A feature catcher with excellent deblurring effects

XINYUE ZHANG, the College of Computer Science and Technology, Qingdao University
GUODONG WANG*, the College of Computer Science and Technology, Qingdao University
HAO DONG, the College of Computer Science and Technology, Qingdao University
BINING ZENG, the College of Computer Science and Technology, Qingdao University

Camera shake and object movement are the two prime causes of blurred images. Efficient feature extraction is crucial for deblurring. Although the existing methods have achieved remarkable achievements in the deblurring task, there is still room for improvement in effects. In this paper, we propose an efficient architecture called the feature catcher network(FCN). In this multi-stage FCN architecture, the following design allows us to achieve improvements in performance. Firstly, we propose to apply different calculated trust ratios to the output results of different stages before calculating losses and then carry out the cumulative evaluation to update parameters for backpropagation. Secondly, we have improved Transformer to create a query-key mechanism that is effect-friendly to the deblurring task. Thirdly, we propose a multi-stage attention block to make up for the loss of information in high-level feature extraction. And the enhanced feature extraction block is employed to capture detailed information to ensure a greater degree of image recovery. Fourthly, besides considering detailed features and high-level features at the same stage, we also construct residual supplements for blurry images in the raw information mechanism. The experimental results on several datasets demonstrate that our model(FCN) outperforms state-of-the-art methods in terms of deblurring effect. The code and models will be available at https://github.com/XinyueZhangqdu/FCN.

CCS Concepts: • **Networks** → **Network architectures**.

Additional Key Words and Phrases: image deblurring, multi-stage, multi-patch

ACM Reference Format:
Xinyue Zhang, Guodong Wang*, Hao Dong, and Bining Zeng. 2022. A feature catcher with excellent deblurring effects. In *2022 5th International Conference on Big Data Technologies (ICBDT) (ICBDT 2022), September 23–25, 2022*. Qingdao, China, 10 pages. https://doi.org/10.1145/3565291.3565338

1 INTRODUCTION

With the development of the deep convolutional neural network, the robustness of the deblurring domain has been continuously enhanced, which makes the established model have amazing performance on the test sets with variability. However, the reasons for the lack of clarity are complex, such as motion blur caused by camera shake. Poor visual effects sometimes suffer from the natural environment. The complexity adds to the difficulty of building a model to solve the problem. Recently, many excellent methods have used multi-stage, multi-patch, and multi-scale to build models. However, we find that these approaches ignore the cooperation of local information and global information. So there is room for further improvement in terms of effect. Monotonously increasing the depth of the network will not improve the effect linearly. We believe that when the deep network obtains complex global information(like texture), it should combine with global information(like color and shape) to recover effectively. In this article, we build a novel

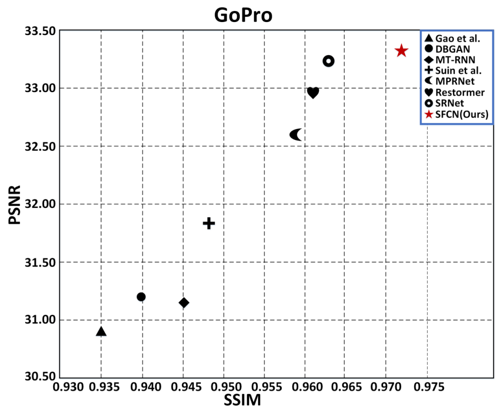


Fig. 1. This figure plots the PSNR and SSIM effects of some recent architectures on the GoPro test set [8]. It is obvious that the effect of our proposed architecture (FCN) exceeds that of other recent architectures.

