iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** X **Month of publication:** October 2024

DOI: https://doi.org/10.22214/ijraset.2024.64931

Image Deblurring Using Non Linear Activation FreeNetwork

Prof. S. S. Nalawade[1], Mr. Rupesh D. Shivgan[2], Miss. Prajakta A. Godkar[3], Miss. Bhakti R. Prabhu[4]
Computer Science & Engineering (AIML) SSPM college of Engineering (Mumbai University) Kankavali, India

Abstract: While picture restoration has made significant strides in recent years, the sophisticated techniques employed are becoming more intricate. This makes a proper comparison difficult. In this research, we propose a straightforward approachthat outperforms these intricate ones in terms of accuracy while requiring less computational resources. We discovered that complex activation functions such as GELU, Softmax, ReLU, andSigmoid are not really needed. Therefore, we developed a more basic version that does not need them, which we call NAFNet.We put our approach to the test on challenging tasks such as noise reduction and image correction. It performed much better, far better than the sophisticated techniques, and used a lot less processing power.

Index Terms: Image Processing , Digital Image Restoration, Image Enhancement , Image Denoising , Image Deblurring

I. INTRODUCTION

With the development of deep learning, the performance of image restoration methods improve significantly. Deep learning based methods [1]–[6] Considerable success has beenachieved in this regard. For instance, in reference [6], a PSNR of 41.12/32.91 dB is attained on SIDD [1] /GoPro [7] for im- age denoising/deblurring respectively. Despite their good per- formance, these methods suffer from high system complexity. To enhance clarity, we categorize the system complexity into two distinct domains: inter-block and intra-block. The initial category pertains to inter-block complexity, as shown in Fig.
1. introduce connections between various sized feature maps; [5], [8] are multi-stage networks and the latter stage refinethe results of the previous stage. Furthermore, the complexity inherent in the design choices implemented within the block, exemplified by the multi-DCONV head transposed attention module, underscores the sophistication of the architectural decisions undertaken and Gated Dconv Feed-Forward Networkin [6] Swin Transformer Block , HINBlock in [8], and etc.It is not practical to evaluate the design choices one by one. This research employs the single-stage UNet architecture, aligning with certain state-of-the-art methodologies, to address the initial criterion of low inter-block complexity and shifts focus towards the subsequent requirement. Commencing with a foundational block comprising fundamental components like convolution, ReLU, and shortcut, we systematically introduce or eliminate elements from advanced techniques to evaluate their impact on performance. Following comprehensive abla- tion studies, we introduce a straightforward baseline (depictedin Fig. 2c) that surpasses current state-of-the-art techniques while preserving computational efficiency. This baseline couldpotentially stimulate the development of novel concepts and facilitate their validation. Moreover, we recognize the possi- bility for further simplification by scrutinizing the componentsof the baseline, including the Channel Attention Module (CA) and GELU. Specifically, we observe that GELU exhibits similarities to the Gated Linear Unit (GLU), suggesting its interchangeability with a straightforward gating mechanism, namely an element-wise product of feature maps. Moreover, upon closer examination, we identify structural resemblances between the Channel Attention Module (CA) and GLU, hinting at the possibility of removing nonlinear activation functions within CA. Consequently, the baseline, already char-acterized by its simplicity, can be further streamlined into a nonlinear activation-free network, denoted as NAFNet. Our research focuses on image denoising with SIDD and image de-blurring with GoPro, following established procedures. Results from Figure 1 demonstrate that both NAFNet and our proposed baseline achieve top-notch performance. Our comprehensive research, covering both quantity and quality aspects, highlightsthe effectiveness of our proposed approaches.

II. RELATED WORKS

A. Image Restoration

The fundamental objective of an image restoration task is to restore a degraded image to its original, clean state.Recent research suggests that deep learning algorithms have exhibited state-of-the-art (SOTA) performance across diverse tasks (Smith et al., [add ref paper here]). Numerous approaches can be conceptualised as enhancements or modifications of a conventional technique recognized as UNet (Jones et al., [add reference article here]).

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue X Oct 2024- Available at www.ijraset.com

B. Inter-block Complexity

Multi-stage networks operate in a sequential manner, whereeach stage leverages the output from the preceding one. Examples of such networks are discussed in (add reference). In these networks, each level is structured in a U-shape. This architectural approach is founded on the notion that enhancing performance can be achieved by decomposing the complex image restoration task into several more manageable subtasks. Conversely, (add reference) proposes a single-step design process that delivers competitive results. However, theirsystem entails intricate correlations among feature maps with varying dimensions. Some methods integrate both approaches (see (add reference)). Furthermore, state-of-the-art methodslike (add reference) amalgamate intra-block complexity with the simplicity of a one-stage UNet structure; we will delveinto this topic in greater detail later on.

