpha} [\tfrac{2\sigma_{\mathcal{X}\mathcal{Y}}+c_2}{\sigma_{\mathcal{X}}^2+\sigma_{\mathcal{Y}}^2+c_2}]^{\beta} [\tfrac{\sigma_{\mathcal{X}\mathcal{Y}}+c_3}{\sigma_{\mathcal{X}}\sigma_{\mathcal{Y}}+c_3}]^{\gamma} \cdot \end{aligned} \tag{12}$$

where where $\mu_x, \mu_y, \sigma_x, \sigma_{,}\sigma_{xy}$ are the local means, standard deviations, and cross-covariance for images $\mathcal{X}, \mathcal{Y}$. The parameter $\alpha = \beta = \gamma = 1$,$c_3 = c_1/2$ are set by default in MATLAB 2018b. Thus, Eq (4) can be written as:

$$SSIM(\mathcal{X}, \mathcal{Y}) = \frac{(2\mu_{\mathcal{X}}\mu_{\mathcal{Y}} + c_1)(\sigma_{\mathcal{X}\mathcal{Y}} + c_2)}{(\mu_{\mathcal{X}}^2 + \mu_{\mathcal{Y}}^2 + c_1)(\sigma_{\mathcal{X}}^2 + \sigma_{\mathcal{Y}}^2 + c_2)} \tag{13}$$

where the smaller SSIM, the better reconstruction quality.

Erreur Relative Globale Adimensionnelle de Synth‘ese (ERGAS) [6]

$$ERGAS(\mathcal{X}, \mathcal{Y}) = \frac{100}{c} \sqrt{\frac{1}{S} \sum_{i=1}^{S} \frac{MSE(\mathcal{X}^i, \mathcal{Y}^i)}{\mu_{\mathcal{Y}^i}^2}}. \tag{14}$$

where the smaller ERGAS ,the better reconstruction quality.

Spectral Angle Mapper (SAM).

$$SAM(\mathcal{X},\mathcal{Y})=\frac{1}{MN}\sum_{j=1}^{MN} arccos\frac{y_j^T x_j}{||y_j||_2||x_j||_2}. \tag{15}$$

where the smaller SAM, the less spectral distortion.

Universal Image Quality Index(UIQI)[7]

$$UIQI(\mathcal{X}^i,\mathcal{Y}^i)=\frac{1}{M}\sum_{j=1}^{M}\frac{\sigma_{\mathcal{X}_j^i\mathcal{Y}_j^i}}{\sigma_{\mathcal{X}_j^i}\sigma_{\mathcal{Y}_j^i}}\frac{2\mu_{\mathcal{X}_j^i}\mu_{\mathcal{Y}_j^i}}{\mu_{\mathcal{X}_j^i}+\mu_{\mathcal{Y}_j^i}}\frac{2\sigma_{\mathcal{X}_j^i}\sigma_{\mathcal{Y}_j^i}}{\sigma_{\mathcal{X}_j^i}+\sigma_{\mathcal{Y}_j^i}}, \tag{16}$$

represents the UIQI between i^th band of $\mathcal{X},\mathcal{Y}$. Then the UIQI between $\mathcal{X},\mathcal{Y}$ would be:

$$UIQI(\mathcal{X},\mathcal{Y})=\frac{1}{S}\sum_{i=1}^{S} UIQI(\mathcal{X}^i,\mathcal{Y}^i). \tag{17}$$

where the larger UIQI, the better reconstruction quality.

3 Experiment setup

3.1 DCT bases

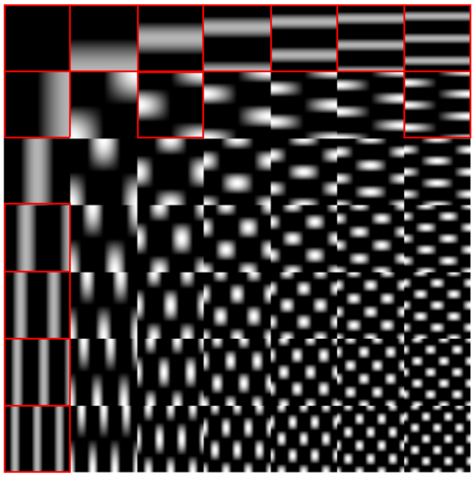

Fig. 2. Visualization of the DCT bases. The ones marked with red boxes are the DCT bases we used in the experiment.

As shown above, for the DCT bases used in our paper implementation, we followed the FcaNet[5] and selected the DCT bases according to the two-step selection method. We also divide the input features into 16 groups along the channels in our method and assign each group the DCT basis corresponding to one of the frequency components circled in the red box.

3.2 The Selection of g in Channel Shuffle

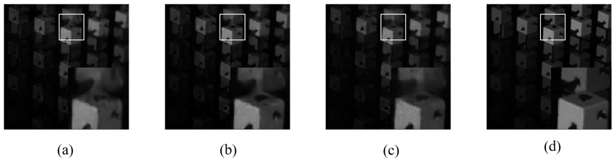

Fig. 3. Visualization comparison of the selection of g in channel shuffle. (a) $g = 2$; (b)$g = 8$; (c)$g = 16$; (d)Groud Truth; Zoom for better view.