architecture called FCN to improve the accuracy of deblurring. We extract multi-stage edge information of the vague images, multi-patch is applied for acquisition of input details, and multi-scale fuse the information of different stages and patches. In order to reduce the possibility of losing important information in the top layer features as much as possible, we adopt the pre-processing block, and bridge regrets by making finer use of original layer features. In the middle of the model, we put forward the enhancement feature extraction module to efficiently combine different layers in detail. In the real world, if we are observing objects and want to intuitively know nothing important content, we will be combined with usual experience from the perspective of a number of different thinking. Therefore, the processing of the attention mechanism module is equivalent to capturing the information of the picture from different angles and optimizing it in combination with the updated parameters of the backpropagation. In order to provide global information at the later stage of image restoration and effectively assemble high-level abstract information, we use the query-key mechanism to capture global information after conducting the pre-processing module. Then, the feature representations from previous stages are integrated into the multi-stage attention mechanism to obtain clear images. In addition, we find that in the training process, if the size of the initial receptive field is reduced, the optimization of the model can be promoted. At the same time, the training speed will be reduced. It's worth sacrificing some time to get parameters that better match the model. After many experiments, we find that the accuracy of the model will be improved if the attention ratio of each stage is properly adjusted before backpropagation. In summary, our main contributions are as follows:

- ★ We propose that before calculating the loss for backpropagation, the attention ratio should be applied to adjust the attention ratio for the characteristics of different stages.
- ★ We embed a processing block and an enhanced feature extraction module. On the one hand, we use the processing block to obtain the underlying edge information and detail information, and on the other hand, we embed the enhanced feature extraction module attention to obtain the feature information from different perspectives.
- ★ An effective query-key mechanism is created. The use of this mechanism makes the network can better capture global information so that shape and color information in the blurred images can be captured.

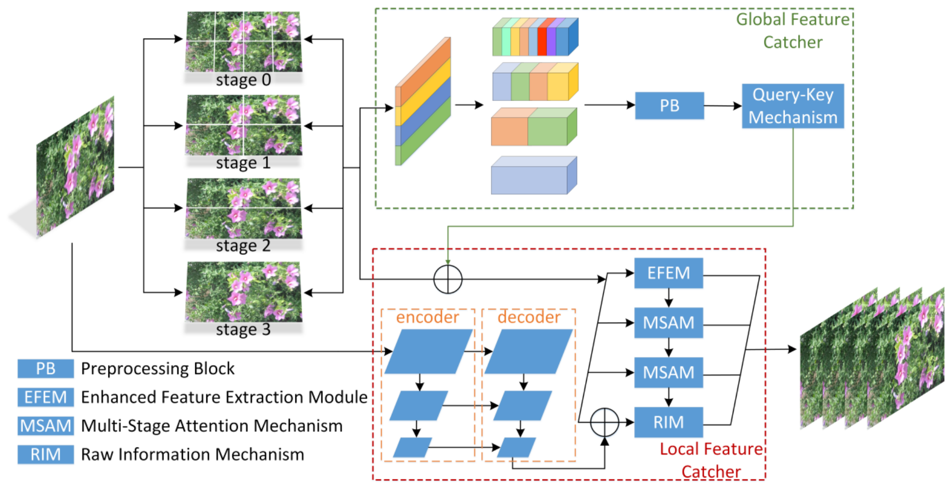


Fig. 2. The overall structure diagram of FCN proposed by us. The architecture can be roughly divided into a global feature catcher that is used to capture global features and a local feature catcher that has access to local information. The whole architecture follows the design idea of multi-patch.

- ★ We have designed attention mechanisms called the multi-stage attention mechanism and raw information mechanism that can effectively integrate complex information from multiple stages.
- ★ Through the confirmation of multiple datasets, our model achieves the best effect in the field of deblurring(as shown in Fig. 1).

  Our experiment performs excellently on the dataset of deblurring, and most importantly, we are the first to propose the use of the attention ratio to adjust output attention in different stages before calculating the loss. The main methods of multi-stage, multi-patch, and multi-scale are combined in our model, which achieves a good accuracy.

2 RELATED WORK

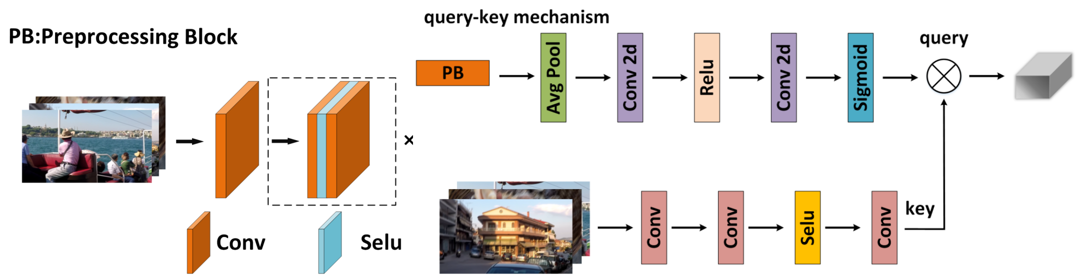
Due to the rapid development and improvement of GPUs, more and more models based on the CNN method have been used recently. The following are some relevant work introductions about multi-stage, and attention mechanisms.