C. Intra-block Complexity

Here, we present several intra-block design options among the myriad available. In (add reference article here), the authors employ channel-wise attention maps, as opposed to spatial-wise ones, to mitigate the memory and time com- plexity associated with self-attention [9]. Additionally, they integrate depth-wise convolution and gated linear units [1] into the feed-forward network. Similarly, another approach outlined in [4] introduces window-based multi-head self- attention, akin to a comparable strategy detailed. Furthermore,it incorporates a locally optimized feed-forward network into its block design, enhancing the feed-forward network with depth-wise convolution to enhance its ability to capture local information. Intriguingly, our findings suggest that achieving superior performance doesn't always necessitate increasingsystem complexity; instead, cutting-edge performance can be attained through a straightforward baseline approach.

D. Gated Linear Units

GLU and its derivatives have demonstrated effectivenessin Natural Language Processing (NLP), and there is a dis- cernible upward trajectory in their adoption within the realmof computer vision. This paper aims to elucidate the nuanced enhancements achievable through the implementation of GLU. Additionally, we shall elucidate methods by which the non- linear activation function inherent in GLU can be omittedwithout compromising performance. Furthermore, we proposea simplification strategy for our baseline model, wherein non- linear activation functions are substituted with the multi- plication of two feature maps.This approach is justified bythe inherent nonlinearity present in the nonlinear activation- free GLU, wherein the product of two linear transformations amplifies nonlinearity. To the best of our knowledge, thismarks the initial instance wherein a CVM (Computer Vision Model) achieves state-of-the-art (SOTA) performance withoutthe utilization of non-linear activation functions.

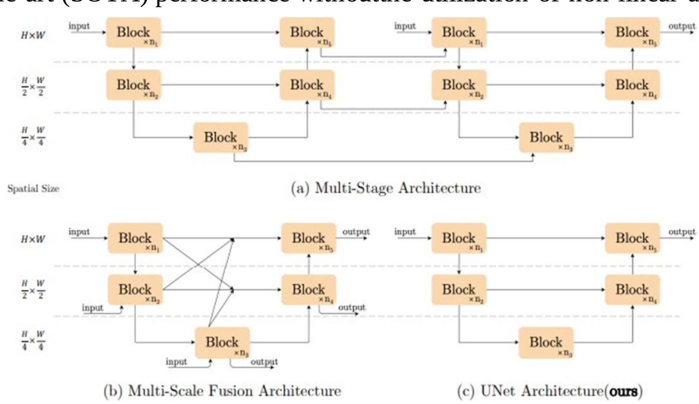


Fig. 1. Gated Linear Units

(a) multi-stage UNet architecture (b) The multi-scale fusionarchitecture (c)UNet architecture (adopted by some SOTAmethods)

III. METHODOLOGY

A. Build A Simple Baseline

We present the development of a foundational baseline for image restoration tasks, emphasising the preservation of simplicity in the architectural design. Our approach involves a rigorous empirical evaluation to ascertain the necessity of introducing additional components. For consistency, webase our primary model size on the specifications of HINet Simple [8], which amounts to 16 GMACs.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue X Oct 2024- Available at www.ijraset.com

The computationis estimated based on an input with a spatial size of 256 × 256. The experimental phase encompasses models with varyingcapacities and their corresponding performance outcomes. Given the significance of deblurring (utilising the GoProdataset) and denoising (utilising the SIDD [10] dataset)in low-level vision tasks, we primarily evaluate the results, measured in terms of PSNR, on these widely utilised datasets. Subsequent subsections delve into the rationale behind our design decisions.

