End-to-End Learning for Joint Image Demosaicing, Denoising and Super-Resolution Supplementary Material

Wenzhu Xing and Karen Egiazarian
Computational Image Group, Tampere University, Finland
{wenzhu.xing, karen.eguiazarian}@tuni.fi
https://github.com/xingwz/End-to-End-JDNDMSR

1. Image demosaicing and super-resolution: additional comparison of the joint solutions

A noise-free version of the proposed JD_ND_MSR we call as JD_MSR. A comparison between three joint demosaicing and super-resolution methods is shown in Table 1. One can see from this table that the proposed combined solution JD_MSR outperforms two sequential solutions. Note that the proposed network $JD_ND_MSR^+$ is initialized by the learned parameters of the trained JD_MSR model.

Table 1: Quantitative comparison of different solutions of joint demosaicing and super-resolution using datasets Kodak and McMaster. The scale factor is 2. The best results are shown in bold.

Pipeline	McMaster		Kodak	
	cPSNR	SSIM	cPSNR	SSIM
DJDD[1]→VDSR[2]	31.67	0.9590	31.08	0.9404
DJDD* →VDSR*	31.37	0.9562	30.91	0.9395
JD_MSR	**32.32**	**0.9632**	**31.36**	**0.9440**

2. Additional settings

Comparison on cost functions. The patch size during training was 32×32, and the number of epochs is 50 with 1000 training steps and 200 validation steps. The batch size is 16. The learning rate is 0.001 for first 10 epochs, then falls to 0.0001 for the remaining epochs.

Comparison with State-of-the-Art. For each training epoch, the mini-batch size is 16, and the patch size is 64×64. For the optimization of network parameters, we use Adam with $\beta_1 = 0.9, \beta_2 = 0.999$ and the learning rate is initialized to 0.001. The training continues 250,000 iterations.

Ablation study. In this comparison, all models are trained for 50 epochs with 1000 training iterations per epoch, 100 validation iterations per epoch, and 32×32 patch size. For the pre-processing of training data (DF2K), the scale factor is 2 and the noise level is randomly sampled from [0, 20] out of 255. The evaluation is done by calculating cPSNR values on McM dataset with the noise level 10.

3. Analysis of cost functions

In our paper, we have used MAE to further optimize our original JD_ND_MSR model. MAE as a cost function for deep CNN has been used in many papers. See, e.g. [4], where authors have proven that MAE loss function can give an upper bound for regression errors to ensure a DNN robustness against additive noise, resulting in the performance advantage of DNN-MAE over DNN-MSE.

4. Limitations and potential improvements

Although our JD_ND_M and JD_ND_MSR are trained with images corrupted by noise level from 0 to 20, they still are capable of processing images with higher noise level, *e.g.* $\sigma > 20$. The cPSNR curves are shown in Fig. 1. However, as the noise level increases, more and more details in the resulting images are eliminated along with the noise (Fig. 3-4).

5. Efficiency comparison with State-of-the-Art

We evaluate efficiency of the proposed $JD_ND_MSR^+$ and TENet on the Nvidia Tesla P100 GPU. As it is shown in Table 2, our $JD_ND_MSR^+$ model has not only lighter (has a smaller size) but also faster (needs less time to compute) than TENet.

Table 2: Efficiency comparison on four datasets.

Models	Parameter (MB)	McMaster	Kodak	B100	Urban100
TENet	81.3	0.70s	1.07s	0.42s	2.14s
$JD_ND_MSR^+$	78.2	0.64s	0.87s	0.32s	1.83s

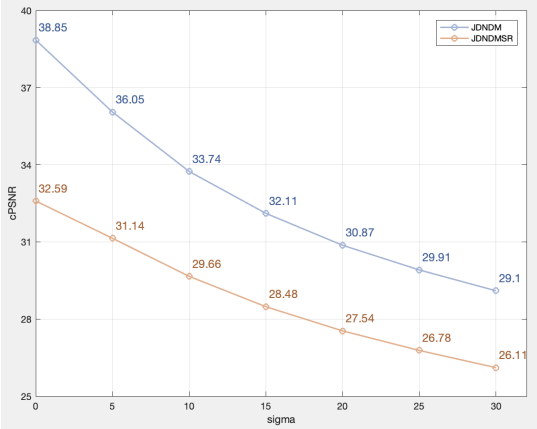


Figure 1: cPSNR comparison JD_ND_M and JD_ND_MSR at different noise level. The metric is averaged across McMaster dataset.