From the data presented in the paper, we can see that the channel shuffle operation is always effective in improving the construction quality regardless of the choice of g. Even though, we hope to find a more reasonable and suitable g. However, the trends of PSNR and SSIM metrics vary with g. As we mentioned in our paper, we determined the choice of g by observing the experimental results, and we did not show the basis of our choice in our paper due to the limitation of the length of the paper, and we will give a comparison here.

Scene02 contains more edges and textures than others, so it is more obvious to see the difference in the method. From the zoom in Fig.3, $g = 16$ when the edge of the junction of light and dark distinction is more precise, and $g = 2$ when there is a specific adhesion.

4 Additional visualization

In the paper, limited by the length, we only offer a comparison of two scenes. Here we will give the comparison for the rest scenes, and considering that the results of the iterative method are not satisfactory enough, we only give the comparison based on the learning method here, with DGSMP[2], PnP-DIP[4], TSAnet[3].

References

1. Choi, I., Jeon, D.S., Nam, G., Gutierrez, D., Kim, M.H.: High-quality hyperspectral reconstruction using a spectral prior. ACM Transactions on Graphics (TOG) **36**(6), 1–13 (2017)
2. Huang, T., Dong, W., Yuan, X., Wu, J., Shi, G.: Deep gaussian scale mixture prior for spectral compressive imaging. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16216–16225 (2021)
3. Meng, Z., Ma, J., Yuan, X.: End-to-end low cost compressive spectral imaging with spatial-spectral self-attention. In: European Conference on Computer Vision. pp. 187–204. Springer (2020)
4. Meng, Z., Yu, Z., Xu, K., Yuan, X.: Self-supervised neural networks for spectral snapshot compressive imaging. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2622–2631 (2021)
5. Qin, Z., Zhang, P., Wu, F., Li, X.: Fcanet: Frequency channel attention networks. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 783–792 (2021)
6. Wald, L.: Quality of high resolution synthesised images: Is there a simple criterion? In: Third conference" Fusion of Earth data: merging point measurements, raster maps and remotely sensed images". pp. 99–103. SEE/URISCA (2000)
7. Wang, Z., Bovik, A.C.: A universal image quality index. IEEE signal processing letters **9**(3), 81–84 (2002)
8. Yasuma, F., Mitsunaga, T., Iso, D., Nayar, S.K.: Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum. IEEE transactions on image processing **19**(9), 2241–2253 (2010)

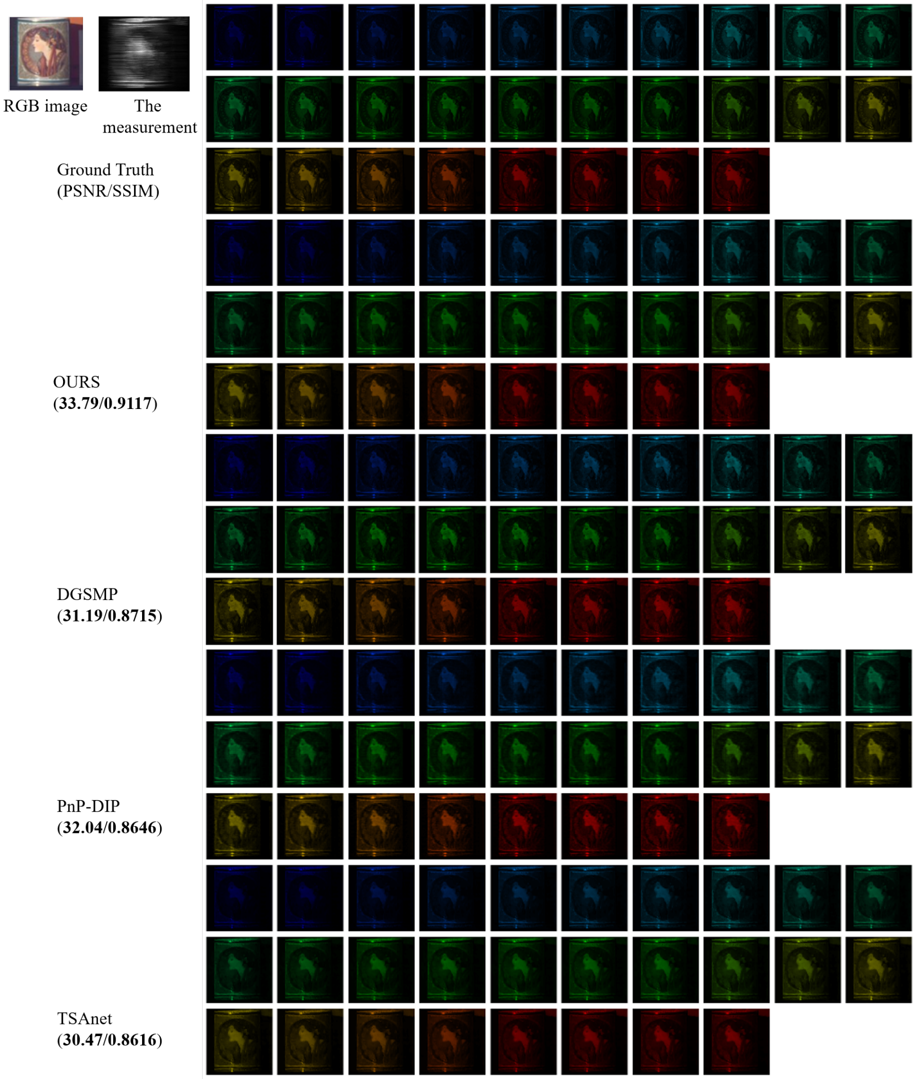

Fig. 4. The visualization results of Scene01, including RGB images, the measurement, ground truth, and the results of our method, DGSMP, PnP-DIP and TSAnet. PSNR(dB) and SSIM are also marked in the figure. Color for a better view.

