Base Frame Selection on Dynamically Exposed Burst

Min Jung Lee[O], Jongmin Lee, Sanghyun Kim, Sunghyun Cho, Minsu Cho
Pohang University of Science and Technology (POSTECH)
{minjlee, ljm1121, sanghyun.kim, s.cho, mscho}@postech.ac.kr

Abstract

Burst image enhancement has been a topic of active research in recent years due to its ability to improve the quality of an image captured in adverse conditions such as motion and low-light scenes. Existing burst benchmarks use the static exposure time for all frames, limiting enhancement potential as dynamically exposed frames offer complementary scene information. In this paper, we introduce a novel approach for burst image enhancement using dynamic exposure times and base frame selection. We propose a new burst image enhancement benchmark using dynamic exposures. Additionally, we propose a learnable module to select the best base frame from the burst sequence, considering the heterogeneous degradation in dynamically exposed bursts. To improve spatial consistency during training, we also introduce a relational contextual bilateral loss. Evaluation on our proposed benchmark demonstrates the effectiveness of our method compared to existing burst image enhancement techniques.

1. Introduction

Burst image enhancement reconstructs a high-quality image from rapid sequence of frames captured by handheld devices [1, 3, 4, 5]. Compared to the existing single image enhancement techniques [6, 7, 8, 9], burst image enhancement obtains a high-quality image in terms of signal-to-noise ratio (SNR) by accumulating aligned light information of the input bursts and a sharp image by aligning bursts while eschewing blurry frames [1, 2, 3, 10, 11, 12, 13, 14, 15, 16, 17, 18]. The popularity of image enhancement using input bursts has increased for capturing motion and low-light scenes, but resulting images often suffer from various issues such as sensor noise, motion blur, and under-/over-exposure. Varying exposure times during burst acquisition can mitigate these image degradations by leveraging complementary information.

Therefore, we propose a novel benchmark called Dynamically Exposed Burst (DEB) for enhancing burst images in challenging conditions. Existing burst image enhancement benchmarks take pictures at static intervals of exposure time, but this approach may not be sufficient to capture all the different types of irradiance in a scene. In contrast, dynamically exposed burst imaging takes pictures at varying intervals of exposure time, which enables the capture of images with a wider range of irradiance. Our DEB leverages the complementary information of the short-exposed and long-exposed frames by combining the sharpness of short-exposed frames with the high signal-to-noise ratio (SNR) of long-exposed frames. Figure 1 shows the enhancement results of trained models with dynamically exposed burst and statically exposed burst images with the same total exposure time.

To improve the quality of burst enhancement, a robust base frame is needed for precise spatial alignment, as it serves as a reference for merging information between bursts. However, severe degradation of the base frame can cause misalignment among the input bursts, resulting in reducing the output image quality. Identifying frames with lesser degradation within the burst can mitigate this issue, allowing for the selection of a more suitable base frame. Despite this, existing methods [1, 2] commonly assume that the first frame of the input bursts is a base frame, even though the first frame may contain severe image degradation. To address this issue, we propose a frame selection module that automatically identifies the most suitable base frame during the burst frame alignment stage of the enhancement

(a) Dynamic (b) Static (c) Ground-truth

Setting of exposures	Aligned PSNR	Aligned SSIM	Aligned LPIPS
Dynamic	**33.841**	**0.934**	**0.110**
Static	32.241	0.910	0.155

Figure 1. Dynamic exposures vs. static exposures. (a) and (b) show the predictions of fine-tuning a DBSR [1] synthetic pretrained model on two different burst datasets: dynamic exposed and statically exposed bursts, respectively. (c) displays the ground-truth frame.

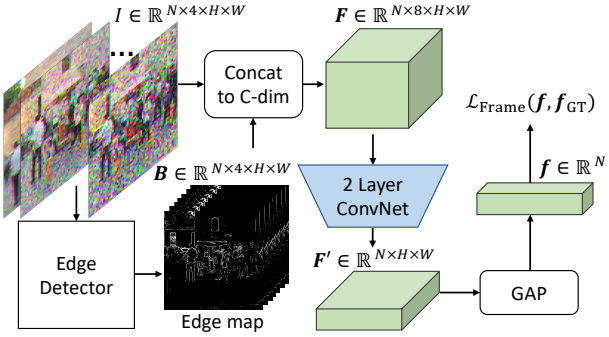

Figure 2. An illustration of the proposed base frame selector. The frame selection vector $\boldsymbol{f}$ is generated from input I and its edge map $\boldsymbol{B}$ for the base frame selection.

pipeline, removing the assumption that the first frame is always the optimal choice. Our base frame selection module can apply to the existing burst enhancement networks in a plug-and-play manner.

