Supplementary Material for "λ-net: Reconstruct Hyperspectral Images from a Snapshot Measurement"

Xin Miao[1,†] Xin Yuan[2,*] Yunchen Pu[3] Vassilis Athitsos[1]
[1]University of Texas at Arlington, TX, USA [2]Nokia Bell Labs, NJ, USA [3]Facebook, CA, USA
xin.miao@mavs.uta.edu xyuan@bell-labs.com pyc40@fb.com athitsos@uta.edu

In this supplementary material (SM), we show the architecture of the U-net in the refinement stage, the results of adding self-attention to different layers of the U-net in the reconstruction stage, and the reconstructed images for all 24 channels of the 16 testing scenes by using different methods. In addition, the reconstructed 24 images of the real data using all four algorithms are also plotted. We further show the robustness of λ-net [4] to noise. Finally, we modified the network to the video compressive sensing system [3, 6] and show the reconstructed videos by different methods.

1. U-net in The Refinement Stage

Table 1. Average PSNR (dB) and SSIM of 16 testing scenes

layers	$feat_{1024}$	$feat_{512}$	$feat_{256}$	$feat_{128}$
PSNR	31.20	31.67	32.29	out of memory
SSIM	0.883	0.889	0.896	out of memory

In Fig. 1, we show the architecture of the U-net in the refinement stage. For clear comparison, we also show the deeper U-net in the reconstruction stage in Fig. 2. Note that the residual learning is used in the refinement U-net .

2. Self-attention on Different Layers

Table 1 shows the results of using self-attention in different layers of the U-net (Fig. 2) in the *reconstruction stage*. We add the self-attention to different layers feature maps in the decoder of the U-net. We use $feat_{num}$ to represent which layer we add the self-attention where num denotes the number of feature maps in that layer before the deconvolution. We can observe that it provides a better result if add the self-attention to higher level feature maps. Due to the limitation of the GPU memory, the highest level feature map that we can add is $feat_{256}$.

3. Reconstructed Hyperspectral Images of Real Data

Figs 4-7 show the 24 reconstructed images using four different algorithms for the real data compared with the truth (captured by a spectrometer) in Fig. 3.

4. Reconstructed Hyperspectral Images for "Real-Mask-In-the-Loop" Simulation Data

In Figs 8-87, we show the 24 frames reconstructed hyperspectral images by using different methods for all 16 testing scenes.

5. Noise Analysis of λ-net

Table 2. Reconstruction PSNR (dB) with various algorithms at different noise levels. In each cell, left (Scene 2), right (Scene 6)

σ	λ-net	GAP-TV [5]	TwIST [1]	DeSCI [2]
0	30.07, 28.09	16.58, 22.81	13.14, 20.11	22.39, 24.75
0.001	30.04, 28.17	16.54, 22.80	13.09, 20.11	21.89, 24.23
0.01	29.88, 27.59	15.49, 21.53	15.91, 20.10	19.95, 22.27
0.1	17.06, 19.40	7.36, 15.00	9.99, 16.14	9.24, 14.15

We have performed the noise analysis of λ-net compared with other algorithms. Due to the long running time of DeSCI, we selected two scenes (Scene 2 and 6) for this analysis. In λ-net, the measurement value is normalized to $[0, 1]$; we added zero-mean white Gaussian noise to the measurement with different variance $\sigma = \{0, 0.001, 0.01, 0.1\}$, where $\sigma = 0$ means noise free. These noisy measurements are sent to λ-net and other algorithms to perform the reconstruction, with results summarized in Table 2. It can be seen that λ-net always provides the best results in different noise levels and thus it is robust to noise. By contrast, GAP-TV is sensitive to noise and when the noise is large, DeSCI cannot improve the PSNR over GAP-TV for Scene 6.

6. Reconstructed Videos

Table 3. Average PSNR (dB) of 48 testing video frames.

method	λ-net	DeSCI [2]	GAP-TV [5]
PSNR	29.04	28.72	20.89

We use 1143 frames traffic video to train our modified λ-net for video CS. Every 8 frames generate a single measurement. Our testing data have 6 measurements, and so 48

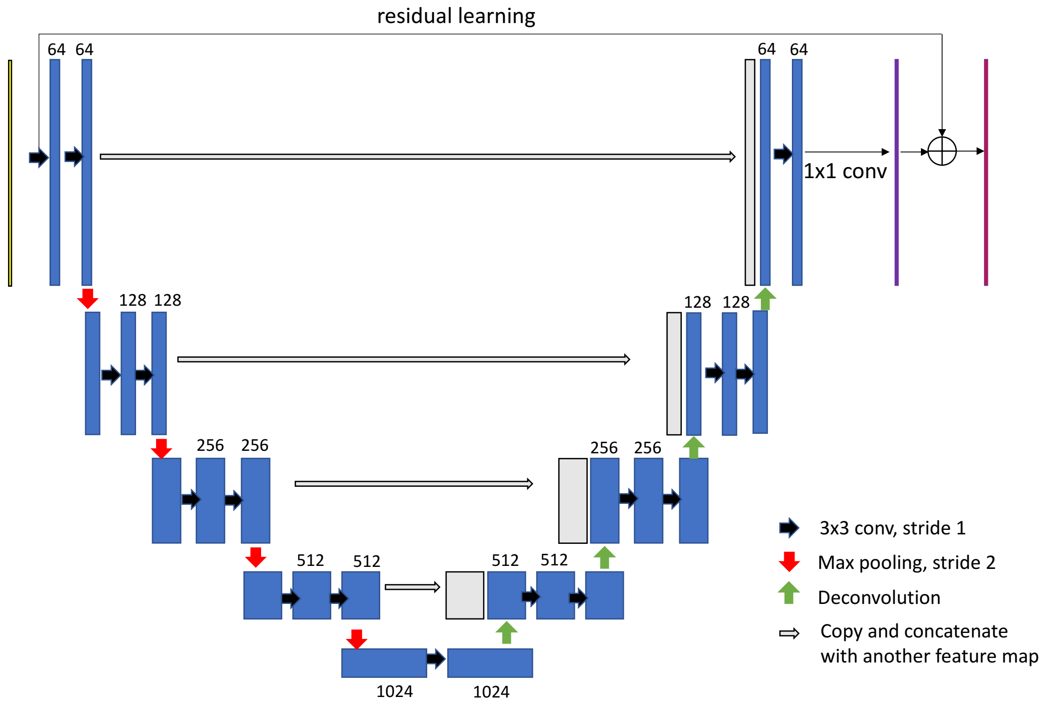

Figure 1. U-net architecture used in the *refinement stage* of our network. The input (far left) is a single frame from the output of the reconstruction stage and the output (far right) is the refined frame. Note that residual learning is used.

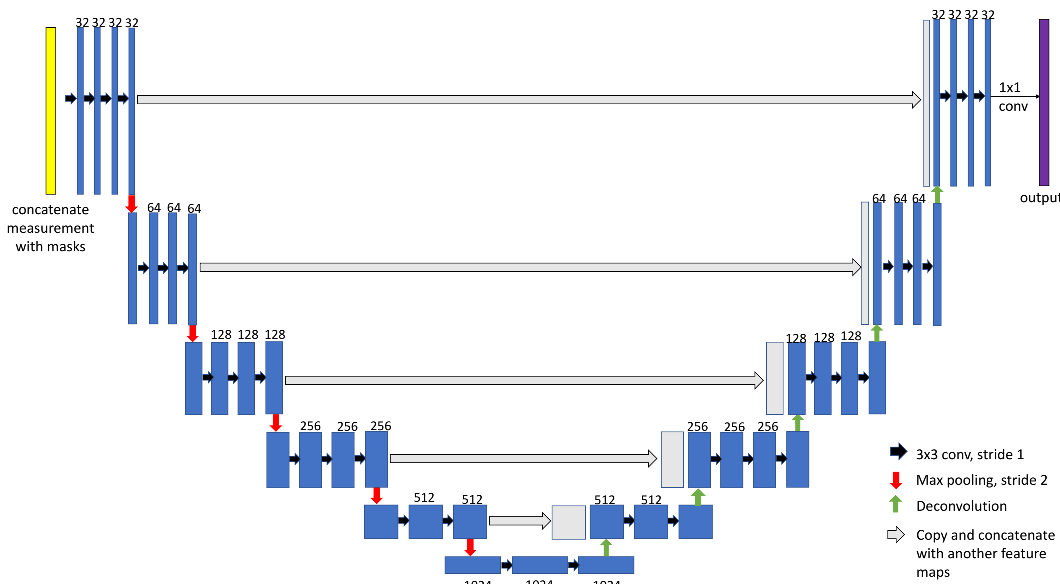

Figure 2. U-net architecture used in the *reconstruction stage* of our network. The input (far left) is the masks plus the measurement captured by SCI camera and the output (far right) is the reconstructed hyperspectral image cube.

frames are reconstructed. The average PSNR for different methods is showed in Table 3. Selected 8 frames reconstructed by several methods are shown in Fig. 88.

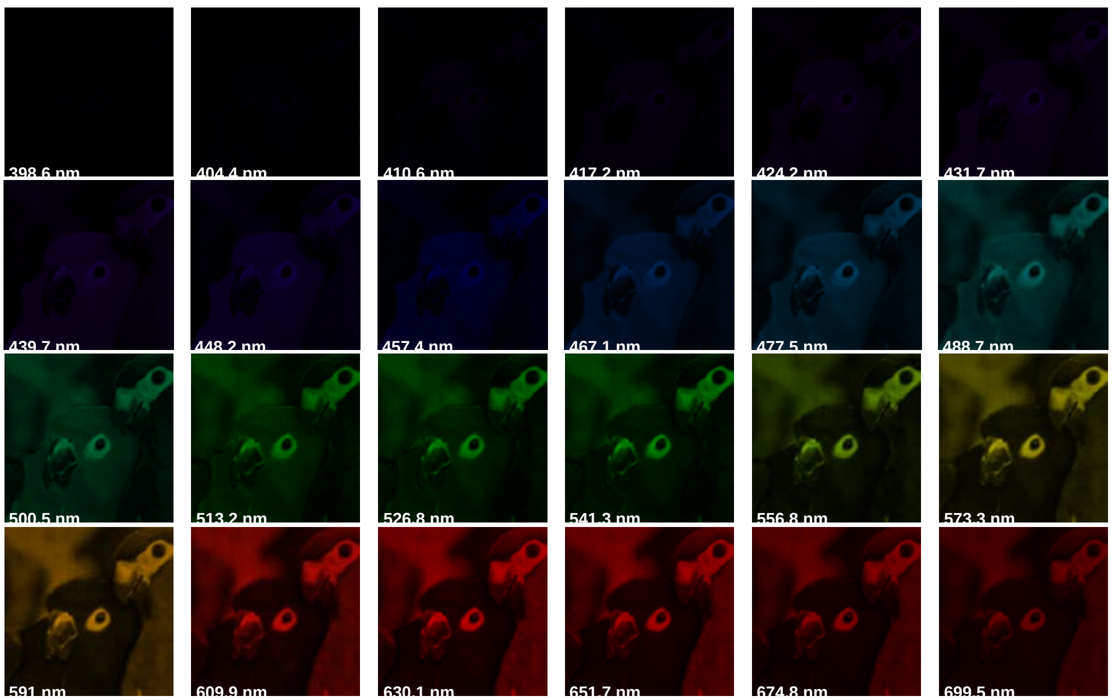

Figure 3. Truth of the real data.

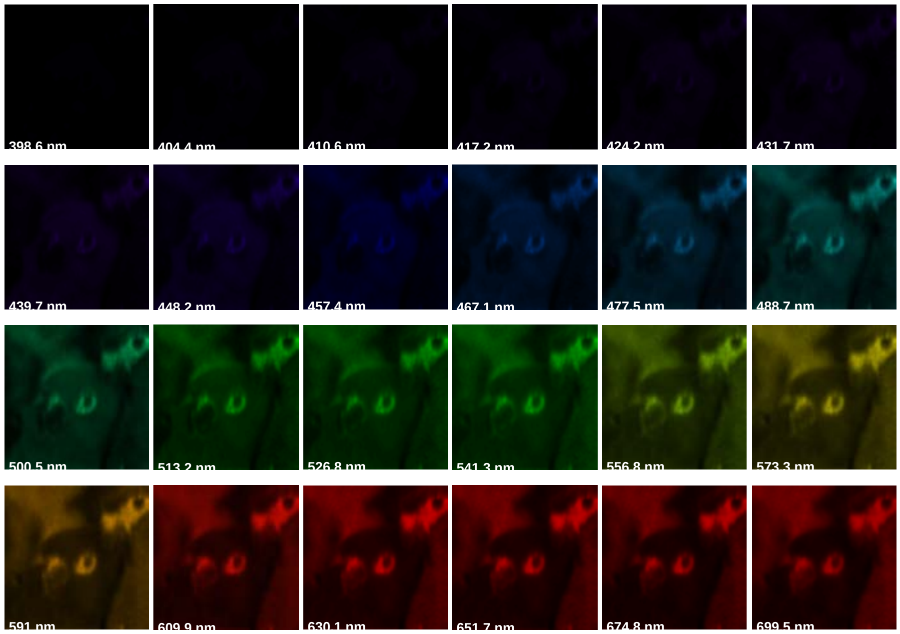

Figure 4. Reconstructed result of λ-net for the real data, PSNR: 25.59dB, SSIM: 0.819.

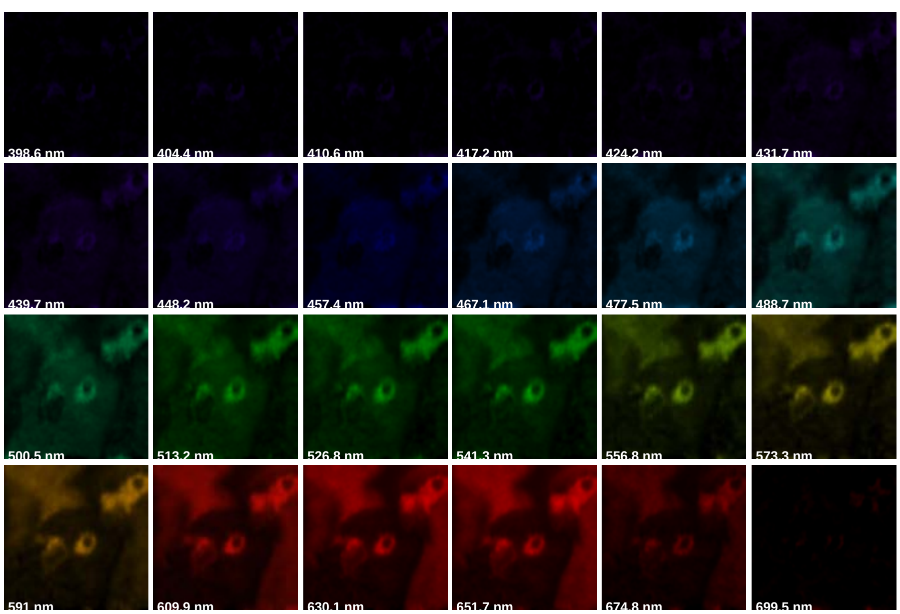

Figure 5. Reconstructed result of GAP-TV for the real data, PSNR: 24.58dB, SSIM: 0.741.

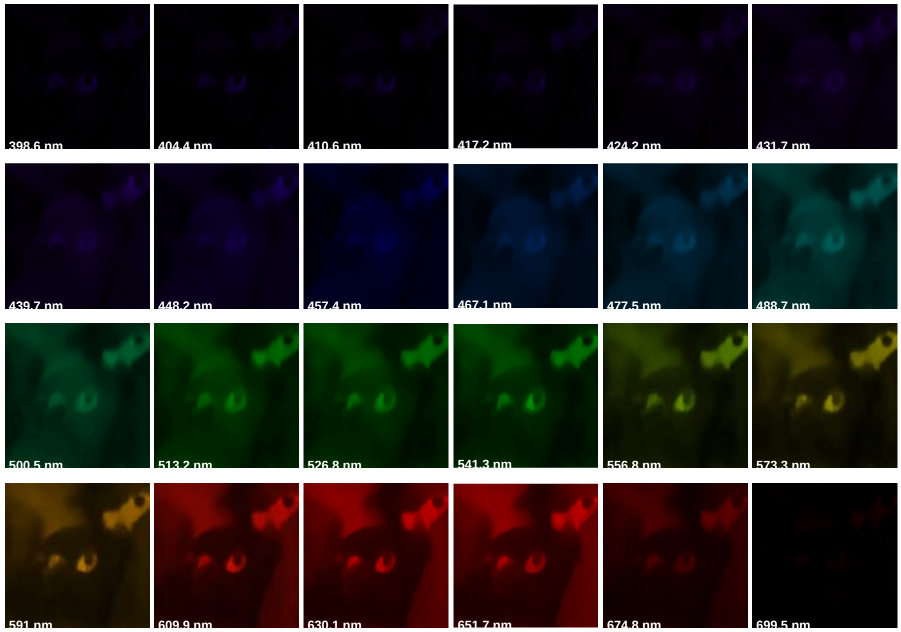

Figure 6. Reconstructed result of TwIST for the real data, PSNR: 24.33dB, SSIM: 0.785.

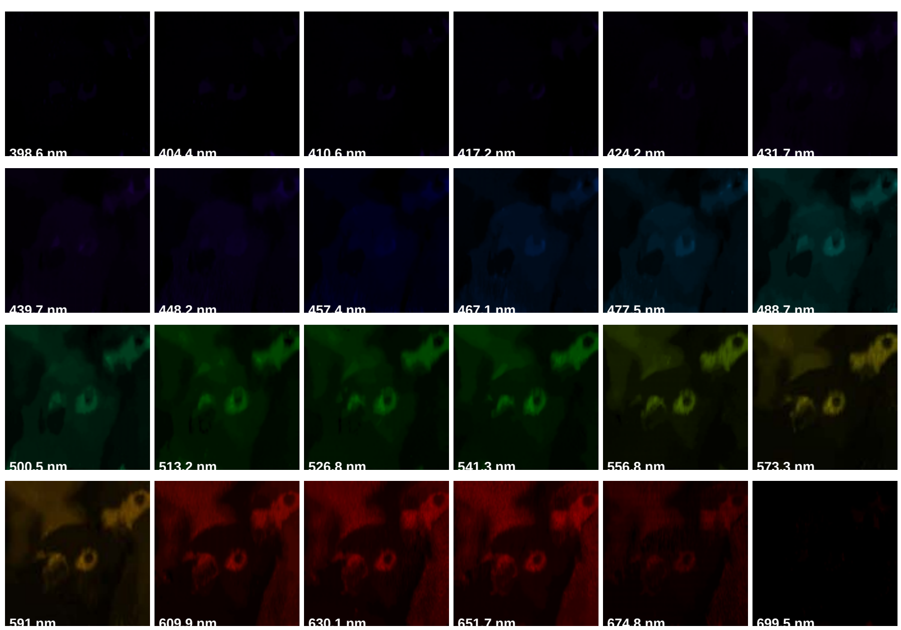

Figure 7. Reconstructed result of DeSCI for the real data, PSNR: 25.13dB, SSIM: 0.797.