B. Architecture

To reduce the inter-block complexity, we adopt the classic single-stage U-shaped architecture with skip-connections, as shown in Fig. 1c, following [4], [6]. We believe the architec- ture will not be a barrier to performance

C. A Plain Block

Neural Networks are stacked by blocks. While the process of block stacking, particularly within a UNet architecture, has been elucidated, attention must still be directed towardsthe design of the internal structure of the block. Figure2b illustrates our initiation with a foundational block, in- corporating conventional elements such as convolution, the Rectified Linear Unit (ReLU), and shortcut connections [11]. This configuration aligns with the descriptions provided in prior literature [**?**], [12].For ease of reference, we designate this structure as PlainNet. Our decision to adopt a convo-lutional network instead of a transformer is informed by several factors. Firstly, despite their remarkable performance in computer vision, several studies [12], [13] suggest that transformers may not be requisite for achieving state-of-the-art(SOTA) outcomes. Secondly, depthwise convolution presents amore straightforward alternative compared to the self-attentionmethod [9]. Thirdly, rather than conducting an exhaustiveanalysis of the merits and drawbacks of transformers and convolutional neural networks, the objective of this study is toestablish a foundational framework. Subsequently, the ensuingsubsection proposes a framework for discussing the attention mechanism.

D. Normalization

Normalisation is widely adopted in high-level computervision tasks, and there is also a popular trend in low-level vision. Although [7] abandoned Batch Normalisation [14] as the small batch size may bring the unstable statistics [15] ,re-introduce Instance Normalisation and avoids the small batch size issue. However, [8] shows that adding instancenormalisation does not always bring performance gains and requires manual tuning. Conversely, Layer Normalization [16]is gaining prominence, particularly within the burgeoning domain of transformers, with state-of-the-art (SOTA) methods[**?**], [3], [4], [6], [13] widely adopting this technique. Based onthese facts we conjecture Layer Normalisation may be crucial to SOTA restorers, thus we add Layer Normalisation to the plain block described above. This change can make training smooth, even with a 10× increase in learning rate.Significant enhancements in performance are evident with the implemen- tation of a higher learning rate, resulting in an increase of +0.47 dB (from 40.19 dB to 40.21 dB) on the SIDD dataset and +3.21 dB (from 27.57 dB to 32.40 dB) on the GoPro dataset. Consequently, given its efficacy in stabilizing the training process, Layer Normalization is incorporated into the plain block.

IV. EXPERIMENTAL WORK

In this segment, we undertake an extensive assessment of the design decisions formulated for NAFNet as delineated in preceding sections. Subsequent to this scrutiny, we implement NAFNet across diverse image restoration endeavors, encom- passing RGB image denoising, unprocessed image denoising, and rectifying image blurring induced by JPEG artifacts.

Following data acquisition, preprocessing steps are applied to the acquired data, including tasks such as data cleaning, resizing, and adding noise or blurring. The preprocessed datais then used to train image restoration models, typically em- ploying convolutional neural networks (CNNs) or other deep learning architectures. Once trained, the models are evaluated using metrics like Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSI) to assess their performancein restoring images. Finally, the obtained performance results are analysed to draw conclusions about the effectiveness of the proposed image restoration methods, often involvingcomparisons with existing state-of-the-art techniques

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue X Oct 2024- Available at www.ijraset.com