We additionally introduce a new shift-invariant perceptual loss called the Relational Contextual Bilateral loss (RCB), which modifies the spatial distance term in the contextual bilateral loss. The contextual bilateral loss (CoBi) [19] added a regularization term for spatial structure in the contextual loss [20] to overcome inaccurate loss signals for false matches, but it penalized distance in absolute coordinates, which was inappropriate for large misalignments. To address this limitation, we calculate the spatial distance in relative rather than absolute coordinates using the average flow from source to target.

2. Methods

2.1 Burst Frame Selection

The objective of this frame selector is to find a *robust* base frame within the input bursts, serving as the best target image for alignment with other burst images. The architecture, illustrated in Figure 2, takes an input burst $I \in \mathbb{R}^{N\times4\times H\times W}$ to generate an edge map $\mathbf{B} \in \mathbb{R}^{N\times4\times H\times W}$. An integrated feature map $\mathbf{F} \in \mathbb{R}^{N\times8\times H\times W}$ is then constructed by concatenating I and $\mathbf{B}$ in the channel direction. Subsequently, the channel dimension of $\mathbf{F}$ is collapsed using a convolution block, followed by spatial collapse through global average pooling (GAP), resulting in the frame selection vector $\mathbf{f} \in \mathbb{R}^{N}$.

We first obtain an edge map to measure image sharpness by counting the number of edge pixels, as a sharp image is good to use as a reference image for alignment [21]. The output edge magnitude map $\mathbf{G} \in \mathbb{R}^{N\times4\times H\times W}$ is computed as the gradient magnitude using Sobel filters [22]. The final edge map $\mathbf{B}$ is obtained by thresholding the edge magnitude map as follows: $\mathbf{B}_i = 1$, if $\mathbf{G}_i > \tau$, 0 otherwise, where i is a flattened index of the tensor $\mathbf{B}$ and τ is a threshold value. We set $\tau = 0.007$.

We obtain an integrated feature map $\mathbf{F}$ by concatenating the edge map $\mathbf{B}$ and the input I to the channel dimension. This operation is expressed as $\mathbf{F} = \mathbf{B} \oplus I$, where $\oplus$ denotes the concatenation operation

(a) source image P (b) target image Q

	$\mathbb{D}(P,Q)$	$\mathbb{D}'(P,Q)$
CoBi Loss	0.3467	0.9118
RCB Loss	0.3467	0.0013

(c) average of feature distances

True $(f_x, f_y) = (16.0, 16.0)$

Pred $(\bar{f}_x, \bar{f}_y) = (12.7, 14.0)$

(d) average displacement

Figure 3. Toy experiment of $\mathcal{L}_{RCB}$. Q is the image shifted by (16,16) pixels from P. (c) shows the average values computed from feature distance term $\mathbb{D}(P,Q)$ and spatial distance term $\mathbb{D}'(P,Q)$ in $\mathcal{L}_{CoBi}$ and $\mathcal{L}_{RCB}$. (d) shows the difference between an average of predicted displacement $(\bar{f}_x, \bar{f}_y)$ computed from $\mathbb{D}(P,Q)$ and the true shift (f_x, f_y).

along the channel dimension. The integrated feature map $\mathbf{F}$ is transformed into a vector by collapsing both the spatial dimension and the channel dimension. We first collapse the channel dimension using 2-layer convolution denoted as η: $\mathbf{F}' = \eta(\mathbf{F})$. η consist of two 1×1 convolutional layers with ReLU, where the first layer serves as channel attention for the integrated feature map, and the second layer handles dimension reduction. We get the final frame selection vector $\mathbf{f} \in \mathbb{R}^{N}$ by applying spatial global average pooling:

$$\mathbf{f} = \frac{\sum_{h\in H}\sum_{w\in W}\mathbf{F}'_{:,h,w}}{H\times W}, \quad (1)$$

where each value in the vector $\mathbf{f}$ represents the score of a base frame candidate. During inference, we apply the argmax operation to the frame selection vector $\mathbf{f}$ to determine the base frame index as follows: $\hat{k} = \text{argmax}(\mathbf{f})$, where $\hat{k}$ is the predicted base frame index.