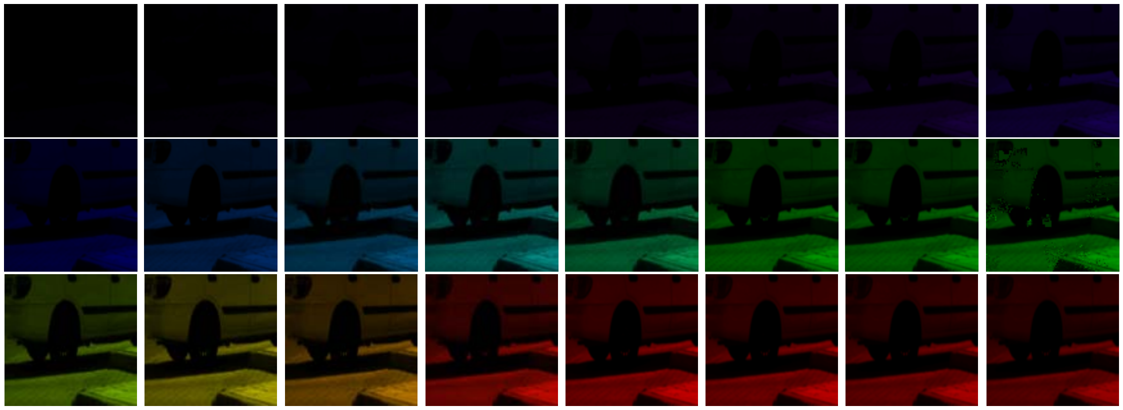

Figure 8. Truth of Scene 1.

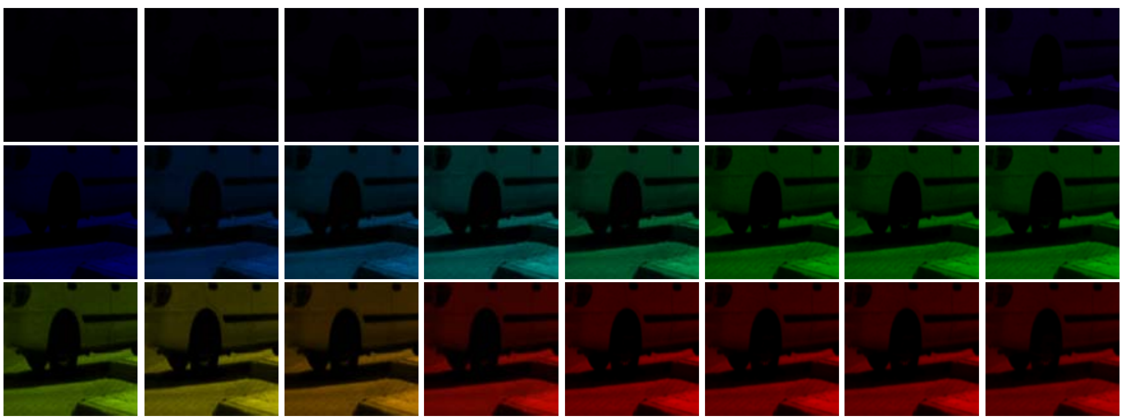

Figure 9. Reconstructed result of λ-net for Scene 1, PSNR: 36.29, SSIM: 0.925.

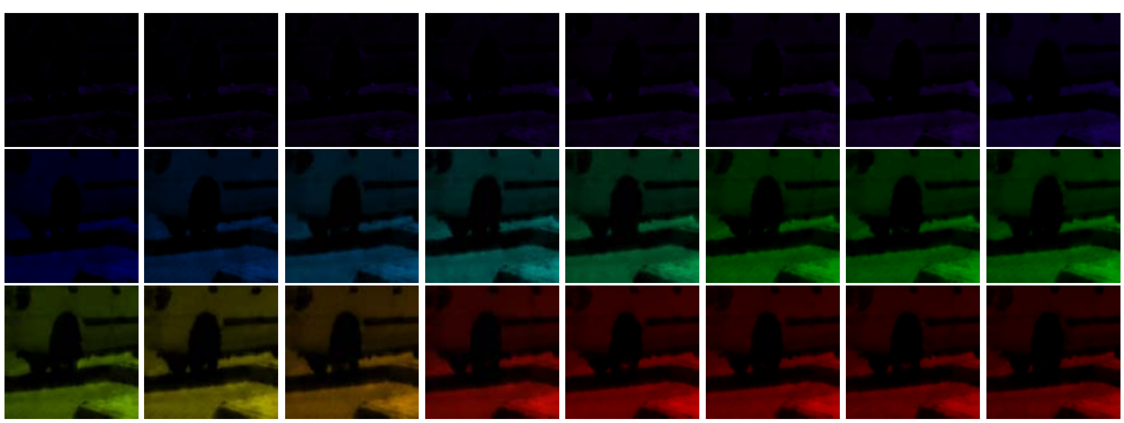

Figure 10. Reconstructed result of GAP-TV for Scene 1, PSNR: 29.48, SSIM: 0.800.

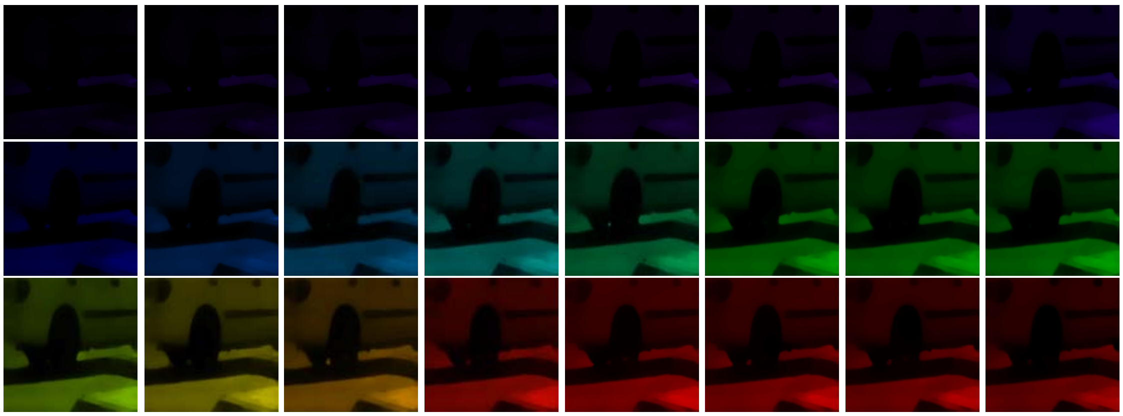

Figure 11. Reconstructed result of TwIST for Scene 1, PSNR: 26.77, SSIM: 0.772.

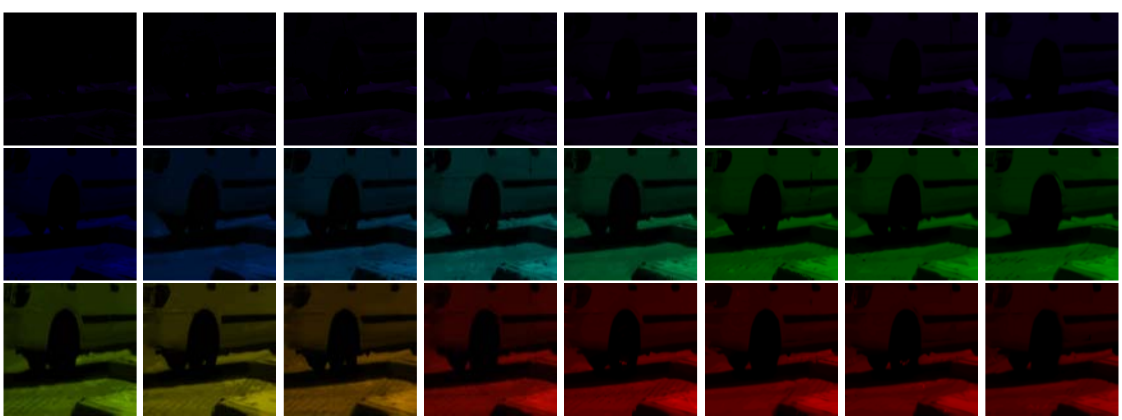

Figure 12. Reconstructed result of DeSCI for Scene 1, PSNR: 31.51, SSIM: 0.896.

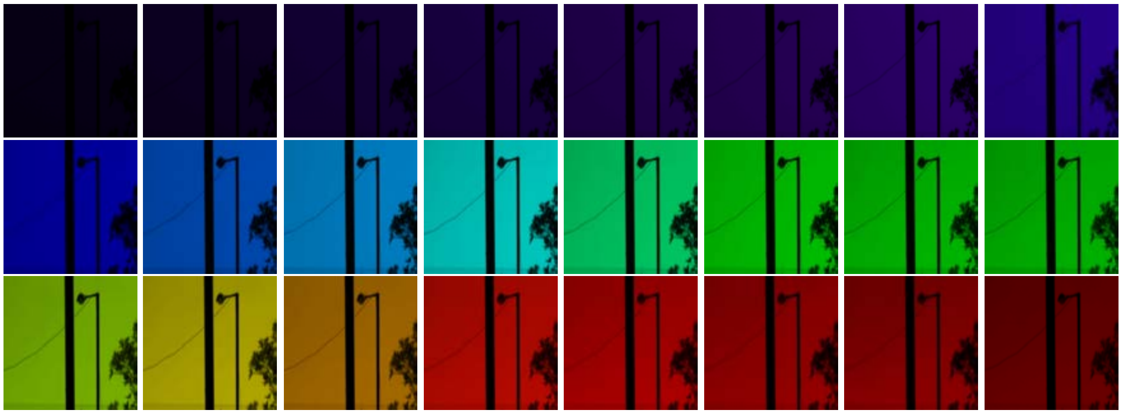

Figure 13. Truth of Scene 2.

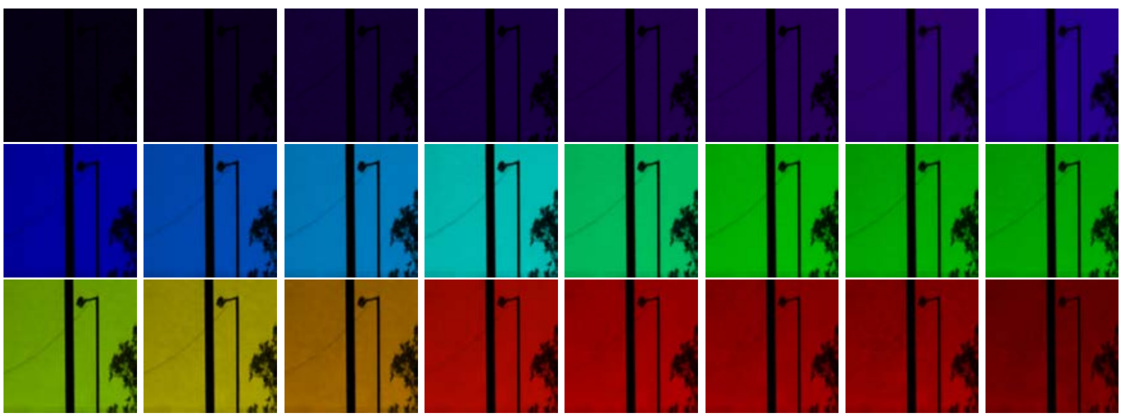

Figure 14. Reconstructed result of λ-net for Scene 2, PSNR: 30.07, SSIM: 0.929.

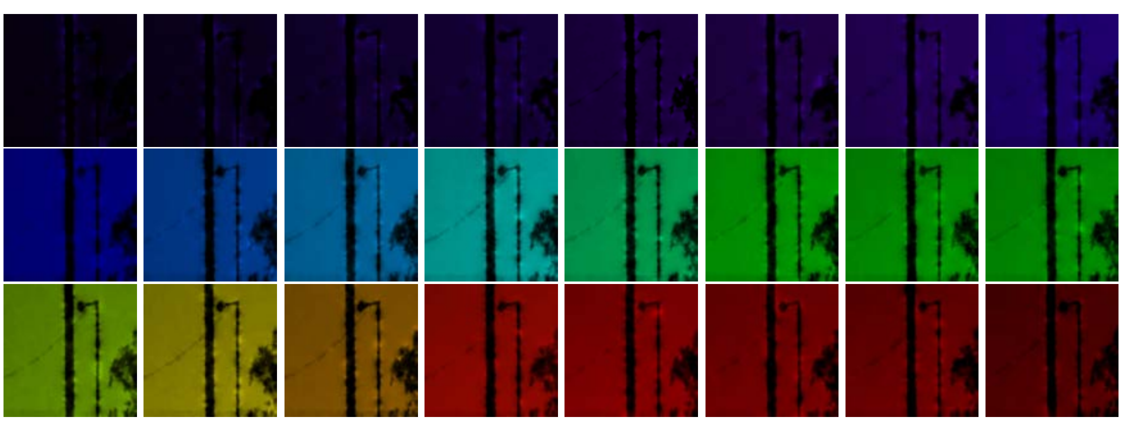

Figure 15. Reconstructed result by GAP-TV for Scene 2, PSNR: 16.58, SSIM: 0.805.

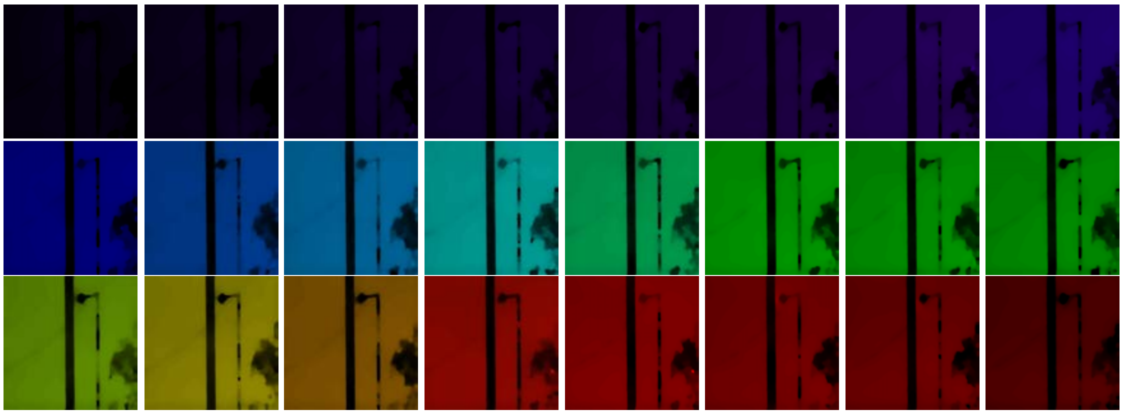

Figure 16. Reconstructed result by TwIST for Scene 2, PSNR: 13.14, SSIM: 0.753.

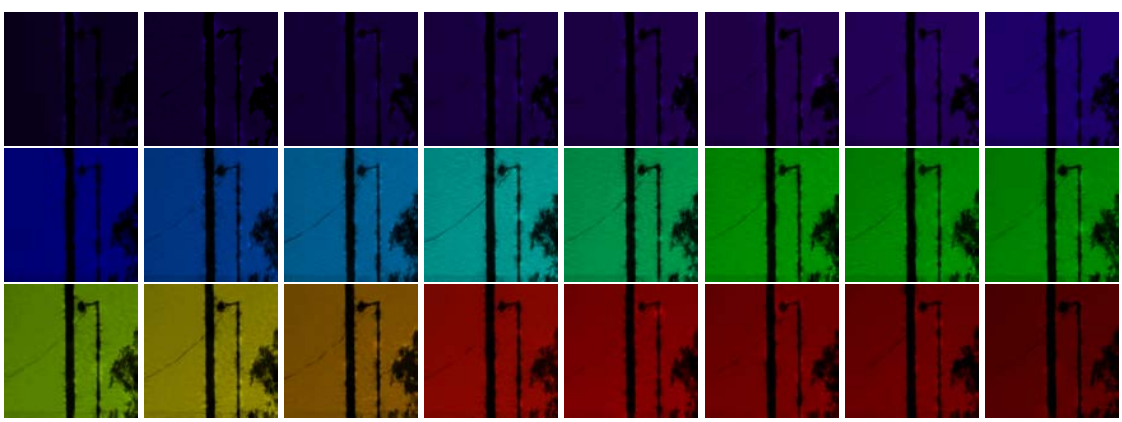

Figure 17. Reconstructed result by DeSCI for Scene 2, PSNR: 22.39, SSIM: 0.806.

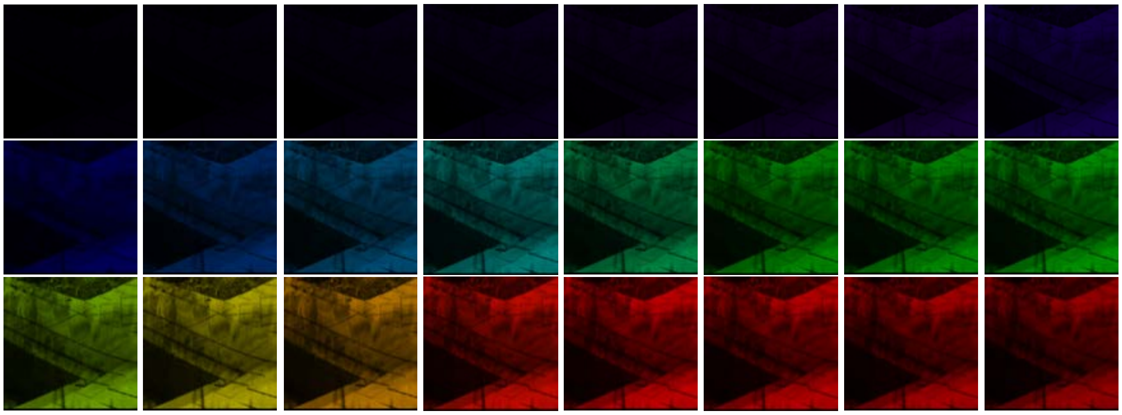

Figure 18. Truth for Scene 3.