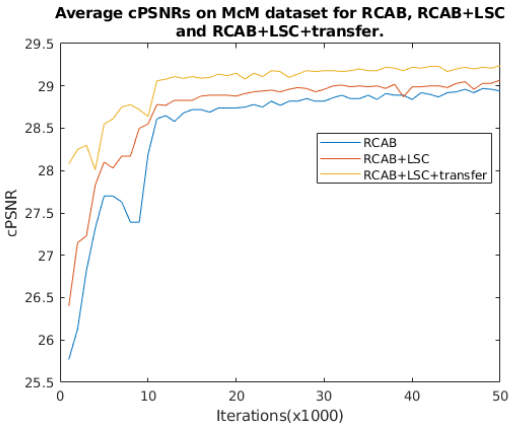


Figure 2: The convergence curves of JD_ND_MSR with long skip connection and transfer learning.

6. Additional ablation study

Fig. 2 shows that the additional LSC improves the performance of the network. In addition, we exploit the transfer learning, which transfers the well-learned parameters from the pre-trained noise-free model JD_MSR (Section 1). The curves (yellow and red lines) in Fig. 2 prove that this kind of easy-to-hard transfer learning strategy not only improves the performance of network, but also supports a better starting point (at least 1.5 dB higher cPSNR).

7. More qualitative results

We provide more examples to compare different joint solutions of the mixture problem, image demosaicing, denoising and super-resolution. As shown in Fig. 5 and Fig. 6, our proposed JD_ND_MSR and $JD_ND_MSR^+$ can generate more accurate images with less color artifacts and blur than other methods.

In Fig. 7-9, we give some examples of processed images for the qualitative comparison of the proposed method and the state-of-the-art joint demosaicing, denoising and super-resolution, TENet [5]. These figures demonstrate that TENet generates color artifacts (see the last row in Fig 7 and Fig 9 and the second and third rows in Fig 8) and blur (see the first and last rows in Fig 8 and the first two rows in Fig 9) in resulting images. In contrast, the proposed $JD_ND_MSR^+$ can eliminate noise and artifacts while retaining more accurate details and textures.

In addition to this, we show that the proposed method outperforms state-of-the-art also for the case of the scaling factor equal to 1, i.e. for the joint denoising and demosaicing problem. The processed images for the visual comparison between JD_ND_M (which is a special case of the proposed $JD_ND_MSR^+$ for SR scale factor equal to 1) and the state-of-the-art methods are shown in Fig. 10-11. From these illustrations, one can see that the DJDD [1] method will cause blur and color artifacts which appear to be more severe by increasing a noise level. The method proposed by Kokkinos [3] causes blur in the high frequency region (even for noise-free data) and fails to remove a noise. Analysis of the processed images by the proposed JD_ND_M and the state-of-the-art methods, allow us to conclude that images reconstructed by JD_ND_M have significantly less artifacts compared to the results of application of the state-of-the-art methods.

References

[1] Michaël Gharbi, Gaurav Chaurasia, Sylvain Paris, and Frédo Durand. Deep joint demosaicking and denoising. *ACM Transactions on Graphics (TOG)*, 35(6):191, 2016. 1, 2, 4, 9, 10

[2] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1646–1654, 2016. 1, 4

[3] Filippos Kokkinos and Stamatios Lefkimmiatis. Deep image demosaicking using a cascade of convolutional residual denoising networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 303–319, 2018. 2, 9, 10

[4] Jun Qi, Jun Du, Sabato Marco Siniscalchi, Xiaoli Ma, and Chin-Hui Lee. On mean absolute error for deep neural network based vector-to-vector regression. *IEEE Signal Processing Letters*, 27:1485–1489, 2020. 1

[5] Guocheng Qian, Jinjin Gu, Jimmy S Ren, Chao Dong, Furong Zhao, and Juan Lin. Trinity of pixel enhancement: a joint solution for demosaicking, denoising and super-resolution. *arXiv preprint arXiv:1905.02538*, 2019. 2

[6] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017. 4

Ground truth | $\sigma = 0$ | $\sigma = 5$ | $\sigma = 10$

$\sigma = 15$ | $\sigma = 20$ | $\sigma = 25$ | $\sigma = 30$

Figure 3: Qualitative comparison $JD_N D_M$ at different noise level.