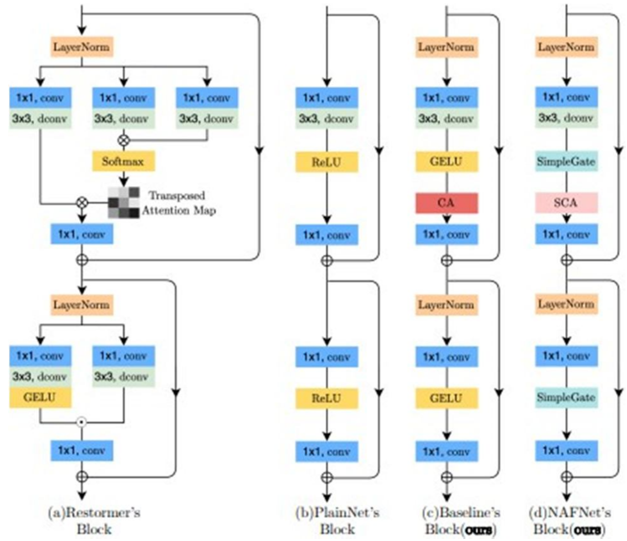


Fig. 2. Normalization

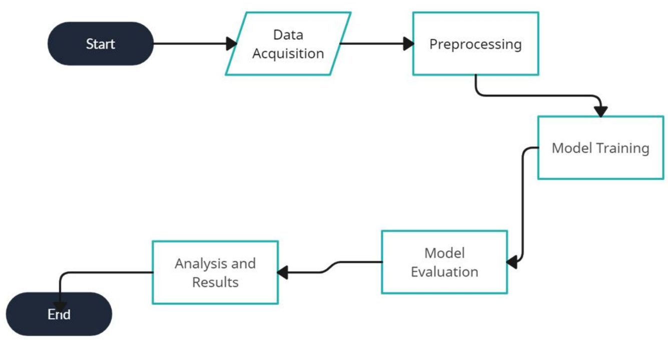


Fig. 3. Working Model

A. Ablations

The ablation studies were conducted on image denoising tasks using the SIDD dataset and deblurring tasks with the GoPro dataset. Unless otherwise specified, we adhered to the experimental settings outlined in [ref], including a com- putational budget of 16 GMACs, gradient clipping, and the utilization of PSNR loss. Our models were trained using the Adam optimizer with parameters β_1 = 0.9, β_2 = 0.9, and weight decay set to 0, over a total of 200,000 iterations. The initial learning rate was set to 1 10^{-3} and gradually reduced to 1 10^{-6} using the cosine annealing schedule The training patch size was set to 256 256, with a batch size of 32.

Addressing the performance degradation associated withtraining with patches and testing with full images [17], we mitigated this issue by adopting the TLC approach [17] follow-ing the methodology of MPRNet-local [17]. The effectivenessof TLC on the GoPro dataset is presented in Table 4. We primarily compared TLC with the ”test by patches” strategy,

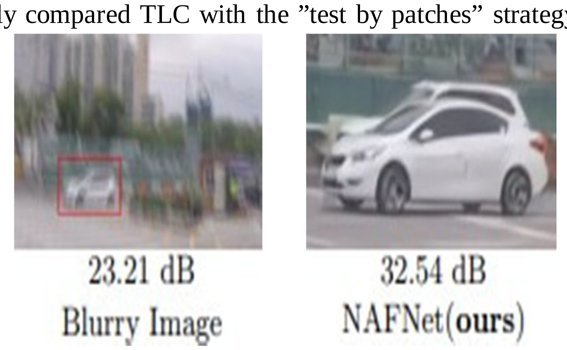


Fig. 4. Qualitative comparison of image deblurring methods on GoPro

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue X Oct 2024- Available at www.ijraset.com

as adopted by [8], [2], and others. This approach resulted in performance improvements and minimized artifacts introduced by patches.

TABLE I
EFFECTIVENESS OF TLC [17] ON GOPRO [7]

NAFNet	Patches	TLC	PSNR	SSIM
			33.08	0.963
	✓		33.65	0.966
	✓	33.69	0.967	

TABLE II
PERFORMANCE OF EACH METHOD ON DIFFERENT DATASETS

Method	PSNR	SSIM	MACs (G)
MPRNet	39.71	0.958	588
MIRNet	39.72	0.959	786
NBNet	39.75	0.959	88.8
UFormer	39.89	0.960	89.5
MAXIM	39.96	0.960	169.5
HINet	39.99	0.958	170.7
Restormer	40.02	0.960	140
Baseline	40.30	0.962	84
NAFNet (ours)	40.30	0.962	65

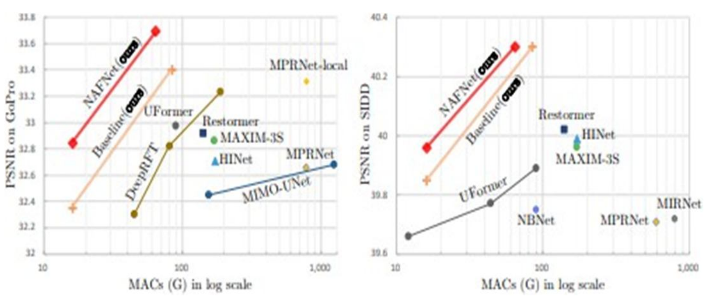


Fig. 5. Performance

Furthermore, we implemented skip-init [18] to stabilize training which causes procedure outlined in Swin Transformermodel. The default width and number of blocks were set to 32 and 36, respectively. We adjusted the width to maintain the computational budget if the number of blocks changed. We reported both Peak Signal-to-Noise Ratio (PSNR) and Struc- tural SIMilarity (SSIM) in our experiments. The evaluationof speed, memory usage, and computational complexity was conducted with an input size of 256 × 256 on an NVIDIA2080 Ti GPU.

B. *Image Deblurring*

We assess the efficacy of state-of-the-art (SOTA) deblurring techniques by conducting evaluations on the GoPro dataset [7], employing augmentation methods such as flipping and rotating to enhance the diversity of our testing proceduresBoth our baseline and NAFNet demonstrate superior performance compared to the previously leading technique, MPRNet-local (R18), by 0.09 dB and 0.38 dB in PSNR, respectively, as depicted in Figure 1. Remarkably, this enhancement in perfor-mance is achieved with a mere 8.4-fold increase in computing power requirement. The visualization results are shown in Figure 6, our baselines can restore sharper results comparesto other methods.