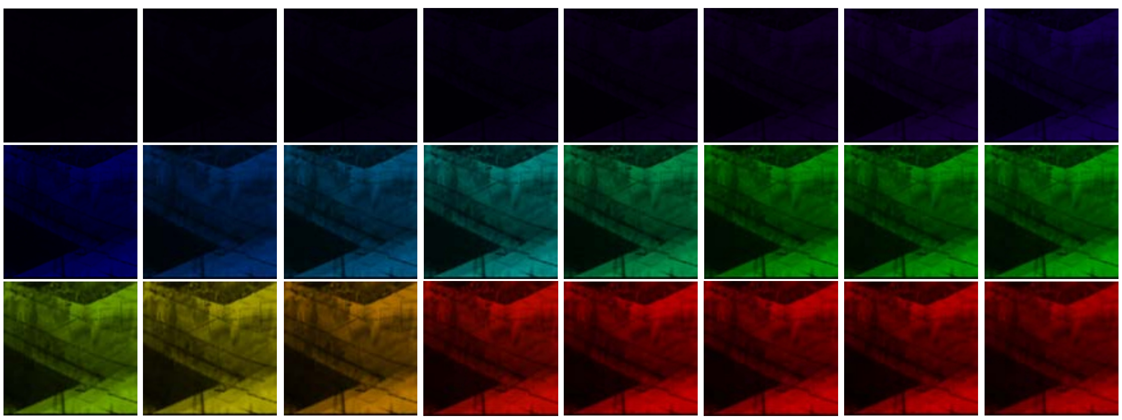

Figure 19. Reconstructed result by λ-net for Scene 3, PSNR: 34.19, SSIM: 0.940.

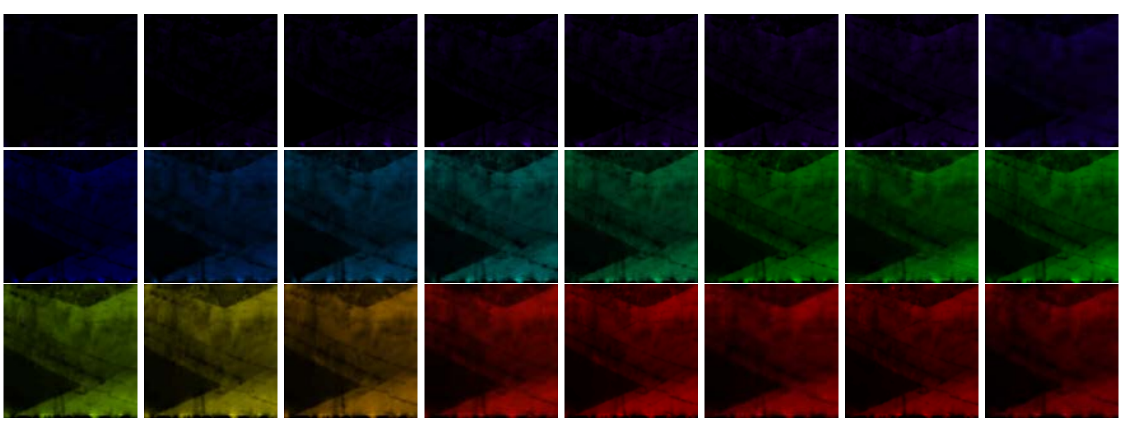

Figure 20. Reconstructed result by GAP-TV for Scene 3, PSNR: 21.48, SSIM: 0.769.

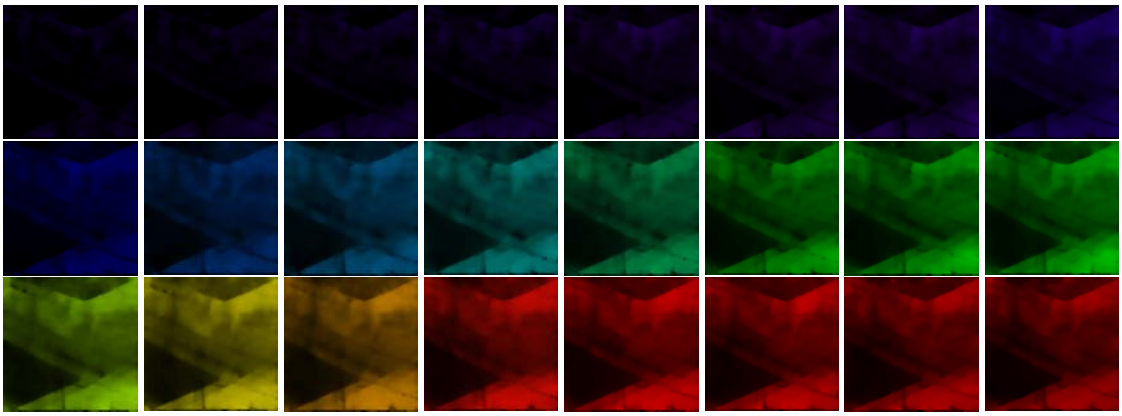

Figure 21. Reconstructed result by TwIST for Scene 3, PSNR: 23.66, SSIM: 0.738.

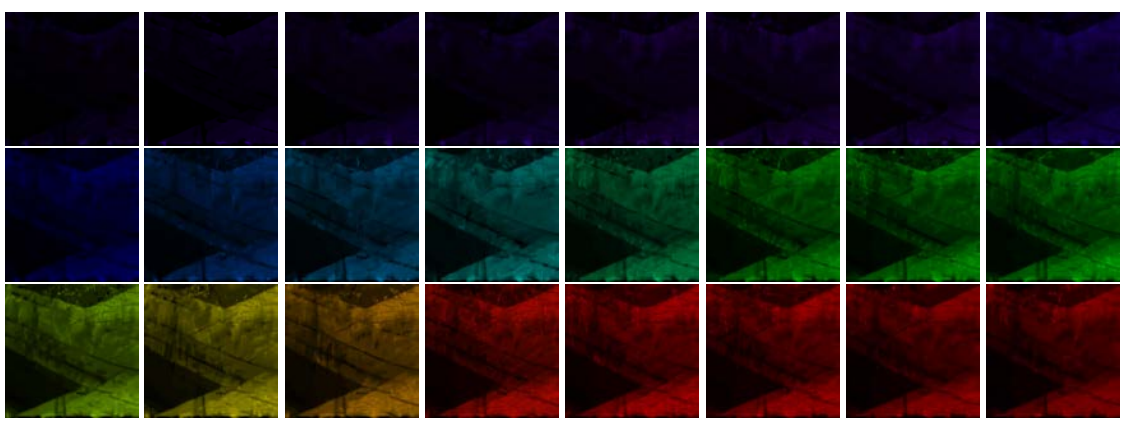

Figure 22. Reconstructed result by DeSCI for Scene 3, PSNR: 24.92, SSIM: 0.822.

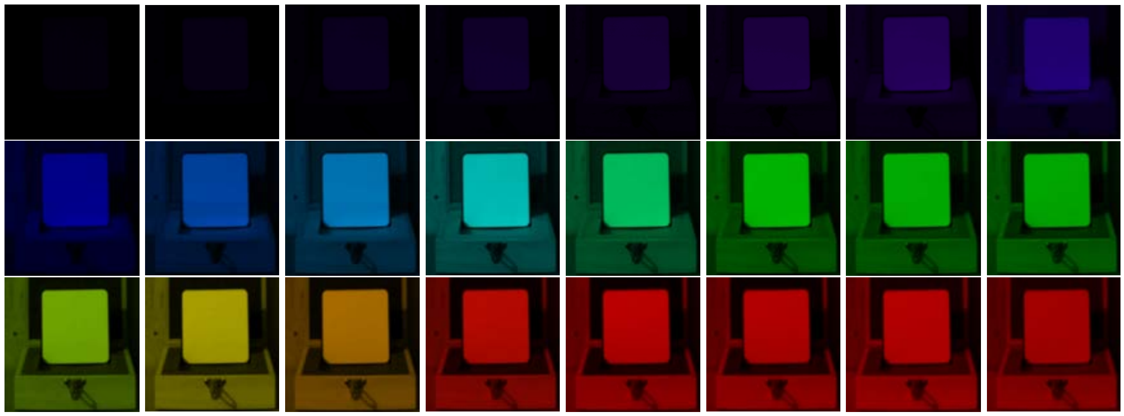

Figure 23. Truth for Scene 4.

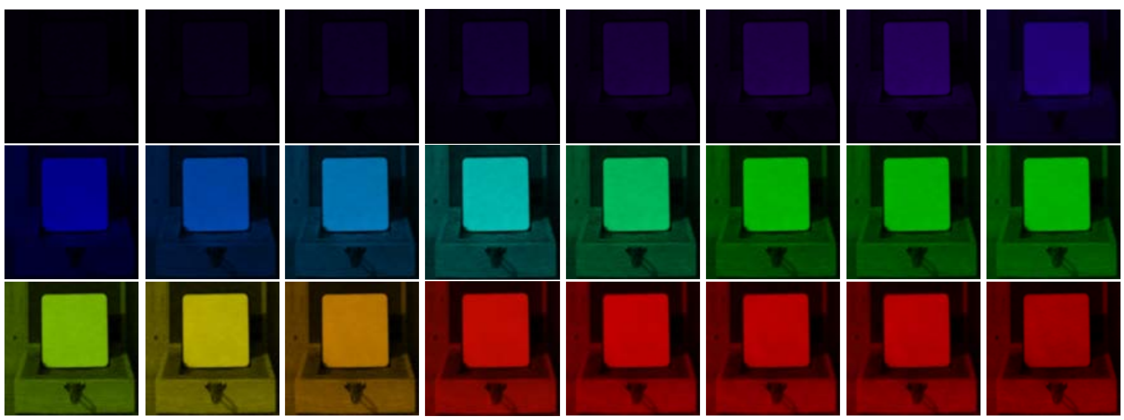

Figure 24. Reconstructed result by λ-net for Scene 4, PSNR: 28.90, SSIM: 0.899.

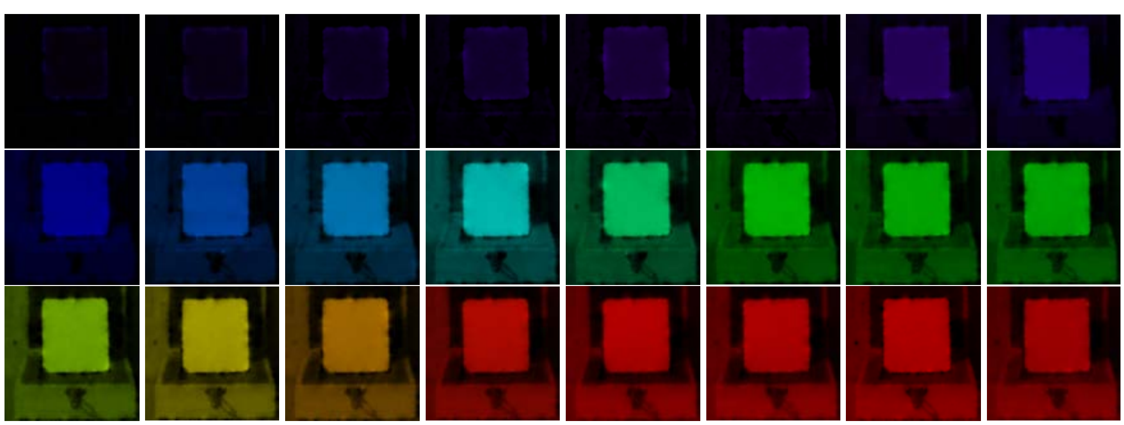

Figure 25. Reconstructed result by GAP-TV for Scene 4, PSNR: 26.49, SSIM: 0.822.

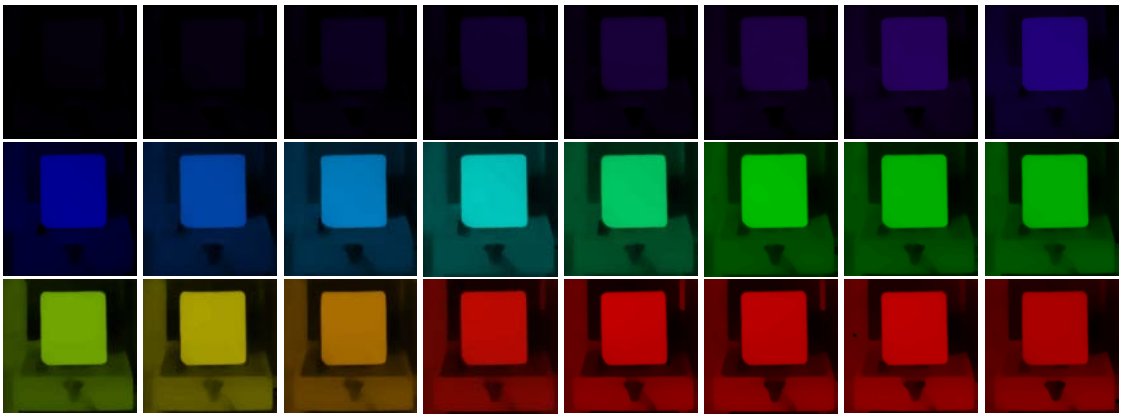

Figure 26. Reconstructed result by TwIST for Scene 4, PSNR: 26.08, SSIM: 0.861.

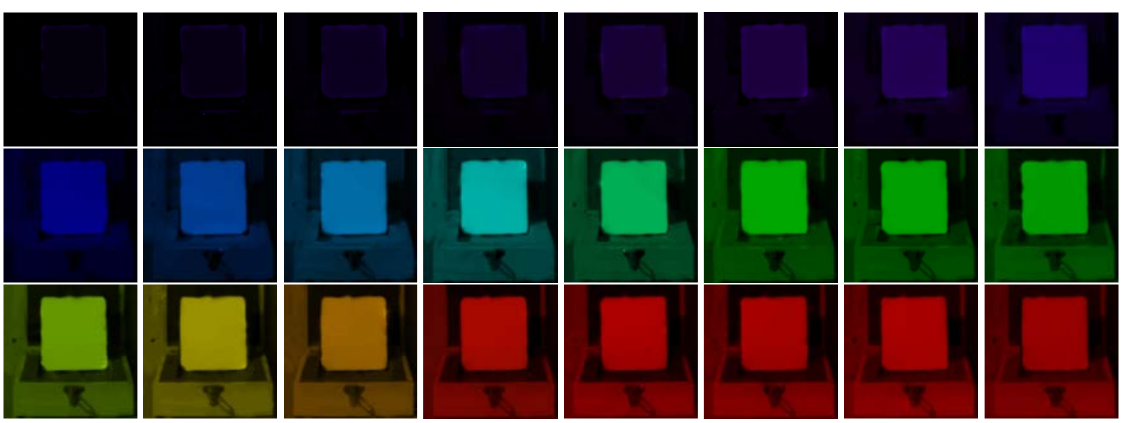

Figure 27. Reconstructed result by DeSCI for Scene 4, PSNR: 29.78, SSIM: 0.907.

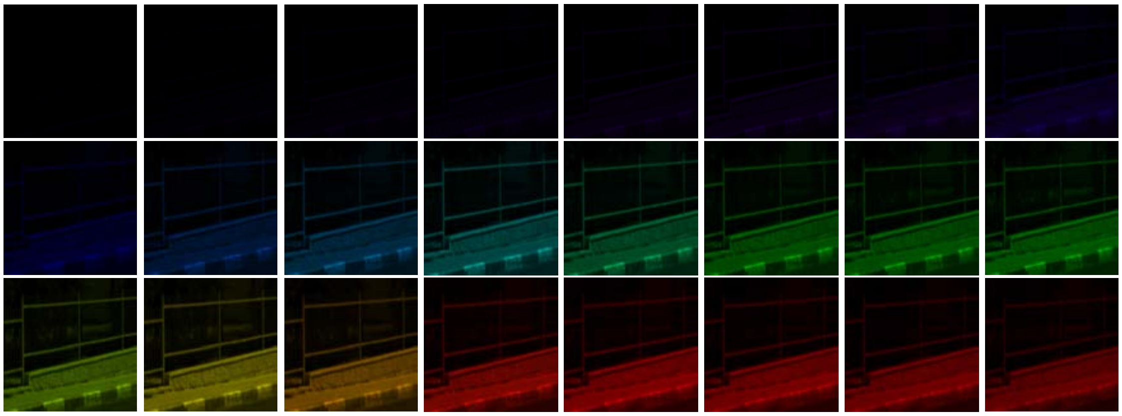

Figure 28. Truth for Scene 5.

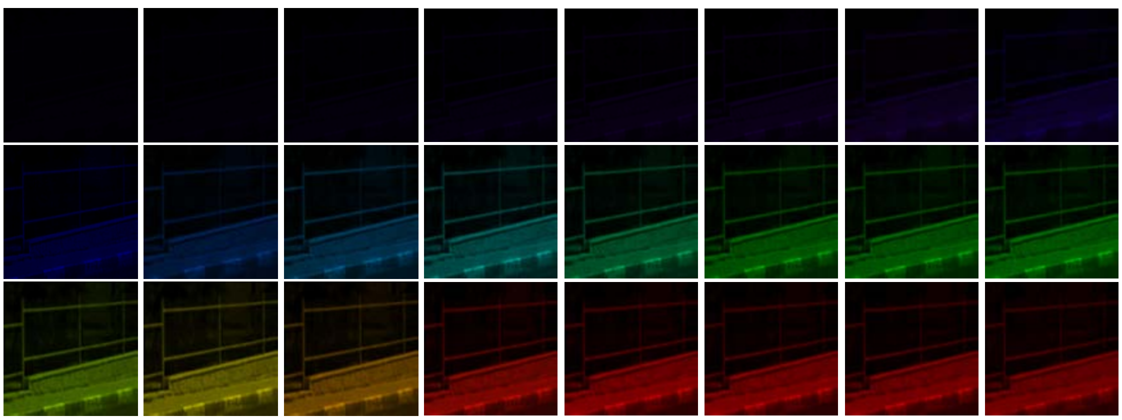

Figure 29. Reconstructed result by λ-net for Scene 5, PSNR: 34.58, SSIM: 0.890.

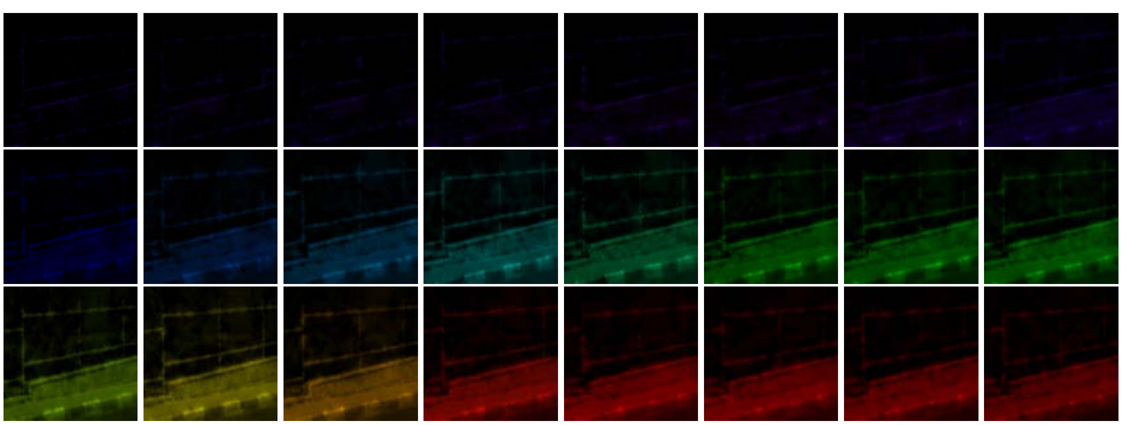

Figure 30. Reconstructed result by GAP-TV for Scene 5, PSNR: 26.63, SSIM: 0.688.

