SAGAN: Adversarial Spatial-asymmetric Attention for Noisy Nona-Bayer Reconstruction

S M A Sharif[1]
sma.sharif.cse@ulab.edu.bd

Rizwan Ali Naqvi [2,*]
rizwanali@sejong.ac.kr

Mithun Biswas[1]
mithun.bishwash.cse@ulab.edu.bd

[1] Rigel-IT, Bangladesh

[2] Sejong University, South Korea

* Corresponding author

Abstract

Nona-Bayer colour filter array (CFA) pattern is considered one of the most viable alternatives to traditional Bayer patterns. Despite the substantial advantages, such non-Bayer CFA patterns are susceptible to produce visual artefacts while reconstructing RGB images from noisy sensor data. This study addresses the challenges of learning RGB image reconstruction from noisy Nona-Bayer CFA comprehensively. We propose a novel spatial-asymmetric attention module to jointly learn bi-direction transformation and large-kernel global attention to reduce the visual artefacts. We combine our proposed module with adversarial learning to produce plausible images from Nona-Bayer CFA. The feasibility of the proposed method has been verified and compared with the state-of-the-art image reconstruction method. The experiments reveal that the proposed method can reconstruct RGB images from noisy Nona-Bayer CFA without producing any visually disturbing artefacts. Also, it can outperform the state-of-the-art image reconstruction method in both qualitative and quantitative comparison. Code available: https://github.com/sharif-apu/SAGAN_BMVC21.

1 Introduction

The past decade has experienced a revolutionary takeover in mobile photography. Explicitly, the alleviation in computation photography and innovation on mobile hardware allows the original equipment manufacturers (OEMs) to provide handy experiences to the mobile photographers. However, the perceptual quality of smartphone cameras still incorporates notable drawbacks due to the smaller sensor size and unable to deliver professional-grade image quality in stochastic lighting conditions [13, 28]. Contrarily, enlarging the sensor size of the mobile cameras always remains a strenuous process. Explicitly, the compact nature of mobile devices holds back the OEMs to perceive a substantial push in the sensor size. To address such an inevitable dilemma, many OEMs have leveraged pixel enlarging techniques known as pixel-binning with non-Bayer CFA patterns [5, 17, 20, 28]. Among such non-Bayer CFA patterns, Nona-Bayer has illustrated widespread practicability over its Bayer counterparts.

Typically, a Nona-Bayer CFA pattern comprises three consecutive homogenous pixels in the vertical and horizontal direction, as shown in Fig. 1. Notably, such a CFA pattern allows the sensing hardware to combine homogenous pixels into a bigger pixel to gather up to three times higher light intensity in stochastic lighting conditions. Apart from the improving low-light performance, a Nona-Bayer CFA concedes the practicability of higher resolution sensors in mobile devices and allows to produce high definition contents (i.e., 8K videos) with a natural bokeh effect. Hence, most recent flagship smartphones like Samsung S20 Ultra, Note 20 Ultra, S21 Ultra, Xiaomi Mi 11 Ultra, etc., have utilized such Nona-Bayer CFA on top of the 108-megapixel image sensor to deliver a versatile photography experience to enthusiastic mobile photographers.

Despite numerous advantages, reconstructing an RGB image from a Nona-Bayer CFA is a challenging task. It is worth noting, the distance of homogenous pixels between two recurring Nona-Bayer CFA patterns are three-time larger than a typical Bayer CFA (please see Fig. 1. Subsequently, any complex composition like text with a distinct background that appears between two consecutive patterns can produce visual artefacts. Moreover, the substantial sensor noise along with artefact-prone CFA pattern makes the reconstruction process notably complicated [28].

Figure 1: Comparison between CFA patterns. (a) Nona-Bayer CFA. (b) Bayer CFA.

We found even the state-of-the-art deep image reconstruction methods (i.e., joint demosaicing and denoising (JDD), non-Bayer reconstruction methods) illustrate notable shortcomings in reconstructing RGB images from a noise-contaminated Nona-Bayer CFA pattern. In most instances, the existing methods tend to produce structural distortion and false colour artefacts, as shown in Fig. 2.
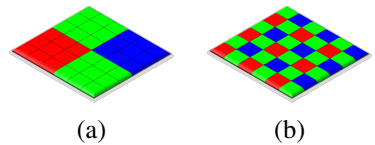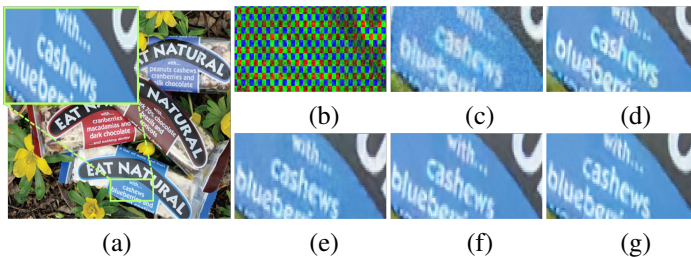