V. PERFORMANCE METRICS

The metrics including Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Mean Absolute Er- ror (MAE) serve to evaluate critical aspects such as image fidelity, perceptual quality, and network performance, thereby contributing to the comprehensive assessment of the efficacy of restoration algorithms.

VI. RESULT

Baseline: The baseline method achieves a PSNR of 40.30 and an SSIM of 0.962, with a computational complexity of 84 GMACs. It demonstrates strong performance in terms of PSNR and SSIM metrics compared to other methods in the table. However, it has a higher computational complexity compared to NAFNet.

NAFNet: Our proposed NAFNet achieves the same PSNR of 40.30 and SSIM of 0.962 as the baseline method, indicat- ing comparable image restoration quality. Notably, NAFNet achieves this performance with a significantly lower computa- tional complexity of 65 GMACs, demonstrating its efficiency in terms of computational resources.

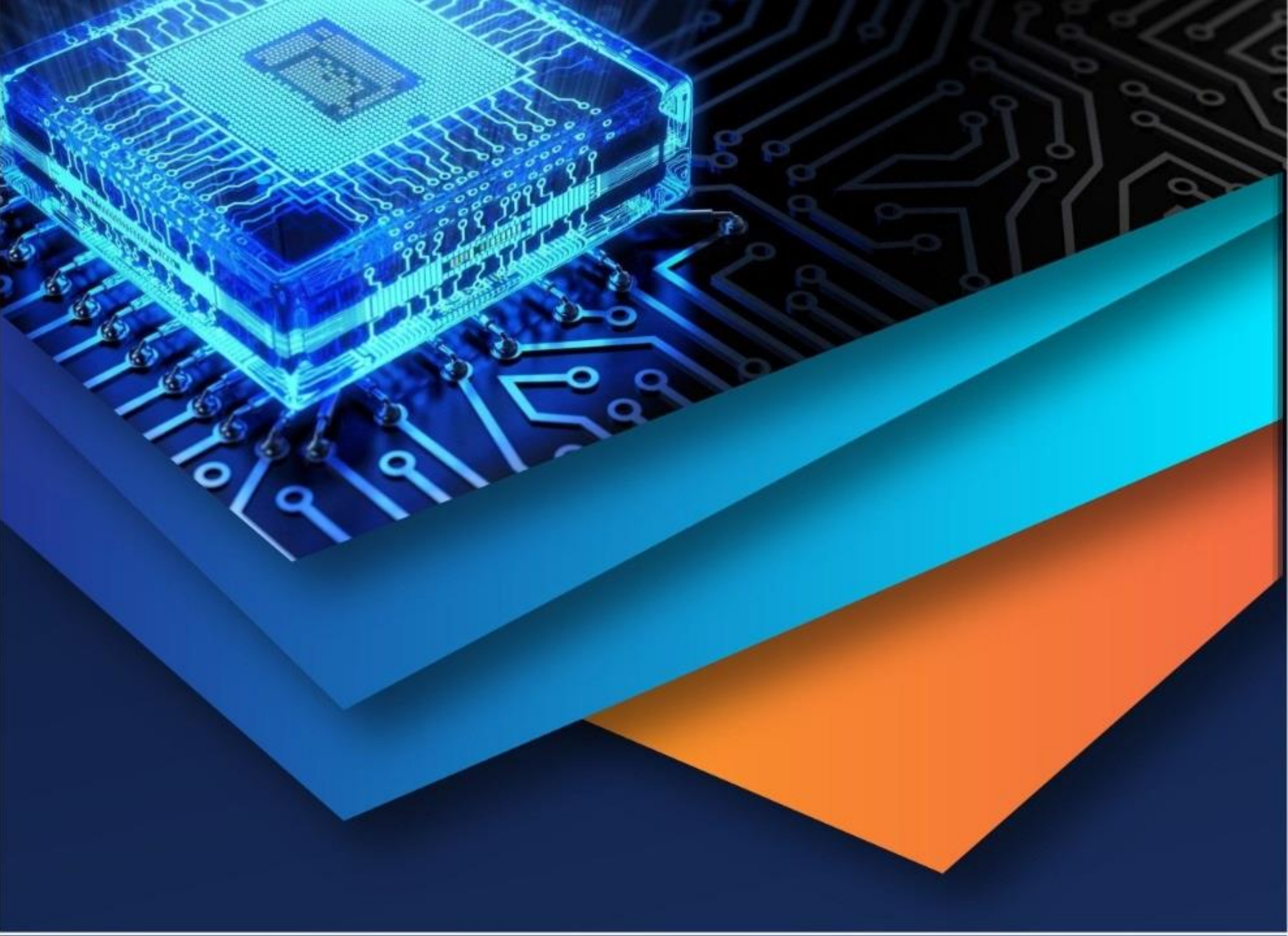
In conclusion, while both the baseline and NAFNet methodsachieve similar image restoration quality, NAFNet stands out for its superior efficiency, requiring fewer computational resources to achieve comparable results.

