Self-Supervised Image Denoising

Sogang University
Vision & Display Systems Lab, Dept. of Electronic Engineering

Presented By
나송주

Outline

- Introduction
  - Image Denoising 소개 및 필요성
  - Background
    - Noise2Noise (ICML 2018)
    - Noise2Void (CVPR 2019)

- Paper Review
  - Spatially Adaptive Self-Supervised Learning for Real-World Image Denoising (CVPR 2023)
  - Iterative Denoiser and Noise Estimator for Self-Supervised Image Denoising (ICCV 2023)
  - Self-supervised Image Denoising with Downsampled Invariance Loss and Conditional Blind-Spot Network (ICCV 2023)

Introduction

- Image Denoising 소개 및 필요성
  - Problem formulation
    - Image denoising은 noisy image $\boldsymbol{x} = \boldsymbol{s} + \boldsymbol{n}$에서 noise $\boldsymbol{n}$를 제거하여 image signal $\boldsymbol{s}$를 복원하는 task
    - Noisy image는 다음과 같은 joint distribution을 따른다고 볼 수 있음. 또한 noise는 일반적으로 zero-mean을 가진다고 가정하며 따라서 noisy image의 기대값 (평균)은 image signal이 됨

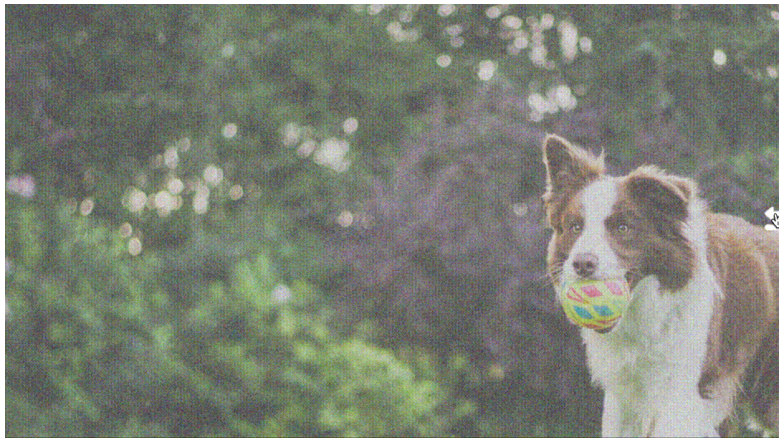
$$p(\boldsymbol{s}, \boldsymbol{n}) = p(\boldsymbol{s})p(\boldsymbol{n}|\boldsymbol{s})$$

$$\mathbb{E}[\boldsymbol{n}_i] = 0,$$

$$\mathbb{E}[\boldsymbol{x}_i] = \boldsymbol{s}_i.$$

    - Denoiser $f(\cdot;\theta)$를 학습 시킬 때 가장 널리 사용되는 방법은 noisy-clean paired data를 사용하여 mean-squared-error (MSE)를 최소화하는 방법

$$\underset{\theta}{\operatorname{argmin}}\, \mathbb{E}_{x,s}\left[\|f(x;\theta) - s\|_2^2\right]$$