Figure 2: Noisy Nona-Bayer reconstruction with state-of-the-art image reconstruction methods at $\sigma = 30$. (a) Ground-truth RGB Image. (b) Noisy Nona-Bayer Input. (c) Deepjoint [8]. (d) Kokkinos [19]. (e) DPN [17]. (f) BJDD [28]. (g) **SAGAN (Ours)**

To address the deficiencies of existing works, we propose a novel learning-based JDD method for Nona-Bayer reconstruction. To the best concern, this is the first work in the open literature that introduces an end-to-end deep model for reconstructing RGB images from a noisy Nona-Bayer CFA pattern. Our proposed method incorporates a novel spatial-asymmetric attention module to reduce visual artefacts from reconstructed RGB images. Our proposed module learns attention over the vertical and horizontal transformation of a Nona-Bayer CFA and combines it with large-kernel global attentions. Additionally, we proposed an adversarial (a.k.a generative adversarial network (GAN) [9]) guidance with our spatial-asymmetric attention for producing visually plausible images. We denoted our proposed

method as spatial-asymmetric attention GAN (SAGAN) in the rest of the paper. The practicability of the proposed method has been extensively studied with the benchmark dataset and compared with state-of-the-art deep reconstruction methods. The major contributions of the proposed method have been summarized as follows: 1) Proposes and illustrates the practicability of an end-to-end deep network for performing image reconstruction from challenging noisy Nona-Bayer CFA pattern images. 2) Proposes a novel spatial-asymmetric attention module to reduce visual artefacts and combined it with adversarial training to produce plausible images. 3) Compare and outperform existing learning-based reconstruction methods in both qualitative and quantitative comparison.

2 Related Works

The related works of our proposed method have briefly described in this section.

Joint demosacing and denoising. Noise suppression with reconstructing RGB images from CFA patterns have illustrated a significant momentum in recent years. In practice, such JDD manoeuvres can significantly improve the perceptual quality of final reconstructed images. In the early days, JDD was mostly performed with optimization-based strategies [1, 30]. However, in recent time, deep learning has takeover the limelight from its traditional counterparts by learning JDD from a convex set of data samples.

In recent work, [8] trained an end-to-end deep network to achieve state-of-the-art performance in Bayer JDD. Later, [19] combined deep residual denoising with a majorization-minimization technique to perform JDD on the same CFA pattern. Similarly, [21] also proposed a deep method with density-map and green channel guidance to outperform their previous JDD methods. Apart from the Bayer JDD, a recent study [28] proposed a deep network to perform JDD on Quad Bayer CFA. Notably, [28] has illustrated that visual attention with perceptual optimization can significantly accelerate the performance of non-Bayer JDD.

Non-Bayer Reconstruction. Quad Bayer CFA shared similar characteristics as a Nona-Bayer CFA and widely used in recent smartphones cameras. A recent study [17] has proposed a duplex pyramid network for reconstructing the Quad Bayer CFA pattern. Similarly, [16] proposed to learn an under-display camera pipeline exclusively for Quad Bayer CFA.

Attention Mechanism. The concept of attention mechanisms intends to focus on the important features as similar to the human visual system. In the past decade, many works have incorporated novel attention mechanisms for accelerating different vision tasks. In recent work,[12] proposed a squeeze-and-excitation network for achieving channel-wise attention for accelerating image classification. [32] proposed a residual attention network for having 3D attention over intermediate features. Later, [33] proposed a lightweight convolutional block attention module to accelerate the learning process of feed-forward networks. Similarly, [36] proposed a convolutional attention mechanism to learn dynamic feature attentions. It is worth noting, none of the existing methods has exploited the visual attention on asymmetric manner. In this study, we depicted that such spatial-asymmetric attention can significantly improve the performance of low-level vision task, explicitly the Nona-Bayer reconstruction.

3 Method

This section describes the proposed method as well as our SAGAN architecture.
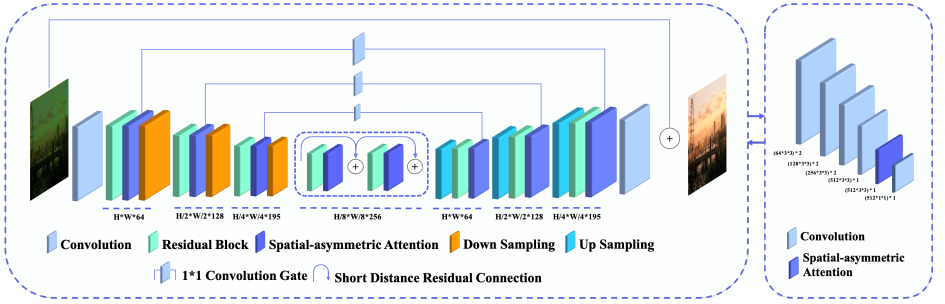
Figure 3: Overview of the proposed method. Our SAGAN comprises a novel spatial-asymmetric module and guided by adversarial training.