2.1 Multi-stage

Existing multi-stage [4, 5, 17] technologies, either using a coder block, are effective at disseminating situational information in the encoding phase, but do a poor job of preserving spatial details. Or use a single-scale pipeline, which does a good job of spatial accuracy, but is not so good for semantic information. We build a model to address the above two problems in the multi-stage image restoration work.

2.2 Attention Mechanism

In the field of computer vision, the attention mechanism [3, 15] is widely used. Transformer is a representative application of the attention mechanism today. Different from the traditional convolution neural network(CNN) design,



(a) PB (b) query-key mechanism

Fig. 3. (a)Processing block (b)Illustration of query-key mechanism. With the help of calculating feature correlation, the ability to capture the global features is significantly improved.

the application of Transformer [13, 16] in a neural network can better capture global representations. Methods that have been itemized show their effectiveness in global information but are unreliable in preserving local details.

3 METHODS

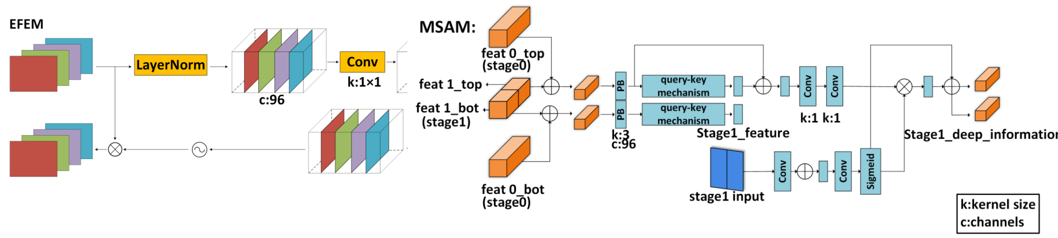
The framework we proposed for blurry image recovery mainly includes four stages. Images of different patch sizes are processed in each stage. As shown in Fig. 2, images with patch sizes of 8,4,2,1 are processed from stage0 to stage3. In the following sections, some modules embedded in the global feature catcher and local feature catcher are introduced in detail.

3.1 Processing Block

The schematic diagram of the processing block is shown in Fig. 3(a). As the first module of the global feature catcher, this module first goes through a 3×3 convolution kernel for convolution to retain the basic information of the lowest layer of the image. It is then transmitted to the query-key mechanism we designed. The number of input and output channels of this block is 96. In the module looking for details, we use Selu [2] to replace the ReLU activation function [6], so that the corresponding processing and changes can be obtained even when the value is less than 0. This operation avoids the phenomenon of overfitting in the training process. Also in this module, we use multiple stacks of the same module. As the layer number of networks increases, the features at the bottom are gradually extracted to highly abstract ones, and the features at different levels become more abundant. At the same time, the problems of gradient disappearance and explosion are effectively avoided.

3.2 Query-key Mechanism

We make a bold attempt to combine the query-key mechanism(shown in Fig. 3(b)) used in tracking with deblurring. The Query-Key mechanism in Single Object Tracking (SOT), which tracks the object of the current frame through the object characteristics of the previous frame, has great potential to establish a simple joint detection and tracking MOT paradigm. However, because the query-key method fails to detect newly emerging objects, little research has been done. In our module, we acquire more advanced image semantic information as the query information, and the initial message of the image as the key value to be retrieved. This approach is also inspired by Transformer, which correlates high-level image information with previous image information by obtaining the same number of queries and key value



(a) EFEM (b) MASM

Fig. 4. (a)Structure of enhanced feature extraction module. In order to better capture the correlation degree of feature information between channels, LayerNorm normalization function is adopted. (b)Multi-stage attention mechanism. The high-level abstract feature representation of adjacent stages is integrated to supplement the details lost in the deep network.

groups as the number of image chunks at each stage. With the design in Fig. 3 (b), some values between zero and one are stored in the query maps. Some further abstraction of the images is stored in key maps. The values stored in the query maps effectively represent the importance of each pixel region for deblurring. We effectively saved the global features we needed by associating them with the information in the key maps.

