

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

S M A Sharif¹

sma.sharif.cse@ulab.edu.bd

Rizwan Ali Naqvi²

rizwanali@sejong.ac.kr

Mithun Biswas¹

mithun.bishwash.cse@ulab.edu.bd

¹ Rigel-IT, Bangladesh

² Sejong University, South Korea

Supplementary

1 Overview

This supplementary explains architectural details of our proposed SAGAN and illustrates subjective and objective results of additional experiments. The overall supplement has structured as follows:

- Section 2 describes implementation details.
- Section 3 illustrates quantitative results on benchmark datasets.
- Section 4 compares our SAGAN performance with Bayer-RGB images.
- Section 5 demonstrates the performance of SAGAN on real-world noisy inputs.
- Section 6 concludes our finding.

2 Network Details

Our SAGAN has designed as a fully-convolutions network. We denoted kernel size as K , stride as S , padding as P , batch normalization as BN, activation function as ACT, skip connection as SC, gated connection as GC, and our proposed Spatial-asymmetric attention module as SAM throughout this section.

2.1 Generator

Table. 1 illustrates the details of our proposed SAGAN generator.

Level	Block	Tensor Dimension	Configuration
Input	Conv	In: $H \times W \times 3$ Out: $H \times W \times 64$	$K = 3 \times 3, S = 1, P = 1,$ ACT = Leaky ReLU
Level 1 (Encoder)	Residual Block 1	In: $H \times W \times 64$ Out: $H \times W \times 64$	$K = 3 \times 3, S = 1, P = 1$
	SAM 1	In: $H \times W \times 64$ Out: $H \times W \times 64$	$K = 3 \times 3, S = 1, P = 1,$ ACT = Leaky ReLU
	Gate 1	In: $H \times W \times 64$ Out: $H \times W \times 64$	$K = 1 \times 1, S = 1, P = 1$
	Down Sampling 1	In: $H \times W \times 64$ Out: $H/2 \times W/2 \times 128$	$K = 3 \times 3, S = 2, P = 1,$ ACT = Leaky ReLU
Level 2 (Encoder)	Residual Block 2	In: $H/2 \times W/2 \times 128$ Out: $H/2 \times W/2 \times 128$	$K = 3 \times 3, S = 1, P = 1$
	SAM 2	In: $H/2 \times W/2 \times 128$ Out: $H/2 \times W/2 \times 128$	$K = 3 \times 3, S = 1, P = 1,$ ACT = Leaky ReLU
	Gate 2	In: $H/2 \times W/2 \times 128$ Out: $H/2 \times W/2 \times 128$	$K = 1 \times 1, S = 1, P = 1$
	Down Sampling 2	In: $H/2 \times W/2 \times 128$ Out: $H/4 \times W/4 \times 192$	$K = 3 \times 3, S = 2, P = 1,$ ACT = Leaky ReLU
Level 3 (Encoder)	Residual Block 3	In: $H/4 \times W/4 \times 192$ Out: $H/4 \times W/4 \times 192$	$K = 3 \times 3, S = 1, P = 1$
	SAM 3	In: $H/4 \times W/4 \times 192$ Out: $H/4 \times W/4 \times 192$	$K = 3 \times 3, S = 1, P = 1,$ ACT = Leaky ReLU
	Gate 3	In: $H/4 \times W/4 \times 192$ Out: $H/4 \times W/4 \times 192$	$K = 1 \times 1, S = 1, P = 1$
	Down Sampling 3	In: $H/4 \times W/4 \times 192$ Out: $H/8 \times W/8 \times 256$	$K = 3 \times 3, S = 2, P = 1,$ ACT = Leaky ReLU
Level 4 (Middle)	Residual Block 4	In: $H/8 \times W/8 \times 256$ Out: $H/8 \times W/8 \times 256$	$K = 3 \times 3, S = 1, P = 1$
	SAM 4	In: $H/8 \times W/8 \times 256$ Out: $H/8 \times W/8 \times 256$	$K = 3 \times 3, S = 1, P = 1,$ ACT = Leaky ReLU, SC = D. Sam. 3
Level 5 (Middle)	Residual Block 5	In: $H/8 \times W/8 \times 256$ Out: $H/8 \times W/8 \times 256$	$K = 3 \times 3, S = 1, P = 1$
	SAM 5	In: $H/8 \times W/8 \times 256$ Out: $H/8 \times W/8 \times 256$	$K = 3 \times 3, S = 1, P = 1,$ ACT = Leaky ReLU, SC = : SAM 4
Level 6 (Decoder)	Upsampling 1	In: $H/8 \times W/8 \times 256$ Out: $H/4 \times W/4 \times 192$	Scailing factor = 2
	Residual Block 6	In: $H/4 \times W/4 \times 192$ Out: $H/4 \times W/4 \times 192$	$K = 3 \times 3, S = 1, P = 1,$ ACT = Leaky ReLU
	SAM 6	In: $H/4 \times W/4 \times 192$ Out: $H/4 \times W/4 \times 192$	$K = 3 \times 3, S = 2, P = 1,$ ACT = Leaky ReLU, GC = : Gate 3
Level 7 (Decoder)	Upsampling 2	In: $H/4 \times W/4 \times 192$ Out: $H/2 \times W/2 \times 128$	Scailing factor = 2
	Residual Block 7	In: $H/2 \times W/2 \times 128$ Out: $H/2 \times W/2 \times 128$	$K = 3 \times 3, S = 1, P = 1,$ ACT = Leaky ReLU
	SAM 7	In: $H/2 \times W/2 \times 128$ Out: $H/2 \times W/2 \times 128$	$K = 3 \times 3, S = 2, P = 1,$ ACT = Leaky ReLU, GC = : Gate 2
Level 8 (Decoder)	Upsampling 3	In: $H/2 \times W/2 \times 128$ Out: $H \times W \times 64$	Scailing factor = 2
	Residual Block 8	In: $H \times W \times 64$ Out: $H \times W \times 64$	$K = 3 \times 3, S = 1, P = 1,$ ACT = Leaky ReLU
	SAM 8	In: $H \times W \times 64$ Out: $H \times W \times 64$	$K = 3 \times 3, S = 2, P = 1,$ ACT = Leaky ReLU, GC = : Gate 1
Output	Conv	In: $H \times W \times 64$ Out: $H \times W \times 3$	$K = 3 \times 3, S = 2, P = 1,$ ACT = tanh, SC = : input