3.1 Network Design

Fig. 3 illustrates the overview of the proposed SAGAN architecture. The proposed method has been designed as a deep network incorporating novel spatial-asymmetric attention module along with adversarial training. Our generative method (S) learns to translate a Nona-Bayer mosaic pattern ($\mathbf{I_N}$) as $S : \mathbf{I_N} \to \mathbf{I_R}$. Here, ($\mathbf{I_R}$) present the reconstructed RGB image as $\mathbf{I_R} \in [0, 1]^{H \times W \times 3}$. H and W represent the height and width of the input mosaic patterns and output RGB images.

3.1.1 Spatial-asymmetric Attention Module

The proposed spatial-asymmetric attention module intends to reduce visual artefacts from the reconstructed RGB images. As Fig. 4 depicts, we leverage asymmetric convolution operations [7, 22] to extract a sequence of vertical and horizontal feature maps. Later, we utilized spatial attention [33] over the extracted horizontal and vertical features to perceive a pixel-level feature suppression/expansion as follows:

$$\mathbf{F_V} = \tau(\mathbf{C_S}([\mathbf{Z_A}(\mathbf{A_V}(\mathbf{X})); \mathbf{Z_M}(\mathbf{A_V}(\mathbf{X}))])) \quad (1)$$

$$\mathbf{F_H} = \tau(\mathbf{C_S}([\mathbf{Z_A}(\mathbf{A_H}(\mathbf{X})); \mathbf{Z_M}(\mathbf{A_H}(\mathbf{X}))])) \quad (2)$$

Here, $\mathbf{A}(\cdot)$, $\mathbf{C}(\cdot)$, and τ represent the asymmetric convolution operation, square convolution and sigmoid activation, respectively. Additionally, $\mathbf{Z_A}$ and $\mathbf{Z_M}$ present the average pooling and max pooling, which generates two 2D feature maps as $\mathbf{X_A} \in \mathbb{R}^{1 \times H \times W}$, $\mathbf{X_M} \in \mathbb{R}^{1 \times H \times W}$ and concatenated into a single 2D feature map.

An aggregated bi-directional attention over a given feature map has obtained as:
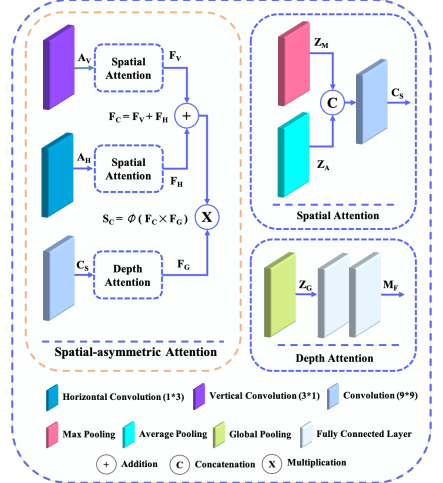

Figure 4: Overview of proposed spatial-asymmetric attention module. Our proposed block aims to substantially reduce visual artefacts, which typically arises due to Nona-Bayer CFA.

$$\mathbf{F_C} = \mathbf{F_V} + \mathbf{F_H} \quad (3)$$

Apart from the asymmetric attention, we also appropriated a squeeze-extractor descriptor [12] to learn depth-wise attention on a globally extracted feature map as follows:

$$\mathbf{F_G} = \mathbf{M_F}(\mathbf{Z_G}(\mathbf{C_S}(\mathbf{X}))) \tag{4}$$

Here, $\mathbf{M_F}$ and $\mathbf{Z_G}$ present consecutive fully connected layers and global pooling operations.

Notably, our spatial-asymmetric module incorporated large-kernel convolution (i.e., 9×9) operations. Here, these square convolution operations are intended to learn global image correction by exploiting larger reception fields [26]. We obtained the final output of the spatial-asymmetric attention module as follows:

$$\mathbf{S_A} = \phi(\mathbf{F_C} \times \mathbf{F_G}) \tag{5}$$

Here, ϕ denotes the leaky ReLU activation function.

3.1.2 SAGAN Generator

The proposed SAGAN generator has been designed as well-known U-Net architecture with convolutional features gates [1, 14, 29]. Our SAGAN generator utilized multiple feature depth levels (i.e., 64, 128, 192, and 256) for feature encoding-decoding. Each feature level of the proposed generator comprises a residual block and a spatial-asymmetric attention block. Here, the residual blocks are intended to accelerate denoising performance, while spatial-asymmetric attention blocks are intended to reduce visual artefacts. We obtained downsampling and upsampling using square convolutional (with stride = 2) and pixel-shuffles upsampling operations. Apart from that, our SAGAN generator also comprises two consecutive middle blocks with a short distance residual connection. Additionally, it connects multiple layers of encoder-decoder with 1×1 convolutional feature gates. Here, the short distance residual connection and the convolutional gates help our SAGAN converge with informative features.

3.1.3 SAGAN Discriminator

The architecture of our SAGAN discriminator has been designed as a stacked convolutional neural network (CNN). The first seven layers of the proposed discriminator are 3×3 convolution layers, which we normalized with batch normalization and activated with a swish activation. The convolutional layers are followed by a spatial-asymmetric attention module and 1×1 convolutional output layer with sigmoid activation. Here, every $(2n-1)^{th}$ layer of our SAGAN discriminator reduces its spatial dimension by incorporating a stride = 2.