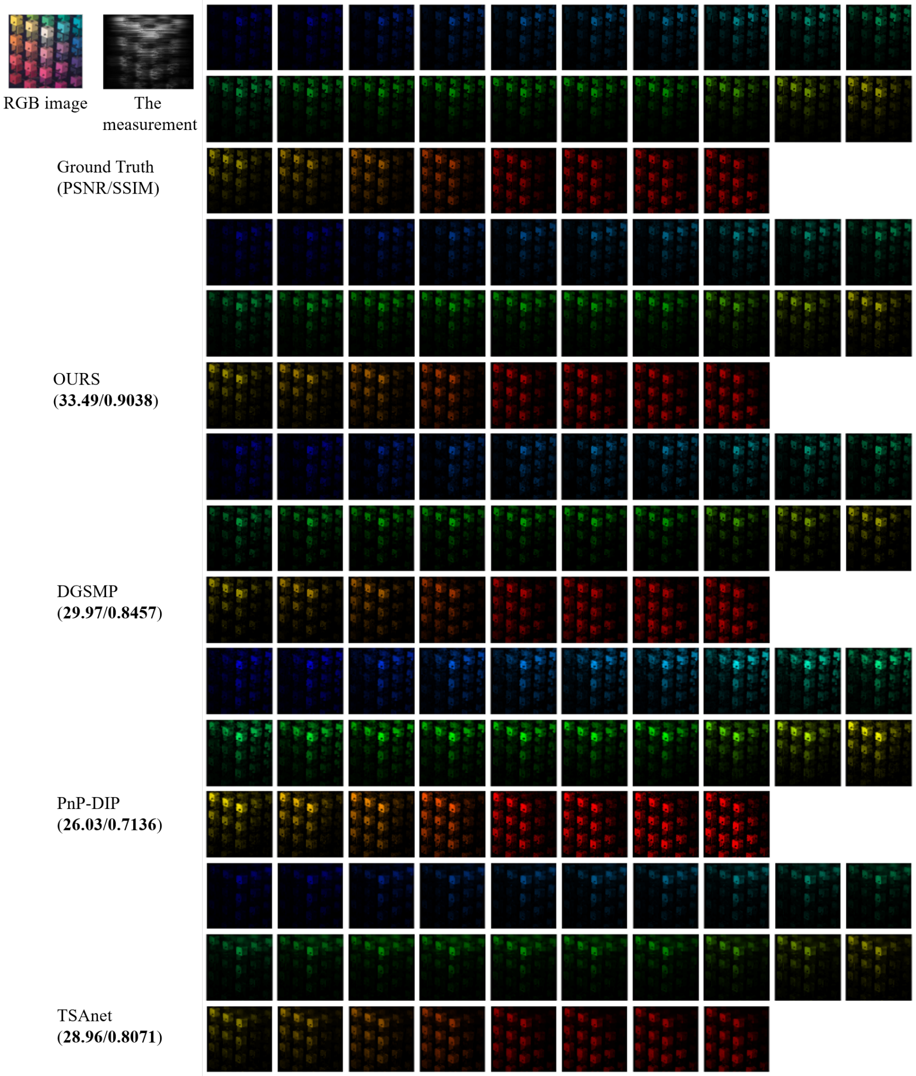

Fig. 5. The visualization results of Scene02, including RGB images, the measurement, ground truth, and the results of our method, DGSMP, PnP-DIP and TSAnet. PSNR(dB) and SSIM are also marked in the figure. Color for better view.