Table 1: Architectural details of proposed SAGAN generator.

2.2 Discriminator

Table. 2 illustrates the details of our proposed SAGAN discriminator.

Level	Block	Tensor Dimension	Configuration
Input	Conv	In: $H \times W \times 3$ Out: $H \times W \times 64$	$K = 3 \times 3$, $S = 1$, $P = 1$, BN, ACT = swish
Layer 1	Conv2	In: $H \times W \times 64$, Out: $H/2 \times W/2 \times 128$	$K = 3 \times 3$, $S = 2$, $P = 1$, BN, ACT = swish
Layer 2	Conv3	In: $H/2 \times W/2 \times 128$ Out: $H/2 \times W/2 \times 128$	$K = 3 \times 3$, $S = 1$, $P = 1$, BN, ACT = swish
Layer 3	Conv4	In: $H/2 \times W/2 \times 128$ Out: $H/4 \times W/4 \times 256$	$K = 3 \times 3$, $S = 2$, $P = 1$, BN, ACT = swish
Layer 4	Conv5	In: $H/4 \times W/4 \times 256$ Out: $H/4 \times W/4 \times 256$	$K = 3 \times 3$, $S = 1$, $P = 1$, BN, ACT = swish
Layer 5	Conv6	In: $H/4 \times W/4 \times 256$ Out: $H/8 \times W/8 \times 512$	$K = 3 \times 3$, $S = 2$, $P = 1$, BN, ACT = swish
Layer 6	Conv7	In: $H/8 \times W/8 \times 512$ Out: $H/8 \times W/8 \times 512$	$K = 3 \times 3$, $S = 1$, $P = 1$, BN, ACT = swish
Layer 7	SAM	In: $H/8 \times W/8 \times 512$ Out: $H/16 \times W/16 \times 512$	$K = 3 \times 3$, $S = 2$, $P = 1$, ACT = Leaky ReLU
Output	Conv	In: $H/16 \times W/16 \times 512$, Out: $H/16 \times W/16 \times 512$	$K = 1 \times 1$, $S = 2$, $P = 1$, BN, ACT = swish
	Average Pooling	Flatten	ACT = sigmoid

