

Supplementary Material of "Iterative Residual CNNs for Burst Photography Applications"

1. Hyperparameters

In Table 1 we provide the values of all the hyperparameters that we used to train our models. The ProxNet was pre-trained on the single frame Gaussian denoising task using the Berkeley segmentation dataset (BSDS) [9], which consists of 500 color images. All the images were randomly cropped into patches of size 180×180 pixels. The patches were perturbed with noise of standard deviation $\sigma \in [0, 15]$ and the network was optimized to minimize the Mean Square Error. While pre-training of the denoiser is not necessary in our approach, we experimentally found out that a pre-trained denoiser typically speeds up the training time of our Iterative Neural Network. This can be attributed to the fact that a pre-trained denoiser can serve as a proper initialization for the proximal network (ProxNet).

2. Computational Cost

Representative execution times of our burst image demosaicking approach for images of various sizes is provided in Fig. 2. These results refer to the average execution time of 10 runs on an NVIDIA TitanX. From this figure is clear that increasing the number of frames inside the burst yields only a small increase in the computational cost. The reason is that our proximal network accepts as an input a single image and not the entire sequence of the burst. The burst frames are only used to compute the gradient of the data fidelity term, which is then combined with the output of the proximal network from the previous iteration to form the input for the current iteration. The gradient of the data fidelity term can be efficiently computed in parallel and, thus, an increase of the number of frames does not affect linearly the computation time. On the contrary, the networks introduced in [1, 4] are applied separately on each frame of the burst and thus the involved computation cost increases linearly with the size of the burst.

3. Failure Cases

The main limitation of our network lies in its dependency on the ECC estimation of the warping matrix [2]. This estimation method can be rather inaccurate especially when there is a strong presence of noise. Therefore, when the

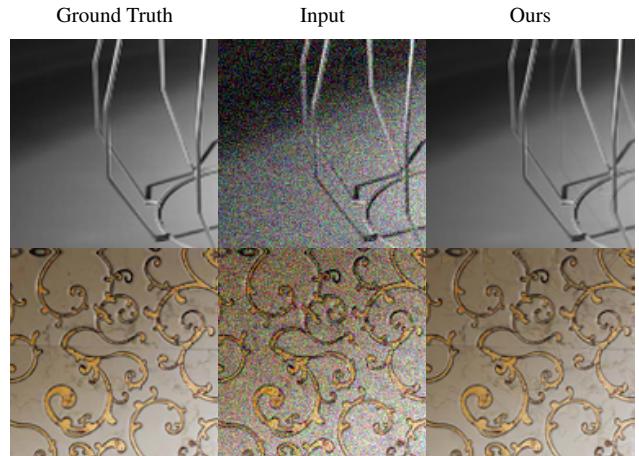


Figure 1: Failure cases of our approach on burst Gaussian denoising where the homography estimation between the burst frames is imprecise.

estimated homography matrix is imprecise, our network inevitably will introduce ghosting artifacts to the final result, such as those depicted in Fig. 1. Indeed, in these examples, while we observe that the images have been denoised successfully, due to the imprecise warping estimation, false edges appear in the restored images.

4. Additional Burst Denoising Results

In Fig. 4, we provide additional qualitative comparisons for the burst Gaussian denoising task. From these results we clearly observe that our denoising approach yields superior restorations, where fine details have been retained while the noise has been efficiently suppressed. We note that this is not the case for ResDNet and VBM4D, where the results either suffer from the appearance of noise artifacts or fine details have been entirely wiped out.

	Burst Gaussian Denoising	Burst Noise-free Demosaicking	Burst Demosaicking
Initial Learning Rate	0.01	0.003	0.003
Iterations K	10	10	10
TBPTT k	5	5	5
Batch Size	10	10	10
Noise Estimation	True	False*	True
s_{max}	$\ln(2)$	$\ln(15)$	$\ln(2)$
s_{min}	$\ln(1)$	$\ln(1)$	$\ln(1)$
Epochs	300	300	300

Table 1: Hyper-parameters used during training for each trained model.*For the case of noise-free demosaicking we consider the presence of low-level noise of fixed standard deviation equal to $\sigma = 1$.