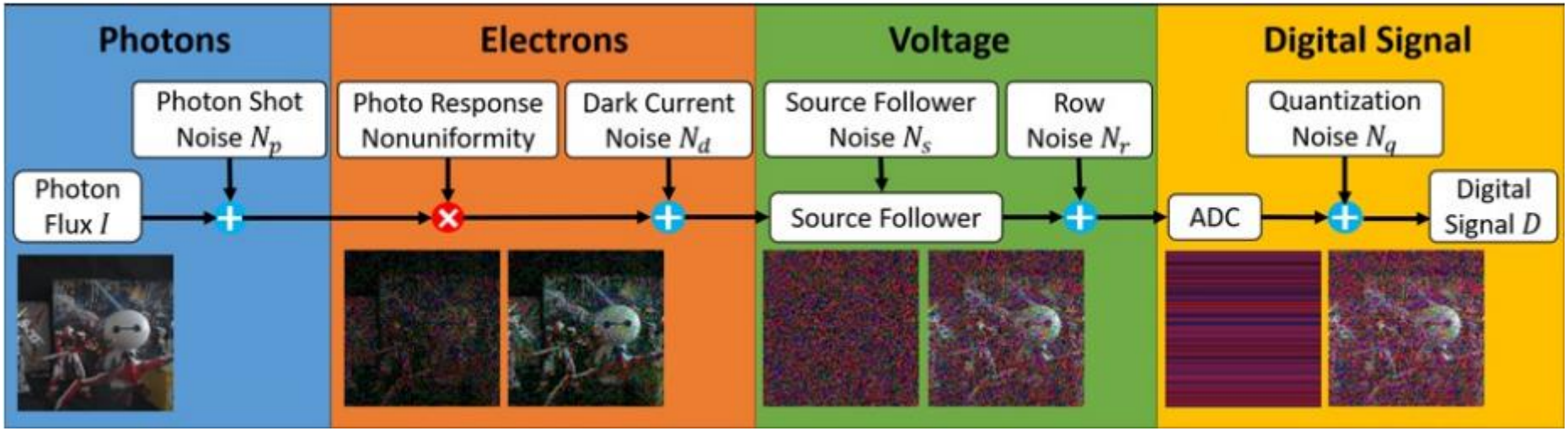


<Image denoising 예시>

서강대학교 SOGANG UNIVERSITY
VDS LAB

Introduction

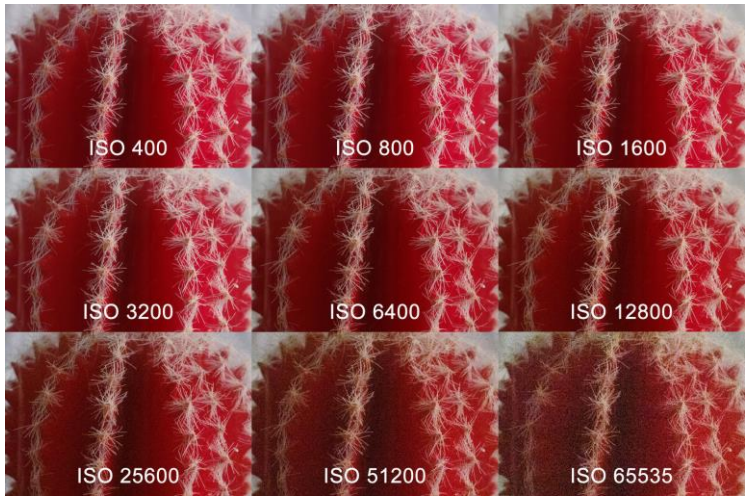
- Image Denoising 소개 및 필요성
  - Various noise sources in camera pipeline
    - 카메라의 광학 센서에 입력되는 광자의 세기를 전압으로 변환하고 이를 디지털 이미지로 변환하는 과정에서 노이즈는 필연적으로 발생
    - 흔히 접하는 광학 카메라 뿐만 아니라 센서를 통해 물리적인 값을 이미징 하는 과정에서도 (예시: 현미경, 적외선 카메라 등) 노이즈가 발생
    - 또한 noise source 마다 발생하는 noise의 형태가 다르다는 특징을 가짐
    - 이처럼 영상 취득 장치에서 필연적으로 노이즈가 발생하며 denoising은 이러한 현상에 해결 방안으로써 오랜 기간 연구되어 옴



<Real noise in camera pipeline>

Introduction

- Image Denoising 소개 및 필요성
  - Challenges on image denoising
    - 다양한 형태의 noise
      - 영상을 취득하는 환경, 카메라 등에 따라 noise의 형태가 변함
      - 특정 조건에서 취득된 noisy-clean image pair를 사용하여 모델을 학습시켜도 다른 형태의 noise가 입력되면 성능이 하락하는 문제가 존재
    - 데이터 취득의 어려움
      - 일반적인 딥러닝 기반의 denoising 방법은 많은 양의 noisy-clean image pair를 필요로 함
      - 극한 환경에서 취득되는 영상은 clean image 얻는 것 자체가 불가능 (예시: 우주 탐사선에서 취득된 영상, 의료 영상 등)