Table 2: Architectural details of proposed SAGAN discriminator.

3 Quantitative Evaluations

To study the feasibility of a learning-based deep method for noisy Nona-Bayer reconstruction, we have extensively studied the state-of-the-art methods and our SAGAN with benchmark datasets. This section detail the performance of deep methods in benchmark datasets from sRGB (i.e., BSD100 [9], McM [12], Urban100 [9], Kodak [13], WED [8]) and linear RGB (i.e., MSR demosaicking [9]) colour spaces.

3.1 Comparison on Benchmark Datasets

Table. 3 depicts the performance of existing reconstruction methods on benchmark datasets. It can be seen that our proposed method can illustrate a consistent performance on benchmark datasets. Our proposed method can considerably outperform the existing reconstruction methods in evaluation criteria.

3.2 Ablation Study on Benchmark Datasets

Table. 4 demonstrates the ablation study on our novel components in benchmark datasets. Our SAGAN components have a clear impact on noisy Nona-Bayer reconstruction. Notably, each proposed component helps us to perceive a consistent performance in a diverse range of image samples.

Model	σ	BSD100			WED			Urban100			McM			Kodak			MSR Demosaicking		
		PSNR	SSIM	DeltaE	PSNR	SSIM	DeltaE	PSNR	SSIM	DeltaE	PSNR	SSIM	DeltaE	PSNR	SSIM	DeltaE	PSNR	SSIM	DeltaE
Deepjoint [1]	10	34.05	0.926	2.56	30.64	0.8954	3.38	30.50	0.9094	3.45	30.96	0.8745	3.20	32.02	0.9077	2.98	39.00	0.9464	1.64
Kokkinos [1]		35.17	0.9524	2.31	32.04	0.9273	2.94	32.12	0.9375	3.17	32.34	0.9053	2.72	33.72	0.9378	2.63	39.26	0.9539	1.74
DPN [1]		35.82	0.9585	2.10	32.25	0.9345	2.88	32.64	0.9447	3.02	32.54	0.9124	2.69	34.19	0.9447	2.43	39.84	0.9702	1.50
BIDD [1]		36.21	0.9624	2.20	32.85	0.9386	2.70	33.30	0.9488	2.87	33.07	0.9202	2.55	34.69	0.9498	2.49	41.40	0.9751	1.64
SAGAN		37.56	0.9676	1.69	33.65	0.9455	2.41	33.94	0.9533	2.57	34.11	0.9311	2.20	35.70	0.9541	2.02	43.17	0.9788	1.11
Deepjoint [1]	20	31.99	0.8667	2.90	29.45	0.8494	3.71	29.36	0.8638	3.75	29.81	0.8271	3.52	30.47	0.8403	3.31	36.14	0.8946	1.94
Kokkinos [1]		33.67	0.9284	2.50	30.98	0.904	3.19	31.07	0.9168	3.36	31.40	0.8827	2.94	32.29	0.9081	2.86	38.18	0.9411	1.76
DPN [1]		34.00	0.9336	2.42	31.11	0.9121	3.17	31.48	0.9247	3.26	31.51	0.8913	2.97	32.53	0.9141	2.76	38.39	0.9572	1.68
BIDD [1]		34.29	0.9389	2.53	31.62	0.9173	3.01	31.99	0.9295	3.15	32.05	0.8996	2.82	32.96	0.9207	2.81	39.71	0.9619	1.86
SAGAN		35.32	0.9457	2.03	32.27	0.9256	2.73	32.50	0.9355	2.85	32.87	0.9117	2.49	33.69	0.9263	2.37	41.26	0.9675	1.32
Deepjoint [1]	30	30.17	0.8027	3.38	28.20	0.7965	4.16	28.13	0.8116	4.17	28.59	0.7759	3.96	28.94	0.77	3.81	34.05	0.8407	2.29
Kokkinos [1]		32.40	0.9031	2.75	30.01	0.8794	3.46	30.06	0.8938	3.61	30.53	0.86	3.19	31.07	0.8785	3.13	36.84	0.9203	1.95
DPN [1]		32.55	0.9074	2.74	30.08	0.8881	3.47	30.39	0.9038	3.53	30.61	0.8694	3.24	31.15	0.8832	3.09	36.99	0.9411	1.93
BIDD [1]		32.90	0.9161	2.80	30.58	0.896	3.30	30.88	0.9104	3.41	31.12	0.8791	3.08	31.64	0.8935	3.10	38.27	0.9466	2.03
SAGAN		33.83	0.9246	2.31	31.17	0.9058	3.02	31.34	0.9181	3.12	31.86	0.8925	2.76	32.28	0.9009	2.68	39.59	0.9525	1.57