Ground truth | $\sigma = 0$ | $\sigma = 5$ | $\sigma = 10$

$\sigma = 15$ | $\sigma = 20$ | $\sigma = 25$ | $\sigma = 30$

Figure 4: Qualitative comparison $JD_N D_M SR$ at different noise level.

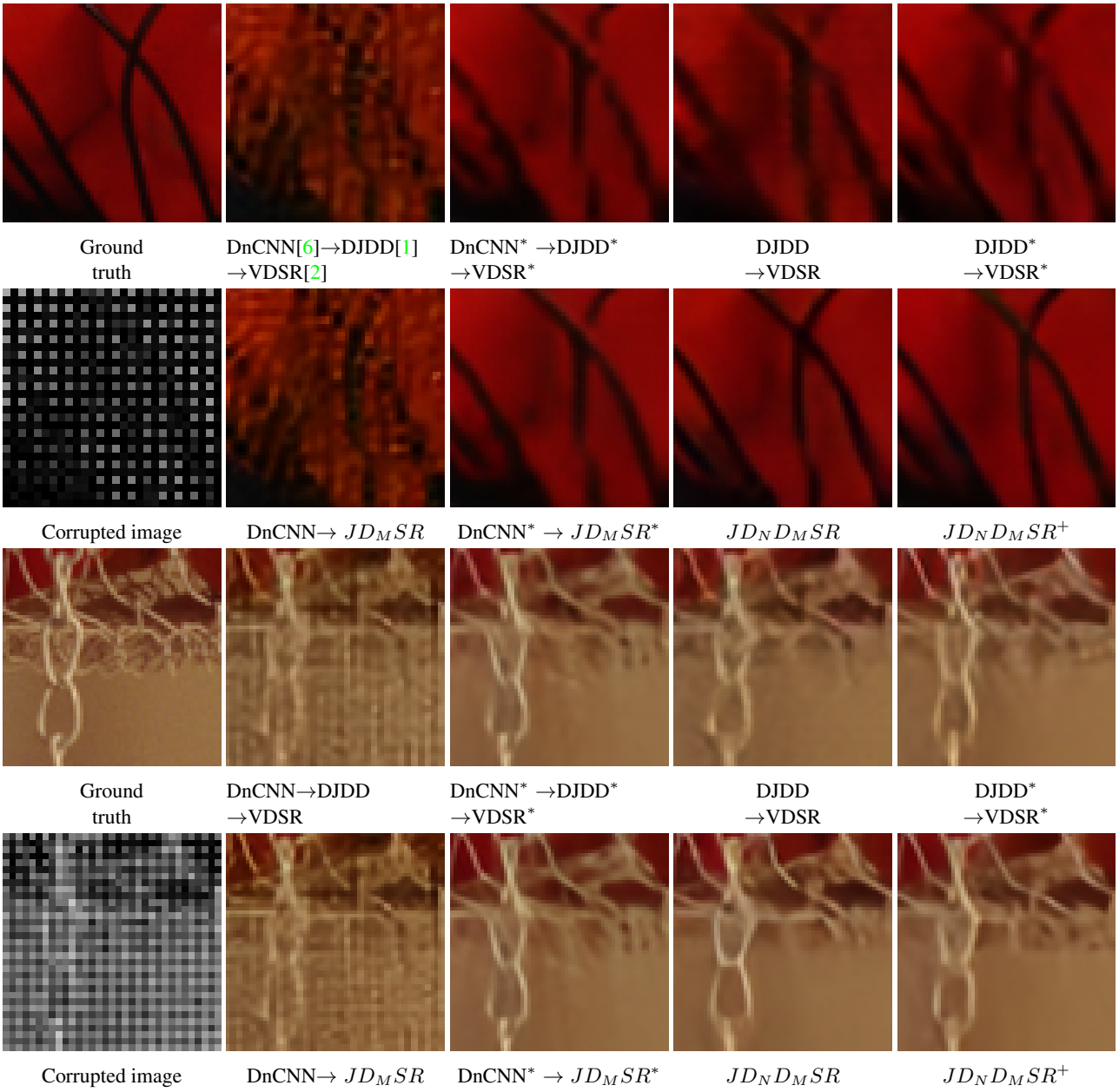


Figure 5: Qualitative comparison between different joint solutions. Scale factor is 2. The noise level is 10. The patches of the upper two rows are from image9 in McMaster. The patches of the last two rows are from image12 in McMaster.