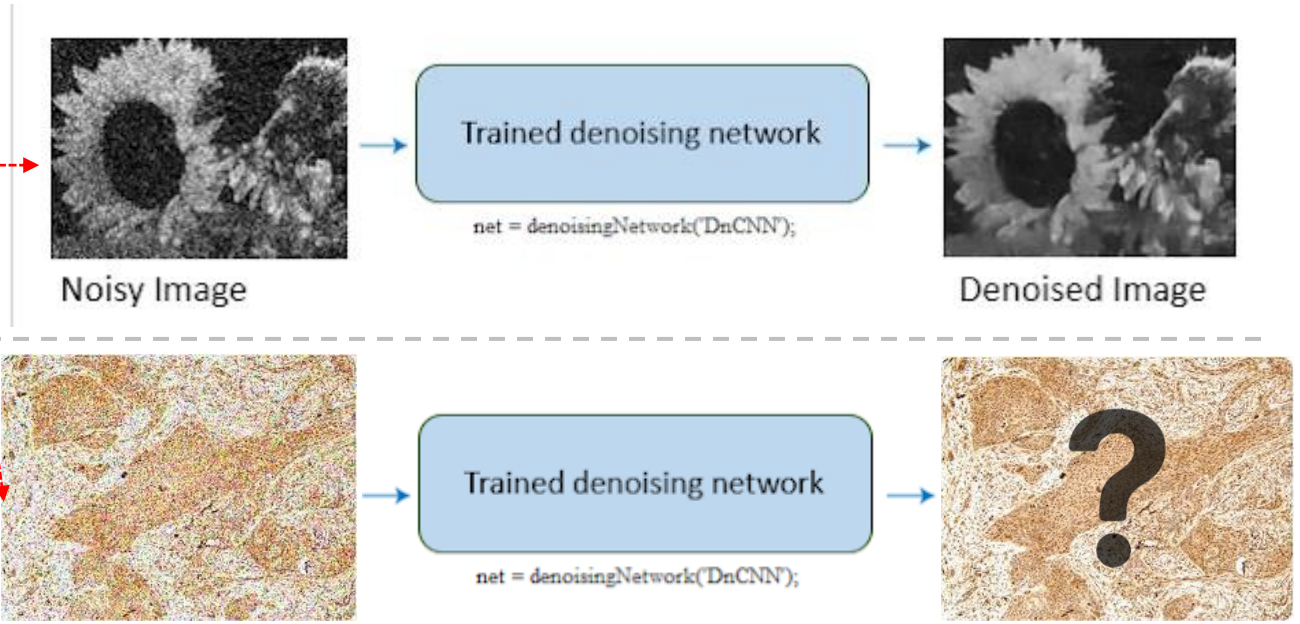
<Camera sensor 민감도에 따른 image 변화>

Introduction

- Background
  - Recent trends in image denoising
    - Real noise dataset을 (예: SIDD, DND 데이터셋) 취득하고 이를 사용하여 학습하는 방법
    - GAN 등을 사용하여 real noise와 유사한 noise를 synthesis하거나 물리적 특성을 분석하여 modeling하는 방법 (예: PNGAN, Zhang *et al*. (ICCV 2023))
    - Noise estimator와 같은 sub-network를 denoiser와 함께 학습시켜 다양한 noise에 대한 강건성을 높이는 방법 (예: CBDNet, Zhou *et al*. (AAAI 2020))
    - **Noisy image만을 사용하여 self-supervised 방식으로 모델을 학습 시키는 방법**