2.2 Loss Functions

Frame selection loss by active learning. We employ hypothesis-based active learning [23, 24] to train our frame selector. A frame index whose frame yields the highest PSNR is given as supervision in the training dataset. We perform N times evaluation, letting each of the N burst frames be the base frame once. A burst index k with the highest value among the N evaluated results is used as the ground-truth base frame index. Then, we convert the index k into a ground-truth one-hot vector $\mathbf{f}^{\text{GT}} \in \mathbb{R}^{N}$. Note that the value of $\mathbf{f}^{\text{GT}}$ for a specific iteration can be changed while the entire burst enhancement network is being trained. We utilize the cross-entropy between the ground-truth frame selection vector $\mathbf{f}^{\text{GT}}$ and the predicted frame selection vector $\mathbf{f}$ with softmax function. The process can be defined as follows:

$$\mathcal{L}_{Frame} = \sum(-\mathbf{f}^{\text{GT}} * \log(\rho(\mathbf{f}))), \quad (2)$$

where ρ is a softmax function.

Relational Contextual Bilateral Loss. We propose the relational contextual bilateral loss (RCB), which addresses large spatial misalignment issues in the spatial distance term

a contextual bilateral loss (CoBi) [19]. When a source image P contains n number of feature points and a target image Q contains m number of feature points, the CoBi loss can be defined as follows:

$$\mathcal{L}_{\text{CoBi}}(P,Q) = \frac{1}{n}\sum_{i}^{n} \min_{j=1,..,m} \mathbb{D}(p_i,q_j) + w_s \mathbb{D}'_{\text{abs}}(p_i,q_j), \quad (3)$$

where $\mathbb{D}(p_i,q_j)$ is a distance matrix between source feature point $p_i \in P$ and target feature point $q_j \in Q$ and $\mathbb{D}'_{\text{abs}}(\text{p}_\text{i},\text{q}_\text{j}) = \sqrt{(\text{x}_\text{i}-\text{x}_\text{j})^2 + (\text{y}_\text{i}-\text{y}_\text{j})^2}$ measures spatial distance. (x_i, y_i) and (x_j, y_j) are the spatial coordinates of feature point p_i and q_j, respectively. w_s denotes the weight of spatial awareness and is determined by measuring the amount of misalignment in the training dataset. However, $\mathbb{D}'_{\text{abs}}(p_i,q_j)$ considers absolute coordinate differences only, potentially leading to incorrect optimization in cases of large misalignment between source and target. To mitigate this, the spatial distance term of $\mathcal{L}_{\text{CoBi}}$ is modified to encode relative spatial distance by compensating for an average displacement. The updated relative spatial distance $\mathbb{D}'_{\text{rel}}$ can be defined as follows:

$$\mathbb{D}'_{rel}(p_i,q_j) = \sqrt{\left((x_i-x_j)-\bar{f}_x\right)^2 + \left((y_i-y_j)-\bar{f}_y\right)^2}, \quad (4)$$

where $\bar{f}_x$ and $\bar{f}_y$ are scalar values of the average displacement between source to target. We can derive $\bar{f}_x$ and $\bar{f}_y$ from the feature distance matrix $\mathbb{D}(P,Q)$ as follows:

$$\bar{f}_x = \frac{1}{n}\sum_{i}(x_{\hat{\jmath}} - x_i), \quad \bar{f}_y = \frac{1}{n}\sum_{i}(y_{\hat{\jmath}} - y_i), \quad (5)$$