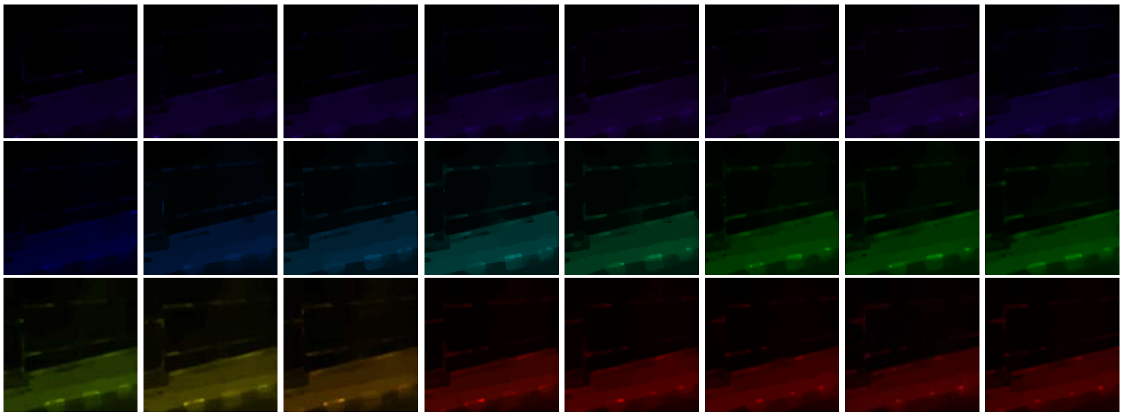

Figure 31. Reconstructed result by TwIST for Scene 5, PSNR: 22.54, SSIM: 0.695.

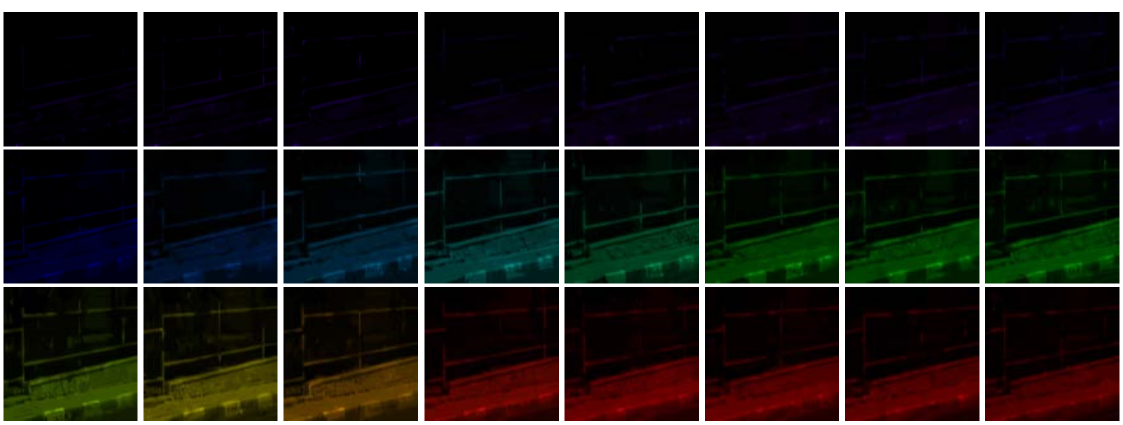

Figure 32. Reconstructed result by DeSCI for Scene 5, PSNR: 29.02, SSIM: 0.844.

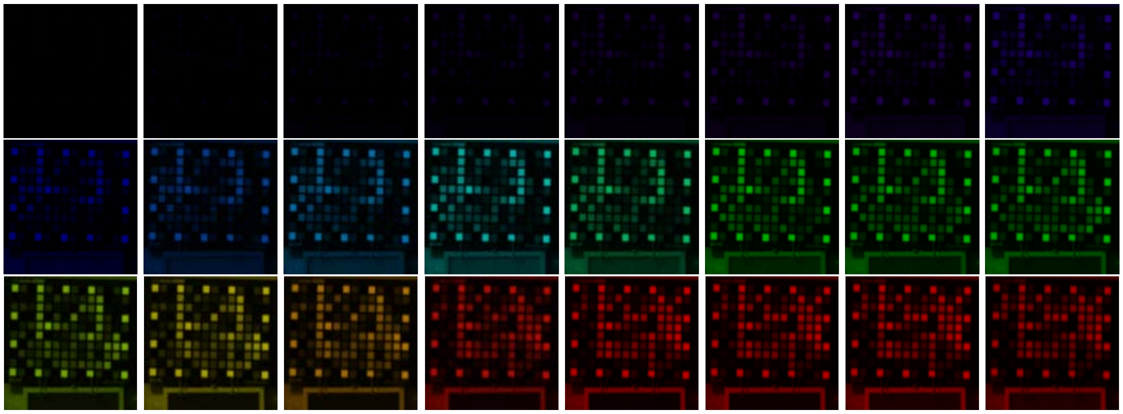

Figure 33. Truth for Scene 6.

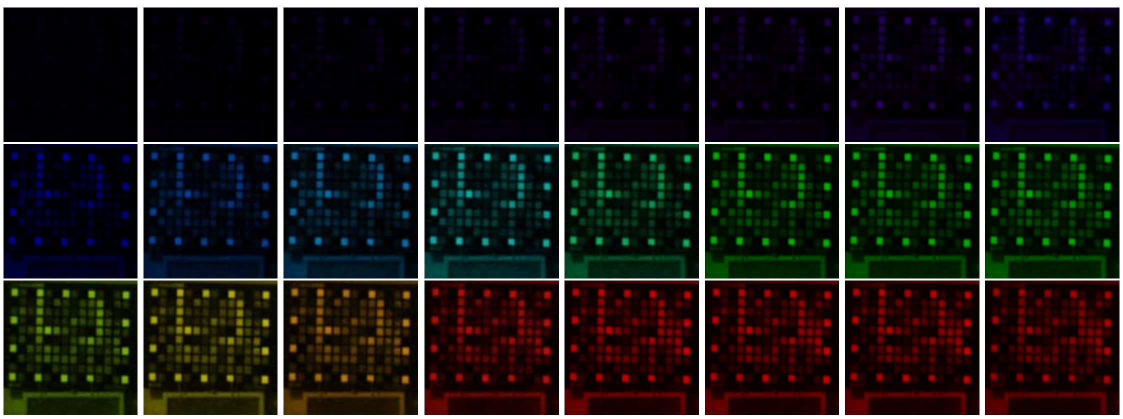

Figure 34. Reconstructed result by λ-net for Scene 6, PSNR: 28.09, SSIM: 0.858.

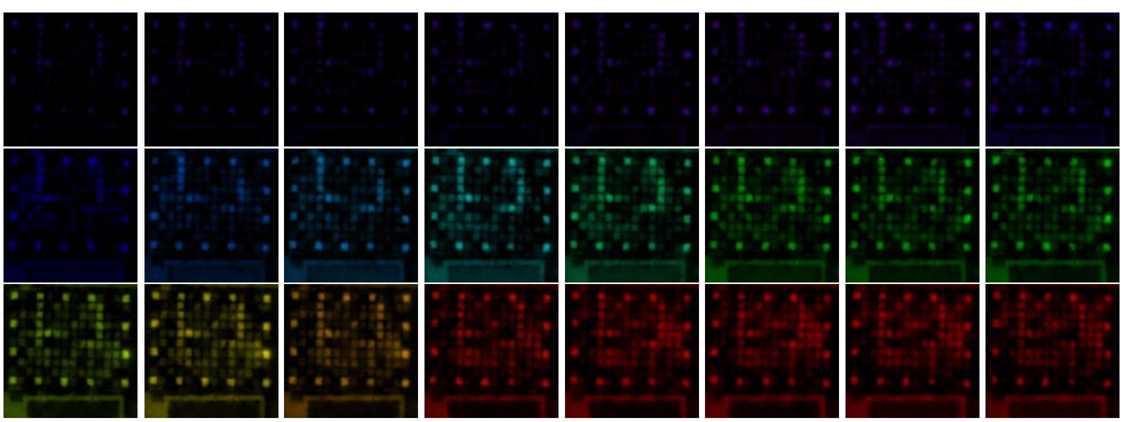

Figure 35. Reconstructed result by GAP-TV for Scene 6, PSNR: 22.81, SSIM: 0.614.

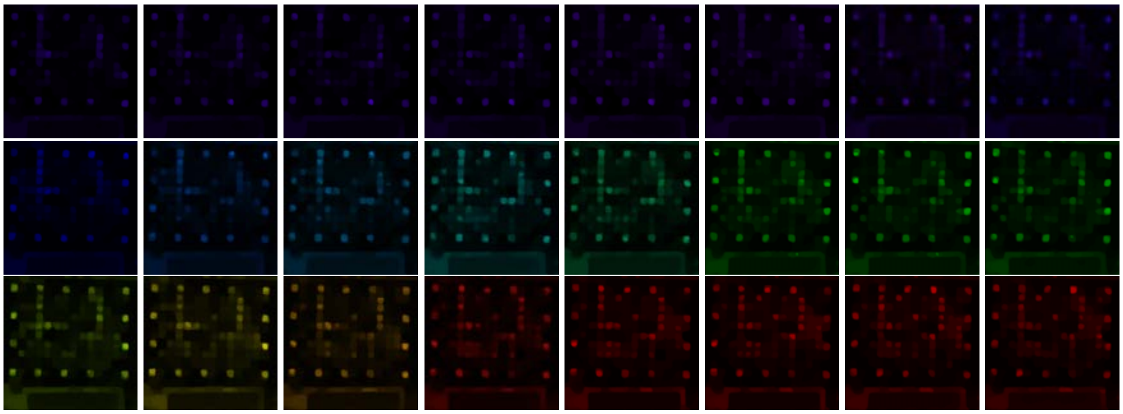

Figure 36. Reconstructed result by TwIST for Scene 6, PSNR: 20.11, SSIM: 0.662.

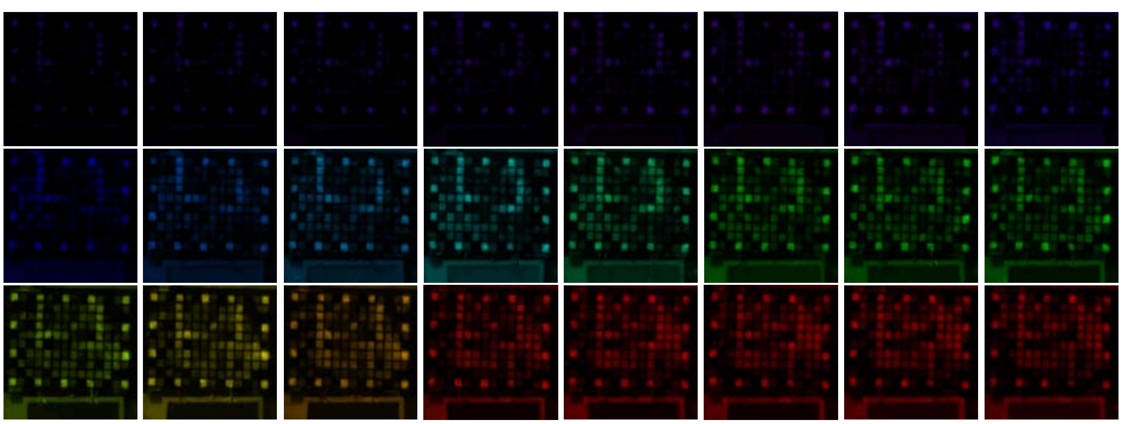

Figure 37. Reconstructed result by DeSCI for Scene 6, PSNR: 24.75, SSIM: 0.797.

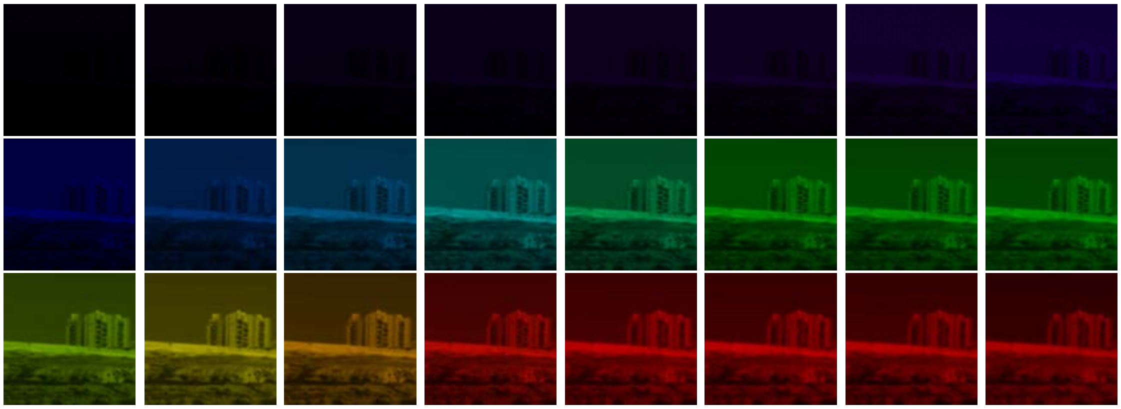

Figure 38. Truth for Scene 7.

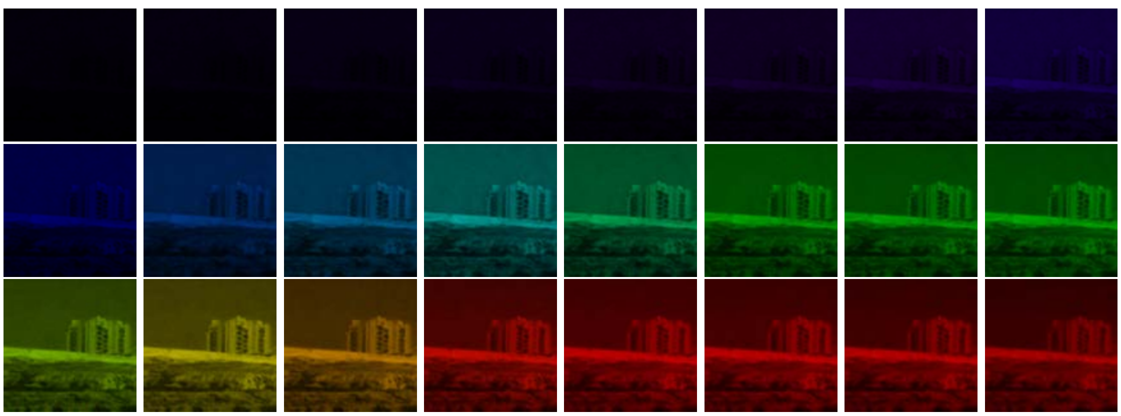

Figure 39. Reconstructed result by λ-net for Scene 7, PSNR: 36.15, SSIM: 0.942.

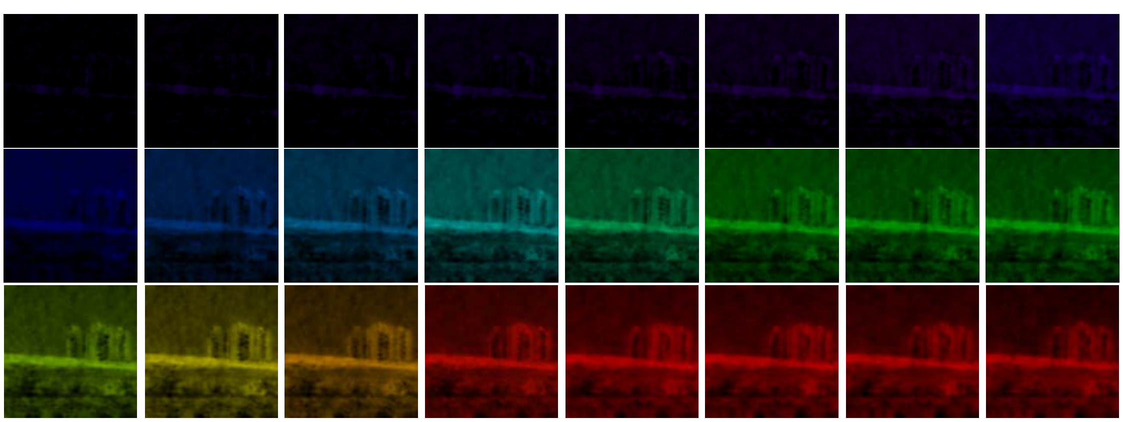

Figure 40. Reconstructed result by GAP-TV for Scene 7, PSNR: 24.95, SSIM: 0.699.

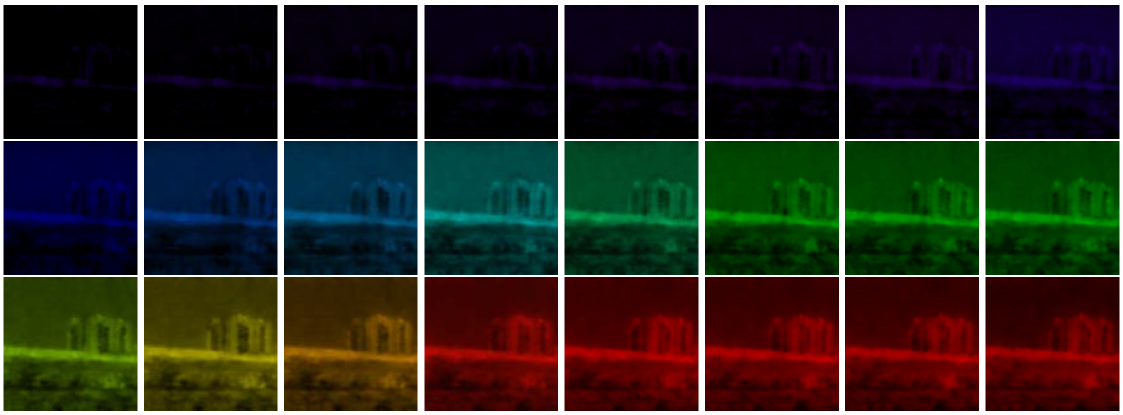

Figure 41. Reconstructed result by TwIST for Scene 7, PSNR: 26.20, SSIM: 0.753.