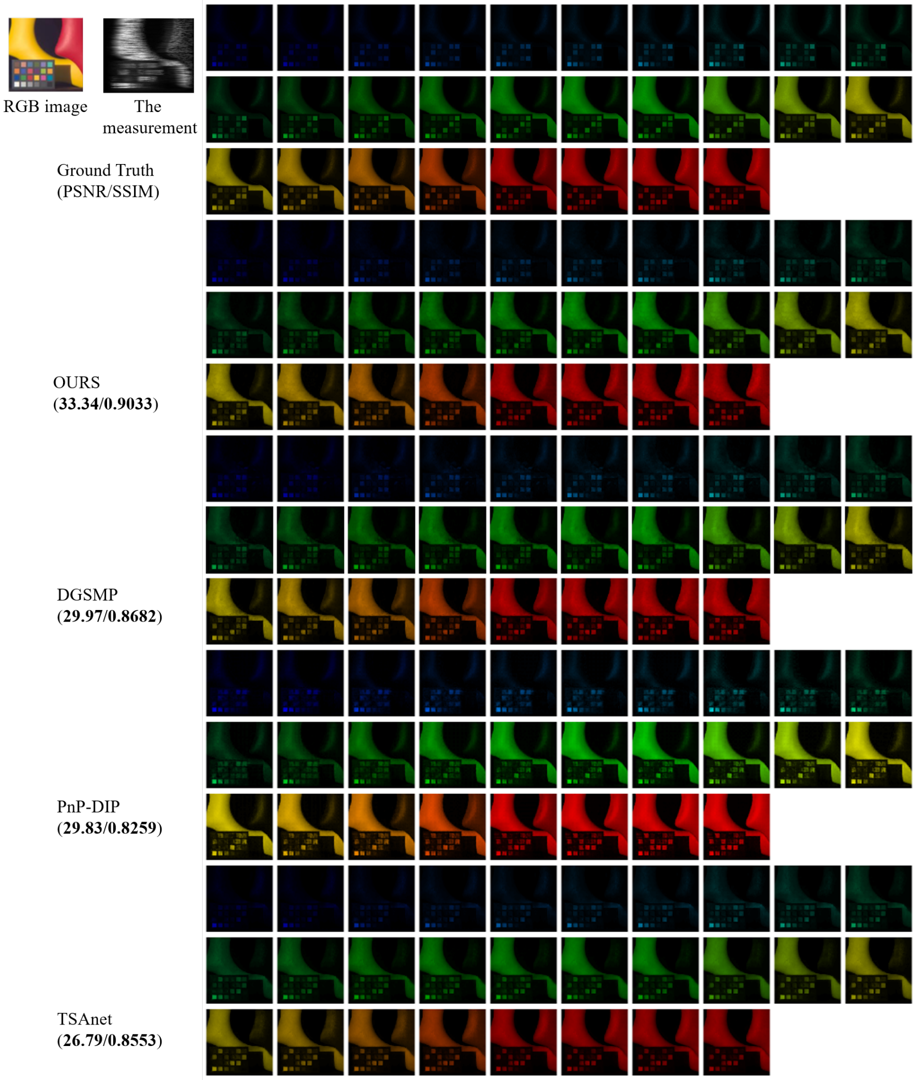

Fig. 6. The visualization results of Scene03, including RGB images, the measurement, ground truth, and the results of our method, DGSMP, PnP-DIP and TSAnet. PSNR(dB) and SSIM are also marked in the figure. Color for better view.

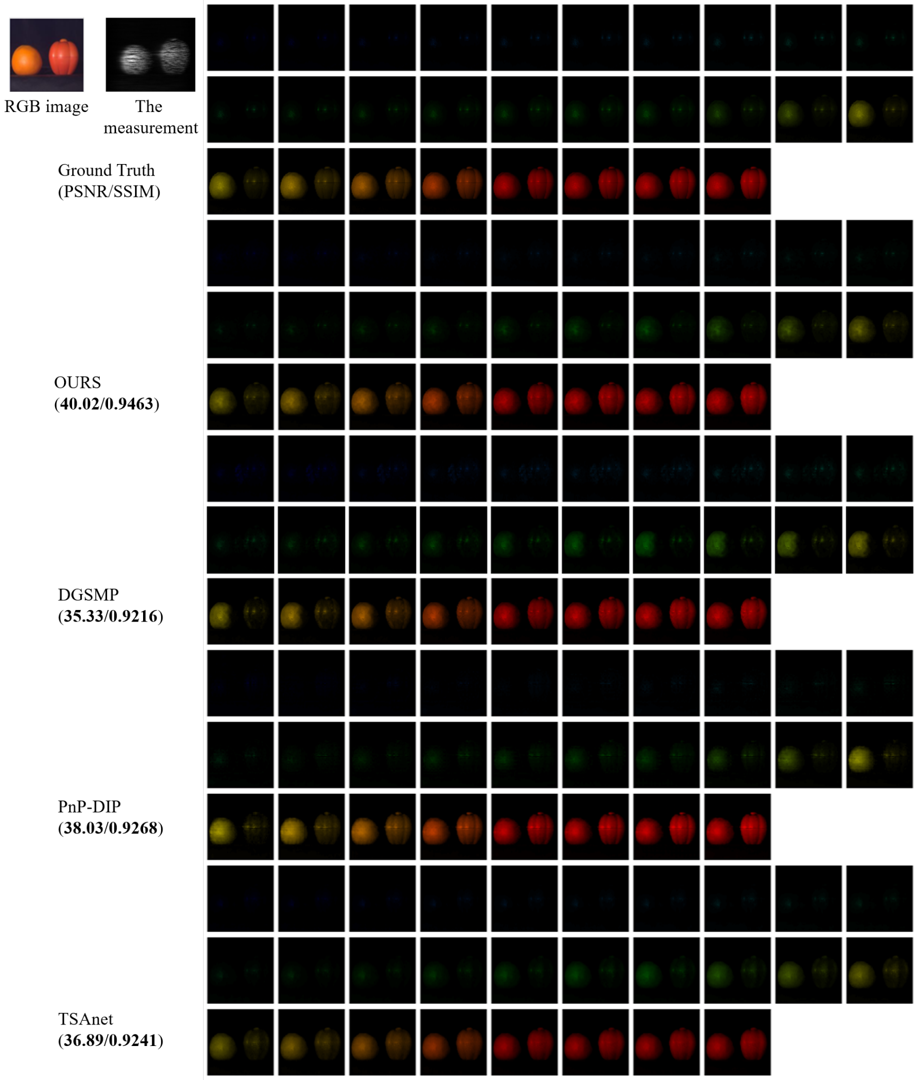

Fig. 7. The visualization results of Scene04, including RGB images, the measurement, ground truth, and the results of our method, DGSMP, PnP-DIP and TSAnet. PSNR(dB) and SSIM are also marked in the figure. Color for better view.