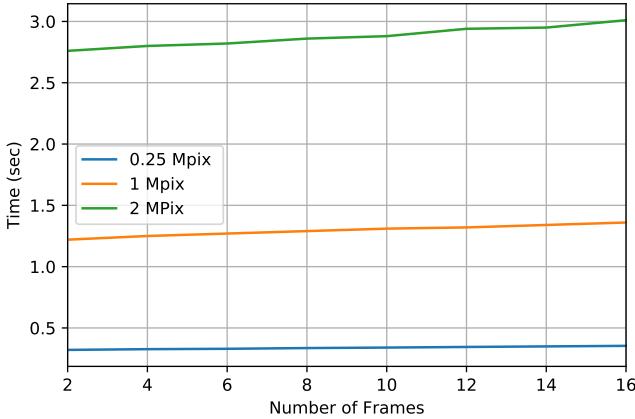


Figure 2: Execution time of burst demosaicking for 0.25, 1 and 2 megapixel images versus the number of used burst frames. Increasing the number of frames yields only a minor increase in computational complexity.

5. Calculation of the Lipschitz Constant

For any pair of variables $\mathbf{z}, \mathbf{x} \in \mathbb{R}^N$ we have

$$\begin{aligned}
& \| \nabla_{\mathbf{x}} f(\mathbf{x}) - \nabla_{\mathbf{z}} f(\mathbf{z}) \|_X = \\
&= \left\| \frac{1}{\sigma^2 B} \sum_{i=1}^B \mathbf{S}_i^T \mathbf{H}^T (\mathbf{H} \mathbf{S}_i \mathbf{x} - \mathbf{y}) - \mathbf{S}_i^T \mathbf{H}^T (\mathbf{H} \mathbf{S}_i \mathbf{z} - \mathbf{y}) \right\|_2 \\
&= \frac{1}{\sigma^2 B} \left\| \sum_{i=1}^B \mathbf{S}_i^T \mathbf{H}^T \mathbf{H} \mathbf{S}_i \mathbf{x} - \mathbf{S}_i^T \mathbf{H}^T \mathbf{H} \mathbf{S}_i \mathbf{z} \right\|_2 \\
&\leq \frac{1}{\sigma^2 B} \sum_{i=1}^B \| \mathbf{S}_i^T \mathbf{H}^T \mathbf{H} \mathbf{S}_i \mathbf{x} - \mathbf{S}_i^T \mathbf{H}^T \mathbf{H} \mathbf{S}_i \mathbf{z} \|_2 \\
&= \frac{1}{\sigma^2 B} \sum_{i=1}^B \| \mathbf{S}_i^T \mathbf{H}^T \mathbf{H} \mathbf{S}_i (\mathbf{x} - \mathbf{z}) \|_2 \\
&\leq \frac{1}{\sigma^2 B} \sum_{i=1}^B \| \mathbf{S}_i^T \mathbf{H}^T \mathbf{H} \mathbf{S}_i \| \| \mathbf{x} - \mathbf{z} \|_2 \\
&\leq \frac{1}{\sigma^2} \| \mathbf{x} - \mathbf{z} \|_2.
\end{aligned}$$