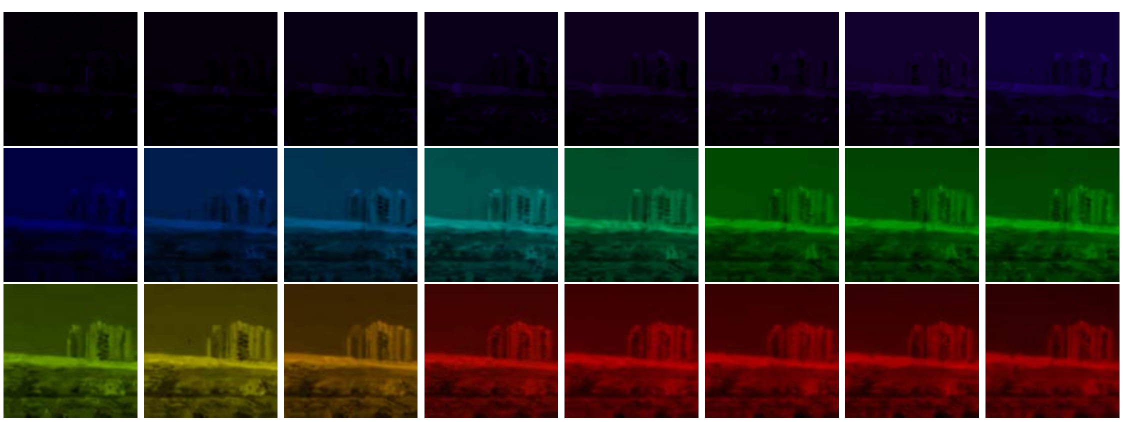

Figure 42. Reconstructed result by DeSCI for Scene 7, PSNR: 29.68, SSIM: 0.881.

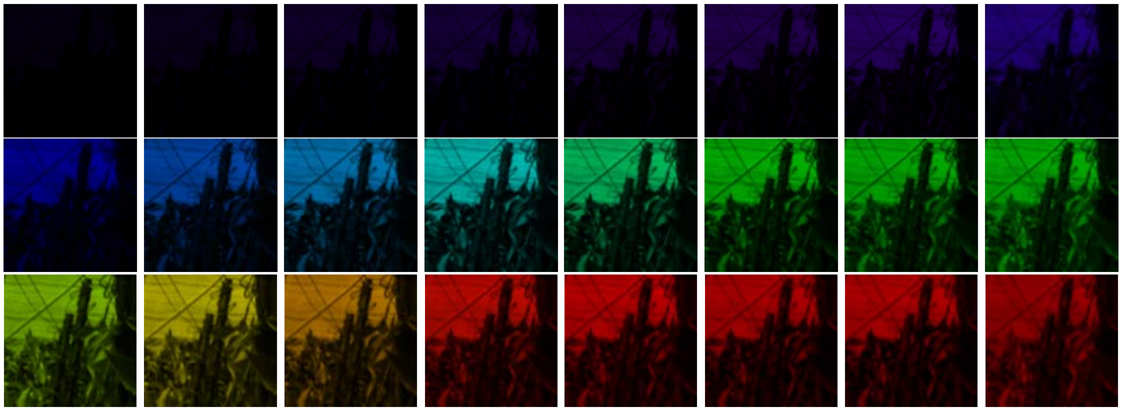

Figure 43. Truth for Scene 8.

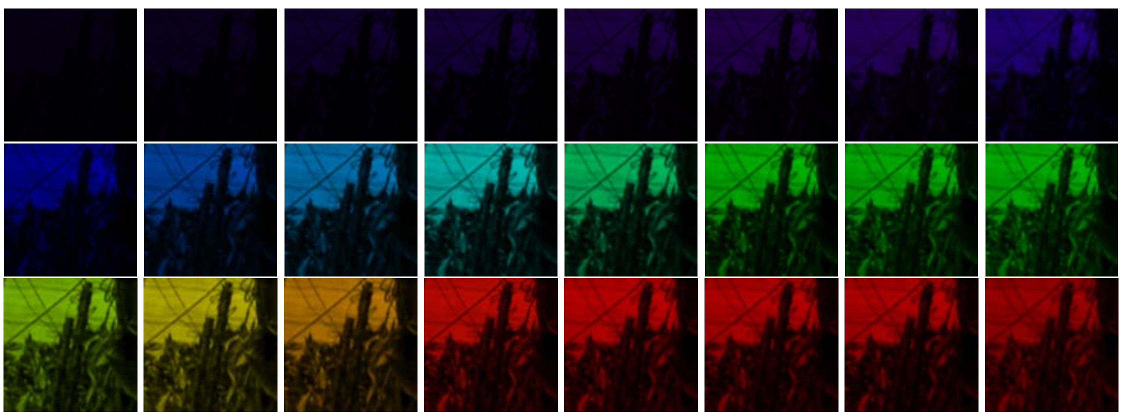

Figure 44. Reconstructed result by λ-net for Scene 8, PSNR: 32.64, SSIM: 0.909.

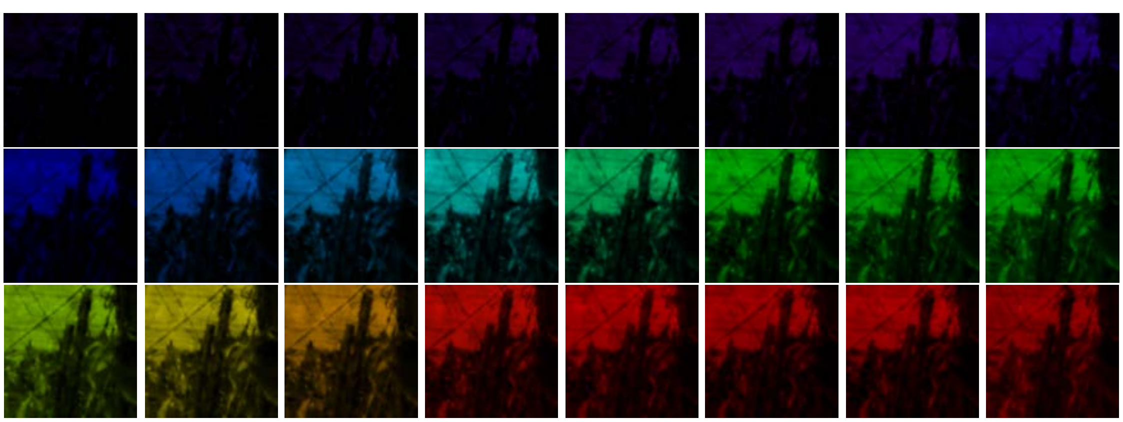

Figure 45. Reconstructed result by GAP-TV for Scene 8, PSNR: 21.26, SSIM: 0.695.

Figure 46. Reconstructed result by TwIST for Scene 8, PSNR: 18.38, SSIM: 0.643.

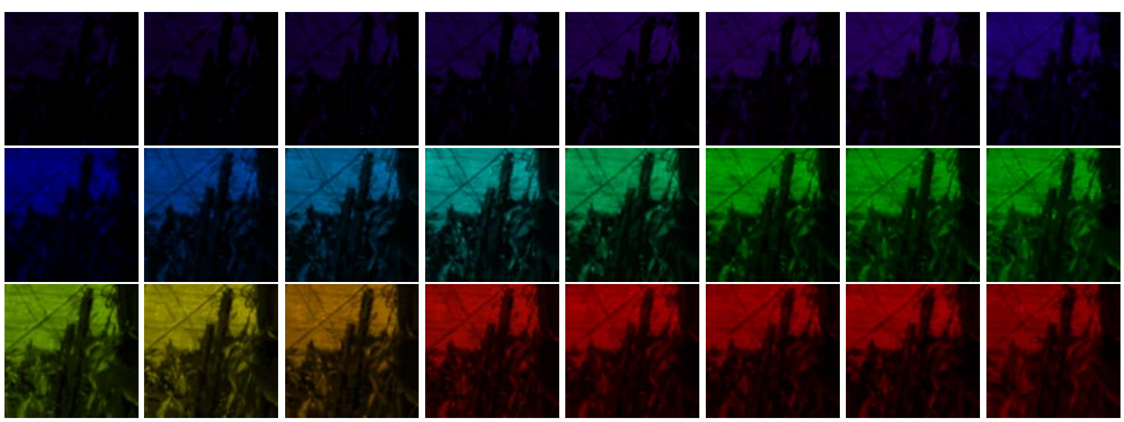

Figure 47. Reconstructed result by DeSCI for Scene 8, PSNR: 25.58, SSIM: 0.823.

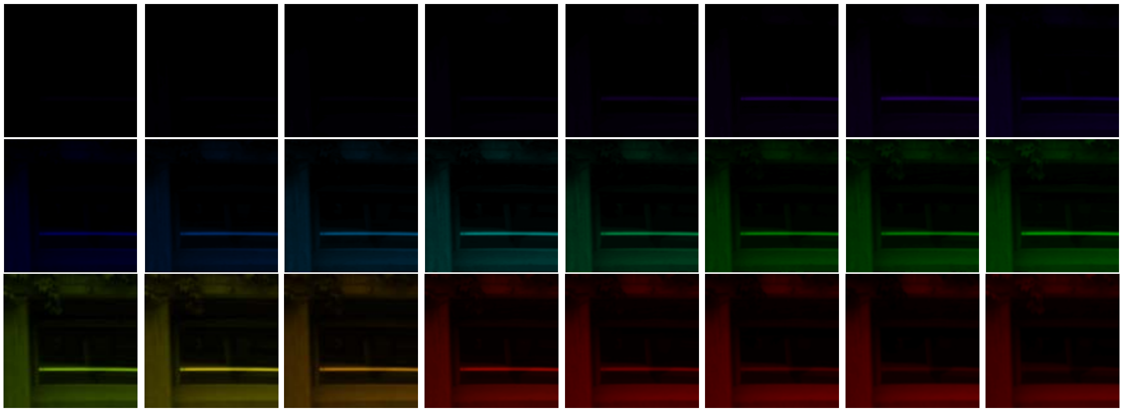

Figure 48. Truth for Scene 9.

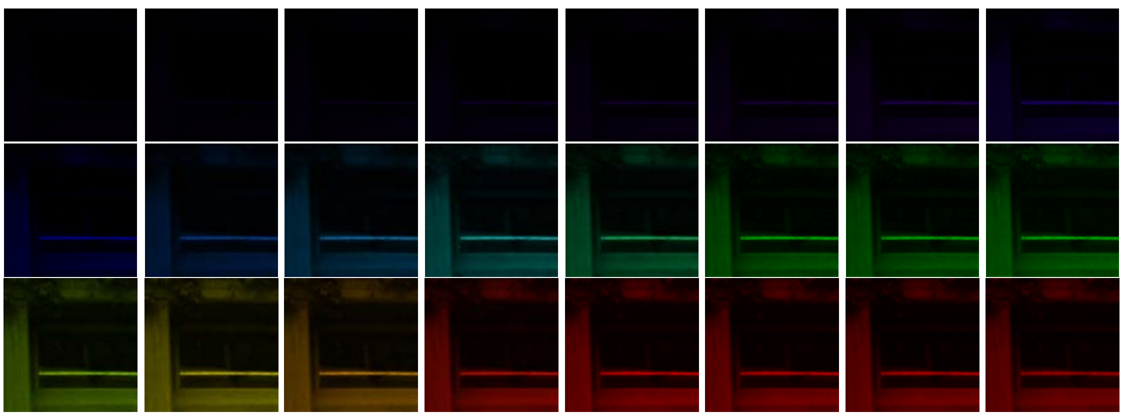

Figure 49. Reconstructed result by λ-net for Scene 9, PSNR: 33.83, SSIM: 0.912.

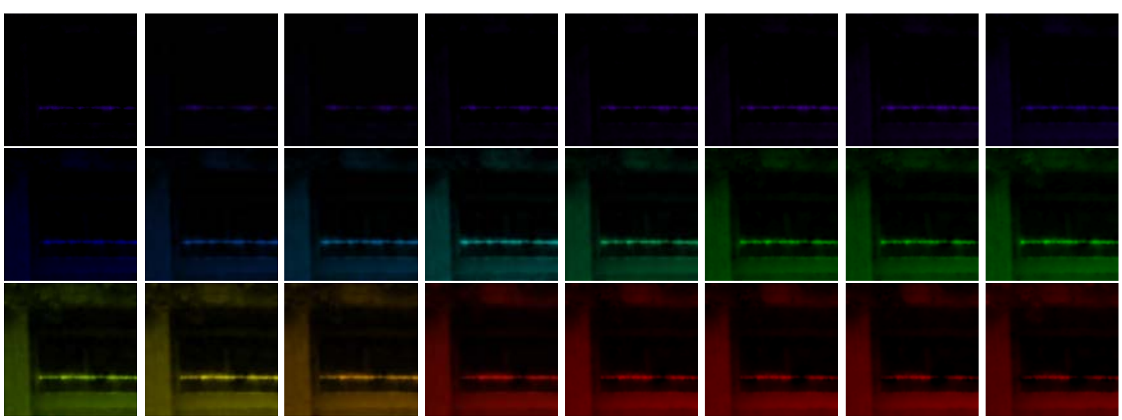

Figure 50. Reconstructed result by GAP-TV for Scene 9, PSNR: 29.94, SSIM: 0.812.

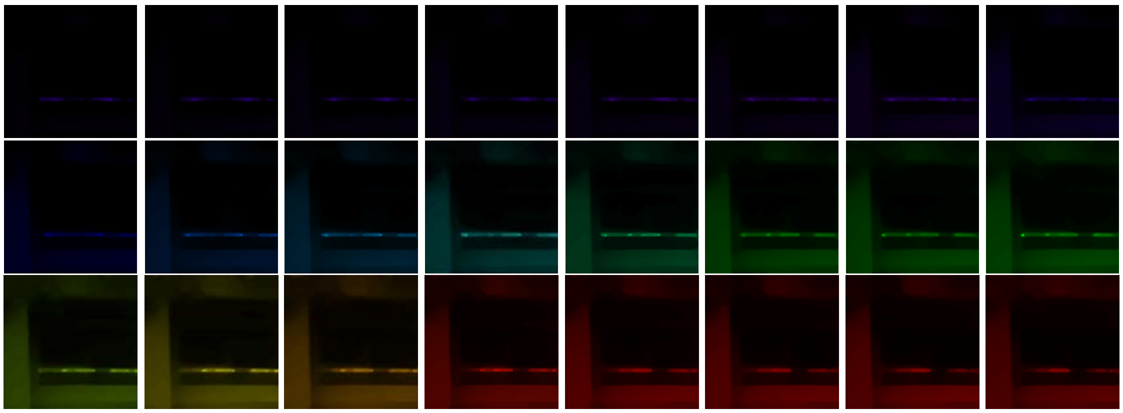

Figure 51. Reconstructed result by TwIST for Scene 9, PSNR: 28.09, SSIM: 0.807.

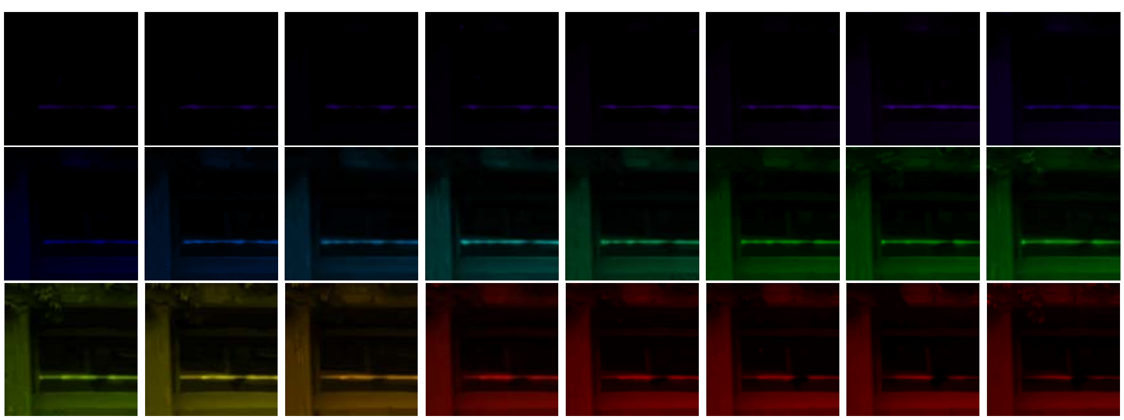

Figure 52. Reconstructed result by DeSCI for Scene 9, PSNR: 32.86, SSIM: 0.937.

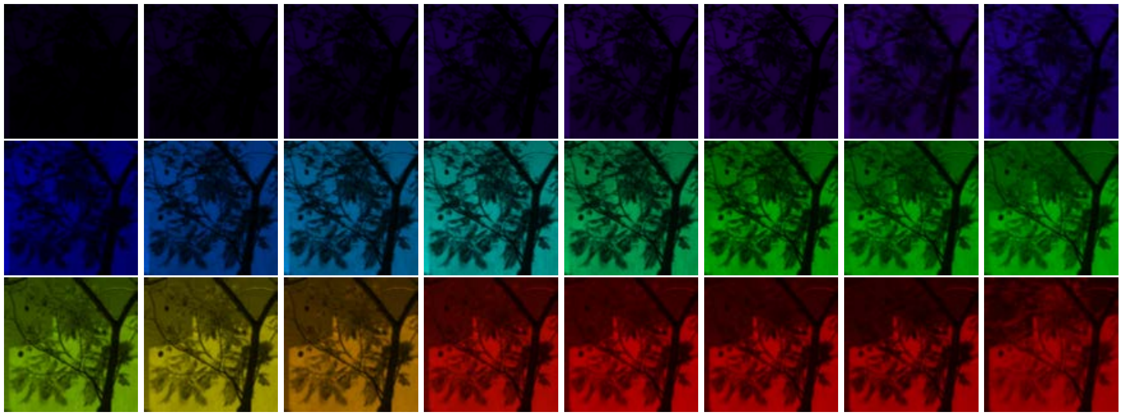

Figure 53. Truth for Scene 10.

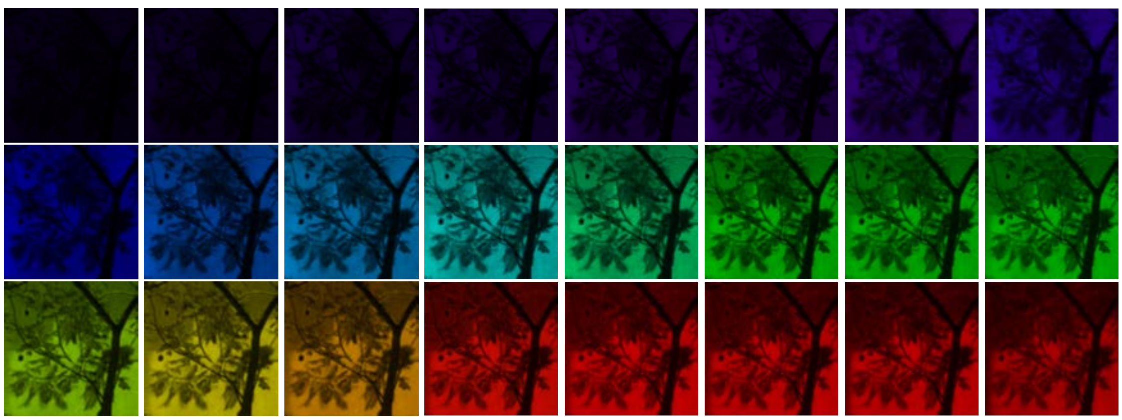

Figure 54. Reconstructed result by λ-net for Scene 10, PSNR: 28.63, SSIM: 0.877.