1. 대부분 방법이 noisy-clean pair를 필요로 함

2. Train – test dataset의 image content가 크게 다르면 성능이 저조할 수 있음



<기존 방법의 한계점 예시>

서강대학교 SOGANG UNIVERSITY
VDS LAB

Introduction

- Background
  - Noise2Noise (N2N)
    - 일반적인 image denoising model은 noisy input $\hat{x}_i$ 와 clean target y_i에 대해 L1 또는 L2 loss를 최소화하는 방법으로 학습됨

$$\underset{\theta}{\operatorname{argmin}} \sum_i L\left(f_\theta(\hat{x}_i),\, y_i\right)$$

    - 특히 L2 loss를 사용하는 경우 아래와 같이 target의 mean으로 수렴함

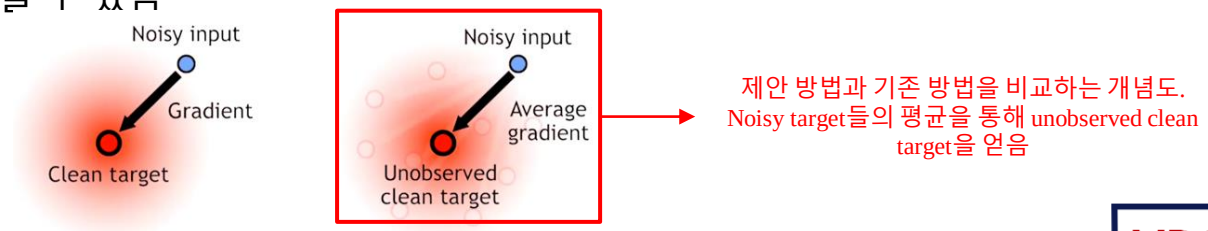
$$MSE = \mathbb{E}\left(\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \mid \mathbf{y}\right) = \mathbb{E}\left(\|\mathbf{x} - f(\mathbf{y})\|_2^2 \mid \mathbf{y}\right) = \int \|\mathbf{x} - f(\mathbf{y})\|_2^2 p(\mathbf{x}|\mathbf{y}) d\mathbf{x}.$$

$$\frac{d}{df(\mathbf{y})} \int \|\mathbf{x} - f(\mathbf{y})\|_2^2 p(\mathbf{x}|\mathbf{y}) d\mathbf{x} = -2 \int (\mathbf{x} - f(\mathbf{y})) p(\mathbf{x}|\mathbf{y}) d\mathbf{x} = 0.$$

$$\int \mathbf{x} p(\mathbf{x}|\mathbf{y}) d\mathbf{x} = \int f(\mathbf{y}) p(\mathbf{x}|\mathbf{y}) d\mathbf{x} = f(\mathbf{y}) \int p(\mathbf{x}|\mathbf{y}) d\mathbf{x} = f(\mathbf{y}).$$

$$f_{MMSE}(\mathbf{y}) = \int_{\mathbf{x}} \mathbf{x} p(\mathbf{x}|\mathbf{y}) d\mathbf{x} = \mathbb{E}\left(\mathbf{x}|\mathbf{y}\right).$$

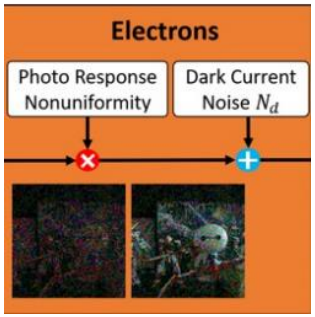
    - Denoising뿐만 아니라 대부분의 image restoration task에서 input과 target은 1:1 mapping이 아니므로 모델은 가능한 target들을 평균 낸 결과에 수렴함
    - 따라서 target이 반드시 clean image일 필요가 없고, expectation이 clean image 이기만 하면 denoising network의 target으로 사용될 수 있음