3.2 Optimization

The proposed SAGAN has been optimized with a multi-term objective function. For a given training set $\{\mathbf{I_N}^t, \mathbf{I_G}^t\}_{t=1}^P$ consisting of N image pairs, the training process aims to minimize the objective function describes as follows:

$$W^* = \arg\min_W \frac{1}{P} \sum_{t=1}^P \mathcal{L}_S(\mathrm{S}(\mathbf{I_N}^t), \mathbf{I_G}^t) \tag{6}$$

Here, $\mathcal{L}_S$ represent the proposed SAGAN loss, and W presents the parameterised weights of the SAGAN generator.

Reconstruction Loss. Our proposed SAGAN loss comprises an L1-norm as standard reconstruction loss as follows:

$$\mathcal{L}_R = \| \mathbf{I_G} - \mathbf{I_R} \|_1 \tag{7}$$

Here, $\mathbf{I_R}$ presents the reconstructed RGB output of S and $\mathbf{I_G}$ presents the ground-truth RGB image.

Perceptual Colour Loss (PCL). Apart from the reconstruction loss, we leverage a perceptual colour loss [28] to perceive a consistent colour accuracy across different colour spaces. Here, the perceptual colour loss is obtained as follows:

$$\mathcal{L}_C = \Delta E\left(\mathbf{I_G}, \mathbf{I_R}\right) \tag{8}$$

Here, ΔE represents the CIEDE2000 colour difference [23], which has calculated by comparing reconstructed image ($\mathbf{I_R}$) and the ground-truth image ($\mathbf{I_G}$).

Adversarial Loss. The proposed SAGAN leverages adversarial training to produce natural colour while retaining texture information. Therefore, discriminator maximise a loss $\mathbb{E}_{X,Y}$ as: $\mathbb{E}_{X,Y}\left[\log D(X,Y)\right]$. Contrarily, our SAGAN generator aims to minimize the generator loss as follows:

$$\mathcal{L}_G = -\sum_t \log D(\mathbf{I_R}, \mathbf{I_G}) \tag{9}$$

SAGAN loss. We perceived SAGAN loss by adding individual losses as follows:

$$\mathcal{L}_S = \mathcal{L}_R + \mathcal{L}_C + \lambda_G.\mathcal{L}_G \tag{10}$$

Here, λ_G presents the adversarial regulators, which has been tuned arbitrarily as $\lambda_G = 1e-4$ for stabilizing our adversarial training.

4 Experiments and Results

The practicability of the proposed SAGAN has verified with dense experiments. This section details the experiments and discusses the results.

4.1 Experiment Setup

We extracted a total of 741,968 non-overlapping 128×128 image patches from DIV2K [4], Flickr2K [51], HDR+ [10] datasets to learn noisy image reconstruction. We presumed that JDD performed after non-linear mapping and was independent of additional ISP tasks similar to previous works [8, 19, 28]. Subsequently, we sampled sRGB images according to the CFA pattern and contaminated the sampled images with random noise as $(\mathcal{N}\mathbf{I_N}|\sigma)$. Here, σ represents the standard deviation of the noise distribution, which is generated by $\mathcal{N}(\cdot)$ over a sampled input $\mathbf{I_N}$. We evaluated our method in both sRGB and Linear RGB colour spaces. To evaluate our method in sRGB space, we combined multiple sRGB benchmark dataset including BSD100 [25], McM [34], Urban100 [6], Kodak [55], WED [24] into an unified dataset. Apart from that, we included linear RGB images from MSR demosaicing dataset [15], which has been denoted as Linear RGB in later sections.

Apart from that, we studied our proposed method with real-world noisy data samples. Therefore, we trained our SAGAN with real-world noisy sampled images from Smartphone Image Denoising Dataset (SIDD) [2, 3]. Also, we developed an android application to capture noisy images with real Nona-Bayer hardware. Later, We incorporated a Samsung Galaxy Note 20 Ultra hardware (i.e., 108MP Nona-Bayer sensor) to collect Nona-Bayer captures for evaluating our SAGAN on real-world scenarios.

We implemented our SAGAN in the PyTorch [27] framework. The generator and discriminator of the proposed network have optimized with an Adam optimizer [18], where hyperparameters are set as $\beta_1 = 0.9$, $\beta_2 = 0.99$, and learning rate $= 5e-4$. We trained our generator and discriminator jointly for 200,000 steps with a constant batch size of 16. It took around 120 hours to converge with given data samples. We employed a graphical Nvidia Geforce GTX 1060 (6GB) graphical processing unit (GPU) to conduct our experiments.

4.2 Comparison with State-of-the-art Methods

The performance of SAGAN has been studied with different CFA Patterns (i.e., Nona-Bayer and Bayer CFA patterns) and compared with state-of-the-art reconstruction methods. We included deep Bayer joint demosaicking and denoising methods like Deepjoint [8], Kokkinos [19], Non-Bayer JDD method like BJDD [28], and Quad Bayer reconstruction method like DPN [17] for the comparison. For fair comparisons, we trained and tested the reconstruction methods with identical datasets. The performance of the compared methods has cross-validated with three different noise levels, where the standard deviation of noise distribution was set as $\sigma = (10, 20, 30)$. Later, we summarized the performance of deep models with standard evaluation metrics like PSNR, SSIM, and DeltaE2000.

