Exploring Object Recognition and Image Denoising in Demanding Unreal and Virtual Environments

Pranav Premdas
School of Electrical & Computer Engineering
Georgia Institute of Technology
Atlanta, USA
ppremdas3@gatech.edu

M. Sait Kilinc
School of Electrical & Computer Engineering
Georgia Institute of Technology
Atlanta, USA
mklinc3@gatech.edu

***Abstract*—In this paper, we perform advanced image processing techniques to tackle the complexities associated with denoising and improving images sourced from diverse datasets with various types of noises and challenging conditions, including CURE-OR, CURE-TSR, CURE-TSD, SSID, and Set-12. Our denoising strategies encompass a hybrid methodology that combines Anisotropic diffusion, Multi-Axis Multilayer perceptron, and Improved adaptive gamma correction for enhancement purposes. This was followed by image quality assessment using performance metrics such as PSNR, SSIM, CW-SSIM, UNIQUE, MS-UNIQUE, CSV, and SUMMER which allows us to systematically assess the effectiveness of these methods in handling various types of noise. The primary objective of the study is to offer insights into the strengths and limitations inherent in each technique, thereby contributing to the progress of image processing in both challenging real-world and simulated environments.**

Keywords—denoising, enhancement, Image quality assessment, object recognition, dataset

I. Introduction

In recent studies, there has been a growing interest in understanding the impact of challenging conditions on image recognition performance [1]. For instance, investigations into the recognition performance of the Google Cloud Vision API revealed vulnerabilities under conditions such as Gaussian and impulse noise, highlighting the susceptibility of this API to tested challenging scenarios[2, 3]. Similarly, assessments of deep neural networks demonstrated their sensitivity to degradation caused by factors like blur, noise, contrast, and compression, with a particular emphasis on the vulnerabilities associated with blur and noise[4-6].

Despite these valuable insights, existing studies have primarily focused on simulated challenging conditions and often overlooked potential imperfections in the image acquisition process [7, 8]. Moreover, studies on adversarial examples have demonstrated vulnerabilities, but these examples are deliberately crafted to bias the model and may not accurately reflect realistic challenging scenarios [9-12]. Previous work in traffic sign recognition and detection systems, while addressing realistic challenging conditions, has limited applicability to a specific field [13, 14].

To address these gaps, our paper takes a comprehensive approach to analyze various denoising and enhancement methods and the impact on generic object recognition tasks [15-17]. We utilize diverse datasets, including CURE-OR, CURE-TSR, CURE-TSD, SSID, and Set-12, to conduct a benchmark analysis of existing solutions in real-world scenarios. Our inquiry utilizes a flexible strategy, combining the effectiveness of Anisotropic diffusion, Multi-Axis Multilayer perceptron, and Improved adaptive gamma correction. The combination of approaches provides a nuanced examination of improving image quality. Our work aims to improve image quality and evaluate the adaptation of different approaches across diverse noise types. This is done via rigorous examination using standard measures like as PSNR, SSIM, CW-SSIM, UNIQUE, MS-UNIQUE, CSV, and SUMMER. This study explores the complexities of denoising and enhancement, providing significant insights to the subject.

II. Methods

A. Denoising Algorithm

In this study, we pursued a comprehensive understanding of denoising and enhancement algorithm robustness under diverse and challenging conditions. To achieve this, we carefully selected a set of datasets that collectively represented a broad spectrum of real-world and simulated scenarios. The chosen datasets, namely CURE-OR, CURE-TSR, CURE-TSD, SSID, and Set-12, were integral to our goal of evaluating algorithmic performance in various environments.

1. Improved adaptive gamma correction

This algorithm was developed based on the principles illustrated in [18]. C. Gang et. al. 2018 use the global brightness average to determine the type of brightness distortion, i.e., the dimmed and bright images, as it offers computational efficiency and simplicity. However, this method lacks spatial distribution information while determining the brightness distortion type and can be influenced by outliers. To overcome this, in this paper we have introduced a more robust way of determining the brightness distortion type and level. We analyze the histogram and the CDF which provides a detailed representation of the distribution of the pixel and allows for fine-grain analysis to set the threshold of different exposure types and level.