where $(x_{\hat{\jmath}}\ y_{\hat{\jmath}})$ is a predicted nearest neighbor coordinate of p_i. The predicted nearest neighbor coordinate can be derived as $\text{x}_{\hat{\jmath}} = \hat{\jmath} \bmod \text{W}$, and $y_{\hat{\jmath}} = \hat{\jmath}\ \left\lfloor \frac{\hat{\jmath}}{W} \right\rfloor$, where W is the width of target Q. The predicted nearest neighbor index $\hat{\jmath}$ is derived as $\hat{\jmath} = \arg\min_{\text{j}} \left(\mathbb{D}(\text{p}_\text{i},\text{q}_\text{j})\right)$. Note that we normalize $\bar{f}_x$ and $\bar{f}_y$ using the width and height of the input. Finally, the proposed RCB loss is defined as:

$$\mathcal{L}_{RCB}(P,Q) = \frac{1}{n}\sum_{i}^{n} \min_{j=1,..,m} \mathbb{D}(p_i,q_j) + \mathbb{D}'_{\text{rel}}(p_i,q_j), \quad (6)$$

where D'_{rel} compensates for incorrectly penalizing the distance, even though they were correctly matched. Figure 3 (c) shows $\mathcal{L}_{RCB}$ penalizes less than $\mathcal{L}_{CoBi}$ by computing the relative coordinate difference in the spatial distance term $\mathbb{D}'(P,Q)$ if the flow field obtained from $\mathbb{D}(P,Q)$ is consistent. Figure 3 (d) shows that $(\bar{f}_x, \bar{f}_y)$ obtained from $\mathbb{D}(P,Q)$ can compensate for incorrect penalizing to nearly reconstruct the true shift (f_x, f_y). We use the perceptual loss to train the enhancement modules as follows:

$$\mathcal{L}_{Percept} = \mathcal{L}_{\text{AlignedL1}} + 0.001 * \mathcal{L}_{RCB}, \quad (7)$$

where $\mathcal{L}_{\text{AlignedL1}}$ is an aligned L1 loss as in [1, 2]. The total loss is $\mathcal{L}_{Total} = \mathcal{L}_{\text{Percept}} + 0.01 * \mathcal{L}_{Frame}$.

3. Benchmark

We propose a new synthetic benchmark, Dynamically Exposed Burst (DEB), which simulates heterogeneous image noise and motion blur with natural misalignments from camera movement of video datasets.

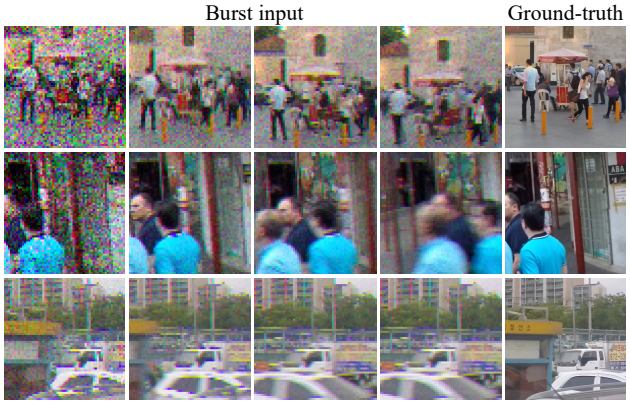

Figure 4. Visualization of DEB. The left four columns display a subset of 14 input bursts, while the rightmost column is the ground-truth image. From the left to the right in burst, the exposure time increases.

We first convert 120 images to RAW images through invert tone mapping and invert gamma compression [25, 26], then inject shot noise following the Poisson distribution with $\lambda \in [1.5^2, 1.5^{10}]$ [16, 27]. Given the various types of cameras, noise is changeable. Therefore, we generate the noisy image by employing the different range of λ within $[1.5^2, 1.5^{10}]$ for the effect that the images look like they are taken with different cameras. After that, we synthesize the motion blur by averaging the frames gradually [28], which increases blur and decreases noise, similar to increasing exposure time [21, 29]. Finally, the bursts go through four steps in the unprocessing pipeline for four channels of RAW space [25]: inverse gain, inverse CCM, downsampling, and mosaicing. Note that the input video frames naturally contain shifting between frames for hand motion, and this is sufficient to simulate the misalignment of the synthetic burst frames. We finally obtain a synthetic RAW burst sequence with heterogeneous image noise and motion blur.

The DEB consists of 1,128 synthetic dynamic exposed RAW bursts sourced from GoPro [30]. Each burst sequence consists of burst size N of 14 RAW frames, which center-cropped to 640×640 pixels. The source frames downscaled by a downsampling factor $s = 4$, resulting in $4 \times 80 \times 80$ input bursts. Ground-truth images are the source image of the first frame, sized at $3 \times 640 \times 640$ pixels. Figure 4 demonstrates the effect of the shorter exposure, which is sharp but noisy, while the images on the right demonstrate the effect of the longer exposure, which is high-SNR but blurry.

4. Experiment

4.1 Experimental settings

We train models from scratch on the Dynamically Exposed Burst (DEB) dataset. The dataset is split into a ratio of 728:200:200 for train, validation, and test sets, respectively. Burst frames are randomly shuffled during training to ensure adaptability to general shooting environment. We use AlignedPSNR, AlignedSSIM, and AlignedLPIPS as evaluation metrics following to [1, 2]. We perform early

Table 1. Results on DEB. We train all models from scratch using the DEB training set. '+ours' denotes the results with our frame selector and the proposed loss functions on top of the baseline model.

	Aligned PSNR ↑	Aligned SSIM ↑	Aligned LPIPS ↓
DBSR [1]	19.837	0.545	0.550
+ours	**25.113**	**0.753**	**0.539**
BIPNet [2]	29.836	0.883	0.389
+ours	**32.116**	**0.918**	**0.293**

Table 2. Ablation study of the loss functions. All the case we include $\mathcal{L}_{\text{AlignedL1}}$. In this experiment, we train our frame selector with BIPNet [2] on the DEB.

	Aligned PSNR ↑	Aligned SSIM ↑	Aligned LPIPS ↓
$\mathcal{L}_{Frame} + \mathcal{L}_{RCB}$	32.116	0.918	0.293
$\mathcal{L}_{RCB}$	32.055	0.917	0.303
$\mathcal{L}_{CoBi}$ [19]	27.882	0.847	0.497

stopping with the AlignedPSNR value of the validation set.

4.2 Quantitative Results

We evaluate the effectiveness of our frame selector by adding on two existing burst image enhancement methods, DBSR [1] and BIPNet [2], on the DEB. Table 1 shows a quantitative comparison with the existing burst image enhancement models and the results of including our frame selector in them. Compared to the trained model of the two representative burst image enhancement methods [1, 9], our model shows a consistent improvement by incorporating our frame selector, $\mathcal{L}_{Frame}$ and $\mathcal{L}_{RCB}$ into the existing models. Furthermore, existing burst image enhancement methods are designed without considering the heterogeneous degradation situation, which leads to subpar performance. On the other hand, our method selects a robust base frame among the input burst frames captured in the heterogeneous degradations, which improves the performance to facilitate the alignment and fusion among burst frames. In addition, our RCB Loss improves the training stability by correcting the wrong training signal in the large misalignment between the predicted image and the GT image.

4.3 Ablation study

Table 2 shows an ablation study of the loss functions to validate each component. In rows 1 and 2, we confirm that our frame selection loss improves performance on the target task while learning the submodule. This shows that the ground-truth base frame supervision obtained actively by the model through alternation is effective for generalization rather than overfitting the training data. In rows 2 and 3, our RCB loss improves performance through more stable training compared to the existing shift-invariant perceptual loss [19]. This is because our RCB loss compensates for large shifts between the prediction and ground-truth image generated by the input bursts better than the existing loss [19]. Note that large misalignments are inevitably caused by handshake and object motion as the acquisition time of the input burst increases.

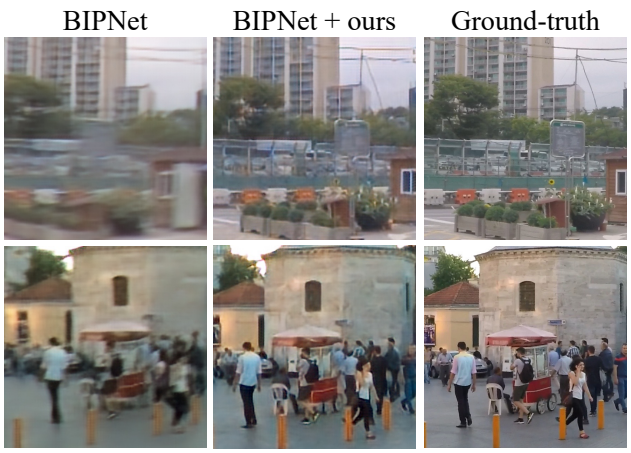

Figure 5. Qualitative comparison on DEB. Our frame selector with BIPNet [2] enhances high-frequency image details by merging effectively from multiple frames, while BIPNet [2] predicts blurry images.

4.4 Qualitative Results

In this subsection, we present the qualitative comparison to verify the effectiveness of our method. Fig 5 shows a qualitative comparison of our frame selector with BIPNet [2]. The results on DEB show well-reconstructed sharp and complex edges of the construction site and the street with pedestrians by our frame selector compared to the sole BIPNet [2]. Overall, our frame selector helps to reconstruct high-frequency details such as sharp edges and texture, while predicted images of sole BIPNet [2] are blurry and noisy.

5. Conclusion

We have proposed a novel burst image enhancement approach using dynamic exposures and base frame selection. We also have introduced a new benchmark and learnable module for selecting the best base frame, along with the relational contextual bilateral loss to improve image quality. Our evaluations on synthetic datasets demonstrated the effectiveness of our proposed method. We believe that our proposed method can contribute to the advancement of burst image enhancement techniques and pave the way for further research in this field.

Acknowledgement

This research was funded from the Government of Korea (Ministry of Science and ICT) in 2021 and supported by the National Research Foundation of Korea (NRF-2021R1A2C3012728), the National Institute of Information and Communication Technology (No. 2019-0-01906: Graduate School of Artificial Intelligence (POSTECH)), and Samsung Electronics (IO201208-07822-01).

Reference

[1] G. Bhat, M. Danelljan, L. Van Gool, and R. Timofte, "Deep burst super-resolution," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2021 (pp. 9209-9218).

[2] A. Dudhane, S.W. Zamir, S. Khan, F.S. Khan, and M.H. Yang, "Burst image restoration and enhancement," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2022 (pp. 5759-5768).

[3] S.W. Hasinoff, D. Sharlet, R. Geiss, A. Adams, JT Barron, F. Kainz, J. Chen, and M. Levoy, "Burst photography for high dynamic range and low-light imaging on mobile cameras," ACM Transactions on Graphics (ToG). 2016 Nov 11;35(6):1-2.

[4] O., Liba, K. Murthy, Y.T Tsai, T. Brooks, T. Xue, N. Karnad, Q. He, J.T. Barron, D. Sharlet, R. Geiss, and S.W. Hasinoff, "Handheld mobile photography in very low light." ACM Trans. Graph., 2019., 38(6), pp.164-1.

[5] B. Wronski, I. Garcia-Dorado, M. Ernst, D. Kelly, M. Krainin, C.K. Liang, M. Levoy, and P. Milanfar, "Handheld multi-frame super-resolution," ACM Transactions on Graphics (ToG), 2019, 38(4), pp.1-18.

[6] C, Dong, C.C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," IEEE transactions on pattern analysis and machine intelligence. 2015 Jun 1;38(2):295-307.

[7] S.W. Zamir, A. Arora, S. Khan, M. Hayat, F.S. Khan, M.H. Yang, and L. Shao, "Learning enriched features for real image restoration and enhancement," In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16 (pp. 492-511). Springer International Publishing.

[8] S.W. Zamir, A. Arora, S. Khan, M. Hayat, F.S. Khan, M.H. Yang, and L. Shao, "Multi-stage progressive image restoration," In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition 2021 (pp. 14821-14831).

[9] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image restoration," IEEE transactions on pattern analysis and machine intelligence 2020, 43(7), pp.2480-2495.

[10] G. Bhat, M. Danelljan, F. Yu, L. Van Gool, and R. Timofte, "Deep reparametrization of multi-frame super-resolution and denoising," In Proceedings of the IEEE/CVF International Conference on Computer Vision 2021 (pp. 2460-2470).