Note that, the first two inequalities follow from the rela-

tion between the norms defined in the \mathbb{R}^N space. To compute an upper bound of the spectral norm $\| \mathbf{S}_i^T \mathbf{H}^T \mathbf{H} \mathbf{S}_i \|$, we exploit the fact that both matrices \mathbf{H} and \mathbf{S} have bounded values between 0 and 1. The aforementioned fact holds since the affine transformation matrix \mathbf{S}_i contains the interpolation coefficients for the weighted average of the pixel intensities and by definition these coefficients sum up to one. Simultaneously, for the problems that we consider, the matrix \mathbf{H} is either identity for the burst denoising problem or binary diagonal matrix for the burst demosaicking problem. This immediately implies that $\| \mathbf{S}_i^T \mathbf{H}^T \mathbf{H} \mathbf{S}_i \| \leq 1$ and, hence, an upper bound of the Lipschitz constant of $\nabla_{\mathbf{x}} f(\mathbf{x})$ will be $L(x) \leq \frac{1}{\sigma^2 B} \sum_{i=1}^B \| \mathbf{S}_i^T \mathbf{H}^T \mathbf{H} \mathbf{S}_i \| \leq 1/\sigma^2$.

6. Additional FlexISP results

In Table 2, we provide the PSNR scores for the synthetic bursts of the FlexISP dataset. Our iterative network outperforms all previous approaches and the improvement in PSNR ranges from 1.2 to 3 dBs. The same conclusion about the superiority of our joint denoising-demosaicking approach, can also be confirmed visually by referring to the results provided in Fig. 3.

References

- [1] M. Aittala and F. Durand. Burst image deblurring using permutation invariant convolutional neural networks. In *The European Conference on Computer Vision (ECCV)*, September 2018. 1
- [2] G. D. Evangelidis and E. Z. Psarakis. Parametric image alignment using enhanced correlation coefficient maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(10):1858–1865, Oct 2008. 1
- [3] M. Gharbi, G. Chaurasia, S. Paris, and F. Durand. Deep Joint Demosaicking and Denoising. *ACM Trans. Graph.*, 35(6):191:1–191:12, Nov. 2016. 3
- [4] C. Godard, K. Matzen, and M. Uyttendaele. Deep burst denoising. In *The European Conference on Computer Vision (ECCV)*, September 2018. 1, 3
- [5] S. W. Hasinoff, D. Sharlet, R. Geiss, A. Adams, J. T. Barron, F. Kainz, J. Chen, and M. Levoy. Burst photography for