3.3 Enhanced Feature Extraction Module

Through the observation of the stability of statistics and gradients, we found that the use of a single batchNorm [10] is likely to cause more oscillations and outliers in the data distribution of the training set and test set. So we incorporated layerNorm into this enhanced feature extraction module. Then, like a regular network, we use the convolution and add the activation function to make the data fit the normal distribution. The schematic diagram of the enhanced feature extraction module is shown in Fig. 4(a). At this stage, we get rid of the convolution kernel with a large receptive field before and use a 1*1 convolution kernel. Under the condition of keeping the width and height of the feature constant, the dimension is changed. It effectively reduces the number of parameters and increases the depth of the model, which improves the performance of the model to a certain extent. In this module, although the number of channels is changed in the middle, the number of characteristic channels output by this module is the same as the number of channels input. On the one hand, the abstract information used in the next stage can be obtained; on the other hand, a short image after deblurring can be output, to keep the final calculation loss to measure the superior quality of the network.

3.4 Multi-Stage Attention Mechanism

This module needs two parts of information as input. On the one hand, it needs information from the previous stage, which is high-level abstract information after the PB module and Query-Key mechanism, and on the other side, it needs features from this stage as input. (see Fig. 4(b)) Compared with the enhanced feature extraction module, the multi-stage attention mechanism module pays more attention to the connection of the context information of different stages in the multi-stage. When considering the region of images, we should also consider its relationship with the surrounding image, to facilitate the recovery of the image information. In the current popular transformer framework, the position-encoding module is used to extract the position information of the image at first, and then combine it with the following information. With the addition of the multi-stage attention mechanism, the network can obtain abstract information without losing fine-grained information. The output of this module includes two parts: the first part is the recovered image at this stage, which can be used for subsequent loss calculation to prepare for optimization; The second part is to provide the characteristic information for the next stage to capture the context information.

3.5 Raw Information Mechanism

This module is used as the last module of stage3. The schematic diagram of the raw information mechanism is shown in Fig. 5. The input to this module has four parts: Part 1: the features processed in this stage by PB and Query-Key Mechanism. Part 2: the stage2_deep_information output from the third stage of MSAM. Part 3: The outputs(orinImg_encoder) of the input blurry images are processed by an encoder [1]. Part 4: The outputs(orinImg_decoder) of the input blurry images are processed by a decoder [1]. By adding this module, not only the details of the image can be captured, but also the original image information can be used to make up for the lack of overall information.

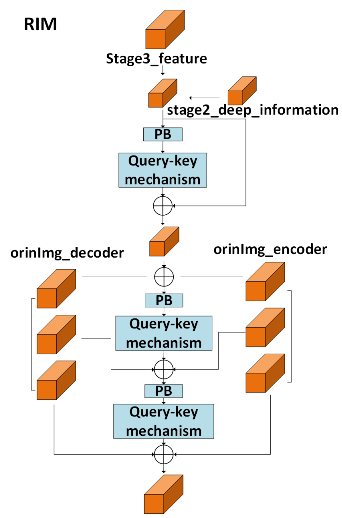


Fig. 5. Illustration of reconstruction information module in our image recovery block. This attention mechanism not only integrates the characteristic information of adjacent stages, but also the information of original images.

3.6 Loss calculation adjustment module

The use of multi-stage in the field of deblurring inspires us to think about whether we should continue to use the previous method of loss calculation. After the multi-stage loss is obtained, we can calculate the mean value for backpropagation. The loss adjustment module we proposed is formally considered from this starting point. By adjusting the attention ratio to the losses at different stages in the later stage of training, the network can more effectively determine the changes in parameters during the backpropagation.

$$Loss = \sum_{s=0}^{4} r_s(\sqrt{||P_s - G||^2} \tag{1}$$

As shown in equation 1, r_s is used to adjust the degree of attention during training at each stage. r_s is a trainable hyperparameter limited between 0 and 1. In addition, P_s is the prediction results of the network. G is used to represent the corresponding ground truths in the dataset.