4.2.1 Noisy Nona-Bayer Reconstruction

We performed an extensive evaluation on challenging noisy Nona-Bayer reconstruction by incorporating quantitative and qualitative comparisons.

Model	σ	sRGB Images			Linear RGB Images		
		PSNR ↑	SSIM ↑	DeltaE ↓	PSNR ↑	SSIM ↑	DeltaE ↓
Deepjoint [8]		31.63	0.9026	3.11	39.00	0.9464	1.64
Kokkinos [19]		33.08	0.9321	2.75	39.26	0.9539	1.74
DPN [17]	10	33.49	0.9390	2.62	39.84	0.9702	1.50
BJDD [28]		34.02	0.9440	2.56	41.40	0.9751	1.64
SAGAN (Ours)		**34.99**	**0.9503**	**2.18**	**43.17**	**0.9788**	**1.11**
Deepjoint [8]		30.22	0.8495	3.44	36.14	0.8946	1.94
Kokkinos [19]		31.88	0.9080	2.97	38.18	0.9411	1.76
DPN [17]	20	32.13	0.9152	2.92	38.39	0.9572	1.68
BJDD [28]		32.58	0.9212	2.86	39.71	0.9619	1.86
SAGAN (Ours)		**33.33**	**0.9290**	**2.49**	**41.26**	**0.9675**	**1.32**
Deepjoint [8]		28.81	0.7913	3.90	34.05	0.8407	2.29
Kokkinos [19]		30.81	0.8830	3.23	36.84	0.9203	1.95
DPN [17]	30	30.96	0.8904	3.21	36.99	0.9411	1.93
BJDD [28]		31.42	0.8990	3.14	38.27	0.9466	2.03
SAGAN (Ours)		**32.10**	**0.9084**	**2.78**	**39.59**	**0.9525**	**1.57**

Table 1: Quantitative Comparison for Noisy Nona-Bayer reconstruction.

Quantitative Comparison. Table. 1 demonstrates the performance of the different learning-based methods for Nona-Bayer reconstruction. The proposed SAGAN outperforms the state-of-the-art methods in both sRGB and linear RGB colour spaces. Also, the performance of our SAGAN is consistent among different noise levels. Apart from suppressing noise, our SAGAN can produce more colour-accurate RGB images with dense structural information.
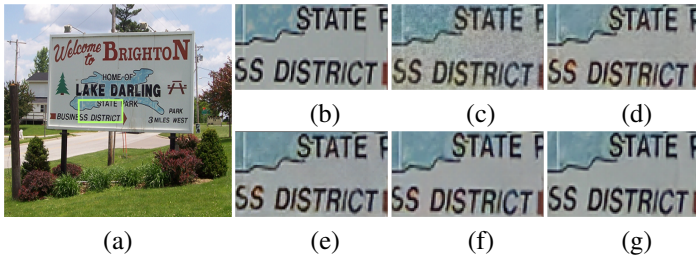
Figure 5: Qualitative Comparison for Noisy Nona-Bayer reconstruction at $\sigma = 30$. (a) Ground-truth RGB Image (full). (b) Ground-truth RGB Image (crop). (c) Deepjoint [8]. (d) Kokkinos [19]. (e) DPN [17]. (f) BJDD [28]. (g) **SAGAN (Ours)**

Qualitative Comparison. We compared the reconstruction methods qualitatively by visualizing their performance. Fig. 5 illustrates the visual comparison between the existing method and our SAGAN. It be seen that the proposed SAGAN can reconstruct more natural-looking plausible images with maximum noise suppression. Our novel SAGAN can substantially reduce the visual artefacts that occur due to the non-Bayer CFA pattern. Notably, our proposed adversarial spatial-asymmetric attention strategies allow us to learn perceptually admissible images as similar to the reference images.

4.2.2 Noisy Bayer Reconstruction

Typically, Nona-Bayer sensors are capable of forming a Bayer pattern by leveraging the pixel-binning technique. Thus, we have studied our method on noisy Bayer reconstruction to confirm its practicability in real-world scenarios.

Quantitative Comparison. Table. 2 illustrates the comparison between state-of-the-art methods for noisy Bayer reconstitution on different noise levels. Notably, our SAGAN outperforms the existing methods for noisy Bayer reconstruction as well.

Model	σ	sRGB Images			Linear RGB Images		
		PSNR ↑	SSIM ↑	DeltaE ↓	PSNR ↑	SSIM ↑	DeltaE ↓
Deepjoint [8]		33.04	0.9262	2.80	37.89	0.9496	1.81
Kokkinos [19]		34.24	0.9412	2.664	38.45	0.9550	1.79
DPN [17]	10	36.51	0.9593	1.88	42.80	0.9790	1.21
BJDD [28]		36.68	0.9561	1.91	43.70	0.9760	1.12
SAGAN (Ours)		**37.07**	**0.9616**	**1.76**	**43.66**	**0.9756**	**1.16**
Deepjoint [8]		31.08	0.8594	3.54	35.30	0.8839	2.53
Kokkinos [19]		32.37	0.9052	3.16	36.77	0.9251	2.19
DPN [17]	20	34.22	0.9316	2.33	40.32	0.9642	1.53
BJDD [28]		34.43	0.9323	2.31	41.10	0.9596	1.42
SAGAN (Ours)		**34.56**	**0.9375**	**2.20**	**42.22**	**0.9715**	**1.22**
Deepjoint [8]		28.99	0.7789	4.49	32.89	0.7997	3.38
Kokkinos [19]		30.27	0.8562	3.85	34.18	0.865	2.80
DPN [17]	30	32.32	0.8983	2.82	38.06	0.9405	1.92
BJDD [28]		32.75	0.9074	2.63	38.21	0.9298	1.75
SAGAN (Ours)		**33.28**	**0.9212**	**2.41**	**40.78**	**0.9613**	**1.32**

Table 2: Quantitative comparison for noisy Bayer reconstruction.

Qualitative Comparison. Fig. 6 visually confirms that the proposed SAGAN can produce plausible images while reconstructing images from noisy Bayer inputs. Also, it can suppress maximum noise by retaining details comparing to its counterparts.

4.3 Nona-Bayer Reconstruction with Real-world Denoising

Real-world sensors are typically surrounded by multiple noise sources and can go beyond a synthesized noise. Hence, we studied our method on real-world noisy images also.
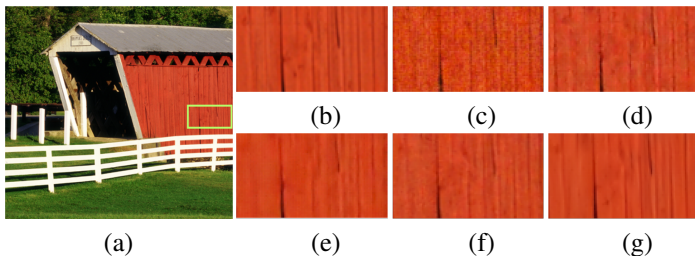
Figure 6: Qualitative Comparison for Noisy Bayer reconstruction at $\sigma = 30$. (a) Ground-truth RGB Image (full). (b) Ground-truth RGB Image (crop). (c) Deepjoint [8]. (d) Kokkinos [19]. (e) DPN [17]. (f) BJDD [28]. (g) **SAGAN (Ours).**

Visual Results. Fig. 7 depicts the performance of proposed SAGAN on real-world denoising with Nona-Bayer reconstruction. It can be seen that our method can handle real-world noise and can produce visually plausible images without any visual artefacts.
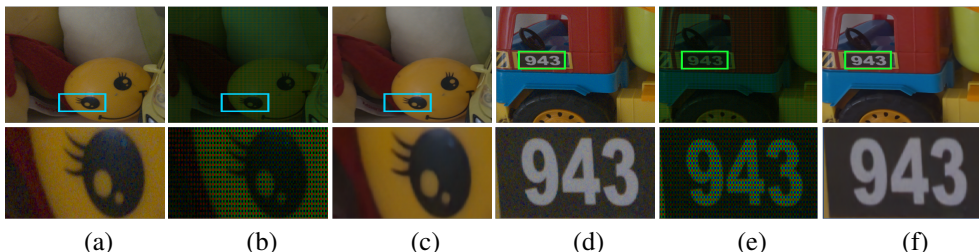

Figure 7: Nona-Bayer reconstruction with real-world noise suppression. (a) & (d) Noisy image (RAW). (b) & (e) Noisy Nona-Bayer (input). (c) & (f) **Reconstructed with SAGAN**.

User Study. We performed a blindfold user study to verify the practicability of our proposed method. Therefore, we showed image pairs comprising our reconstructed and noisy (RAW) image to the random users and asked them to select their preferred image from each image pair. Later, we calculated the mean opinion score (MOS) to summarized user preferences. Our proposed method can substantially score higher MOS, as shown in Table. 3.

Method	RAW (Visualised)	SAGAN (Ours)
MOS ↑	0.13	**0.87**

Table 3: User study for real-world noisy Nona-Bayer reconstruction.

Model	Base	SA	PCL	GAN	sRGB Images			Linear RGB Images		
					PSNR ↑	SSIM ↑	DeltaE ↓	PSNR ↑	SSIM ↑	DeltaE ↓
BaseNet	✓	✗	✗	✗	22.26	0.5800	10.63	24.65	0.6112	9.42
BaseGAN	✓	✗	✗	✓	27.45	0.7730	6.09	26.52	0.6684	8.48
SANWP	✓	✓	✗	✗	31.99	0.9125	2.89	35.25	0.8191	1.87
SAN	✓	✓	✓	✗	32.75	0.9240	2.58	36.59	0.9588	1.42
SAGAN	✓	✓	✓	✓	**33.47**	**0.9292**	**2.48**	**41.34**	**0.9663**	**1.33**

Table 4: Quantitative evaluation on proposed SAGAN. It can be seen that our proposed components have a meaningful impact on the noisy Nona-Bayer reconstruction.

4.4 Ablation Study

The practicability of our proposed spatial-asymmetric attention with adversarial training has been verified with sophisticated experiments. We removed our proposed components like spatial-asymmetric attention block, PCL, and SAGAN discriminator from the network architectures. Later, we incorporated each of them individually and summarized the practicability

of these components with quantitative and qualitative evaluation.

Quantitative Evaluation. Table. 4 illustrates the practicability of our proposed spatial-asymmetric attention and adversarial guidance in both colour spaces. For simplicity, we depicted the mean performance of each model on different noise levels (i.e., $\sigma = 10, 20, 30$). It is visible that our proposed components play a substantial role in Nona-Bayer reconstruction.

Qualitative Evaluation. Fig. 8 illustrates the visual comparison between SAGAN variants. Also, it confirms that our proposed spatial-asymmetric attention can substantially reduce the visual artefacts, while our adversarial training helps us recover texture with natural colours. Additionally, PCL has helped us to maintain a colour consistency across different colour spaces.

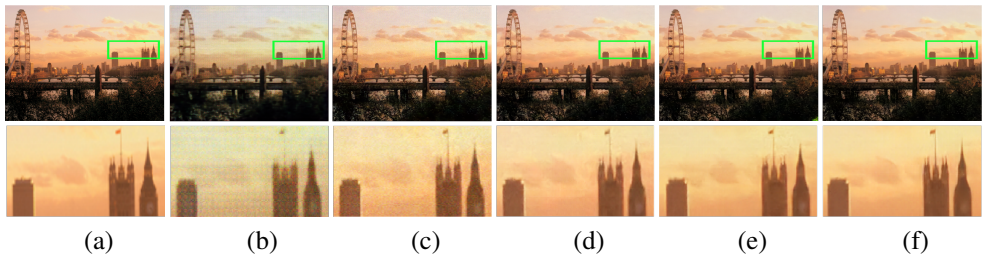| (a) | (b) | (c) | (d) | (e) | (f) |

Figure 8: Qualitative evaluation of our proposed SAGAN at $\sigma = 30$. Our proposed component can substantially reduce visual artefacts and produce perceptually plausible images (best viewed in colour and zoomed). (a) Ground-truth RGB Image. (b) BaseNet. (c) BaseGAN. (d) SANWP. (e) SAN. (f) SAGAN.

4.5 Discussion

Proposed SAGAN comprises a total of 29,448,766 trainable parameters. As being a fully convolutional network architecture, the proposed network can be inference with different resolution images. In our setup, we found it takes less than a second in reconstructing a $1024 \times 1024 \times 3$. It is noteworthy that our proposed SAGAN does not incorporate any pre/post-operations. Therefore, the inference time with similar hardware is expected to remain constant. Despite showing an admissible inference time on a desktop environment, we failed to study SAGAN on a real-world mobile setup due to hardware limitations. Nevertheless, the proposed SAGAN reveals a promising aspect of noisy Nona-Bayer reconstruction through deep learning. Please see the supplementary material for implementation details and more results of our proposed SAGAN.

5 Conclusion

We proposed a novel end-to-end deep method for reconstructing RGB images from a challenging Nona-Bayer CFA. Notably, our proposed method incorporates a novel spatial-asymmetric attention mechanism with adversarial training. We studied the feasibility of our SAGAN on different colour spaces and diverse data samples. Experimental results illustrate that our SAGAN can outperform the existing methods in both quantitative and qualitative comparisons. However, due to hardware constraints, we failed to evaluate the performance of our SAGAN by deploying it into real mobile hardware. It has planned to study the practicability of a deep method like SAGAN for reconstructing images from Nona-Bayer along with Quad-Bayer CFA on real mobile hardware in the foreseeable future.

References

[1] SM A Sharif, Rizwan Ali Naqvi, Mithun Biswas, and Sungjun Kim. A two-stage deep network for high dynamic range image reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 550–559, 2021.

[2] Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. A high-quality denoising dataset for smartphone cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1692–1700, 2018.

[3] Abdelrahman Abdelhamed, Radu Timofte, and Michael S Brown. Ntire 2019 challenge on real image denoising: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.

[4] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *IEEE Conf. Comput. Vis. Pattern Recog. Worksh.*, pages 126–135, 2017.

[5] Sandor L Barna, Scott P Campbell, and Gennady Agranov. Method and apparatus for improving low-light performance for small pixel image sensors, June 11 2013. US Patent 8,462,220.

[6] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3213–3223, 2016.

[7] Xiaohan Ding, Yuchen Guo, Guiguang Ding, and Jungong Han. Acnet: Strengthening the kernel skeletons for powerful cnn via asymmetric convolution blocks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1911–1920, 2019.

[8] Michaël Gharbi, Gaurav Chaurasia, Sylvain Paris, and Frédo Durand. Deep joint demosaicking and denoising. *ACM Trans. Graph.*, 35(6):1–12, 2016.