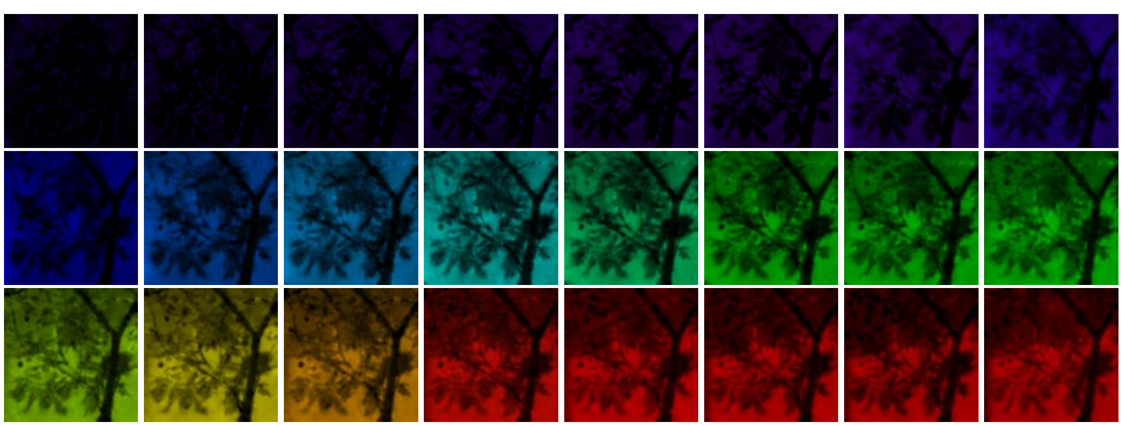

Figure 55. Reconstructed result by GAP-TV for Scene 10, PSNR: 23.04, SSIM: 0.706.

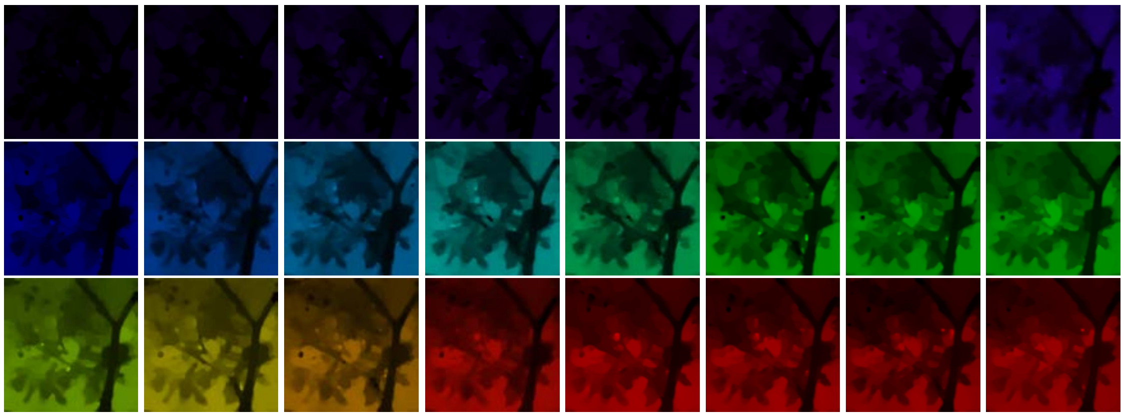

Figure 56. Reconstructed result by TwIST for Scene 10, PSNR: 20.84, SSIM: 0.620.

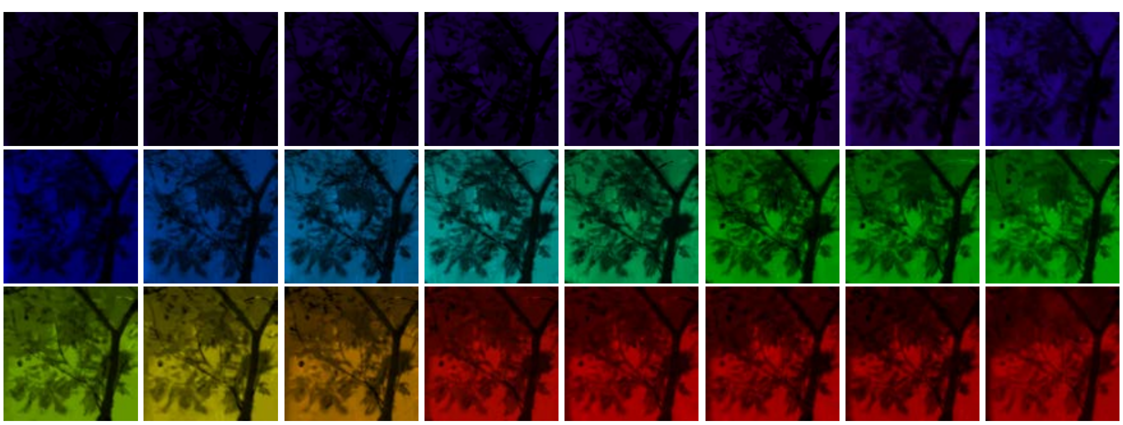

Figure 57. Reconstructed result by DeSCI for Scene 10, PSNR: 24.00, SSIM: 0.748.

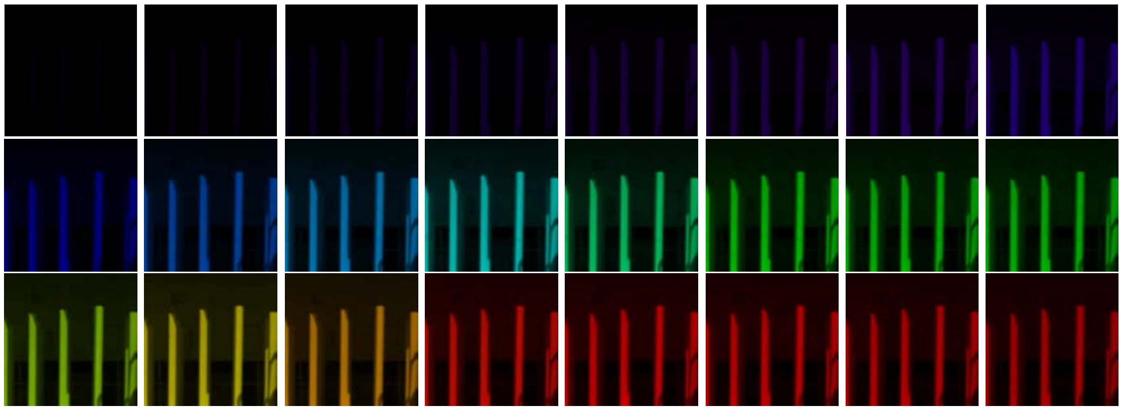

Figure 58. Truth for Scene 11.

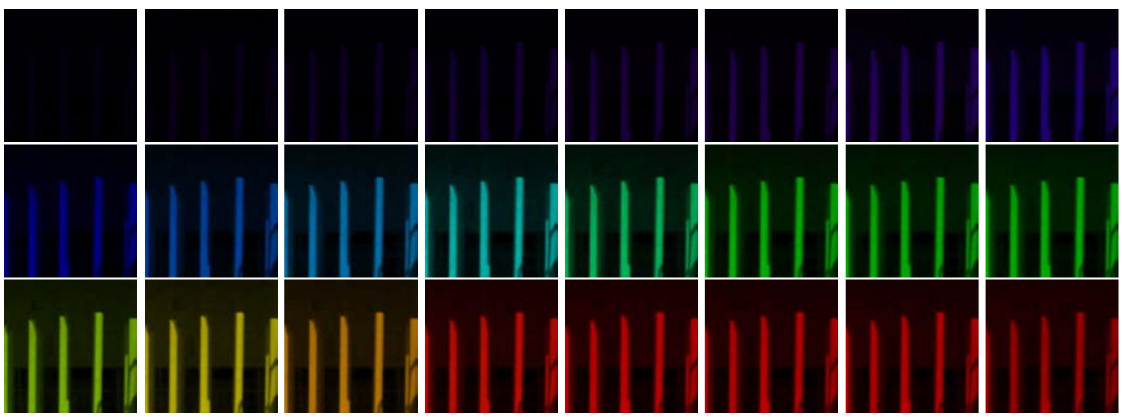

Figure 59. Reconstructed result by λ-net for Scene 11, PSNR: 35.21, SSIM: 0.946.

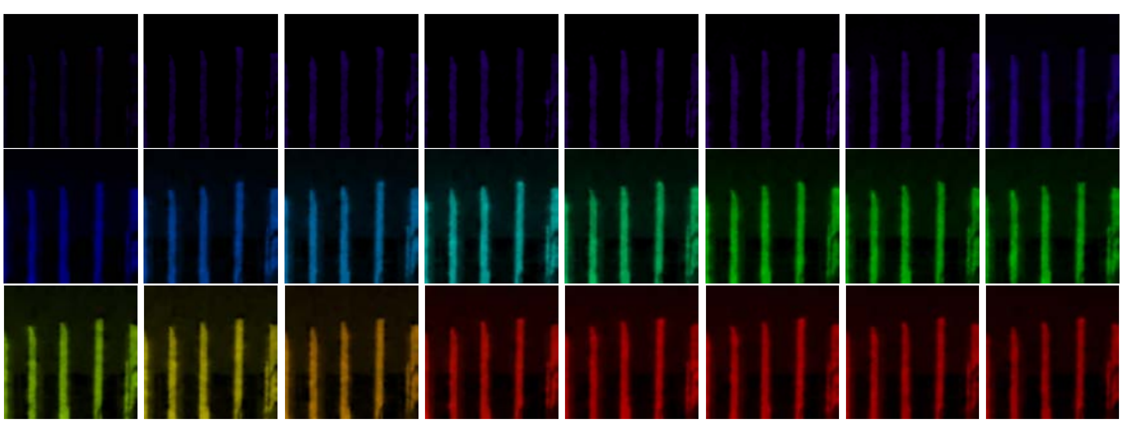

Figure 60. Reconstructed result by GAP-TV for Scene 11, PSNR: 24.07, SSIM: 0.754.

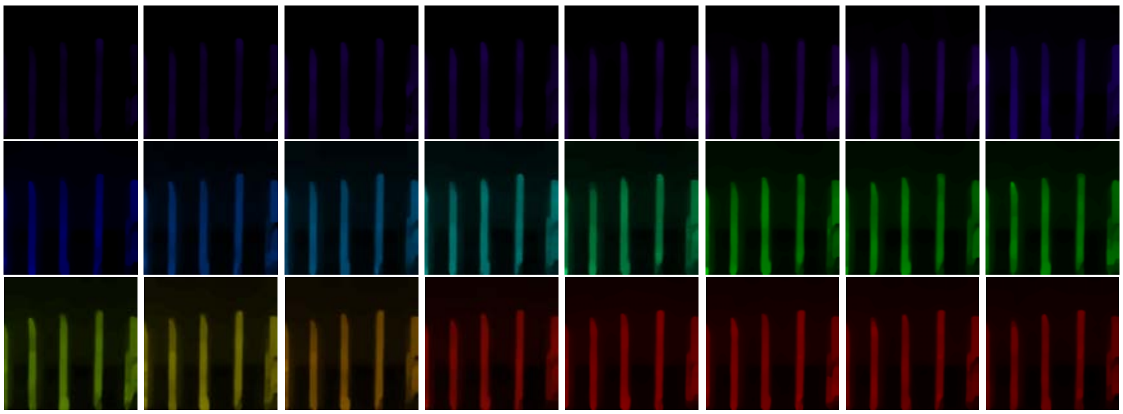

Figure 61. Reconstructed result by TwIST for Scene 11, PSNR: 21.75, SSIM: 0.785.

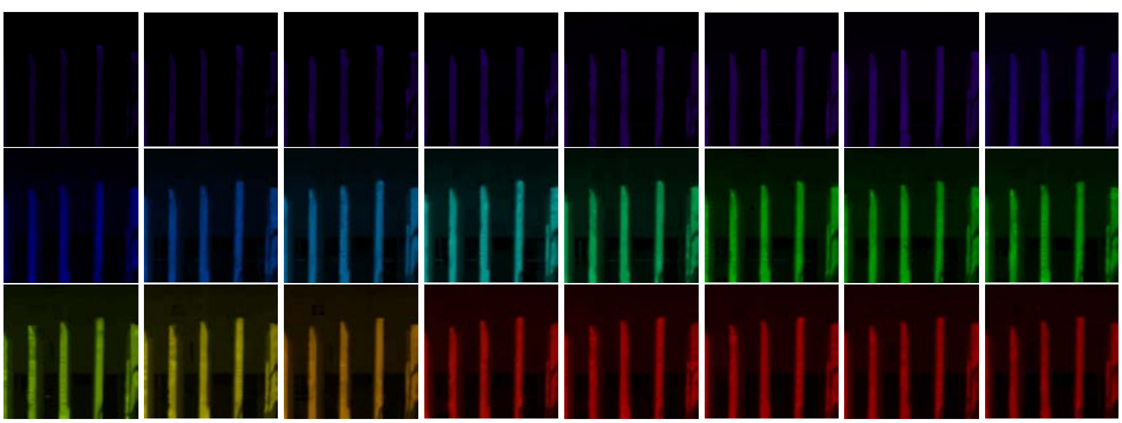

Figure 62. Reconstructed result by DeSCI for Scene 11, PSNR: 28.19, SSIM: 0.912.

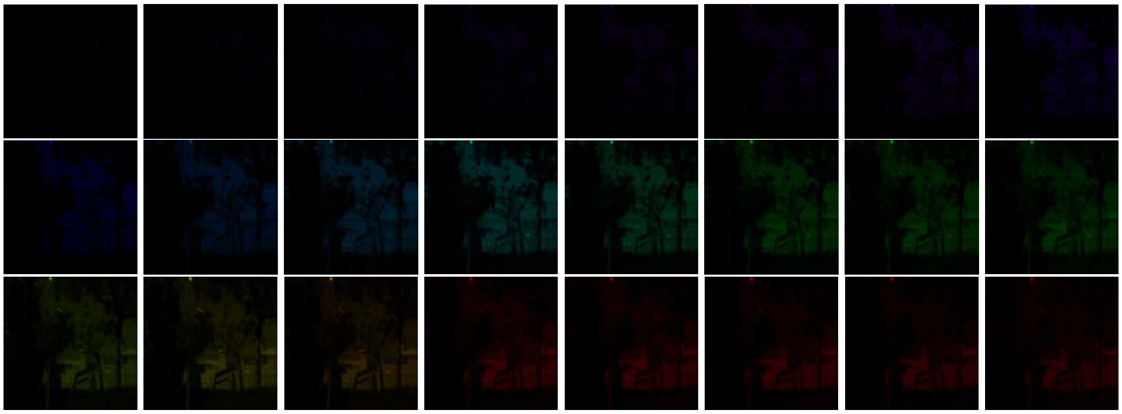

Figure 63. Truth for Scene 12.

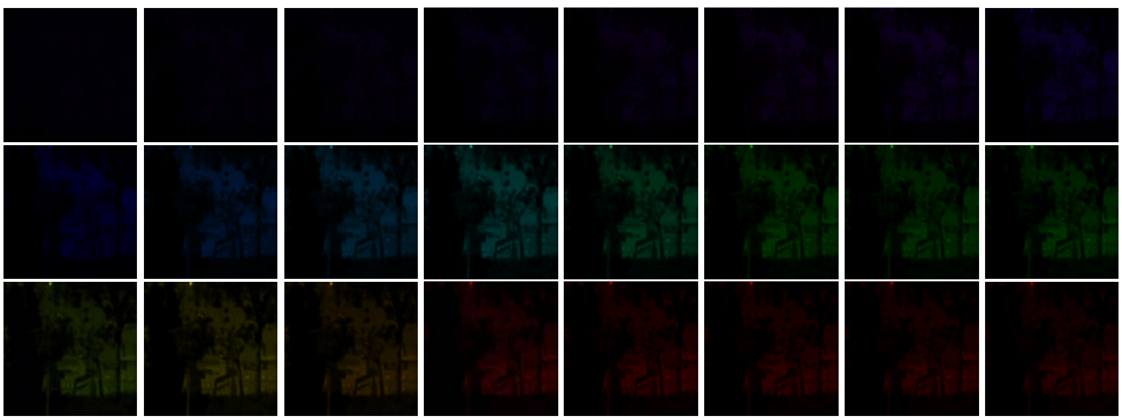

Figure 64. Reconstructed result by λ-net for Scene 12, PSNR: 34.77, SSIM: 0.823.

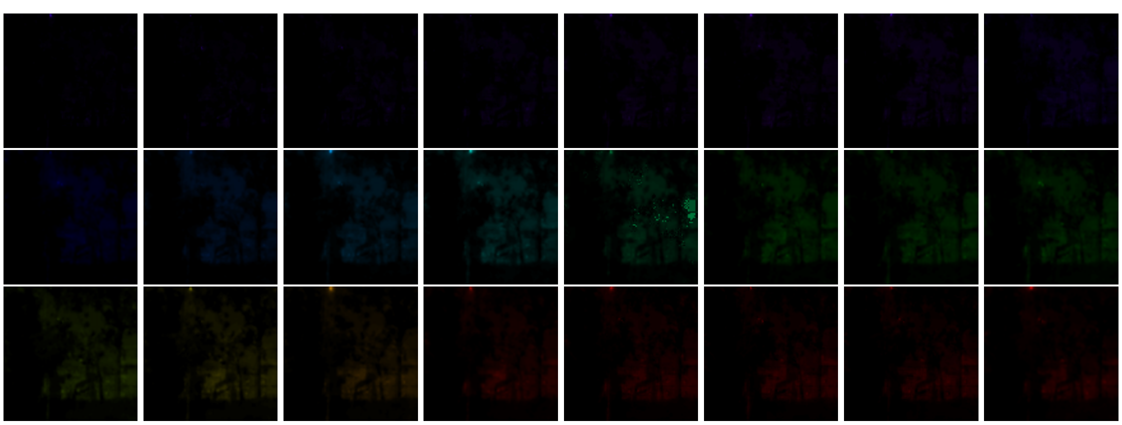

Figure 65. Reconstructed result by GAP-TV for Scene 12, PSNR: 28.99, SSIM: 0.758.