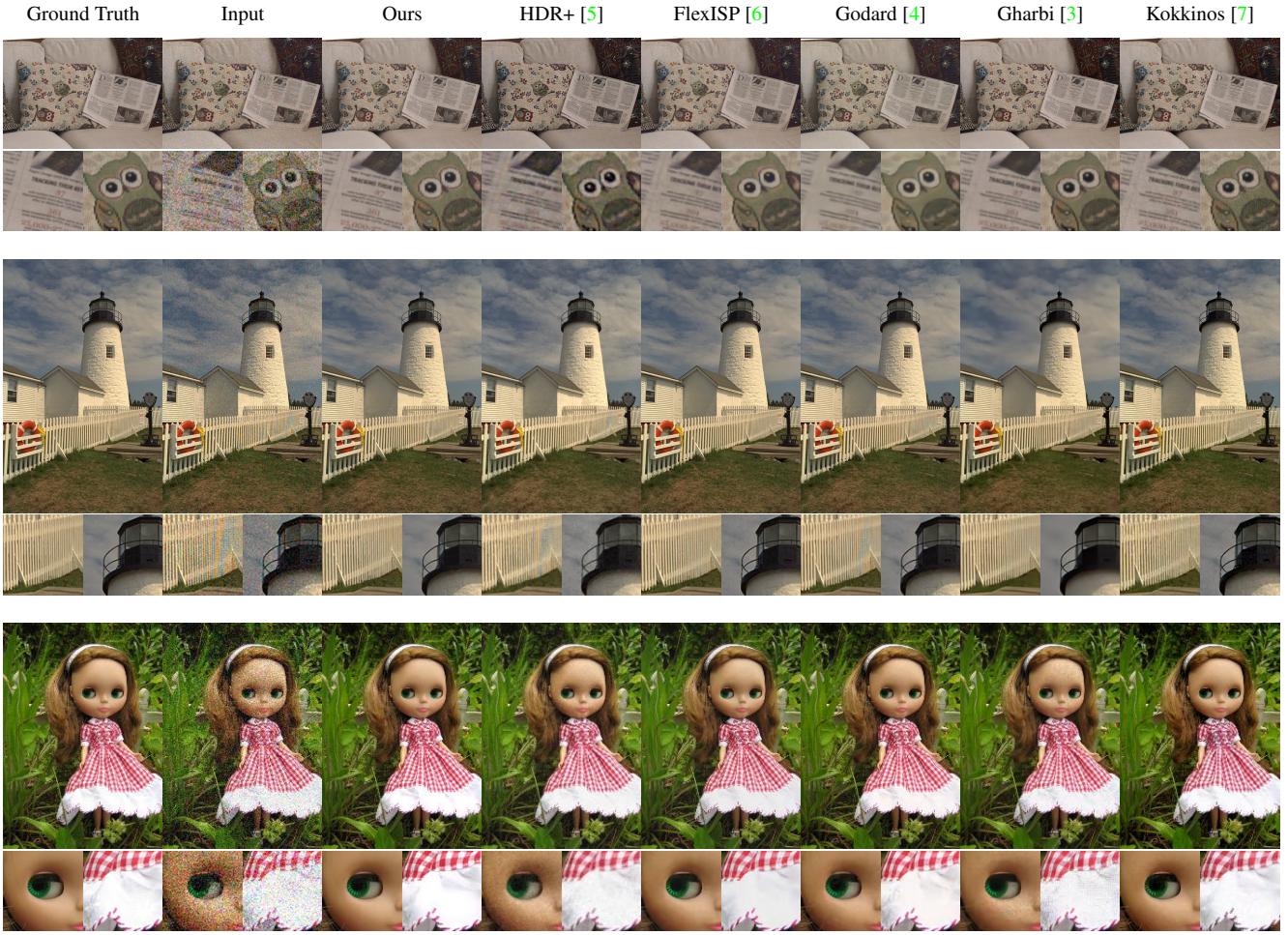


Figure 3: Burst demosaicking results on a real and two synthetic bursts from the FlexISP dataset [6]. Our model successfully restores the missing colors of the underlying images while suppressing noise.

Image	Single Frame			Multi Frame				
	Input	Gharbi	Kokkinos	FlexISP	ProImal	HDR+	Godard	Ours
<i>Flickr Doll</i>	20.25	29.61	28.33	29.41	30.23	26.52	29.39	31.22
<i>Kodak Fence</i>	26.28	32.82	32.93	34.42	-	30.07	34.08	35.24
<i>Living Room</i>	21.74	34.4	34.61	32.27	-	25.79	31.32	37.53

Table 2: Comparisons of several methods for the burst demosaicking task on the synthetic FlexISP dataset.

- high dynamic range and low-light imaging on mobile cameras. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, 35(6), 2016. 3
- [6] F. Heide, M. Steinberger, Y.-T. Tsai, M. Rouf, D. Pajak, D. Reddy, O. Gallo, J. Liu, W. Heidrich, K. Egiazarian, et al. Flexisp: A flexible camera image processing framework. *ACM Transactions on Graphics (TOG)*, 33(6):231, 2014. 3
- [7] F. Kokkinos and S. Lefkimiatis. Deep image demosaicking using a cascade of convolutional residual denoising networks. In *The European Conference on Computer Vision (ECCV)*, September 2018. 3

[8] M. Maggioni, G. Boracchi, A. Foi, and K. Egiazarian. Video denoising using separable 4d nonlocal spatiotemporal transforms. In *Image Processing: Algorithms and Systems IX*, volume 7870, page 787003. International Society for Optics and Photonics, 2011. 4

[9] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pages 416–423 vol.2, 2001. 1