[11] C. Chen, Q. Chen, J. Xu, V. and Koltun, "Learning to see in the dark," In Proceedings of the IEEE conference on computer vision and pattern recognition 2018 (pp. 3291-3300).

[12] M. Deudon, A. Kalaitzis, I. Goytom, M.R. Arefin, Z. Lin, K. Sankaran, V. Michalski, S.E. Kahou, J. Cornebise, and Y. Bengio, "Highres-net: Recursive fusion for multi-frame super-resolution of satellite imagery," 2020, arXiv preprint arXiv:2002.06460.

[13] S. Guo, X. Yang, J. Ma, G. Ren, and L. Zhang, "A differentiable two-stage alignment scheme for burst image reconstruction with large shift," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2022 (pp. 17472-17481).

[14] A.S. Karadeniz, E. Erdem, and A. Erdem, "Burst photography for learning to enhance extremely dark images," IEEE Transactions on Image Processing 2021, 30, pp.9372-9385.

[15] B. Lecouat, J. Ponce, and J. Mairal, "Lucas-kanade reloaded: End-to-end super-resolution from raw image bursts," In Proceedings of the IEEE/CVF International Conference on Computer Vision 2021 (pp. 2370-2379).

[16] B. Mildenhall, J.T. Barron, J. Chen, D. Sharlet, R. Ng, and R. Carroll, "Burst denoising with kernel prediction networks," In Proceedings of the IEEE conference on computer vision and pattern recognition 2018 (pp. 2502-2510).

[17] A.B. Molini, D. Valsesia, G. Fracastoro and E. Magli, "Deepsum: Deep neural network for super-resolution of unregistered multitemporal images," IEEE Transactions on Geoscience and Remote Sensing 2019, 58(5), pp.3644-3656.

[18] D. Zhao, L. Ma, S. Li and D. Yu, "End-to-end denoising of dark burst images using recurrent fully convolutional networks," 2019, arXiv preprint arXiv:1904.07483.

[19] X. Zhang, Q. Chen, R. Ng and V. Koltun, "Zoom to learn, learn to zoom." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2019 (pp. 3762-3770).

[20] R. Mechrez, I. Talmi and L. Zelnik-Manor, "The contextual loss for image transformation with non-aligned data," In Proceedings of the European conference on computer vision (ECCV) 2018 (pp. 768-783).

[21] O. Dahary, M. Jacoby and A.M. Bronstein, "Digital gimbal: End-to-end deep image stabilization with learnable exposure times," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2021 (pp. 11936-11945).

[22] N. Kanopoulos, N. Vasanthavada and R.L. Baker, "Design of an image edge detection filter using the Sobel operator," IEEE Journal of solid-state circuits, 1988, 23(2), pp.358-367.

[23] B.A. Griffin and J.J Corso, "Bubblenets: Learning to select the guidance frame in video object segmentation by deep sorting frames," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2019 (pp. 8914-8923).

[24] L.M. Learning, "Synthesis Lectures on Artificial Intelligence and Machine Learning."

[25] T. Brooks, B. Mildenhall, T. Xue, J. Chen, D. Sharlet and J.T. Barron, "Unprocessing images for learned raw denoising," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2019 (pp. 11036-11045).

[26] K. Wei, Y. Fu, J. Yang and H. Huang, "A physics-based noise formation model for extreme low-light raw denoising," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2020 (pp. 2758-2767).

[27] Y. Zhang, H. Qin, X. Wang and H. Li, "Rethinking

noise synthesis and modeling in raw denoising," In Proceedings of the IEEE/CVF International Conference on Computer Vision 2021 (pp. 4593-4601).

[28] J. Rim, G. Kim, J. Kim, J. Lee, S. Lee and S. Cho, "Realistic blur synthesis for learning image deblurring," In European conference on computer vision 2022 October (pp. 487-503). Cham: Springer Nature Switzerland.

[29] Z. Zhang, R. Xu, M. Liu, Z. Yan and W. Zuo, "Self-supervised image restoration with blurry and noisy pairs," Advances in Neural Information Processing Systems 2022, 35, pp.29179-29191.

[30] S. Nah, T. Hyun Kim and K. Mu Lee, 2017. Deep multi-scale convolutional neural network for dynamic scene deblurring. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3883-3891).