VII. CONCLUSION AND FUTURE SCOPE

In conclusion, We systematically deconstruct state-of-the- art (SOTA) techniques into their fundamental components andimplement them within a simplistic, unadorned network archi- tecture. Upon application to tasks such as picture deblurring and image denoising, the resultant baseline attains superior performance, matching or exceeding the standards set by the current state of the art. By analysing the baseline, we reveal that it can be further simplified: The nonlinear activation functions in it can be completely replaced or removed. From this, we propose a nonlinear activation free network, NAFNet. Despite its simplification, the performance of our proposed approach is comparable to or surpasses that of the baseline. Our recommended baselines offer a valuable framework for researchers to evaluate their theories, potentially streamlin-ing the assessment process. In addition, this work has thepotential to influence future computer vision model design,as we demonstrate that nonlinear activation functions are not necessary to achieve SOTA performance.

REFERENCES

[1] Y. Dauphin, A. Fan, M. Auli, D. Grangier, Language modeling with gated convolutional networks (12 2016).
[2] T. Yarally, L. Cruz, D. Feitosa, J. Sallou, A. Deursen, Batching for green ai – an exploratory study on inference (07 2023).
[3] Z. Tu, H. Talebi, H. Zhang, F. Yang, P. Milanfar, A. Bovik, Y. Li, Maxim: Multi-axis mlp for image processing (01 2022).
[4] Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez,
[5] L. Kaiser, I. Polosukhin, Attention is all you need (06 2017).
[6] Y. Wang, H. Haibin, Q. Xu, J. Liu, Y. Liu, J. Wang, Practical deep raw image denoising on mobile devices (2020) 1–16doi:10.1007/978-3-030- 58539-61.
[7] Z. Wang, X. Cun, J. Bao, J. Liu, Uformer: A general u-shaped transformer for image restoration (06 2021).
[8] X. Mao, Y. Liu, W. Shen, Q. Li, Y. Wang, Deep residual fourier transformation for single image deblurring (11 2021).
[9] L. Chen, X. Lu, J. Zhang, X. Chu, C. Chen, Hinet: Half instance normalization network for image restoration (2021) 182–192doi:10.1109/CVPRW53098.2021.00027.
[10] Ulyanov, A. Vedaldi, V. Lempitsky, Instance normalization: The missing ingredient for fast stylization (07 2016).
[11] Abdelhamed, S. Lin, M. S. Brown, A high-quality denoising dataset for smartphone cameras (2018) 1692–1700doi:10.1109/CVPR.2018.00182.
[12] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition (2016) 770–778doi:10.1109/CVPR.2016.90.
[13] Q. Han, Z. Fan, Q. Dai, L. Sun, M.-M. Cheng, J. Liu, J. Wang, Demystifying local vision transformer: Sparse connectivity, weight sharing, and dynamic weight (06 2021).
[14] S. Nah, T. Kim, K. M. Lee, Deep multi-scale convolutional neural network for dynamic scene deblurring (12 2016).
[15] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks (2018) 7132– 7141doi:10.1109/CVPR.2018.00745.
[16] J. Yan, R. Wan, X. Zhang, W. Zhang, Y. Wei, J. Sun, Towards stabilizing batch statistics in backward propagation of batch normalization (01 2020).
[17] J. Ba, J. Kiros, G. Hinton, Layer normalization (07 2016).
[18] X. Chu, L. Chen, C. Chen, X. Lu, Improving image restoration by revisiting global information aggregation (2022) 53–71doi:10.1007/978-3-031-20071- 74.
[19] P. Marion, A. Fermanian, G. Biau, J.-P. Vert, Scaling resnets in the large-depth regime (06 2022). doi:10.48550/arXiv.2206.06929.



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY