Impact of a Sharpness Based Loss Function for Removing Out-of-Focus Blur

Uditangshu Aurangabadkar⋆, Darren Ramsook†, Anil Kokaram‡

Sigmedia Group, Dept. of Electronic and Electrical Engineering, Trinity College Dublin, Ireland

⋆aurangau@tcd.ie, †ramsookd@tcd.ie, ‡anil.kokaram@tcd.ie

www.sigmedia.tv

Abstract—**Recent research has explored complex loss functions for deblurring. In this work, we explore the impact of a previously introduced loss function – Q which explicitly addresses sharpness and employ it to fine-tune State-of-the-Art (SOTA) deblurring models. Standard image quality metrics such as PSNR or SSIM do not distinguish sharpness from ringing. Therefore, we propose a novel full-reference image quality metric Ω that combines PSNR with Q. This metric is sensitive to ringing artifacts, but not to a slight increase in sharpness, thus making it a fair metric for comparing restorations from deblurring mechanisms. Our approach shows an increase of 15% in sharpness (Q) and up to 10% in Ω over the use of standard losses.**

Index Terms—**Deblurring, Sharpness, Sharpness Metric, Ringing, Loss Function**

I. INTRODUCTION

Image deblurring is the problem of recovering a sharp image $\hat{I}$ from a blurry image G such that it resembles the original image I. The process of blurring an image can be defined as follows:

$$G = I * H_K + \eta \tag{1}$$

where H_K is a blur kernel of size $K \times K$ and $\eta \in \mathcal{N}(0, \sigma^2)$ is the additive Gaussian noise. Depending on whether the details of the blur kernel are given or not, the process can be classified as *non-blind* or *blind*, respectively. The current work focuses on blind deblurring as it is more relevant for real-world restoration tasks.

Classical methods such as the Wiener filter or Richardson-Lucy deconvolution [1], [2] have long been the standard for restoring images. However, with the widespread adoption of Deep Neural Networks (DNNs) for restoration, primarily deblurring, several works [3], [4] have proposed complex architectures that produce State-of-the-Art (SOTA) results in terms of visual quality of the reproduced image.

In practice, the main challenge in deblurring or restoration is to produce a sharper image without going so far as to introduce ringing artifacts. Ringing or *over sharpening* results when high frequency components are boosted too much. A good restoration result is therefore often designed by applying a post-process after the initial deblurring step, removing any artifacts produced in the process. Several works [5]–[7], have now shifted to using a hybrid approach combining classical or iterative deblurring methods with Machine Learning (ML) based post-processors to produce sharper restorations, yet

carefully removing any artifacts produced in the process. The method proposed by Chen et. al. [8] utilizes the Richardson-Lucy deconvolution algorithm as an initial step to remove blur from an image, followed by a Deep Neural Network (DNN) to remove any artifacts introduced in the process.

None of the previous techniques explicitly address sharpness of a restored image while also assessing the onset of ringing. What is needed, therefore, is a new loss function that focuses on enhancing this quality. An example would be to use the loss proposed by Aurangabadkar et. al. [9] that is based on a no-reference metric Q proposed by Zhu and Milanfar [10], first introduced to optimize iterative denoising methods such as BM3D [11] with the goal of producing sharper restored images. In this paper, we combine this sharpness specific loss with the corresponding losses in a number of SOTA models in order to show the positive effect on image quality. This has not been done before.

An unfortunate consequence of using a loss function like Q, is that it increases monotonically with sharpness even when that sharpness manifests as ringing. An image with excessive ringing will therefore exhibit high Q yet show a low Structural Similarity (SSIM) [12] or PSNR with respect to the ground truth reference. Inspired by this observation, we present a novel full-reference image quality metric Ω that combines PSNR with Q, such that it is sensitive to ringing. This new metric is then used to fairly assess the performance of our SOTA models trained with a hybrid loss function. Our primary contributions are as follows.

- An analysis of the incorporation of our sharpness based loss function Q [9] in training SOTA deblurring models.
- A novel full-reference image quality metric that combines PSNR and metric Q, hence properly accounting for ringing.

II. BACKGROUND

It is clear that both the architectures and loss functions have an impact on the performance of DNN based deblurring mechanisms. We therefore consider three examples of architectures that span SOTA performance – ARKNet [9], XY-Deblur [13] and EHNet [14].

ARKNet proposed by Aurangabadkar et. al. [9] is a standard U-Net [15] based architecture comprising of 4 encoder *layers*, where each layer consists of 5 convolutional blocks. Each block, in turn, comprises of a single 3×3 convolution

layer, followed by Batch Normalization [16] and GELU [17] activation. The model contains a total of 4.8 million trainable parameters.

XY-Deblur introduced by Ji et. al. [13] is a single encoder multiple decoder architecture initially intended for restoring images degraded by motion blur. The model leverages the fact that employing multiple decoders allows for decomposing features into directional components, namely horizontal and vertical. The use of shared kernels amongst the decoders allows for improved deblurring performance. These caveats keep the total number of trainable parameters identical to a standard U-Net, viz. 4.2 million, while producing significantly sharper restorations.

EHNet proposed by Ho et. al. [14] is a transformer based architecture that combines Convolutional Neural Networks (CNNs) and transformers to create a hybrid deblurring mechanism. The CNNs allow for efficient local feature extraction, whereas the transformer decoder with dual-attention enable the model to capture spatial and channel-wise dependencies. The model consists of approximately 8.7 million trainable parameters. Detailed diagrams of all the architectures can be found in the supplementary materials. *

A. Loss Functions

One way to address sharpness would be to use a loss function that explicitly increases sharpness in an image. In our previous work [9], we propose a loss function which is based on a no-reference metric Q introduced by Zhu and Milanfar [10].

For measuring Q, an image is first divided into non-overlapping patches of size $k \times k$. The singular values s_1 and s_2 ($s_1 > s_2$), which determine the patch properties are calculated based on the eigen-decomposition of the matrix $\mathbf{G^T G}$, where $\mathbf{G}$ is the gradient matrix. Anisotropic patches (patches with texture) are then selected using an adaptive thresholding mechanism based on eigenvalues of $\mathbf{G^T G}$. Q is then calculated as follows.

$$Q = s_1 \cdot \frac{s_1 - s_2}{s_1 + s_2} \qquad (2)$$

Hence an image with sharper edges shows a large difference between eigenvalues from patches, producing a higher Q than a blurry image.

Each SOTA model discussed above employs a different loss function associated with that work. We wish to incorporate our sharpness based loss Q into these loss functions. Therefore, denoting the existing loss for a particular SOTA model $\mathcal{L}_\varphi(\cdot)$, we deploy a composite loss augmenting this existing loss with our sharpness loss as follows.

$$\mathcal{L} = \mathcal{L}_\varphi(I, \tilde{I}) - \beta \cdot Q(\tilde{I}) \qquad (3)$$

where I and $\tilde{I}$ denote the Ground-Truth (GT) and restored images respectively. Note that $Q(\cdot)$ is a no-reference metric that depends only on the output deblurred image.
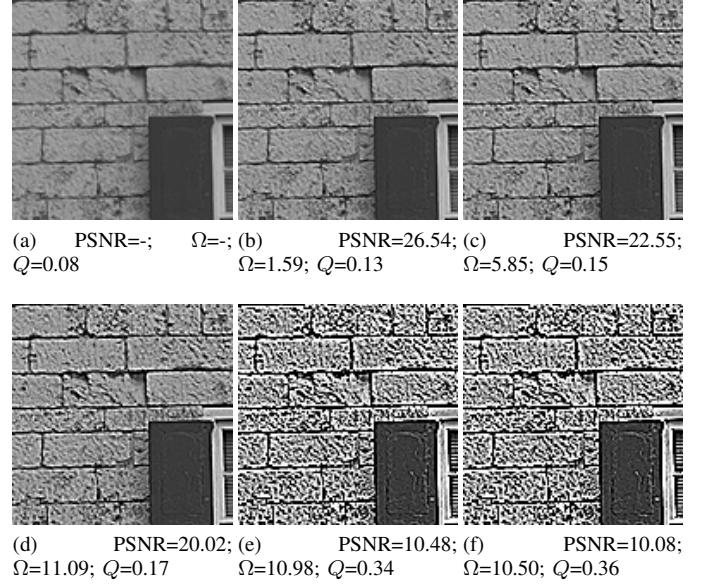
(a) PSNR=-; Ω=-; (b) PSNR=26.54; (c) PSNR=22.55; Q=0.08 Ω=1.59; Q=0.13 Ω=5.85; Q=0.15

(d) PSNR=20.02; (e) PSNR=10.48; (f) PSNR=10.08; Ω=11.09; Q=0.17 Ω=10.98; Q=0.34 Ω=10.50; Q=0.36

Fig. 1: *Sensitivity of Ω to images with different sharpness amounts (γ). Top Row(L–R): Original Image. Image Sharpened with $\gamma = 0.8$ and 1.3. Bottom Row(L–R): Image Sharpened with $\gamma = 2.5$, 11.8 and 13.8. An increase in sharpness leads to an increase in Ω, however with the introduction of ringing Ω decreases.*

$\mathcal{L}_\varphi$ for ARKNet and XY-Deblur was l_1, whereas EHNet employs a combination of l_1 and frequency loss $\mathcal{L}_{freq}$ as follows.

$$\mathcal{L}_{freq} = ||\mathcal{F}(\tilde{I}) - \mathcal{F}(I)|| \qquad (4)$$

where $\mathcal{F}(\cdot)$ is the Fourier transform.

To compare these systems it would be simple just to measure the loss function itself, but each method has a different effective loss function and so we need a fair metric. As stated previously, Q increases with sharpness and also increases with ringing. However, PSNR or SSIM will reduce with ringing (therefore able to highlight that as a detrimental effect) but also reduces with sharpness. Figure 1 shows this phenomenon by applying the Matlab function `imsharpen` with parameter γ to introduce sharpness into an image. We use here a 128×128 crop from an image in the Kodak dataset [18]. We sharpen the original image with five different sharpness amounts $\gamma = 0.8, 1.3, 1.8, 2.5, 11.8, 13.8$. As sharpness ($\gamma$) increases, so does Q but it continues to increase with ringing. PSNR on the other hand reduces with sharpness including when ringing artifacts appear. This effect is made more clear in Figure 2.

What is needed therefore, is a measure that combines the strengths of standard picture quality metrics such as PSNR with Q, so that sharpness and ringing are treated correctly. We therefore propose a metric Ω which is a weighted combination of PSNR and Q. This is discussed next.
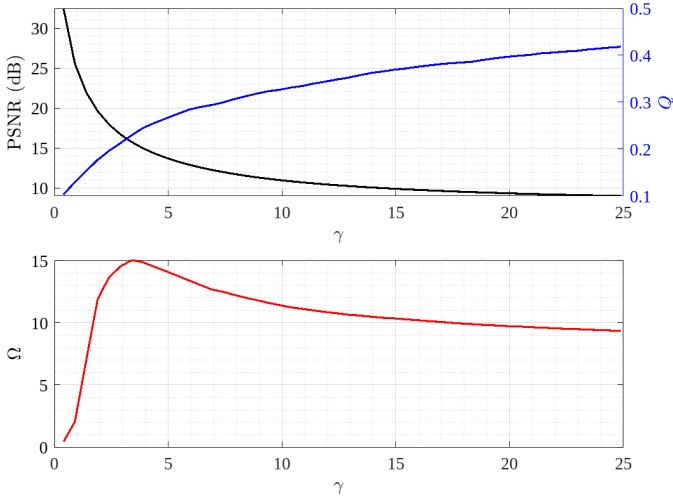
Fig. 2: *Behavior of Ω, Q and PSNR with increasing sharpness amounts (γ).* Top: Comparison of Q (blue) and PSNR (black) w.r.t γ. Bottom: Change in Ω w.r.t γ. As the sharpness increases, Q increases, whereas PSNR decreases. However, Ω increases to a certain point and then begins to decrease as ringing is introduced.

III. PROPERLY ACCOUNTING FOR RINGING

As specified in the previous sections, an image with ringing will have a high Q, whilst having low SSIM or PSNR. We wish to devise a metric which increases with sharpness, but decreases with ringing artifacts in an image. We must note that PSNR behaves approximately correctly, but lacks the knowledge about sharpness, thus leading to a low value even for images that are sharp and have no spurious artifacts. Q, on the other hand, is a reliable indicator of sharpness when PSNR of a restored image is high, but becomes unreliable when PSNR is low. Therefore, we want to combine the two in a non-linear fashion given as follows.

$$\Omega = (1 - \sigma(\alpha)) \cdot P' + \sigma(\alpha) \cdot \tilde{Q} \tag{5}$$

The intuition behind such a metric is based on the idea of giving more weight to PSNR for a patch that has high ringing and Q when a patch has low ringing. The weighting function $\sigma(\cdot)$ is as follows.

$$\sigma(\alpha) = \frac{1}{1 + e^{R(\alpha - \alpha_0)}} \tag{6}$$

The parameters R and α_0 control the amount of *acceptable* sharpness. Based on 13,052 images, we set $R = 5$ and $\alpha_0 = 1.2$. Hence, an increase in sharpness leads to an increase in Ω and when ringing artifacts are produced, Ω decreases. Figure 2 shows a plot of the behavior of Ω PSNR and Q with increasing amounts of γ. An increase in the sharpness amount (γ) initially leads to an increase in Ω. However, as ringing artifacts are produced due to over-sharpening, Ω decreases.

We divide the GT (I) and restored images ($\tilde{I}$) into N non-overlapping patches of size $m \times m$. For each patch from both

images, we measure $Q = Q(I)$ and $\tilde{Q} = Q(\tilde{I})$ respectively to determine the sharpness. To determine the amount of ringing, we measure a deviation ratio α given as follows:

$$\alpha = \frac{|\tilde{Q} - Q|}{Q} \tag{7}$$

A higher α value corresponds to the presence of ringing in the restored patch. We then measure PSNR between the GT (p) and reference patch ($\tilde{p}$), clipping it at 50 dB, to avoid producing a value of ∞ when both patches are identical.

$$P' = \min(PSNR(p, \tilde{p}), 50) \tag{8}$$

Ω for the entire image is the average per patch.

Figure 1 shows the behavior of Ω using a patch size $m = 16$. As opposed to PSNR or Q, where the values are monotonically decreasing and increasing w.r.t the sharpness amounts, Ω increases with increasing sharpness for images with negligible ringing (Top Row, Center and Left). However, over-sharpening the image introduces ringing (Bottom Row), leading to a decrease in Ω values.

IV. EXPERIMENTS

Our primary goal is to evaluate the loss Q comprehensively. Therefore, we employ three different models and three different datasets. We train models with their respective losses $\mathcal{L}_{\varphi}$ for a fixed set of epochs until the best possible combination of weights are found and then fine-tune with our composite loss $\mathcal{L}$. The datasets used and the implementation specifics are mentioned next.

A. Datasets

For training and testing the aforementioned methods, we use three datasets - RealDOF [19], DPDD [20] and ARK [9]. The Real Depth of Field (**RealDOF**) set is a collection of 50 scenes, where each scene comprises of the in-focus image along with its out-of-focus counterpart. Each image is of the size 2320×1536 pixels.

The Dual Pixel Defocus Dataset (**DPDD**) consists of 500 out-of-focus images along with their in-focus images. Each image is of the size 6720×4480 pixels.

The dataset proposed by Aurangabadkar, Ramsook and Kokaram (**ARK**) comprises of three different out-of-focus images separated by the level of blur – low, medium and high, captured by changing the focus ring of the camera by 2 units, w.r.t their in-focus Ground Truth (GT) counterparts. Each collection consists of 305 images of the size 5796×3870 pixels.

B. Implementation Specifics

For training the models, we use 143,104 crops from the DPDD dataset, 38,200 from the RealDOF dataset and 79,000 crops from the ARK dataset. For the purpose of inference, we use 4,996 crops from the RealDOF dataset, 7,898 crops from the DPDD dataset and 188 crops from the ARK dataset, all of which are of the size 128×128. Only the luma (Y) channel was used for deblurring.
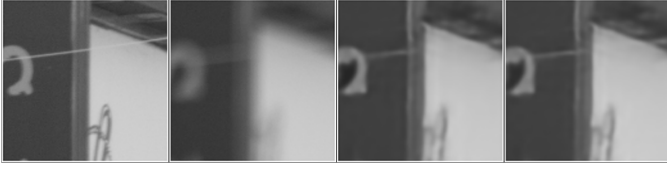
Fig. 3: *Ringing introduced by EHNet on an image from RealDOF dataset*. (L–R) GT image. Blurry image (Ω=1.06). Image restored with $\mathcal{L}_\varphi$ (Ω=1.75). Image restored with $\mathcal{L}$ (Ω=1.72). Using our composite loss leads to the production of slightly over-sharpened edges, thus leading to a slight decrease in Ω value.

The β value in equation 3 was empirically set to 0.1 for training ARKNet, 0.01 and 0.1 for XY-Deblur and EHNet, respectively. Each model was trained with $\mathcal{L}_\varphi$ for 100 epochs and then fine-tuned using $\mathcal{L}$ for another 30 epochs.

V. RESULTS

For evaluating the quality of deblurred images, we use standard metrics such as PSNR, SSIM, alongside those that focus on sharpness – Q, J [21], Ω, as well as, neural metrics such as LPIPS [22]. J is a sharpness metric normalized between [0, 1] that measures the sharpness difference of restored image $\hat{I}$ from the blurry and GT image. A restoration identical with the GT or blurry images will have a J value of 1 or 0 respectively.

As can be seen from Table I, incorporating Q into the training leads to an increase of approximately 15% in sharpness (Q) over using $\mathcal{L}_\varphi$ alone. We perform a paired t-test on the metric values and find the observed difference between means of metrics of images restored using $\mathcal{L}$ as opposed to $\mathcal{L}_\varphi$ are significant at the 5% level ($p < 0.05$), the exception being SSIM, signifying its inability to properly detect sharpness in an image.

Our proposed metric Ω shows an increase of approximately 10% for XY-Deblur and 3% for ARKNet when we use $\mathcal{L}$ as opposed to $\mathcal{L}_\varphi$. averaged over three datasets For EHNet, however, we observe a slight reduction of 1% when we use $\mathcal{L}$, as opposed to $\mathcal{L}_\varphi$, averaged over three datasets. This is possibly because of ringing. An example of this can be seen in Figure 3, where the model produces spurious artifacts around the edges. Therefore, subjective testing must be done to thoroughly understand the correlation between Ω and sharpness in an image.

Figure 4 provides a comparison of images from three different datasets (rows 1 to 3) mentioned in section IV-A deblurred using XY-Deblur with and without Q (columns 4 and 3, respectively). Using Q during training produces noticeably sharper edges.

Figure 5 provides a comparison of the different methods trained with and without Q. Figures denoted by GT and B are the original and blurry images, respectively. All models trained using $\mathcal{L}$ produce much sharper restorations. Restorations generated by the complex transformer model EHNet augmented

Method	Loss	PSNR ↑	SSIM ↑	Q ↑	J ↑	Ω	LPIPS ↓
XY-Deblur [13]		28.091	0.782	1.229	0.642	3.093	0.388
ARKNet [9]	$\mathcal{L}_\varphi$	28.050	0.778	1.132	0.598	3.016	0.388
EHNet [14]		27.992	0.773	1.179	0.605	3.285	0.400
XY-Deblur		27.930	0.780	1.518	0.699	3.409	0.394
ARKNet	$\mathcal{L}$	27.860	0.777	1.283	0.647	3.116	0.393
EHNet		28.065	0.777	1.211	0.622	3.250	0.394

TABLE I: *Comparison of incorporating Q as a loss for 3 deblurring models averaged over 3 different datasets*. We see an increase of about 15% in sharpness (Q) for methods where $\mathcal{L}$ was used (highlighted in red), as compared to those where $\mathcal{L}_\varphi$ was used. In contrast, PSNR, SSIM and LPIPS remain roughly the same, implying that Q as a loss has targeted sharpness effectively.

Fig. 4: *Comparison of restorations from XY-Deblur on 3 datasets*. (L–R) GT image. Blurry image (Ω=3.07,1.47,1.04). Image restored with $\mathcal{L}_\varphi$ (Ω=1.34,1.33,2.24). Image restored with $\mathcal{L}$ (Ω=0.79,2.59,2.13). Models fine-tuned using our composite loss $\mathcal{L}$ produce sharper edges.

with our loss i.e. using $\mathcal{L}$ produce sharper images than other models.

Figure 6 shows the effect of incorporating sharpness with respect to SSIM. Each point shows the average performance over the three datasets used. In each case the use of $\mathcal{L}$ causes an increase in sharpness (i.e. points displaced to the right on the plot) with negligible effect on SSIM. We must highlight that the changes in Q are statistically significant, as opposed to SSIM.

VI. CONCLUSION

We have explored the impact of the loss function Q in SOTA models to produce sharper restorations. We propose a novel full-reference metric Ω that is a weighted combination of PSNR with Q and is sensitive to ringing artifacts. Our approach shows an increase of approximately 15% in sharpness (Q) and up to 10% in Ω values of restored images, as compared to using losses that do not explicitly call out sharpness in images. We acknowledge that there is a scaling issue in
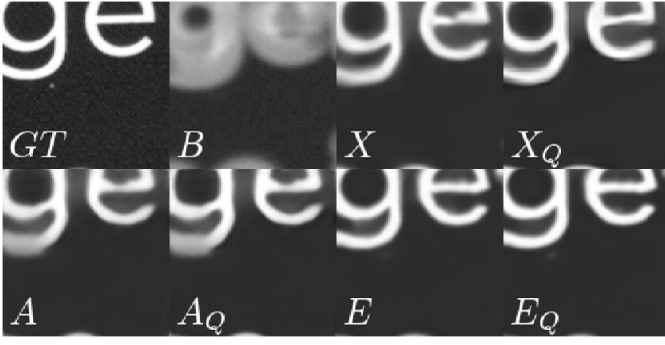
Fig. 5: *Visual comparison of different models without and with Q on an image from the ARK dataset.* XY-Deblur, ARKNet and EHNet trained using $\mathcal{L}_\varphi$ alone are denoted by X, A and E respectively. The methods which were fine-tuned with Q are denoted by X_Q, A_Q and E_Q. All models produce finer textural information when Q is incorporated into training.
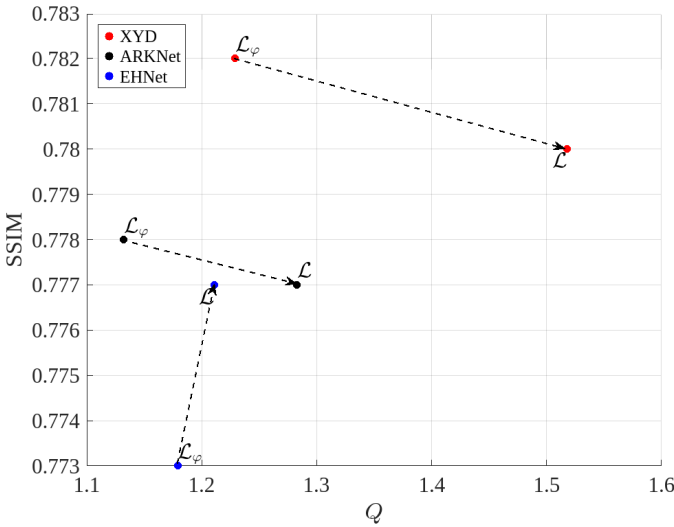

Fig. 6: *Trade-off between SSIM and Q.* Note that while the changes in Q are statistically significant, those in SSIM are not.

Ω. Essentially, Ω mixes decibels (dB) with a dimensionless measure (Q), thus implying that the range of this metric is content specific. While not an issue for optimization, it is an issue for interpretability of the metric. Our future work involves addressing this issue and using Ω as a loss in such networks and performing a large scale subjective study to investigate the correlation between Q, J and Ω with Mean Opinion Scores.

REFERENCES

[1] William Hadley Richardson, "Bayesian-based iterative method of image restoration," *JoSA*, vol. 62, no. 1, pp. 55–59, 1972.

[2] Leon B Lucy, "An iterative technique for the rectification of observed distributions," *Astronomical Journal, Vol. 79, p. 745 (1974)*, vol. 79, pp. 745, 1974.

[3] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas, "Deblurgan: Blind motion deblurring using conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8183–8192.

[4] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 5728–5739.

[5] Filip Šroubek, Tomáš Kerepeckỳ, and Jan Kamenickỳ, "Iterative wiener filtering for deconvolution with ringing artifact suppression," in *2019 27th European Signal Processing Conference (EUSIPCO)*. IEEE, 2019, pp. 1–5.

[6] Santiago López-Tapia, Javier Mateos, Rafael Molina, and Aggelos K Katsaggelos, "Deep robust image restoration using the moore-penrose blur inverse," in *2023 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2023, pp. 775–779.

[7] Jiangxin Dong, Stefan Roth, and Bernt Schiele, "Deep wiener deconvolution: Wiener meets deep learning for image deblurring," *Advances in Neural Information Processing Systems*, vol. 33, pp. 1048–1059, 2020.

[8] Liang Chen, Jiawei Zhang, Zhenhua Li, Yunxuan Wei, Faming Fang, Jimmy Ren, and Jinshan Pan, "Deep richardson–lucy deconvolution for low-light image deblurring," *International Journal of Computer Vision*, vol. 132, no. 2, pp. 428–445, 2024.

[9] Uditangshu Aurangabadkar, Darren Ramsook, and Anil Kokaram, "A sharpness based loss function for removing out-of-focus blur," in *2024 IEEE 26th International Workshop on Multimedia Signal Processing (MMSP)*, 2024, pp. 1–6.

[10] Xiang Zhu and Peyman Milanfar, "Automatic parameter selection for denoising algorithms using a no-reference measure of image content," *IEEE transactions on image processing*, vol. 19, no. 12, pp. 3116–3132, 2010.

[11] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[12] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[13] Seo-Won Ji, Jeongmin Lee, Seung-Wook Kim, Jun-Pyo Hong, Seung-Jin Baek, Seung-Won Jung, and Sung-Jea Ko, "Xydeblur: divide and conquer for single image deblurring," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 17421–17430.

[14] Quoc-Thien Ho, Minh-Thien Duong, Seongsoo Lee, and Min-Cheol Hong, "Ehnet: Efficient hybrid network with dual attention for image deblurring," *Sensors*, vol. 24, no. 20, pp. 6545, 2024.

[15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241.

[16] Sergey Ioffe, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[17] Dan Hendrycks and Kevin Gimpel, "Gaussian error linear units (gelus)," *arXiv preprint arXiv:1606.08415*, 2016.

[18] Rich Franzen, "Kodak lossless true color image suite," *source: http://r0k.us/graphics/kodak*, 1999.

[19] Junyong Lee, Hyeongseok Son, Jaesung Rim, Sunghyun Cho, and Seungyong Lee, "Iterative filter adaptive network for single image defocus deblurring," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 2034–2042.

[20] Abdullah Abuolaim and Michael S Brown, "Defocus deblurring using dual-pixel data," in *European Conference on Computer Vision*. Springer, 2020, pp. 111–126.

[21] Uditangshu Aurangabadkar and Anil Kokaram, "A dictionary based approach for removing out-of-focus blur," in *2024 IEEE International Conference on Image Processing (ICIP)*, 2024, pp. 1494–1499.

[22] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 586–595.