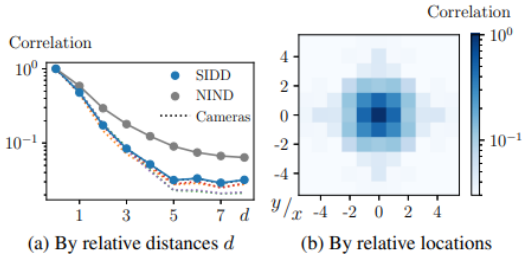


<Noise2Clean과 Noise2Noise의 gradient 개념도>

Introduction

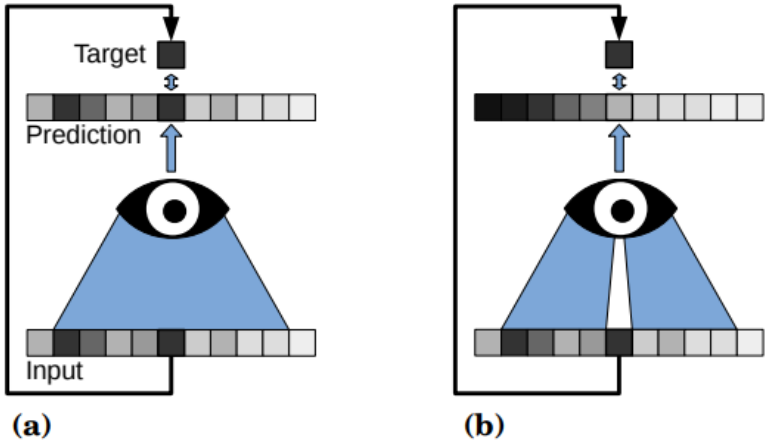
- Background
  - Noise2Void (N2V)
    - N2V는 아래와 같은 통계적 가정에 근거함
      - 1. Image signal s is not pixel-wise independent
      - 2. Noise signal n has zero-mean and is conditionally pixel-wise independent given the signal s





Real noise는 이 조건에 위배됨

- N2V는 input과 GT 모두 noisy training image x ($x = s + n$)로부터 추출하며, receptive field에서 center pixel이 제외된 blind spot network (BSN)구조를 사용함
- 즉, center pixel을 관측하지 못하도록 제한하고 주변 pixel만 관측하여 center pixel을 예측하도록 모델을 학습시킴
  - Noise는 pixel-wise independent하여 주변 픽셀이 center pixel을 예측하는데 아무런 정보를 주지 못하는 반면에 image pixel은 주변 context를 보고 center pixel 유추 가능