Table 3: Qualitative comparison between existing reconstruction methods for Nona-Bayer reconstruction on benchmark sRGB and linear RGB datasets.

Model	σ	BSD100			WED			Urban100			McM			Kodak			MSR Demosaicking		
		PSNR	SSIM	DeltaE	PSNR	SSIM	DeltaE	PSNR	SSIM	DeltaE	PSNR	SSIM	DeltaE	PSNR	SSIM	DeltaE	PSNR	SSIM	DeltaE
BaseNet	10	23.55	0.6623	9.82	22.45	0.6469	9.80	21.60	0.6475	10.04	22.60	0.6129	9.71	23.02	0.63	10.02	25.80	0.6804	8.51
SAN		36.45	0.9651	1.77	32.94	0.9408	2.52	33.14	0.9507	2.66	33.38	0.9240	2.29	35.03	0.95	2.10	38.19	0.9742	1.16
SAGAN		37.56	0.9676	1.69	33.65	0.9455	2.41	33.94	0.9533	2.57	34.11	0.9311	2.20	35.70	0.95	2.02	43.17	0.9788	1.11
BaseNet	20	23.32	0.6021	10.48	22.19	0.5999	10.31	21.40	0.6022	10.59	22.12	0.5657	10.16	22.82	0.56	10.89	24.68	0.6160	9.30
SAN		34.32	0.9416	2.12	31.62	0.9201	2.84	31.84	0.9314	2.94	32.19	0.9038	2.58	33.13	0.92	2.46	36.39	0.9598	1.42
SAGAN		35.32	0.9457	2.03	32.27	0.9256	2.73	32.50	0.9355	2.85	32.87	0.9117	2.49	33.69	0.93	2.37	41.26	0.9675	1.32
BaseNet	30	22.76	0.5329	11.53	21.62	0.5435	11.23	20.95	0.5500	11.51	21.43	0.5091	11.01	22.11	0.48	12.27	23.48	0.5371	10.46
SAN		32.90	0.9184	2.44	30.57	0.8989	3.13	30.75	0.9123	3.22	31.18	0.8827	2.87	31.76	0.90	2.79	35.19	0.9423	1.69
SAGAN		33.83	0.9246	2.31	31.17	0.9058	3.02	31.34	0.9181	3.12	31.86	0.8925	2.76	32.28	0.90	2.68	39.59	0.9525	1.57

Table 4: Ablation study on sRGB and linear RGB benchmark datasets.

4 Qualitative Evaluations

Nona-Bayer CFA patterns are prone to produce visual artefacts while reconstructing RGB images. Therefore, we have extensively studied the performance of our proposed SAGAN for noisy Nona-Bayer reconstruction and compared it with the ground-truth Bayer RGB images.

4.1 Bayer vs. Nona-Bayer Reconstruction

Fig. 1 demonstrates the noisy Nona-Bayer reconstruction performance of our SAGAN on different noise levels. It can be visible that, despite being contaminated with random noises, our proposed SAGAN can reconstruct visually plausible images similar to the RGB Bayer images. Even in the complex structure, our method illustrates a consistent performance and exempt from visual artefacts.

4.2 High Frequency Image Reconstruction

Non-Bayer CFA like Nona-Bayer and Quad Bayer CFAs are prone to produce substantial visual artefacts on high-frequency images [1, 10]. Thus, we studied the practicability of our SAGAN on high-frequency images. Fig. 2 depicts the qualitative results of our proposed SAGAN in high-frequency images on different noise levels. Our proposed method can reconstruct high-frequency images without producing any notable visual artefacts.

5 Real-world Denoising with Nona-Bayer Reconstruction

Real-world sensor noises can considerably different from simulated noises [1, 10]. Hence, we studied our proposed method on real-world noisy images. We trained and tested our



Figure 1: Comparison between our reconstructed images and ground-truth RGB images. In each pair, the left image is reconstructed by SAGN, and the right image refers to the ground-truth RGB image. (a) Reconstruction at $\sigma = 10$. (b) Reconstruction at $\sigma = 20$. (c) Reconstruction at $\sigma = 30$.

method on Smartphone Image Denoising Dataset (SIDD) [14, 15]. It is worth noting, SIDD has been collected with smartphone image sensors and fit best with our target applications. To study the real-world denoising with Nona-Bayer reconstruction, we extracted 205,279

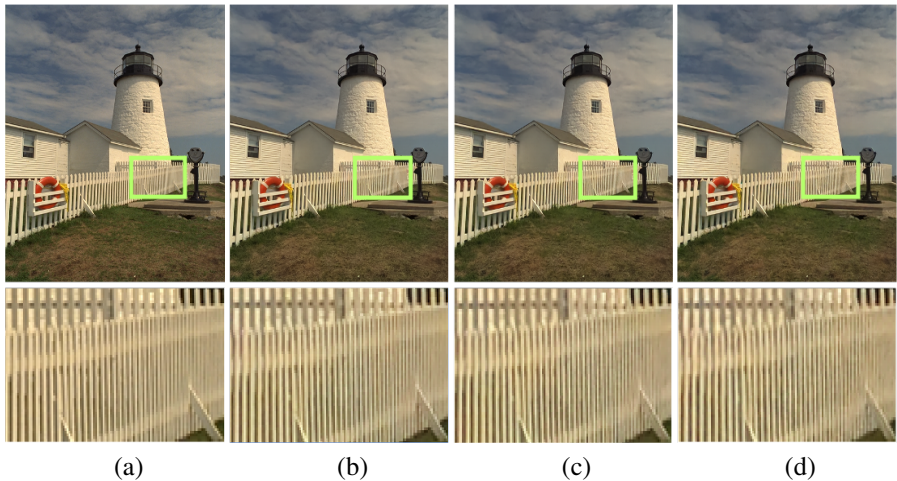


Figure 2: Qualitative evaluation of our proposed SAGAN on the high-frequency image. Our proposed SAGAN can handle high-frequency image scenes without producing any substantial artefacts. (a) Ground-truth RGB image. (b) Reconstruction at $\sigma = 10$. (c) Reconstruction at $\sigma = 20$. (d) Reconstruction at $\sigma = 30$.

image patches (dimension of $128 \times 128 \times 3$) by utilizing 320 scenes from SIDD. We sampled the extracted non-overlapping patches according to the CFA patterns and trained our SAGAN for 200,00 steps (with batch size of 16). Later, we tested our model by incorporating Nona-Bayer inputs with real-world noise. Fig 3 and Fig 4 shows the performance of our SAGAN on real-world noisy well-lit images, and low-light images. Our proposed method can substantially suppress the real-world noises and reconstruct RGB images, even in extreme stochastic lighting conditions.