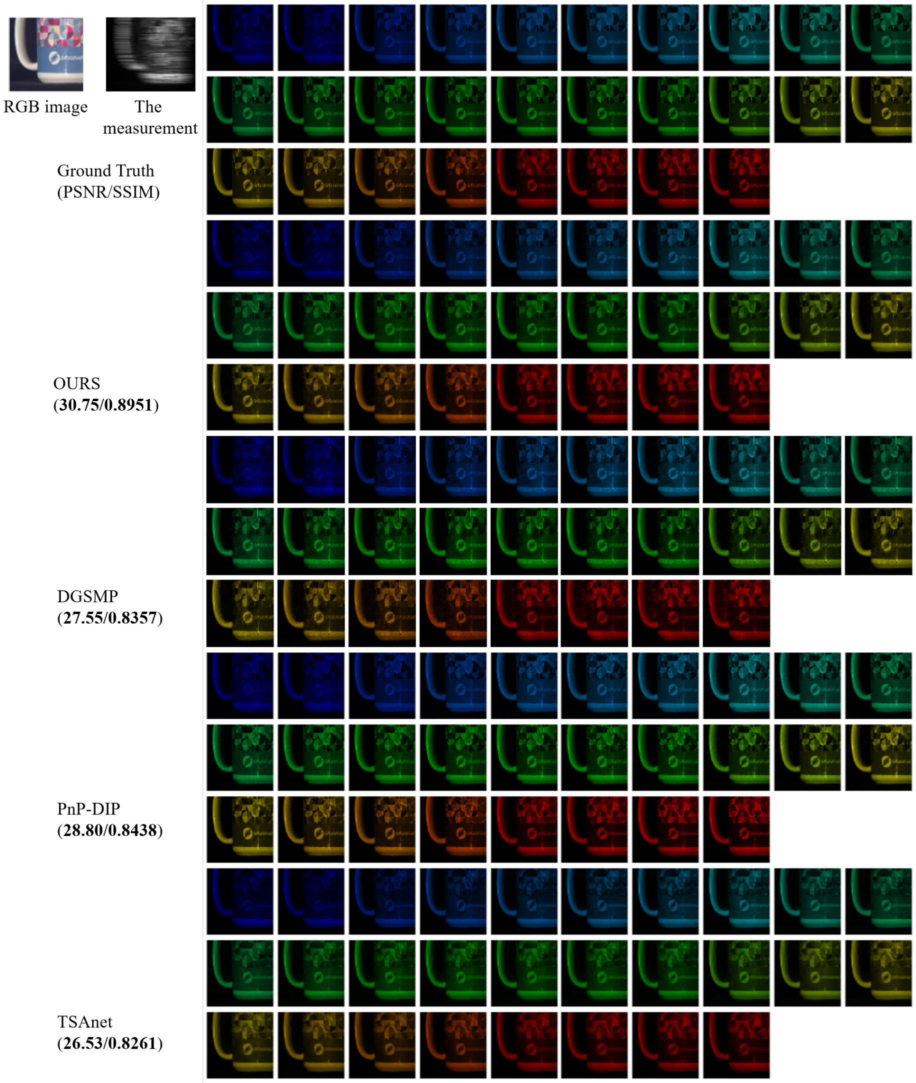

Fig. 8. The visualization results of Scene05, including RGB images, the measurement, ground truth, and the results of our method, DGSMP, PnP-DIP and TSAnet. PSNR(dB) and SSIM are also marked in the figure. Color for better view.

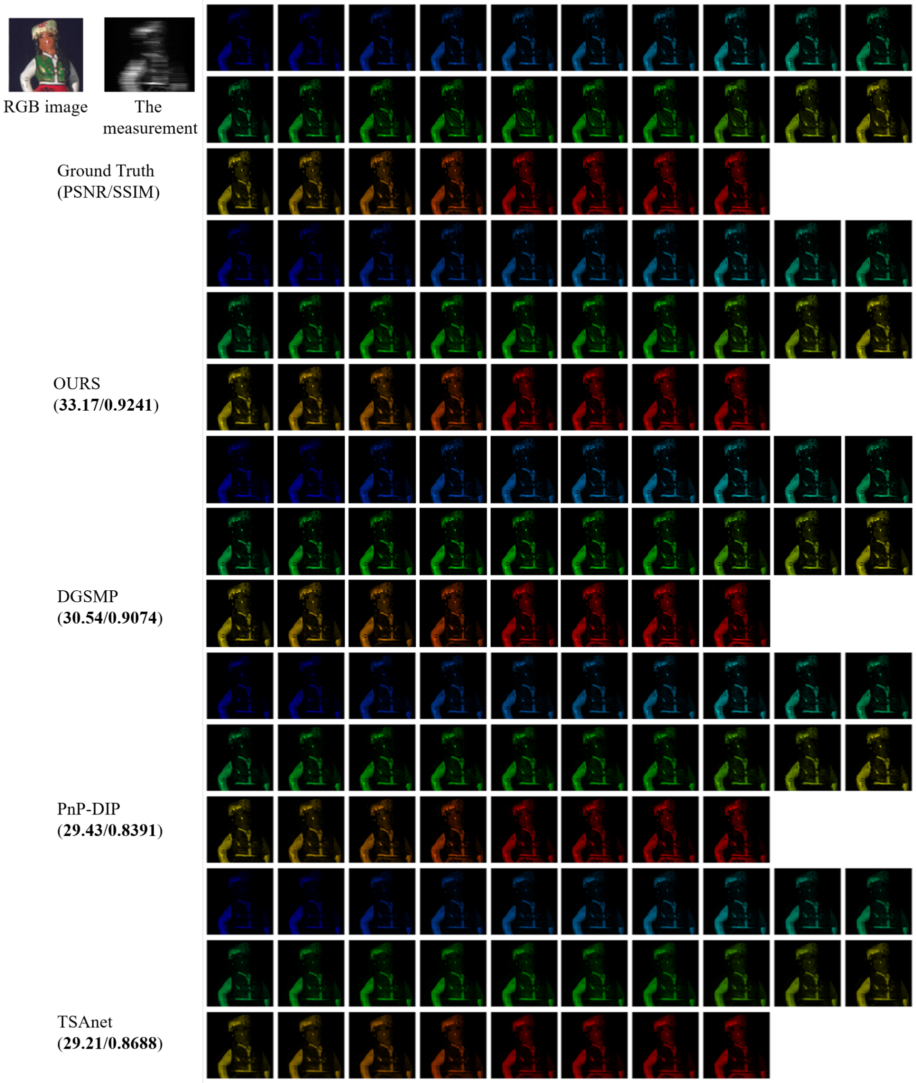

Fig. 9. The visualization results of Scene06, including RGB images, the measurement, ground truth, and the results of our method, DGSMP, PnP-DIP and TSAnet. PSNR(dB) and SSIM are also marked in the figure. Color for better view.

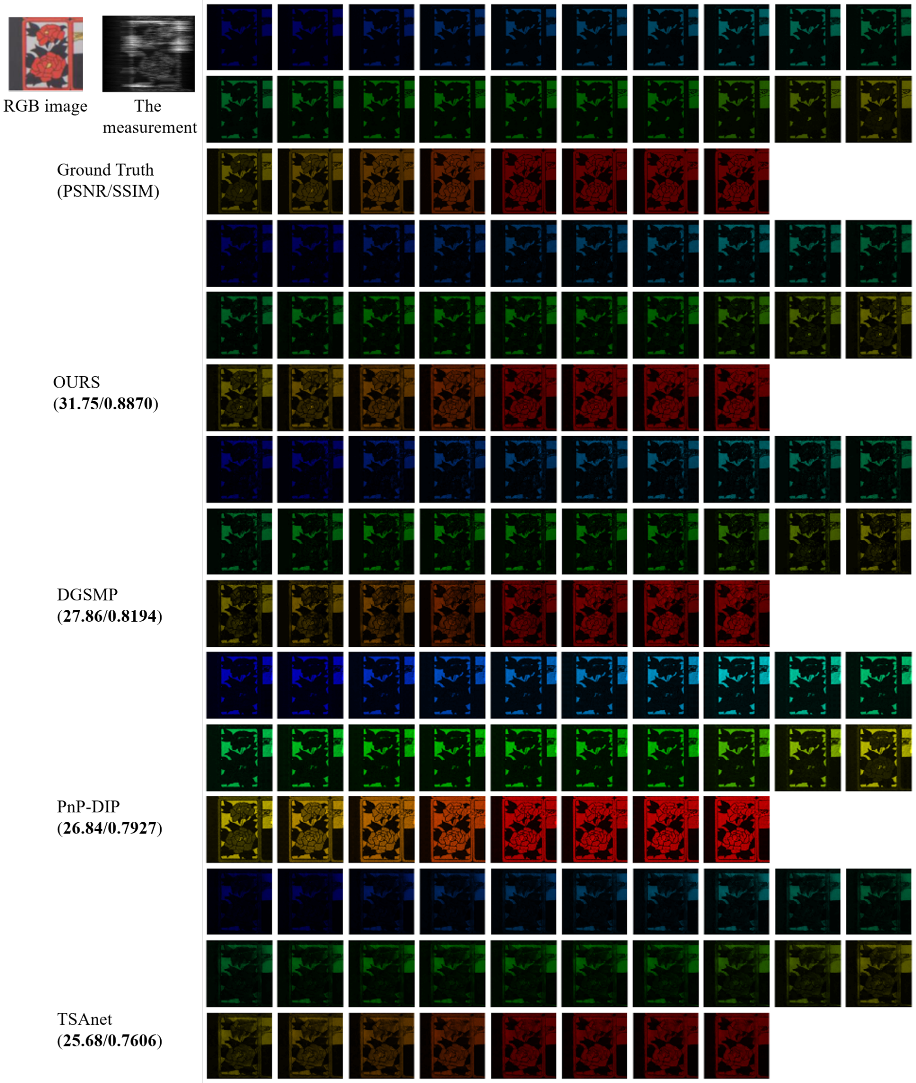

Fig. 10. The visualization results of Scene07, including RGB images, the measurement, ground truth, and the results of our method, DGSMP, PnP-DIP and TSAnetv. PSNR(dB) and SSIM are also marked in the figure. Color for better view.

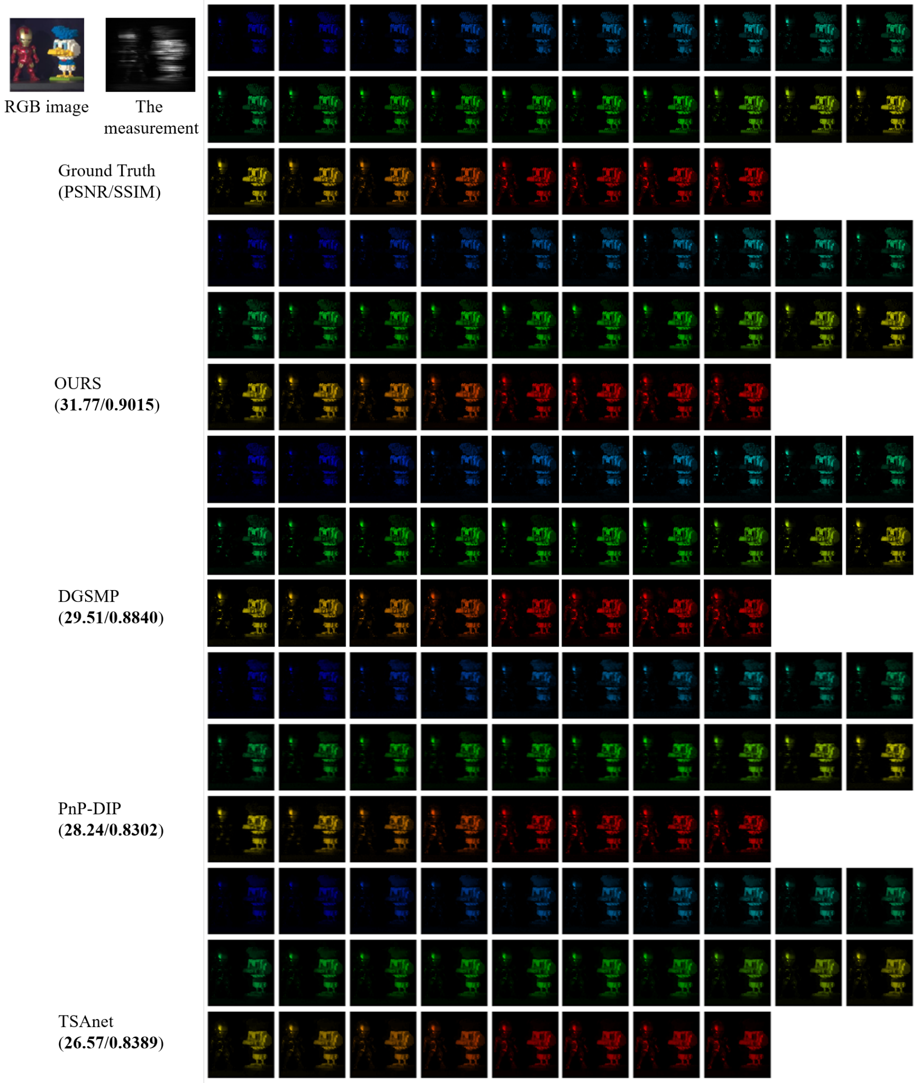

Fig. 11. The visualization results of Scene08, including RGB images, the measurement, ground truth, and the results of our method, DGSMP, PnP-DIP and TSAnet. PSNR(dB) and SSIM are also marked in the figure. Color for better view.

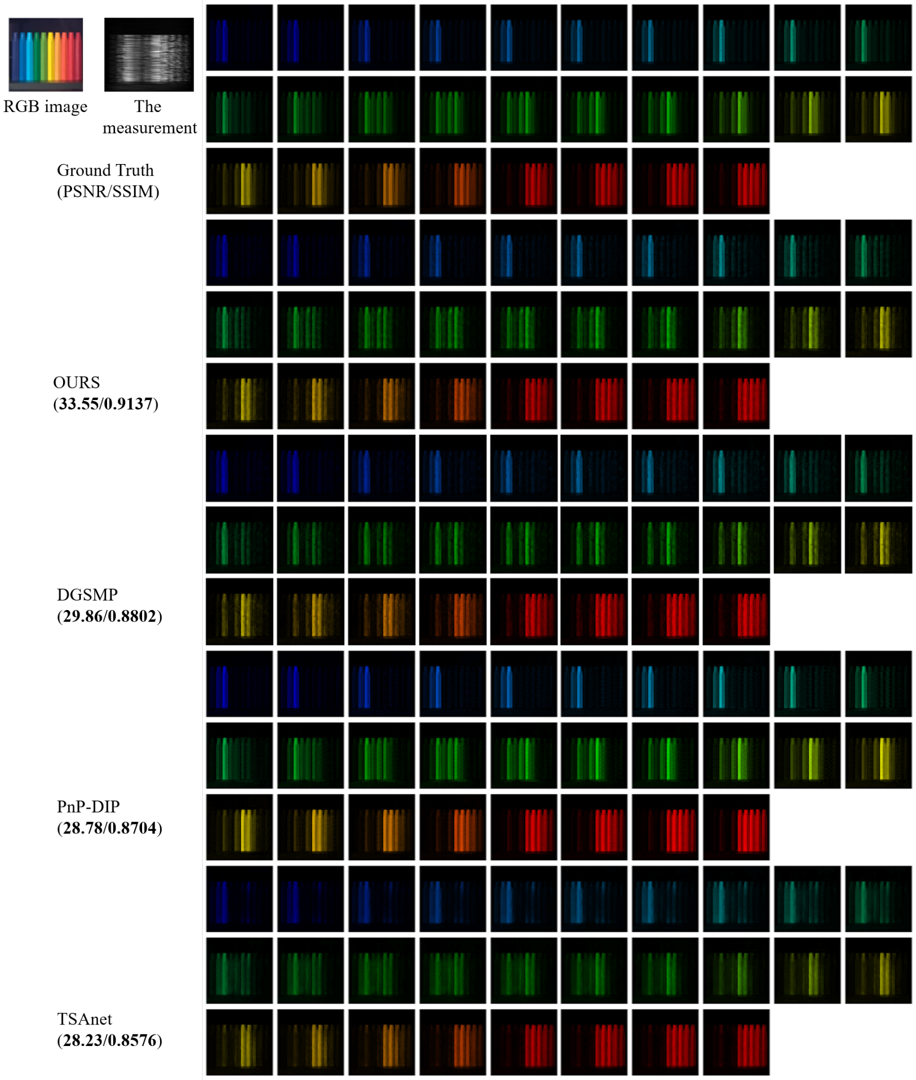

Fig. 12. The visualization results of Scene09, including RGB images, the measurement, ground truth, and the results of our method, DGSMP, PnP-DIP and TSAnet. PSNR(dB) and SSIM are also marked in the figure. Color for better view.

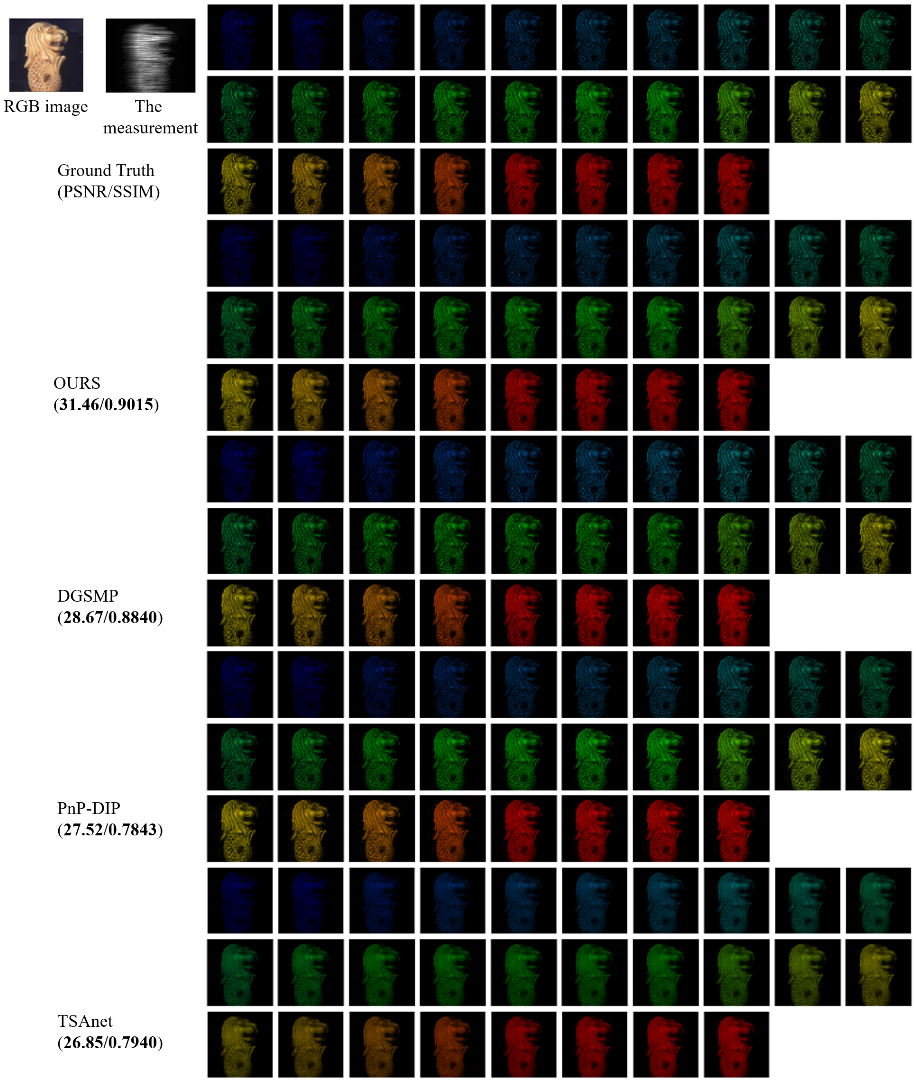

Fig. 13. The visualization results of Scene10, including RGB images, the measurement, ground truth, and the results of our method, DGSMP, PnP-DIP and TSAnet. PSNR(dB) and SSIM are also marked in the figure. Color for better view.