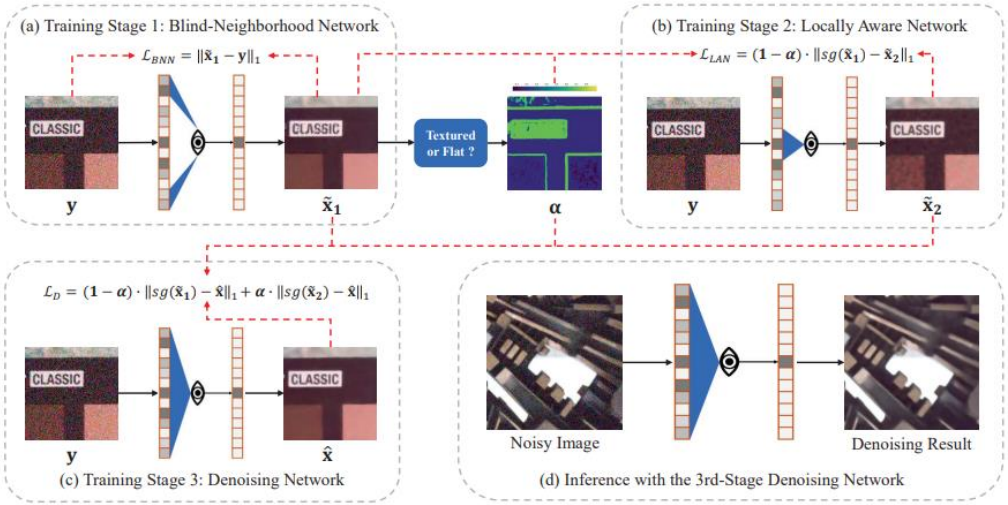
<N2C와 N2V 비교 개념도>

서강대학교 SOGANG UNIVERSITY
VDS LAB

Spatially Adaptive Self-Supervised Learning for Real-World Image Denoising

Paper Review

- Spatially Adaptive Self-Supervised Learning for Real-World Image Denoising
  - Overview
    - N2V는 pixel-wise independent한 noise를 가정하는데, real noise는 pixel 간에 correlation 존재
    - Image의 flat region은 넓은 영역에 걸쳐 같은 image content를 가지므로 noise correlation을 무시할 수 있을 만큼 멀리 있는 pixel까지 blind spot을 넓혀 위 문제 해결 가능
    - 반면 texture region은 원본 image를 복원하는데 있어서 인접 pixel 정보를 반드시 사용해야 함
    - 본 논문은 BNN이라는 flat region에 특화된 네트워크를 먼저 학습 시키고, 이후 BNN으로 잘 복원되지 않는 texture 영역을 LAN이라는 네트워크를 사용하여 복원
    - 마지막으로 학습된 BNN과 LAN을 사용하여 복원된 영상을 일반적인 denoising network의 supervision으로 제공하여 denoising network를 학습시킴



<제안 방법의 전체 개념도>

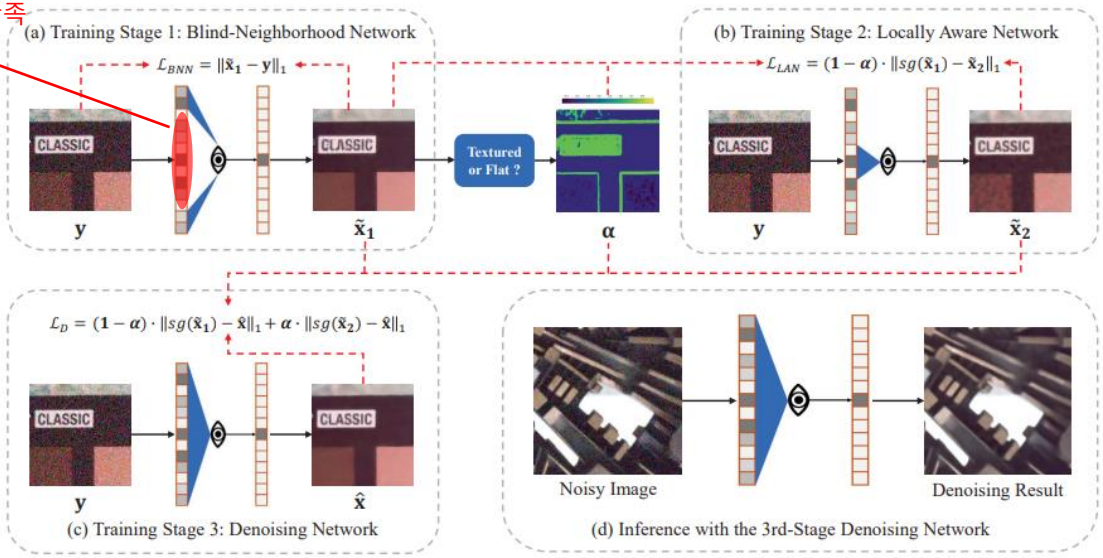
서강대학교 SOGANG UNIVERSITY
VDS LAB

Paper Review

- Spatially Adaptive Self-Supervised Learning for Real-World Image Denoising
  - Method
    - Flat region에서는 인접 pixel뿐만 아니라 noise correlation을 벗어나는 먼 위치에 있는 pixel도 같은 image content를 가지고 있음
    - Noise-correlation이 끊어질 만큼 멀리 있는 pixel까지 blind spot을 늘려서 N2V의 pixel-wise independent 조건을 만족시킬 수 있음
    - 본 논문은 BSN에서 blind spot을 확장한 Blind-Neighbor Network (BNN) 구조를 제안하였으며, 다음과 같은 loss function을 사용하여 BNN을 학습함