4 EXPERIMENT

In this section, some experiments related to this architecture(FCN) are described in detail and the test results are reported.

4.1 Datasets

GoPro dataset [8] contains 3214 pairs of blurry and clear images, among which 2103 pairs of blurry and clear images are used as training sets and 1111 pairs of images are chosen as the testing samples. Both training samples and testing samples are captured at 720×1280 resolution. By averaging different numbers(7-13) of continuous frames, varied blur is produced which is momentous for the form of blurred images.

Hide dataset [11] is specially collected for human-aware motion deblurring which consists of 2025 pairs of blurry and clean images at 720×1280 resolution. Different from other datasets, this dataset is not divided into training sets, and 2025 pairs of data are provided as testing sets.

4.2 Training Details

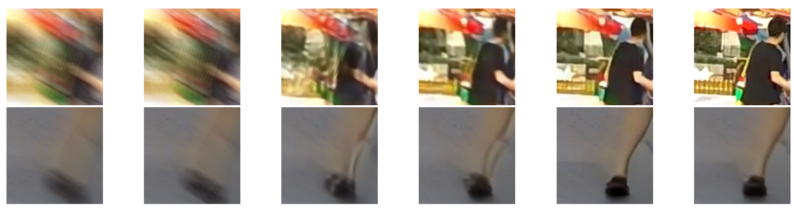
In order to save the training time, we do not use the resolution of the original images in the training. Before the training, we adjust 2103 blurred images and clear ground truth in the GoPro training set [8] to the size of 256×256. Then, as in the previous method, the test experiments are performed on unsized test sets (the resolution on these datasets is 720×1280). The initial learning rate is set at 2e-4. During the training, the learning rate is reduced by half after 200 epochs each time. A total of 600 Epochs are trained. In addition, there is no need to load any pre-training model during the initial training process. The Adam optimizer is chosen as our optimizer. It is worth noting that we do not use the model saved by the last epoch as the test model. During the training of the first 300 epochs, the effects are verified for each training of 50 epochs. We test the model on GoPro test set [8] and record PSNR and SSIM values. In the training process of the last 300 epochs, the effect verification on the GoPro test set [8] and the corresponding PSNR and SSIM values are performed for each training of 15 epochs. At each validation, PSNR and SSIM values recorded by previous tests are compared. If the current validation is superior to the previously recorded data in terms of PSNR and SSIM effectiveness, then the current model is saved as the best model. At the end of the test performed on the test datasets, the loaded pre-training model is the best model mentioned here.

4.3 Comparative Experiments

To verify the effectiveness of our proposed model, seven groups of recent architectures are subjected to the same training strategy as our architecture. After the same training strategy, we test the effect on GoPro test set [8] and the Hide dataset [11]. The results of all the comparative experiments are recorded in Table 1.

Table 1. Deblurring results. Our method is trained only on the GoPro training set [8] and directly applied to the GoPro test set [8] and Hide dataset [11]

Method	GoPro [8]		Hide [11]		Average	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Gao et al. [20]	30.90	0.935	29.11	0.913	30.00	0.924
DBGAN [21]	31.20	0.940	29.09	0.924	30.14	0.932
MT-RNN [9]	31.15	0.945	29.15	0.918	30.15	0.931
Suin et al. [12]	31.85	0.948	29.98	0.930	30.91	0.939
MPRNet [17]	32.66	0.959	30.96	0.939	31.81	0.949
Restormer [16]	32.92	0.961	31.22	0.942	32.07	0.951
SRNet [14]	33.23	0.963	30.07	0.928	31.65	0.945
FCN(Ours)	**33.32**	**0.972**	**32.04**	**0.966**	**32.96**	**0.969**



(a) Input (b) DIDAnet [7] (c) DMPHN [19] (d) MPRNet [18] (e) FCN(Ours) (f) Ground-truth

Fig. 6. Some comparisons for image deblurring on the GoPro dataset [8] and Hide dataset [11]. Compared to the state-of-the-art methods, our model recovers the image much more clearly and sharper.

4.3.1 GoPro. After training our model in the GoPro training set through the training strategy mentioned in Section 4.2, we test the effect of the model on the GoPro test set. Similar to the previous methods, PSNR and SSIM are selected to measure our effectiveness. As can be seen from the data in Table 1, our model exceeds recent architectures in both PSNR and SSIM. In particular, we test Restormer [16] architecture proposed this year on the GoPro dataset. Our architecture (FCN) outperforms Restormer [16] 0.400 and 0.011 on PSNR and SSIM, respectively. In particular, the number of model parameters for the Restormer architecture is 26.097M. However, the number of model parameters of our architecture FCN is only 20.309M. Therefore, we successfully achieve the current state of optimal effect in the deblurring field without adding additional parameters. Some visual comparisons are shown in Fig. 6.

4.3.2 Hide. To verify the generalization capability of the model designed by us, we run tests on the Hide dataset [11]. It is important to note that we did not train the model on the Hide dataset. The model used in this section has only been trained on GoPro's training set [8]. As can be seen from the data in Table 1, no matter the value of PSNR or SSIM, significant improvements are obtained by our model(FCN). Compared with recent SRNet [14], our model improves by 1.97 and 0.038 in PSNR and SSIM respectively.

4.4 Ablation Studies

To verify the importance of different modules to the network, we conduct ablation experiments. The experiments are performed on the GoPro dataset to train the model after removing some modules, and the effect is shown in the table. We combine different modules and train them to produce a model that can be tested. In order to verify the influence of each sub-module on the model as a whole, we conducted ablation experiments on the four modules proposed in this paper. The training strategy used in this ablation experiment was the same as that used in section 4.2. After the

Table 2. Ablation study result of our model

stages	PB	EFEM	MSAM	RIM	PSNR
1	×	√	√	√	28.99
1	√	×	√	√	27.45
1	√	√	×	√	26.19
1	√	√	√	×	26.84
2	×	√	√	√	29.69
2	√	×	√	√	29.32
2	√	√	×	√	28.78
2	√	√	√	×	28.35
3	×	√	√	√	30.97
3	√	×	√	√	30.48
3	√	√	×	√	29.91
3	√	√	√	×	29.33
4	×	√	√	√	31.84
4	√	×	√	√	31.24
4	√	√	×	√	30.71
4	√	√	√	×	30.04

training, PSNR values were tested on Gopro's test set and recorded in Table 2. The "stages" in table 2 represent the number of phases of the adopted architecture. By observing the experimental data, we can draw the following two conclusions :(1) no matter the size of the stage, the MSAM and RIM have the greatest influence on the model. This is enough to prove that the two modules proposed by us have an obvious effect on improving the effectiveness of image restoration. (2) As the number of stages increases, the effect of network architecture is improved. This shows that the complex network design is more conducive to image restoration than the single-stage network.

5 CONCLUSION

In this paper, an architecture that works well in the field of image deblurring is proposed. Compared with the previous method of constantly extracting the local features of the whole image, we deploy the extraction of global features as part of our architecture. At the same time, we adopt the idea of combining multiple stages and multi-patch, so as to give the network the opportunity to observe pictures from different angles. The network can be roughly divided into global feature catcher which is used to extract global features and local feature catcher which has local features extraction. In the global feature catcher, the query-key mechanism that can extract the global feature is embedded. This makes our architecture no longer give equal attention to all areas of the image. The importance of different regions is calculated and the global feature of key regions is extracted. In the local feature catcher, we embed different multi-stage fusion mechanisms, such as the multi-stage attention mechanism and the raw information mechanism. In this way, local features extracted based on different patches can be extracted and fused. At the same time, we fuse abstract information from the original image in the raw information mechanism to supplement the original information in multiple stages. As a result of the above design, our architecture manages to outperform recent architectures on multiple datasets. Several ablation experiments are performed to verify the degree of influence of each module.

6 ACKNOWLEDGEMENT

This work was supported by the Natural Science Foundation of Shandong Province (No. ZR2019MF050) and the Shandong Province colleges and universities youth innovation technology plan innovation team project under Grant (No. 2020KJN011).

REFERENCES

[1] Yuzhu Ji, Haijun Zhang, Zhao Zhang, and Ming Liu. 2021. CNN-based encoder-decoder networks for salient object detection: A comprehensive review and recent advances. *Information Sciences* 546 (2021), 835–857.

[2] Günter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. 2017. Self-normalizing neural networks. *Advances in neural information processing systems* 30 (2017).

[3] Donghyeon Lee, Chulhee Lee, and Taesung Kim. 2021. Wide receptive field and channel attention network for jpeg compressed image deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 304–313.

[4] Ru Li, Junwei Xie, Yuyang Xue, Wenbin Zou, Tong Tong, Ming Luo, and Qinquan Gao. 2022. Enhanced multi-stage network for defocus deblurring using dual-pixel images. In *Thirteenth International Conference on Signal Processing Systems (ICSPS 2021)*, Vol. 12171. SPIE, 162–168.

[5] Jinping Liu, Quanquan Gao, Zhaohui Tang, Yongfang Xie, Weihua Gui, Tianyu Ma, and Jean Paul Niyoyita. 2020. Online monitoring of flotation froth bubble-size distributions via multiscale deblurring and multistage jumping feature-fused full convolutional networks. *IEEE Transactions on Instrumentation and Measurement* 69, 12 (2020), 9618–9633. https://doi.org/10.1109/TIM.2020.3006629

[6] Lu Lu, Yeonjong Shin, Yanhui Su, and George Em Karniadakis. 2019. Dying relu and initialization: Theory and numerical examples. *arXiv preprint arXiv:1903.06733* (2019).

[7] Haoyu Ma, Shaojun Liu, Qingmin Liao, Juncheng Zhang, and Jing-Hao Xue. 2021. Defocus Image Deblurring Network With Defocus Map Estimation as Auxiliary Task. *IEEE Transactions on Image Processing* 31 (2021), 216–226.

[8] Seungjun Nah, Tae Hyun Kim, and Kyoung 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*. 3883–3891.

[9] Dongwon Park, Dong Un Kang, Jisoo Kim, and Se Young Chun. 2020. Multi-temporal recurrent neural networks for progressive non-uniform single image deblurring with incremental temporal training. In *European Conference on Computer Vision*. Springer, 327–343. https://doi.org/10.1007/978-3-030-58539-6_20

[10] Shibani Santurkar, Dimitris Tsipras, Andrew Ilyas, and Aleksander Madry. 2018. How does batch normalization help optimization? *Advances in neural information processing systems* 31 (2018).

[11] Ziyi Shen, Wenguan Wang, Xiankai Lu, Jianbing Shen, Haibin Ling, Tingfa Xu, and Ling Shao. 2019. Human-aware motion deblurring. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 5572–5581. https://doi.org/10.1109/ICCV.2019.00567

[12] Maitreya Suin, Kuldeep Purohit, and AN Rajagopalan. 2020. Spatially-attentive patch-hierarchical network for adaptive motion deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3606–3615. https://doi.org/10.1109/CVPR42600.2020.00366

[13] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. 2022. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 17683–17693.

[14] Jay Whang, Mauricio Delbracio, Hossein Talebi, Chitwan Saharia, Alexandros G Dimakis, and Peyman Milanfar. 2022. Deblurring via stochastic refinement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 16293–16303.

[15] Si Xi, Jia Wei, and Weidong Zhang. 2021. Pixel-guided dual-branch attention network for joint image deblurring and super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 532–540.

[16] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. 2022. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5728–5739.

[17] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. 2021. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 14821–14831. https://doi.org/10.48550/arXiv.2102.02808

[18] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. 2021. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 14821–14831. https://doi.org/10.48550/arXiv.2102.02808

[19] Hongguang Zhang, Yuchao Dai, Hongdong Li, and Piotr Koniusz. 2019. Deep stacked hierarchical multi-patch network for image deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5978–5986. https://doi.org/10.1109/CVPR.2019.00613

[20] Jiawei Zhang, Jinshan Pan, Jimmy Ren, Yibing Song, Linchao Bao, Rynson WH Lau, and Ming-Hsuan Yang. 2018. Dynamic scene deblurring using spatially variant recurrent neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2521–2529.

[21] Kaihao Zhang, Wenhan Luo, Yiran Zhong, Lin Ma, Bjorn Stenger, Wei Liu, and Hongdong Li. 2020. Deblurring by realistic blurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2737–2746. https://doi.org/10.48550/arXiv.2004.01860