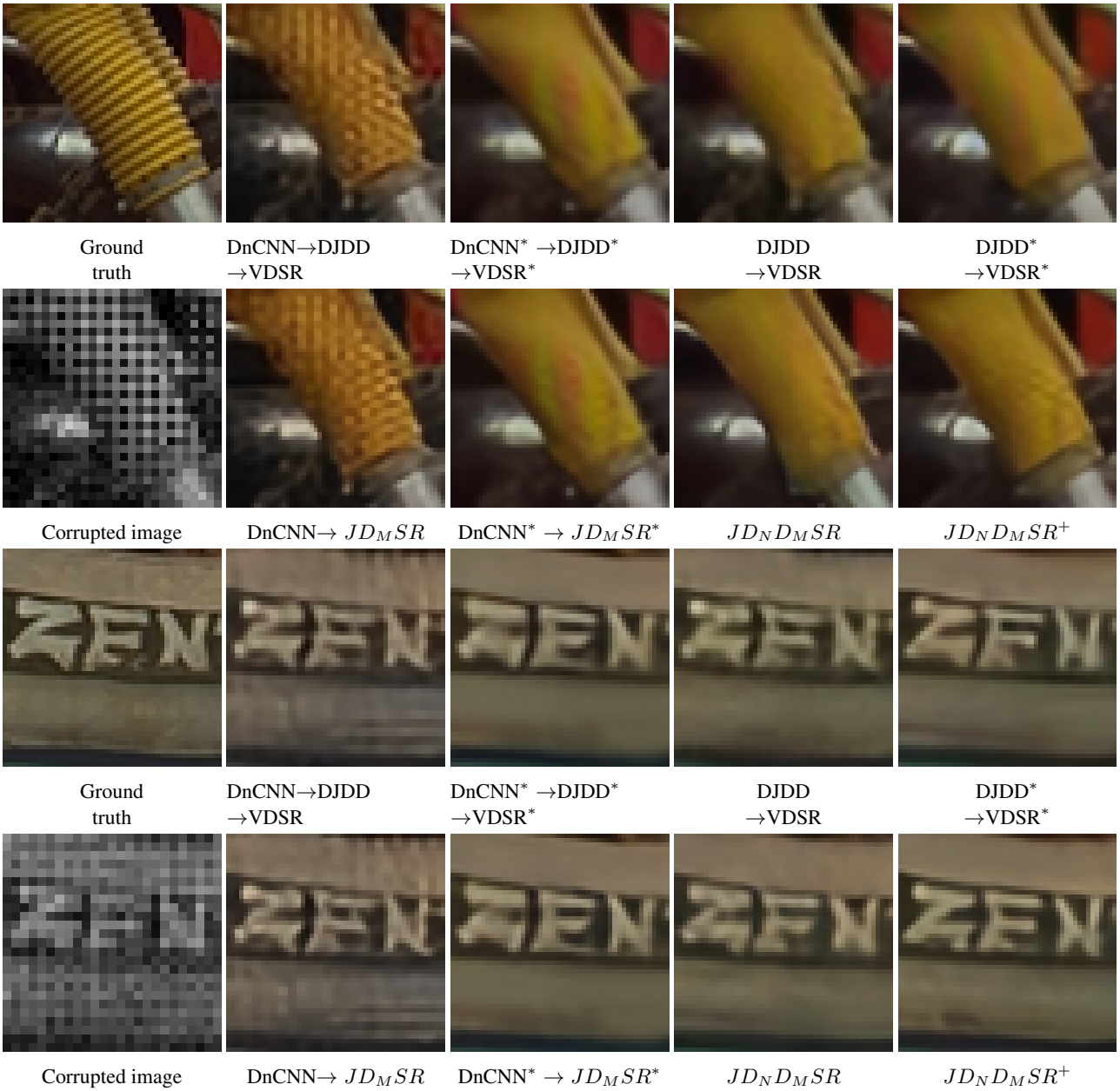
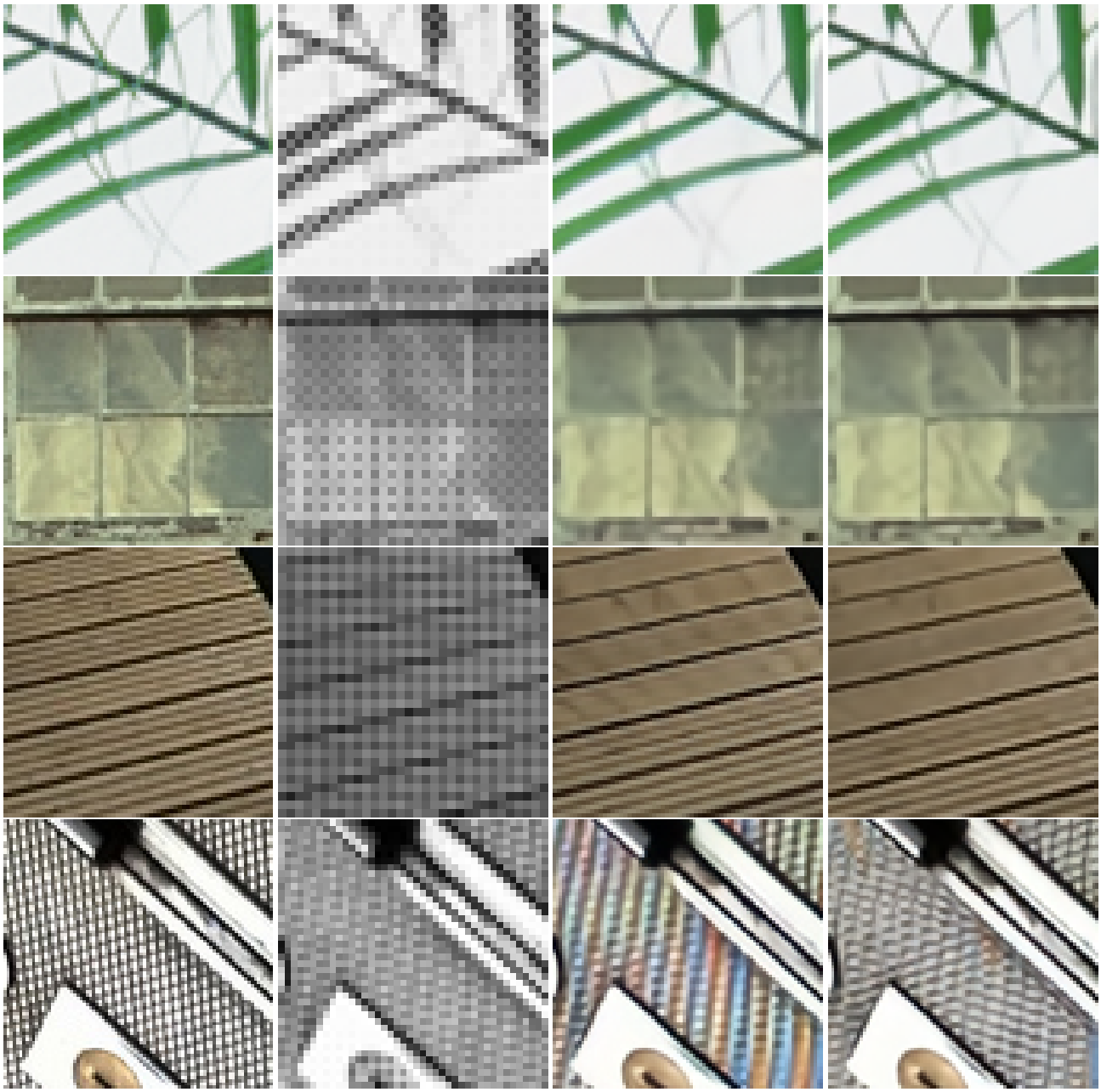
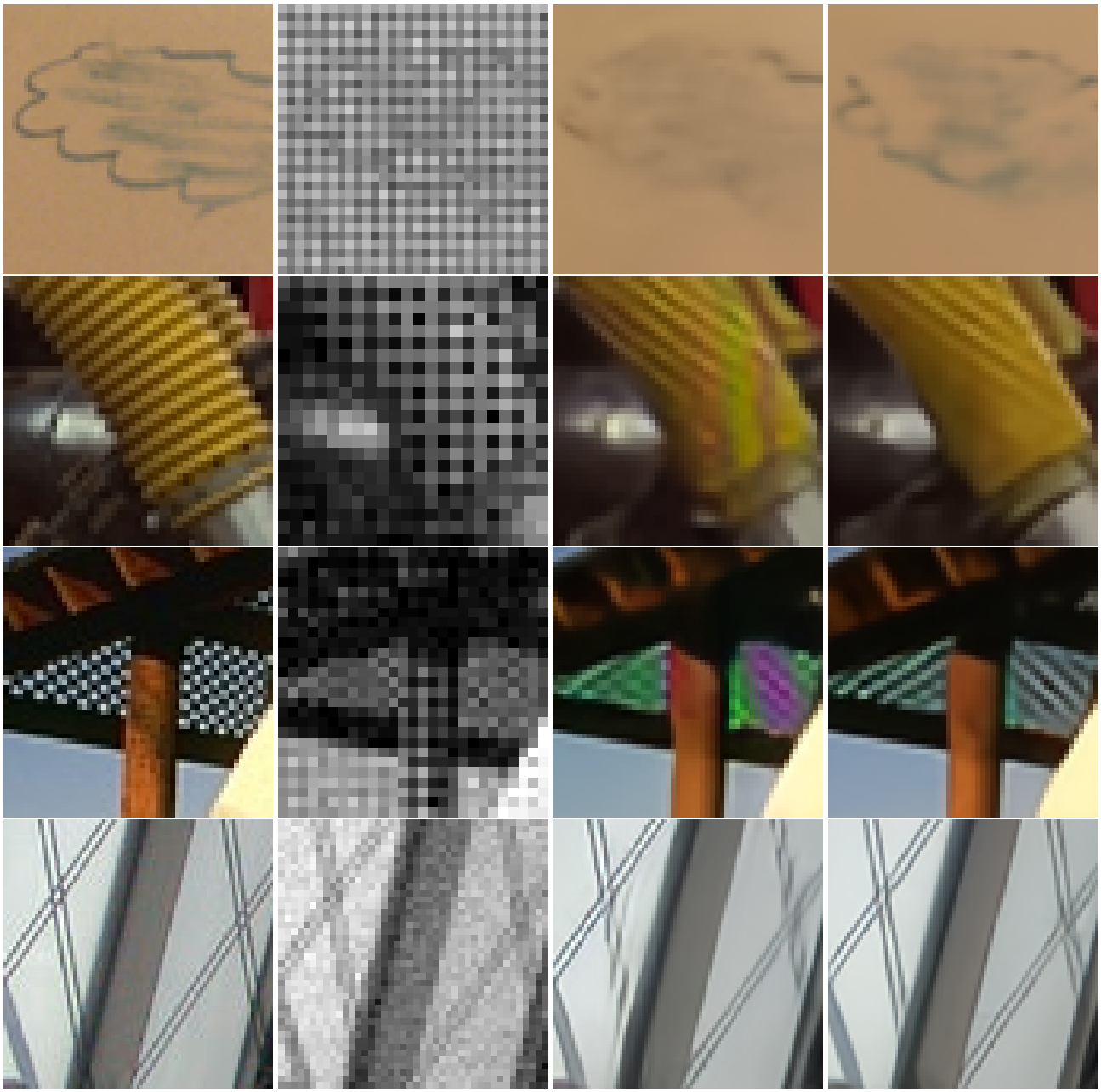


Figure 6: Qualitative comparison between different joint solutions. Scale factor is 2. The noise level is 10. The patches of the upper two rows are from kodim05 in Kodak. The patches of the last two rows are from kodim06 in Kodak.



Ground truth | Corrupted image | TENet | Ours

Figure 7: Qualitative comparison between the SOTA model TENet and the proposed $JD_ND_MSR^+$. Scale factor is 2. The noise level is 0.

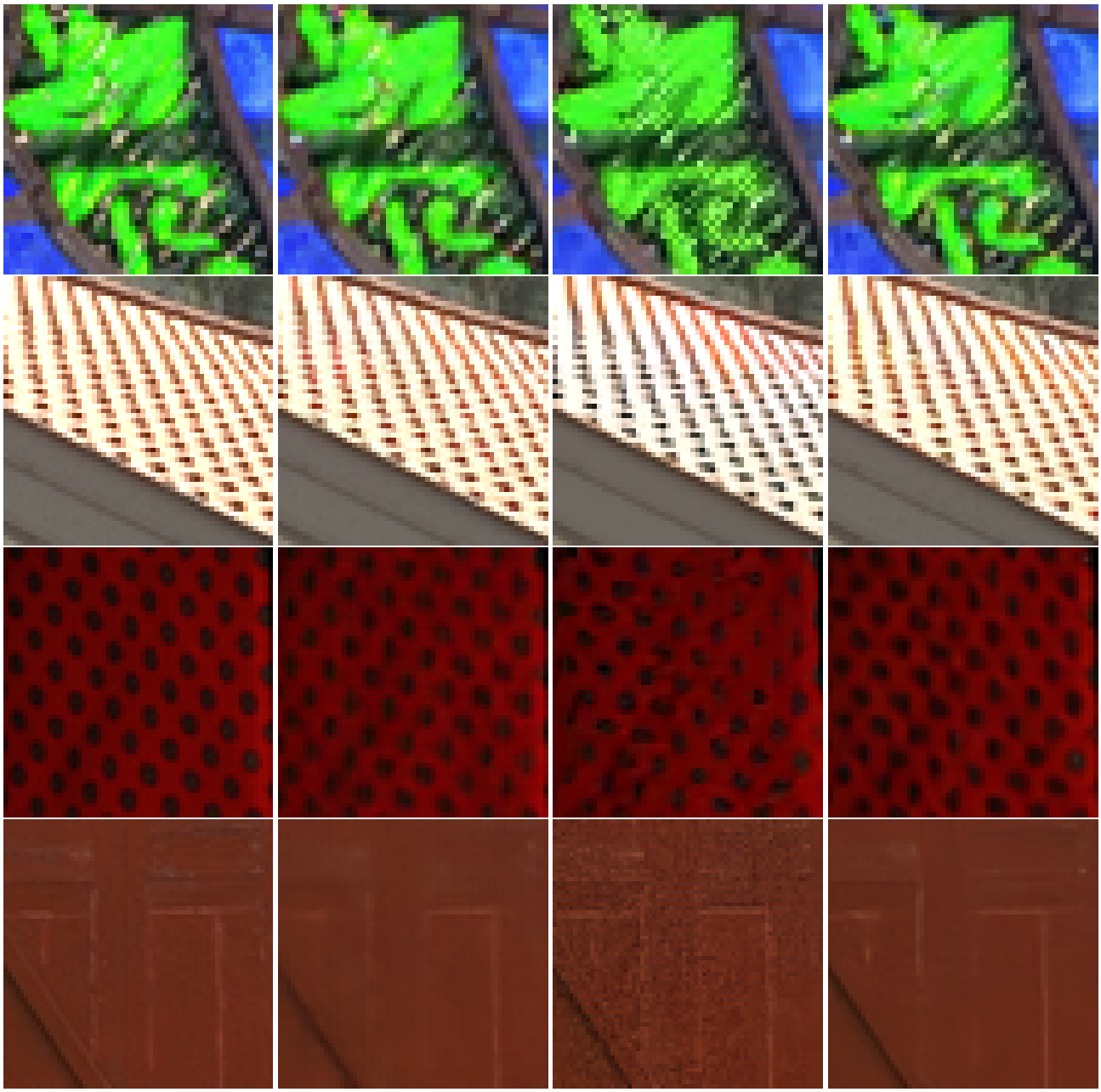


Ground truth | Corrupted image | TENet | Ours

Figure 8: Qualitative comparison between the SOTA model TENet and the proposed $JD_ND_MSR^+$. Scale factor is 2. The noise level is 10.

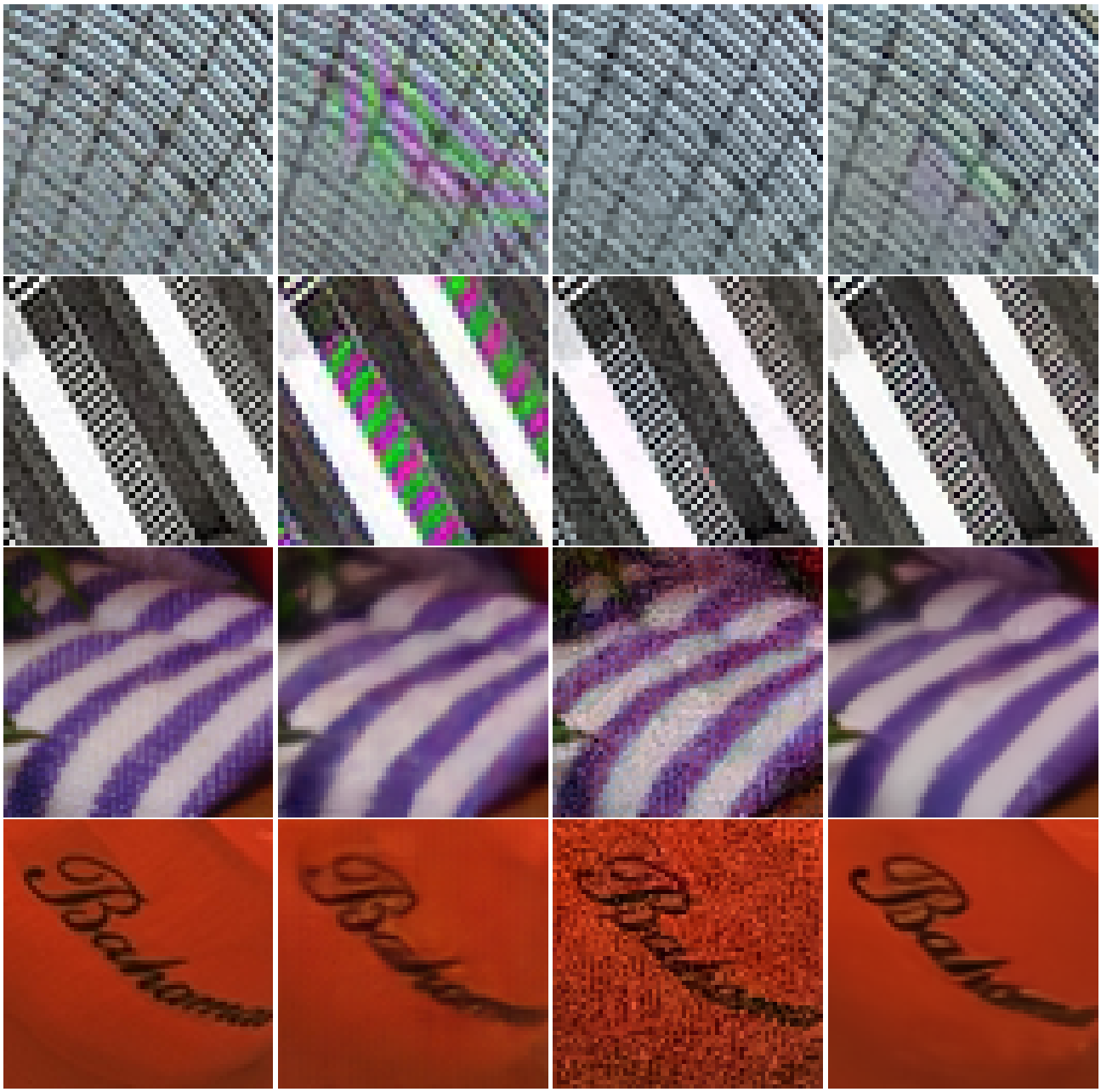
Ground truth | Corrupted image | TENet | Ours

Figure 9: Qualitative comparison between the SOTA model TENet and the proposed $JD_ND_MSR^+$. Scale factor is 2. The noise level is 20.



Ground truth DJDD Kokkinos Ours

Figure 10: Qualitative comparison between our proposed JD_ND_M and state-of-the-art methods, DJDD [1] and kokkinos [3]. The noise level of first two rows is 0. The noise level of last two rows is 5.



Ground truth DJDD Kokkinos Ours

Figure 11: Qualitative comparison between our proposed $JD_N D_M$ and state-of-the-art methods, DJDD [1] and kokkinos [3]. The noise level of first two rows is 10. The noise level of last two rows is 15.