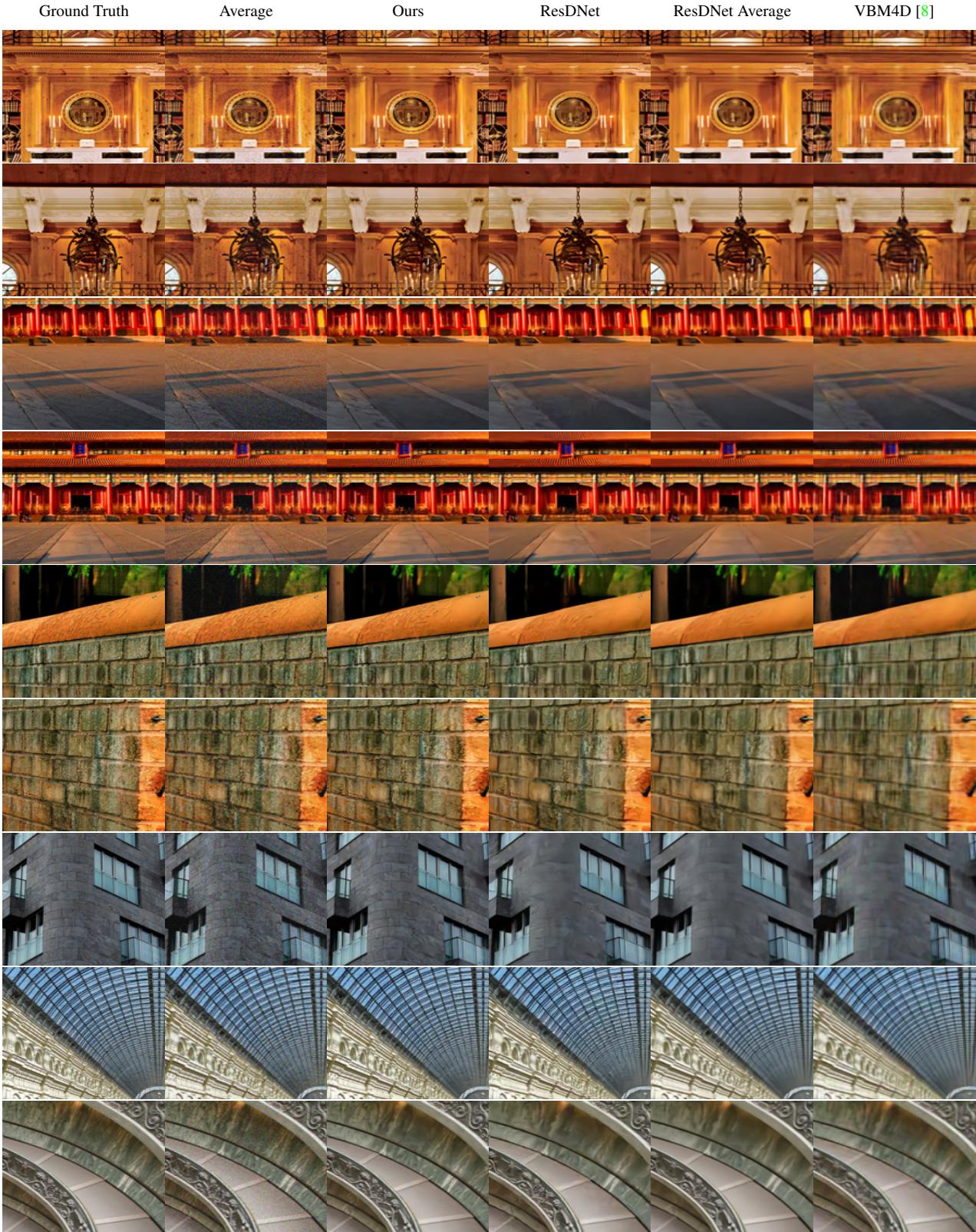


Figure 4: Burst Gaussian denoising with $\sigma = 25$. Our method is able to effectively restore the images and retain fine details, as opposed to the rest of the methods that over-smooth high texture areas. Results best seen magnified on a computer screen.