인접 pixel은 noise끼리 correlation이 있으니
충분히 멀리 있는 pixel만 관측하도록 해서
N2V의 조건을 만족

$$\mathcal{L}_{BNN} = \|\tilde{\mathbf{x}}_1 - \mathbf{y}\|_1$$



<제안 방법의 전체 개념도>

서강대학교 SOGANG UNIVERSITY
VDS LAB

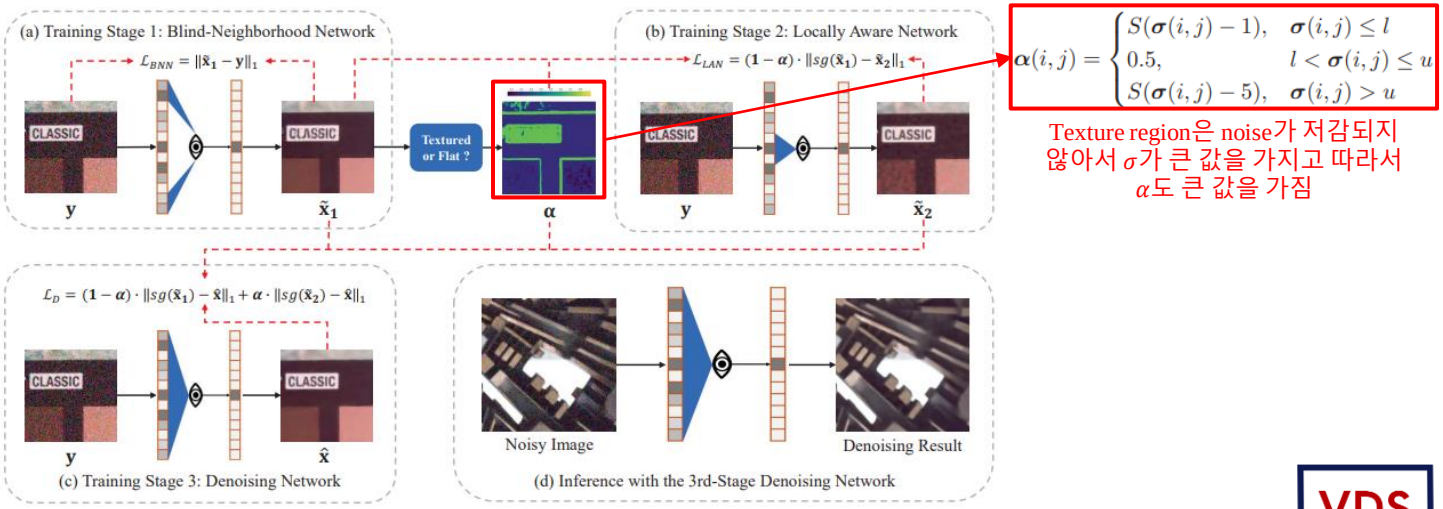
Paper Review

- Spatially Adaptive Self-Supervised Learning for Real-World Image Denoising
  - Method
    - BNN은 flat region에는 잘 동작하지만, texture region은 잘 복원하지 못함
    - Noisy image를 BNN을 사용하여 복원한 후 각 픽셀 위치에서 인접 patch의 표준편차 σ를 계산

$$\boldsymbol{\sigma}(i,j) = std(\tilde{\mathbf{x}}_1(i-\frac{n-1}{2}:i+\frac{n-1}{2}, j-\frac{n-1}{2}:j+\frac{n-1}{2}))$$