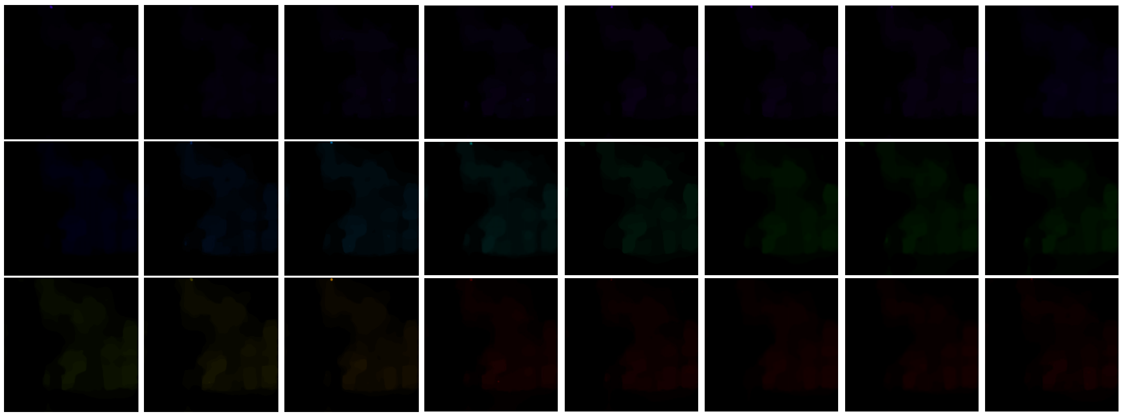

Figure 66. Reconstructed result by TwIST for Scene 12, PSNR: 26.75, SSIM: 0.699.

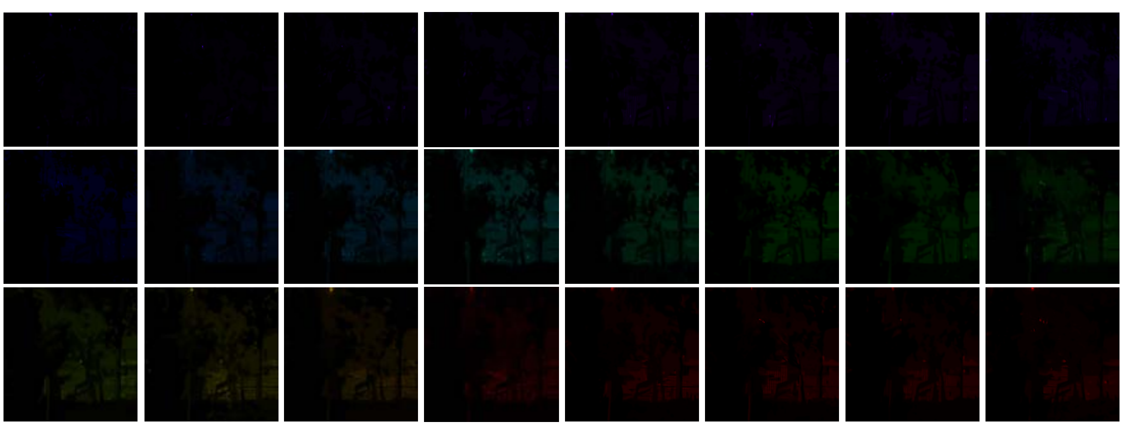

Figure 67. Reconstructed result by DeSCI for Scene 12, PSNR: 31.80, SSIM: 0.863.

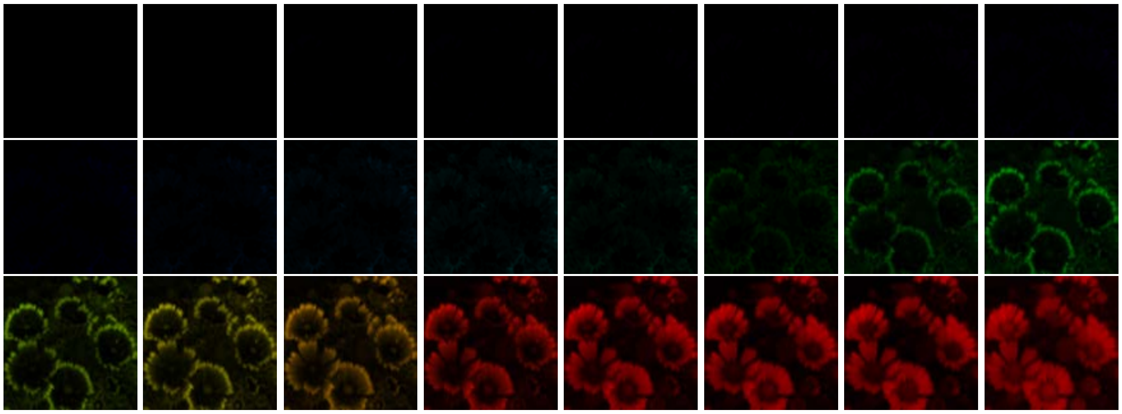

Figure 68. Truth for Scene 13.

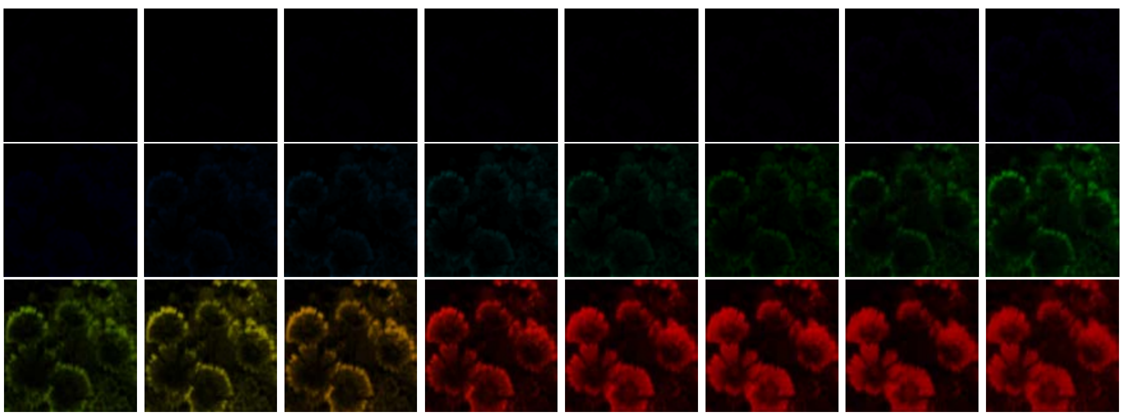

Figure 69. Reconstructed result by λ-net for Scene 13, PSNR: 32.07, SSIM: 0.844.

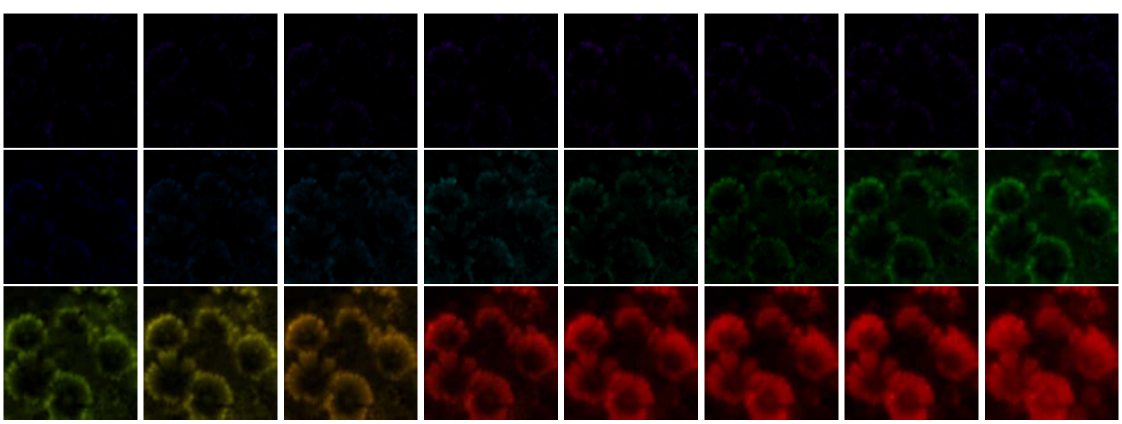

Figure 70. Reconstructed result by GAP-TV for Scene 13, PSNR: 27.57, SSIM: 0.650.

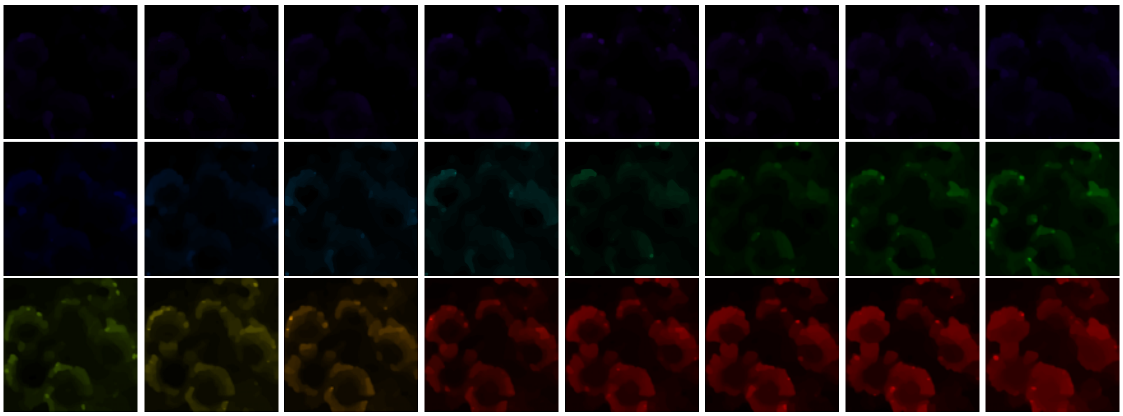

Figure 71. Reconstructed result by TwIST for Scene 13, PSNR: 24.54, SSIM: 0.718.

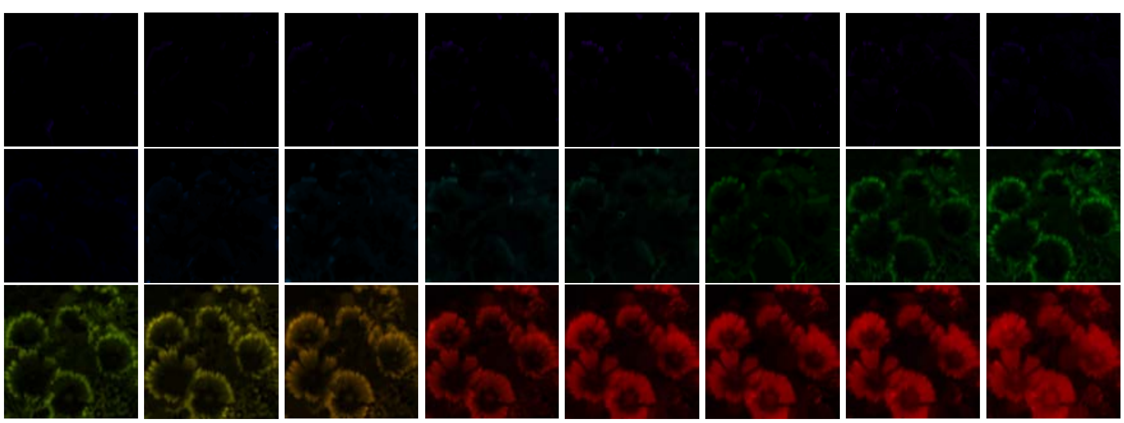

Figure 72. Reconstructed result by DeSCI for Scene 13, PSNR: 30.91, SSIM: 0.823.

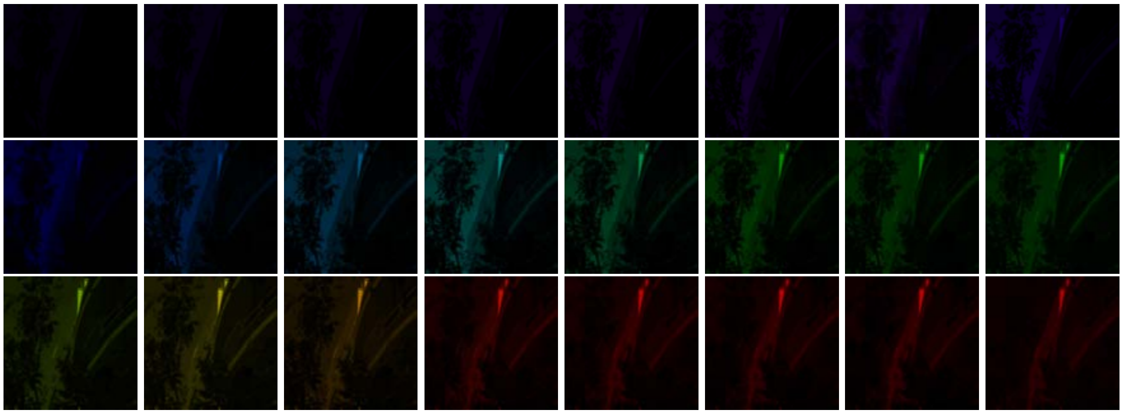

Figure 73. Truth for Scene 14.

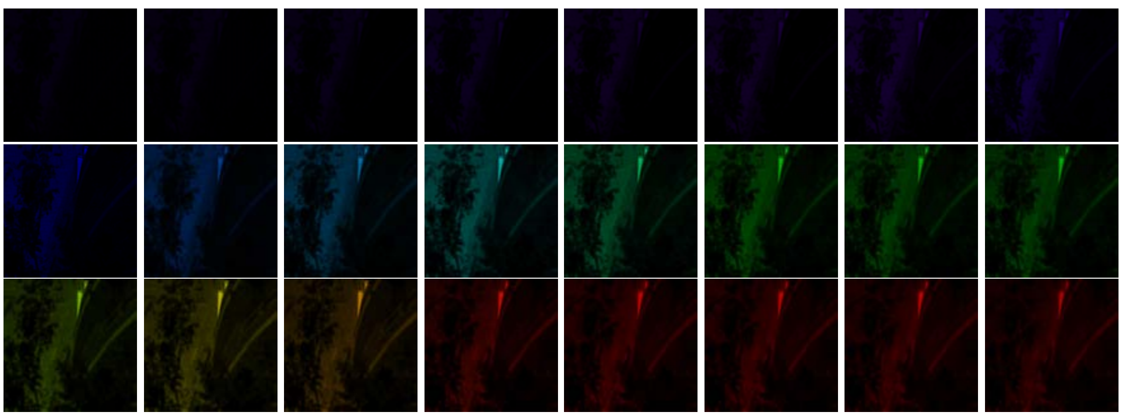

Figure 74. Reconstructed result by λ-net for Scene 14, PSNR: 33.73, SSIM: 0.869.

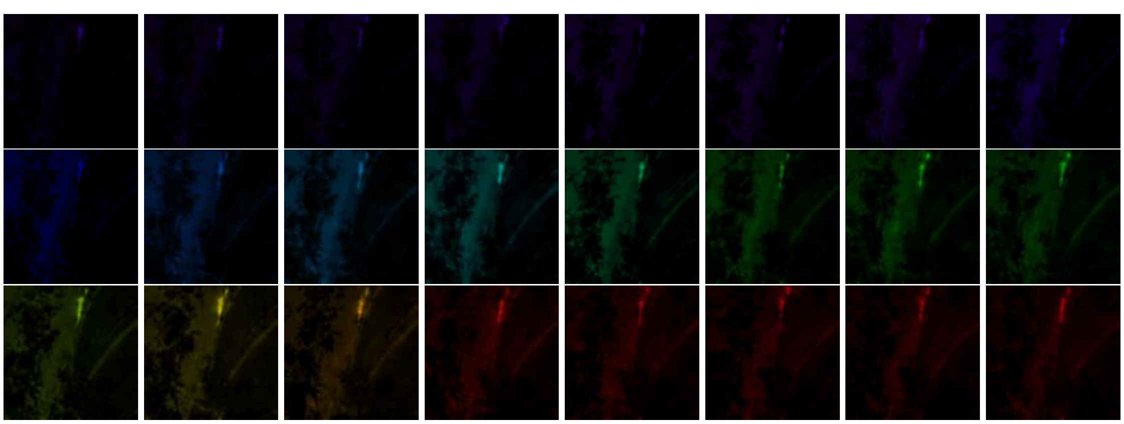

Figure 75. Reconstructed result by GAP-TV for Scene 14, PSNR: 28.54, SSIM: 0.764.

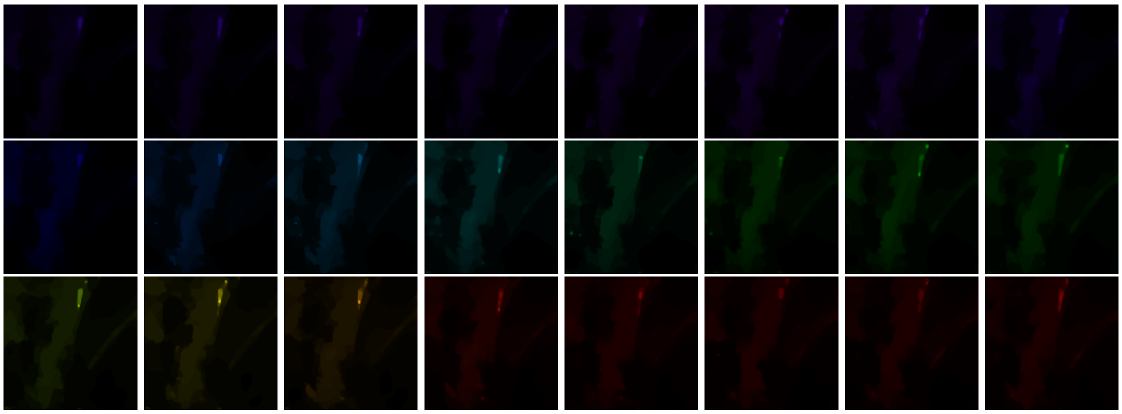

Figure 76. Reconstructed result by TwIST for Scene 14, PSNR: 26.27, SSIM: 0.765.