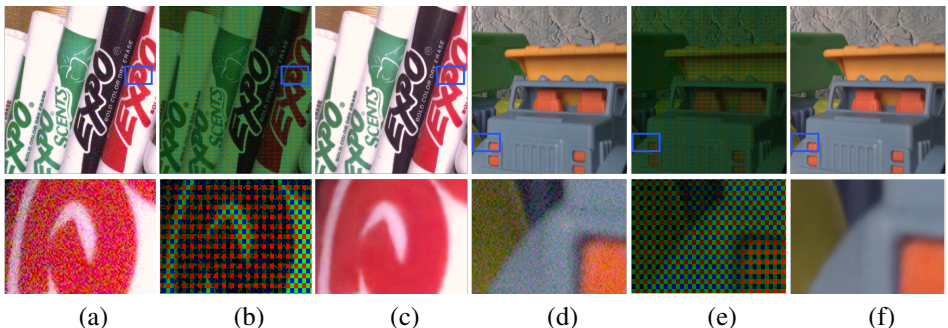


Figure 3: Nona-Bayer reconstruction with real-world noise suppression on well-lit conditions. (a) & (d) Noisy image. (b) & (e) Noisy Nona-Bayer (input). (c) & (f) Reconstructed images obtained by SAGAN.

6 Concluding Remarks

We studied our SAGAN on a diverse range of image samples. Our SAGAN can handle synthesized and real-world sensor noises along with demosaicing without producing any

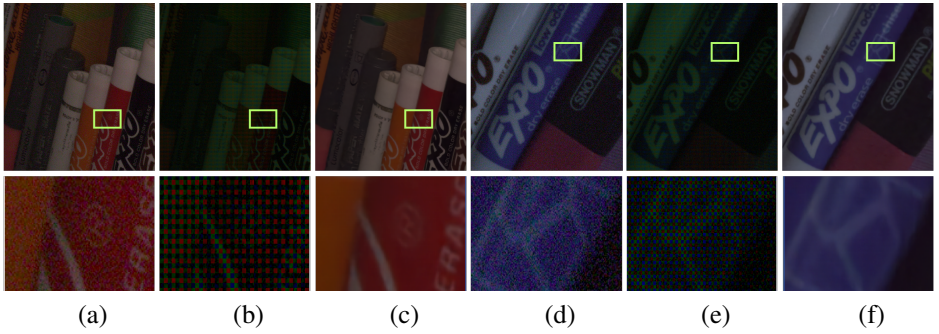


Figure 4: Nona-Bayer reconstruction with real-world noise suppression on low-light conditions. (a) & (d) Noisy image. (b) & (e) Noisy Nona-Bayer (input). (c) & (f) Reconstructed images obtained by SAGAN.

visual artefacts. Overall, the experimental results reveal the practicability of SAGAN for noisy Nona-Bayer reconstruction.

References

- [1] 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.
- [2] 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.
- [3] 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.
- [4] Michaël Gharbi, Gaurav Chaurasia, Sylvain Paris, and Frédo Durand. Deep joint demosaicking and denoising. *ACM Trans. Graph.*, 35(6):1–12, 2016.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.

- [9] 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.
- [10] 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.
- [11] SMA Sharif, Rizwan Ali Naqvi, and Mithun Biswas. Learning medical image denoising with deep dynamic residual attention network. *Mathematics*, 8(12):2192, 2020.
- [12] 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.
- [13] 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.