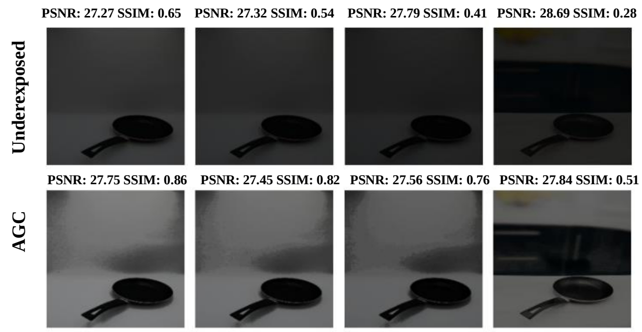


Fig. 1 Evaluation of AGC on level 1-4 of underexposed images from CURE-OR dataset

2. *Anisotropic diffusion*

Anisotropic diffusion is an edge-preserving denoising algorithm used mostly in medical image denoising. In this study, we apply the Perona-Malik diffusion technique to remove noise without removing the details of the image. Unlike isotropic diffusion which has equal diffusion in all direction, anisotropic diffusion adapts to the local structure within the image. The diffusion process is controlled by the gradient of the image, therefore higher the gradient lower the diffusion. This is governed by the following equation:

$$\frac{\partial I}{\partial t} = \nabla(\mathrm{c}(\mathrm{x},\mathrm{y},\mathrm{t})\nabla \mathrm{I}) \text{ and } c = e^{-(\frac{||\nabla I||}{k})^2}$$

Where, $c(x, y, z)$ is the diffusion coefficient, which controls the amount of diffusion and k is the constant that determine s the sensitivity to the edges.

Fig. 2 Evaluation of noise reduction using AD on 3 test image from the SSID dataset

3. *Multi-Axis Multilayer perceptron*

Z. Tu et al. 2022 introduced an efficient and flexible general purpose vision model with a backbone U-net shape architecture with long range interaction enabled by spatially gated MLP. The model available online is not a fine-tunable model, hence for this study we used a pre-trained model to evaluate the performance. There were 5 models for deblurring, dehazing, denoising, deraining and enhancement, to utilize this we categories the noise type into the following categories, enhancement (overexposure, underexposure, darkening, and shadow), blur (Gaussian blur, lens blur, unreal blur), natural (rain and snow), haze and noise (salt & pepper noise, Gaussian noise, noise).

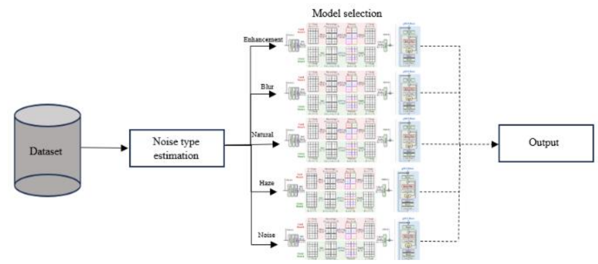


Fig. 3 Model selection and prediction pipeline of MAXIM algorithm

Fig. 4 Evaluation of MAXIM on darkened, Gaussian blur, rain, haze and noisy images from CURE-TSD dataset

B. *Object detection/recogonization*

In this study, we evaluate the impact of denoising on object detection and recogonization.

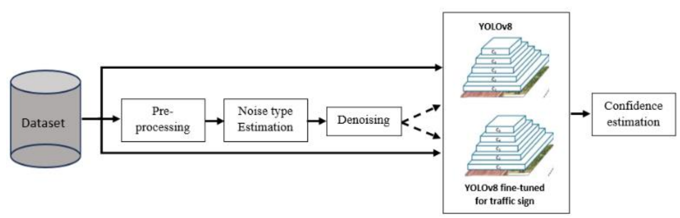
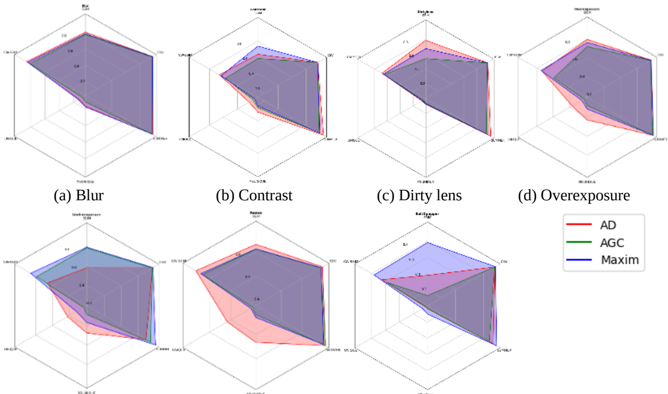


Fig. 5 Object detection performance comparison architecture.

Two models were chosen .i.e., the traffic sign detection model with base YOLOv8 architecture for Object recogonization and pretrained YOLOv8x model for object detection. Finally, the confidence level of the prediction was evaluated on the noisy images and denoised image.

III. Results

In this section, we present the outcomes of both denoising and object recognition processes, employing three distinct methods: Anisotropic Diffusion (AD), Adaptive Gamma Correction (AGC), and Multi-Axis Multi-Layer Perception (MAXIM). These algorithms were tested on same 50 images for each type of noise equally distributed among the noise levels in each of the datasets introduced in the preceding section, and the ensuing results are detailed below. The final results are displayed using the Radar plot to show the performance in terms of PSNR, SSIM, CW-SSIM, UNIQUE, MS-UNIQUE, SUMMER, and CSV. The PSNR was excluded in the radar plot as the range of PSNR was higher than other evaluation metrics (refer Appendix. I for complete table with PSNR).



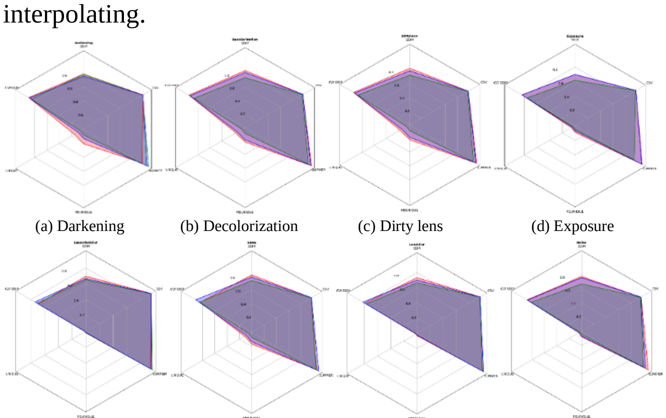
(a) Blur (b) Contrast (c) Dirty lens (d) Overexposure

(e) Underexposure (f) Resize (g) Salt & pepper

Fig. 6 Radar plot of evaluation metrics of AD, AGC, and Maxim on CURE-OR

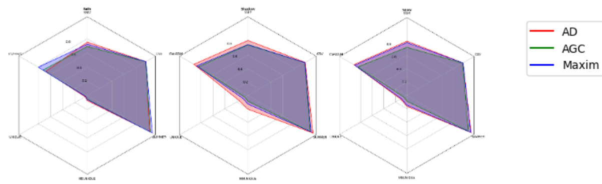
Fig. 6 shows AD performs well when there are texture noises, i.e., Dirty lens and salt & pepper noise present in the image. But, it also perform well on overexposed images. This is mainly because overexposed images often suffer from increased levels of noise due to high brightness level. On the contrary, AGC performs well only on underexposed images, but fails in over exposure, this is due to the clipping effect. Since most pixels are near the maximum in bright images, the application of non-linear gamma correction leads to values exceeding the maximum which are clipped. This can be overcome by performing tone mapping alongside gamma correction. The Maxim model has appreciable performance in all kinds of noise except dirty lens, hence showing robustness in the model architecture.

The CURE-TSR dataset is a particularly challenging dataset mainly due to two reasons .i.e., the vast variation in image size ranging from 10×10 to 128×128 and also the highly challenging conditions of noisy images. All the images in this dataset were resized to 256×256 to maintain consistency while comparing the algorithms as the MAXIM architechture is built for 256×256 input size, this introduces addition artifacts while interpolating.



(a) Darkening (b) Decolorization (c) Dirty lens (d) Exposure

(e) Gaussian Blur (f) Haze (g) Lens Blur (h) Noise

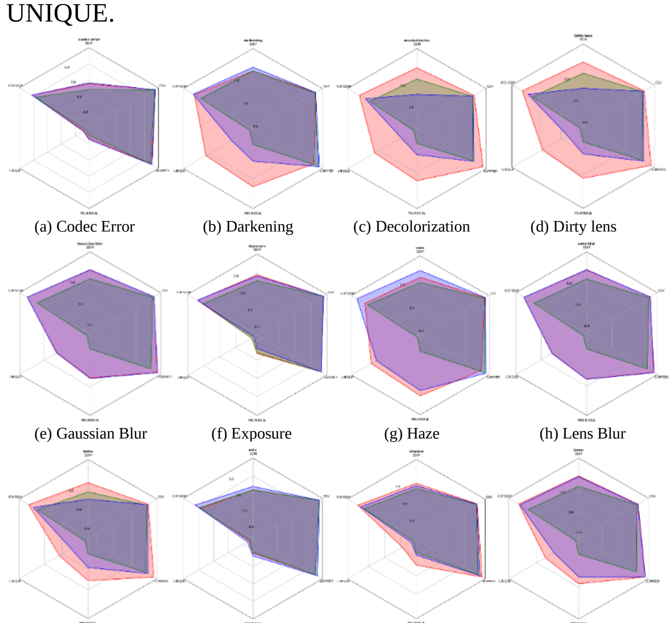


(i) Rain (j) Shadow (k) Snow

Fig. 7 Radar plot of evaluation metrics of AD, AGC, and Maxim on CURE-TSR

The above-mentioned reasons explain the lack of conclusive performance distinction of any algorithm on the different noise types in Fig. 7. This can be overcome by using an ML architecture with dynamic input size, thereby, negating the need for resizing.

The CURE-TSD dataset was evaluated by first separating the video file into separate frames. 50 frames were chosen from each type of noise across all levels. The noteworthy performance here in Fig 8. is of the AD algorithm which has high performance for darkening, decolorization, dirty lens, noise, shadow and snow. This is mainly due to the texture and localized nature of these noises. For instance, dirty lens, snow and shadow noises have only certain regions of noise, therefore the AD smoothens these regions. The AGC has comparable performance on darkening and exposure in terms of SSIM, CW-SSIM, SUMMER and CSV but has a very low UNIQUE and MS-UNIQUE values. This is due to two reasons .i.e., color space transformation, as AGC affects only the luminance channel it might fail to capture the features used in UNIQUE and model mismatch, as AGC is not focused on preserving sharpness which is given higher weight in UNIQUE and MS-UNIQUE.



(a) Codec Error (b) Darkening (c) Decolorization (d) Dirty lens

(e) Gaussian Blur (f) Exposure (g) Haze (h) Lens Blur

(i) Noise (j) Rain (k) Shadow (l) Snow

Fig. 8 Radar plot of evaluation metrics of AD, AGC, and Maxim on CURE-TSD

The SIDD small sRGB dataset folder which consists of 160 images was used to evaluate the performance of the algorithms. It can be seen that MAXIM and AGC have high evaluation metric score across most metrics as the image data is mostly consistent of noisy images under different lighting condition, which as seen previously in other dataset is optimal noise for AGC and MAXIM. But, it is observed that the AD algorithm which is developed to handle texture noise has very low scores in SSIM, CW-SSIM, UNIQUE and MS-UNIQUE. This is mainly because these IQA methods focus on sharpness and structure persistence and AD without appropriate hyper parameter tuning can fail to retain the sharpness in complex images. Hence appropriate hyper parameter tuning might improve the performance of the algorithm.

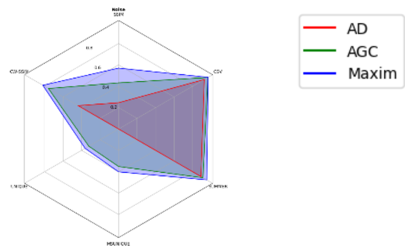


Fig. 9 Radar plot of evaluation metrics of AD, AGC, and Maxim on SIDD

In the Set-12 images, three levels of additive white Gaussian noise were added of σ = 5, 10 and 20. The performance of the algorithm was tested on all 3 noise levels. It can be seen that AD and MAXIM which have previously shown high performance on texture noises have high scores which remain consistent with the observation in Fig. 10. It is again observed that the UNIQUE and MS-UNIQUE scores are low due to failure of the algorithm to preserve sharpness in complex images. The AGC as expected does a poor job as it is mainly used to perform exposure correction.

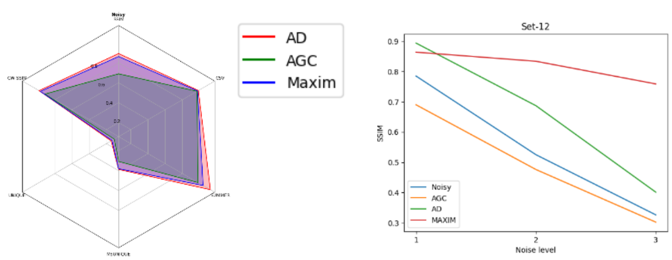
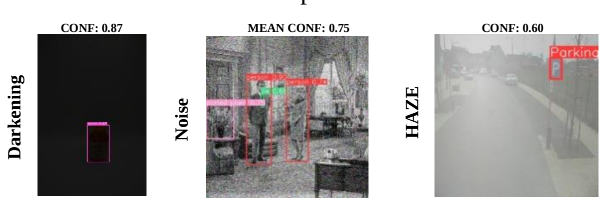


Fig. 10 Evaluation metrics of AD, AGC, and Maxim on CURE-TSD dataset

It can also be observed from Fig. 10 that low level-1 noise AD outperforms MAXIM, showing that for simple texture noises AD works best. But, as the complexity of noise increases MAXIM which is more complex denoising method has high performance compared to AD.

Finally, the effect of denoising was evaluated on object detection and recogonization. For this, the confidence score for prediction was estimated before and after denoising. As seen in Fig. 11 it is seen that when the algorithm is used to tackle the right type of noise .i.e, enhancement for AGC, texture for AD and most noise types for MAXIM, there is a significant increase in the confidence score of the prediction.



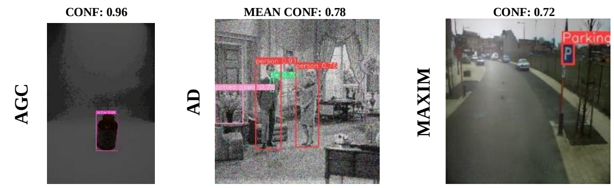


Fig. 11 Evaluation of the effect on object detection/recogonization after applying AGC, AD, and MAXIM on darkened, noisy, haze noise images respectively.

IV. Conclusion

In conclusion, our analysis of the strengths and weaknesses of the three different image denoising and enhancement methods – Adaptive gamma correction, Anisotropic Diffusion, and MAXIM MLP model – across a range of real and unreal noise conditions has provided some valuable insights.

The AGC algorithm showed significant performance in exposure type noises with limited adaptability to other noises. On the contrary, AD showed great adaptability to different types of texture noises while preserving the inherent structure of the image indicated by high SSIM scores. The MAXIM MLP model exhibited great versatility especially in complex noise types due to its complex ML architecture and its ability to characterize complex noise, showing a promising architecture to build on further for denoising and enhancement algorithms.

It is evident from this study that there isn't one denoising algorithm that works best for all types of noises. The charcateristics of the image and the type of noise play an important role in determining the best method for denoising. Therefore, it is important to consider the trade-off while choosing the denosing method.

However, in the future the algorithms presented in this study can be further improved by adding complexity. This can be done by combining two denoising methods, hyper parameter tuning, multiscale denoising, etc, as a preprocessing step to machine vision algorithms like object detection/recogonization for improved performance in real world application.

V. CONTRIBUTION

In this study, Pranav mainly worked on characterizing the noise across datasets and developing the denoising/enhancement algorithms i.e., AD, AGC and MAXIM and worked on a comprehensive way to generate and summarize the performance using various evaluation metrics using the radar plot. Furthermore, integrated the object detection pipeline into the architecture to evaluate the performance. Sait contributed to preparing the denoising algorithm, executing the algorithms, and played a role in preparing the report and poster. GitHub repo (https://github.com/ppremdas3/Object-Detection-and-Image-denoisign-in-demanding-Unreal-and-Real-Environment.git)

References

1. Jayaraman, D., et al., *Objective Quality Assessment of Multiply Distorted Images.* 2012 Conference Record

of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (Asilomar), 2012: p. 1693-1697.
2. Dodge, S. and L. Karam, *Understanding How Image Quality Affects Deep Neural Networks.* 2016 Eighth International Conference on Quality of Multimedia Experience (Qomex), 2016.
3. Saenko, K., et al., *Adapting Visual Category Models to New Domains.* Computer Vision-Eccv 2010, Pt Iv, 2010. **6314**: p. 213-+ DOI: Doi 10.1007/978-3-642-15561-1_16.
4. Temel, D., M. Prabhushankar, and G. AlRegib, *UNIQUE: Unsupervised Image Quality Estimation.* Ieee Signal Processing Letters, 2016. **23**(10): p. 1414-1418 DOI: 10.1109/Lsp.2016.2601119.
5. Singh, A., et al., *BigBIRD: A Large-Scale 3D Database of Object Instances.* 2014 Ieee International Conference on Robotics and Automation (Icra), 2014: p. 509-516.
6. Geusebroek, J.M., G.J. Burghouts, and A.W.M. Smeulders, *The Amsterdam Library of Object Images.* International Journal of Computer Vision, 2005. **61**(1): p. 103-112 DOI: Doi 10.1023/B:Visi.0000042993.50813.60.
7. Torralba, A., R. Fergus, and W.T. Freeman, *80 million tiny images: A large data set for nonparametric object and scene recognition.* Ieee Transactions on Pattern Analysis and Machine Intelligence, 2008. **30**(11): p. 1958-1970 DOI: 10.1109/Tpami.2008.128.
8. Everingham, M., et al., *The PASCAL Visual Object Classes Challenge: A Retrospective.* International Journal of Computer Vision, 2015. **111**(1): p. 98-136 DOI: 10.1007/s11263-014-0733-5.
9. Mogelmose, A., M.M. Trivedi, and T.B. Moeslund, *Vision-Based Traffic Sign Detection and Analysis for Intelligent Driver Assistance Systems: Perspectives and Survey.* Ieee Transactions on Intelligent Transportation Systems, 2012. **13**(4): p. 1484-1497 DOI: 10.1109/Tits.2012.2209421.
10. Bugiel, P., J. Izydorczyk, and T. Sulkowski, *Image Augmentation Techniques for Road Sign Detection in Automotive Vision System.* Artificial Intelligence Methods in Intelligent Algorithms, 2019. **985**: p. 229-242 DOI: 10.1007/978-3-030-19810-7_23.
11. Larsson, F. and M. Felsberg, *Using Fourier Descriptors and Spatial Models for Traffic Sign Recognition.* Image Analysis, 2011. **6688**: p. 238-249.
12. Grigorescu, C. and N. Petkov, *Distance sets for shape filters and shape recognition.* IEEE Trans Image Process, 2003. **12**(10): p. 1274-86 DOI: 10.1109/TIP.2003.816010.
13. Opelt, A. and A. Pinz, *Object localization with boosting and weak supervision for generic object recognition.* Image Analysis, Proceedings, 2005. **3540**: p. 862-871.
14. Li, F.F., R. Fergus, and P. Perona, *Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories.* Computer Vision and Image Understanding, 2007. **106**(1): p. 59-70 DOI: 10.1016/j.cviu.2005.09.012.
15. He, K.M., et al., *Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification.* 2015 Ieee International Conference on Computer Vision (Iccv), 2015: p. 1026-1034 DOI: 10.1109/Iccv.2015.123.
16. Afriyie, Y., B.A. Weyori, and A.A. Opoku, *A scaling up approach: a research agenda for medical imaging analysis with applications in deep learning.* Journal of Experimental & Theoretical Artificial Intelligence, 2023 DOI: 10.1080/0952813x.2023.2165721.
17. Dalal, N. and B. Triggs, *Histograms of oriented gradients for human detection.* 2005 Ieee Computer Society Conference on Computer Vision and Pattern Recognition, Vol 1, Proceedings, 2005: p. 886-893 DOI: DOI 10.1109/cvpr.2005.177.
18. Cao, Gang & Huang, Lihui & Tian, Huawei & Xianglin, Huang & Wang, Yongbin & Zhi, Ruicong. (2018). Contrast enhancement of brightness-distorted images by improved adaptive gamma correction. Computers & Electrical Engineering. 66. 569-582. 10.1016/j.compeleceng.2017.09.012.
19. Z. Tu et al., "MAXIM: Multi-Axis MLP for Image Processing," 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 2022, pp. 5759-5770, doi: 10.1109/CVPR52688.2022.00568.