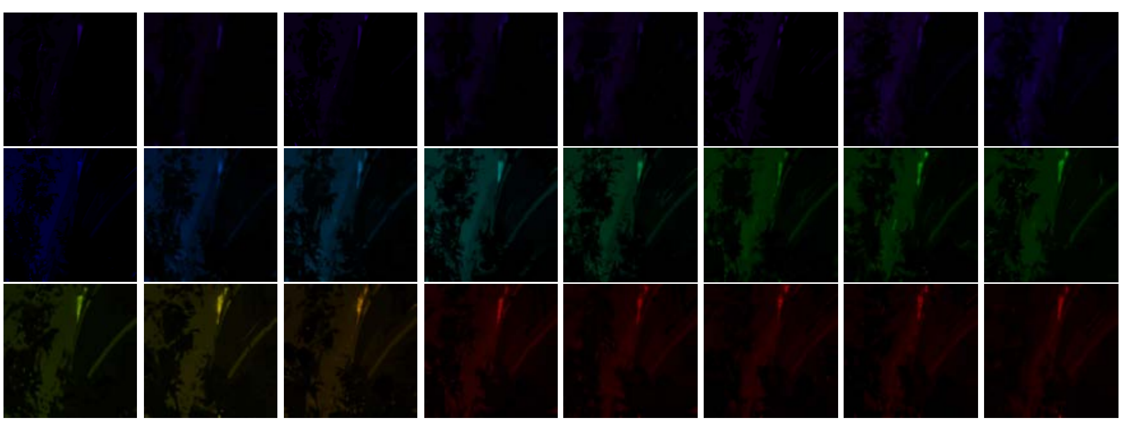

Figure 77. Reconstructed result by DeSCI for Scene 14, PSNR: 29.69, SSIM: 0.852.

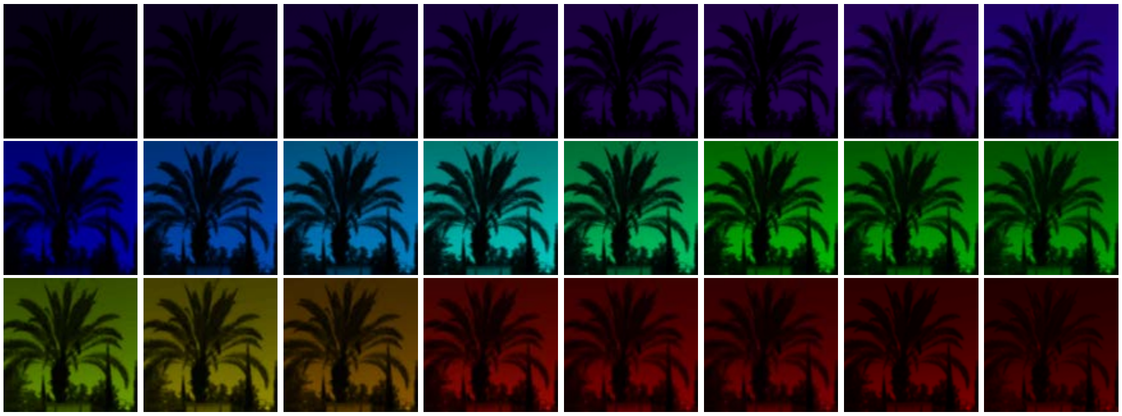

Figure 78. Truth for Scene 15.

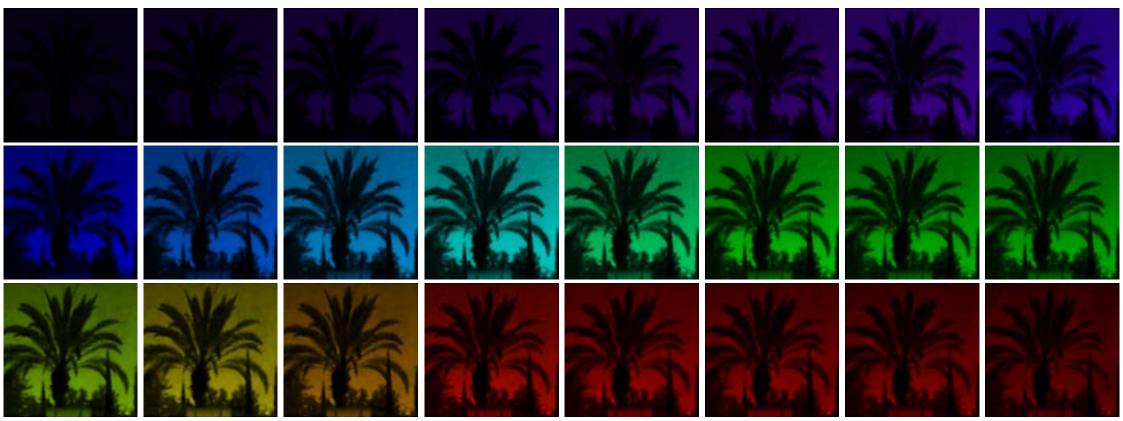

Figure 79. Reconstructed result by λ-net for Scene 15, PSNR: 29.88, SSIM: 0.913.

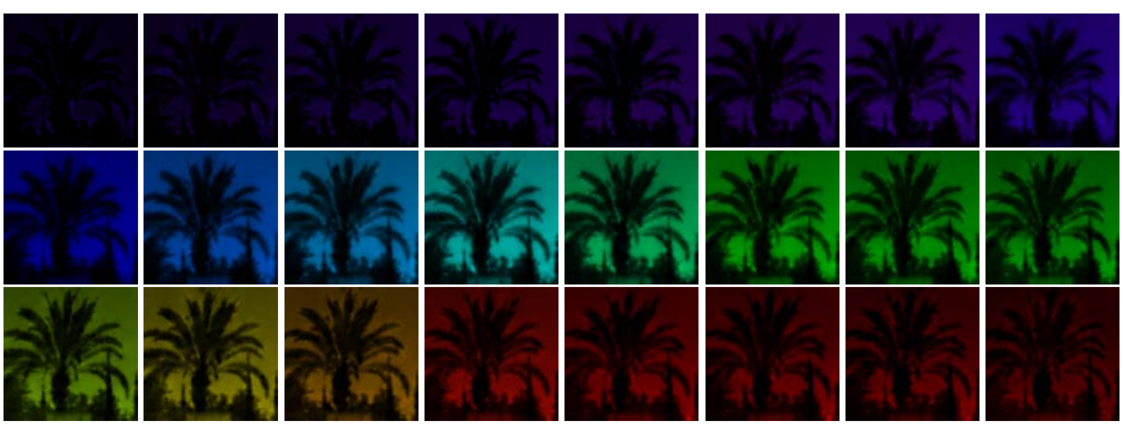

Figure 80. Reconstructed result by GAP-TV for Scene 15, PSNR: 25.80, SSIM: 0.801.

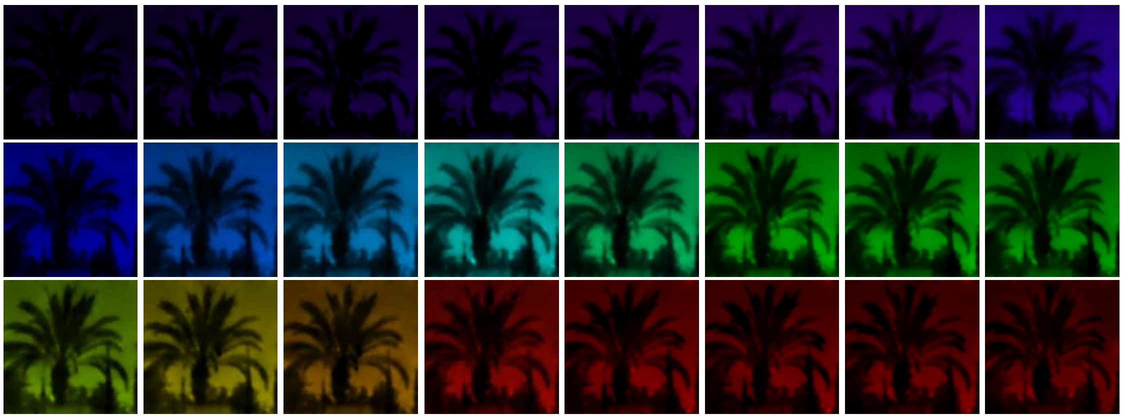

Figure 81. Reconstructed result by TwIST for Scene 15, PSNR: 23.84, SSIM: 0.765.

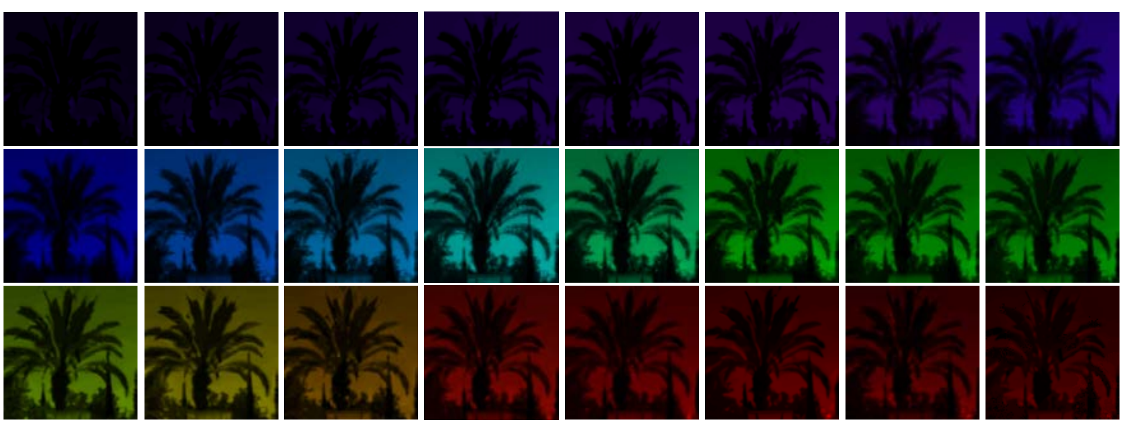

Figure 82. Reconstructed result by DeSCI for Scene 15, PSNR: 27.45, SSIM: 0.864.

References

[1] J. Bioucas-Dias and M. Figueiredo. A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Transactions on Image Processing*, 16(12):2992–3004, December 2007. 1

[2] Y. Liu, X. Yuan, J. Suo, D. Brady, and Q. Dai. Rank minimization for snapshot compressive imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, in press, 2018. 1

[3] P. Llull, X. Liao, X. Yuan, J. Yang, D. Kittle, L. Carin, G. Sapiro, and D. J. Brady. Coded aperture compressive temporal imaging. *Optics Express*, 21(9):10526–10545, 2013. 1

[4] X. Miao, X. Yuan, Y. Pu, and V. Athitsos. λ-net: Reconstruct hyperspectral images from a snapshot measurement. In *IEEE/CVF Conference on Computer Vision (ICCV)*, 2019. 1

[5] X. Yuan. Generalized alternating projection based total variation minimization for compressive sensing. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 2539–2543, Sept 2016. 1

[6] X. Yuan, P. Llull, X. Liao, J. Yang, D. J. Brady, G. Sapiro, and L. Carin. Low-cost compressive sensing for color video and depth. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3318–3325, 2014. 1

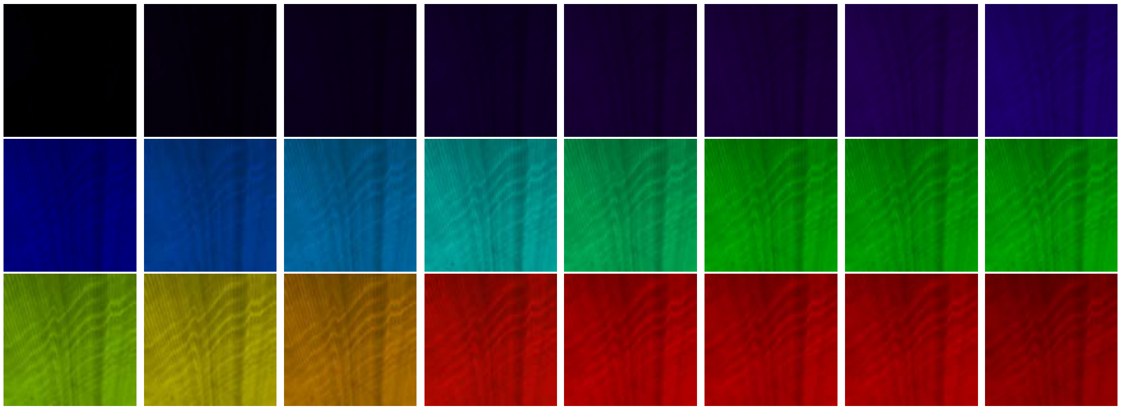

Figure 83. Truth for Scene 16.

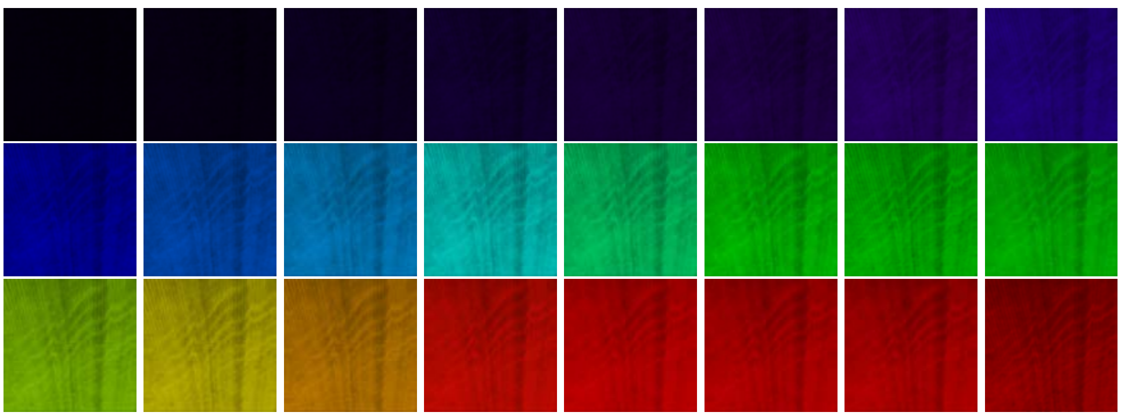

Figure 84. Reconstructed result by λ-net for Scene 16, PSNR: 30.54, SSIM: 0.855.

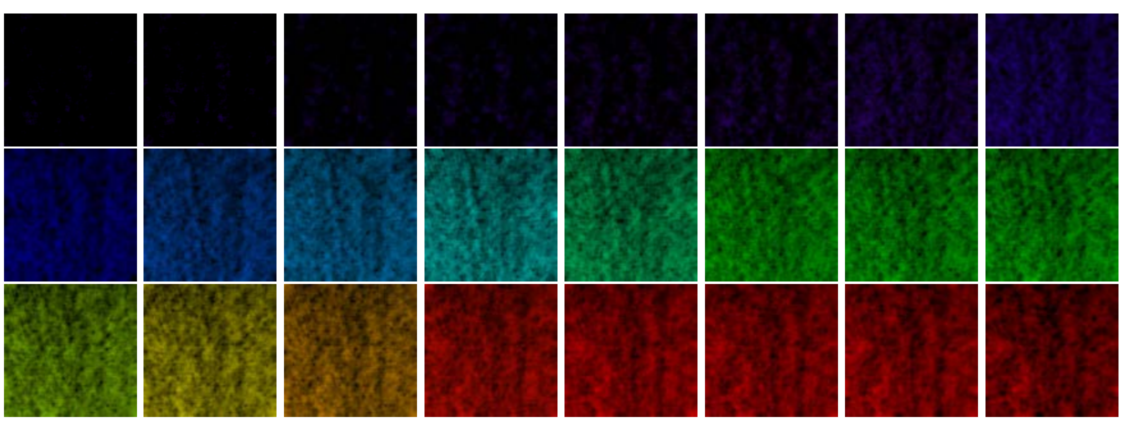

Figure 85. Reconstructed result by GAP-TV for Scene 16, PSNR: 11.99, SSIM: 0.293.

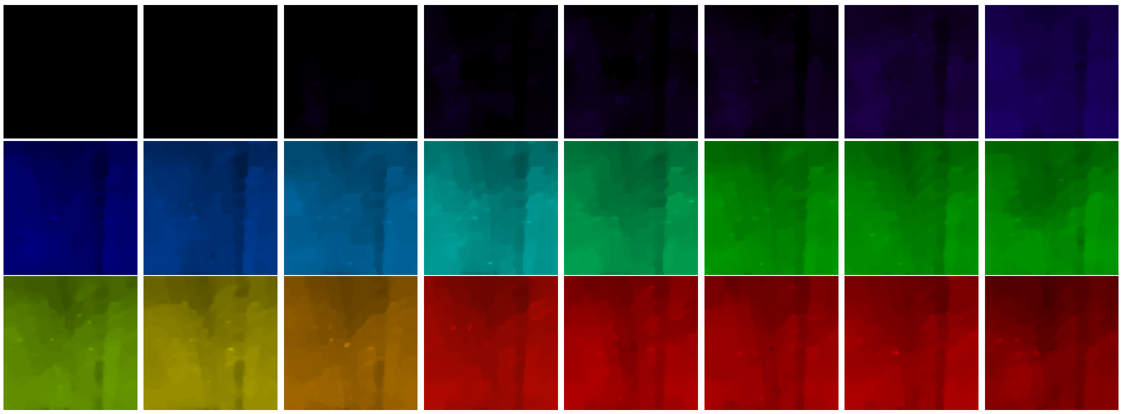

Figure 86. Reconstructed result by TwIST for Scene 16, PSNR: 20.50, SSIM: 0.511.

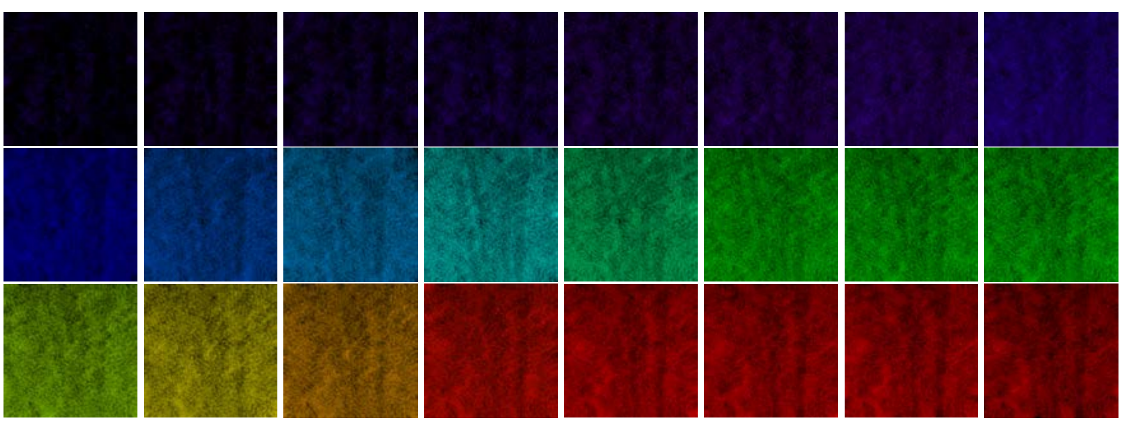

Figure 87. Reconstructed result by DeSCI for Scene 16, PSNR: 19.42, SSIM: 0.305.

Figure 88. Reconstructed result for video datasets.