[9] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *arXiv preprint arXiv:1406.2661*, 2014.

[10] Samuel W Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T Barron, Florian Kainz, Jiawen Chen, and Marc Levoy. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Transactions on Graphics (TOG)*, 35(6):1–12, 2016.

[11] Keigo Hirakawa and Thomas W Parks. Joint demosaicing and denoising. *IEEE Trans. Image Process.*, 15(8):2146–2157, 2006.

[12] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.

[13] Andrey Ignatov, Luc Van Gool, and Radu Timofte. Replacing mobile camera isp with a single deep learning model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 536–537, 2020.

[14] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *IEEE Transactions on Image Processing*, 30:2340–2349, 2021.

[15] Daniel Khashabi, Sebastian Nowozin, Jeremy Jancsary, and Andrew W Fitzgibbon. Joint demosaicing and denoising via learned nonparametric random fields. *IEEE Trans. Image Process.*, 23(12):4968–4981, 2014.

[16] Irina Kim and Eundoo Heo. Under display camera quad bayer raw image restoration using deep learning. *Imaging*, 67:2, 2021.

[17] Irina Kim, Seongwook Song, Soonkeun Chang, Sukhwan Lim, and Kai Guo. Deep image demosaicing for submicron image sensors. *J. Imaging Sci. Techn.*, 63(6):60410–1, 2019.

[18] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[19] Filippos Kokkinos and Stamatios Lefkimmiatis. Deep image demosaicking using a cascade of convolutional residual denoising networks. In *Eur. Conf. Comput. Vis.*, pages 303–319, 2018.

[20] Assaf Lahav and David Cohen. Color pattern and pixel level binning for aps image sensor using 2×2 photodiode sharing scheme, August 10 2010. US Patent 7,773,138.

[21] Lin Liu, Xu Jia, Jianzhuang Liu, and Qi Tian. Joint demosaicing and denoising with self guidance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2240–2249, 2020.

[22] Shao-Yuan Lo, Hsueh-Ming Hang, Sheng-Wei Chan, and Jing-Jhih Lin. Efficient dense modules of asymmetric convolution for real-time semantic segmentation. In *Proceedings of the ACM Multimedia Asia*, pages 1–6. ACM, 2019.

[23] M Ronnier Luo, Guihua Cui, and Bryan Rigg. The development of the cie 2000 colour-difference formula: Ciede2000. *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur*, 26(5):340–350, 2001.

[24] Kede Ma, Zhengfang Duanmu, Qingbo Wu, Zhou Wang, Hongwei Yong, Hongliang Li, and Lei Zhang. Waterloo exploration database: New challenges for image quality assessment models. *IEEE Trans. Image Process.*, 26(2):1004–1016, 2016.

[25] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Int. Conf. Comput. Vis.*, volume 2, pages 416–423, July 2001.

[26] Chao Peng, Xiangyu Zhang, Gang Yu, Guiming Luo, and Jian Sun. Large kernel matters–improve semantic segmentation by global convolutional network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4353–4361, 2017.

[27] Pytorch. PyTorch Framework code. https://pytorch.org/, 2016. Accessed: 2020-11-14.

[28] S M A Sharif, Rizwan Ali Naqvi, and Mithun Biswas. Beyond joint demosaicking and denoising: An image processing pipeline for a pixel-bin image sensor. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 233–242, 2021.

[29] SMA Sharif, Rizwan Ali Naqvi, and Mithun Biswas. Learning medical image denoising with deep dynamic residual attention network. *Mathematics*, 8(12):2192, 2020.

[30] Hanlin Tan, Xiangrong Zeng, Shiming Lai, Yu Liu, and Maojun Zhang. Joint demosaicing and denoising of noisy bayer images with admm. In *IEEE Int. Conf. Image Process.*, pages 2951–2955. IEEE, 2017.

[31] Radu Timofte, Eirikur Agustsson, Luc Van Gool, Ming-Hsuan Yang, and Lei Zhang. Ntire 2017 challenge on single image super-resolution: Methods and results. In *IEEE Conf. Comput. Vis. Pattern Recog. Worksh.*, pages 114–125, 2017.

[32] Fei Wang, Mengqing Jiang, Chen Qian, Shuo Yang, Cheng Li, Honggang Zhang, Xiaogang Wang, and Xiaoou Tang. Residual attention network for image classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3156–3164, 2017.

[33] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018.

[34] Wei Wu, Zheng Liu, Wail Gueaieb, and Xiaohai He. Single-image super-resolution based on markov random field and contourlet transform. *J. Electron. Imaging*, 20(2): 023005, 2011.

[35] Akira Yanagawa, Alexander C Loui, Jiebo Luo, Shih-Fu Chang, Dan Ellis, Wan Jiang, Lyndon Kennedy, and Keansub Lee. Kodak consumer video benchmark data set: concept definition and annotation. *Columbia University ADVENT Technical Report*, pages 246–2008, 2008.

[36] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Freeform image inpainting with gated convolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4471–4480, 2019.