    - 이후 σ를 sigmoid에 통과시켜서 각 픽셀이 얼마나 texture한지 나타내는 α-map을 얻음
    - 한편 texture region은 blind network 구조로는 학습하기 어려우며, supervised 방식의 target이 필요함
    - BNN은 flat region의 noise를 잘 제거할 수 있는 것에 착안하여 일반적인 CNN 구조를 사용한 LAN을 BNN의 출력을 사용하여 다음과 같이 학습시킴

$$\mathcal{L}_{LAN} = (\mathbf{1}-\boldsymbol{\alpha}) \cdot \|sg(\tilde{\mathbf{x}}_1) - \tilde{\mathbf{x}}_2\|_1$$



$$\boldsymbol{\alpha}(i,j) = \begin{cases} S(\boldsymbol{\sigma}(i,j)-1), & \boldsymbol{\sigma}(i,j) \le l \\ 0.5, & l < \boldsymbol{\sigma}(i,j) \le u \\ S(\boldsymbol{\sigma}(i,j)-5), & \boldsymbol{\sigma}(i,j) > u \end{cases}$$

Texture region은 noise가 저감되지 않아서 σ가 큰 값을 가지고 따라서 α도 큰 값을 가짐

<제안 방법의 전체 개념도>

서강대학교 SOGANG UNIVERSITY
VDS LAB

Paper Review

- Spatially Adaptive Self-Supervised Learning for Real-World Image Denoising
  - Method
    - 마지막으로 학습된 BNN과 LAN을 사용하여 일반적인 denoising network를 다음과 같이 학습시킴

$$\mathcal{L}_D = (1-\alpha)\cdot\|sg(\tilde{\mathbf{x}}_1)-\hat{\mathbf{x}}\|_1 + \alpha\cdot\|sg(\tilde{\mathbf{x}}_2)-\hat{\mathbf{x}}\|_1$$

      ⁘ α가 작은 flat region은 BNN의 출력을, α가 큰 texture region은 LAN의 출력을 학습

    - Inference 단계에서는 BNN과 LAN은 사용하지 않고, 학습된 denoising network만을 사용
    - 추가적으로, BNN과 LAN을 단순히 blending하는 방법을 고려할 수 있으나 이는 제안 방법보다 저조한 성능을 가지는 것을 실험을 통해 확인

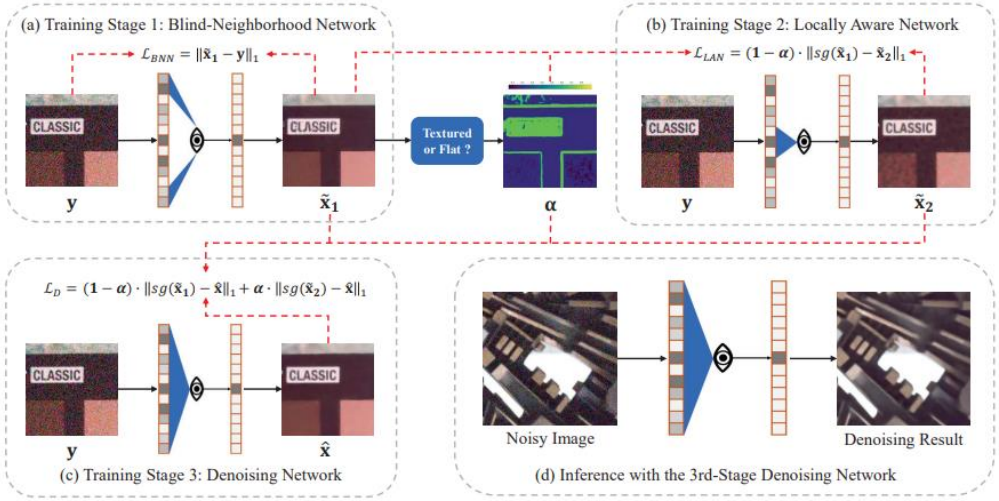


Table 3. Comparison between image-level fusion strategy and our method.

Blending 방법 (d)와 제안 방법 (e)의 성능 비교

Image	$\tilde{\mathbf{x}}_1$	$\tilde{\mathbf{x}}_2$	$(1-\alpha)\cdot\tilde{\mathbf{x}}_1+\alpha\cdot\tilde{\mathbf{x}}_2$	$\hat{\mathbf{x}}$
PSNR	36.37	35.00	36.84	37.39
Time (ms)	16.7	5.9	22.9	4.8

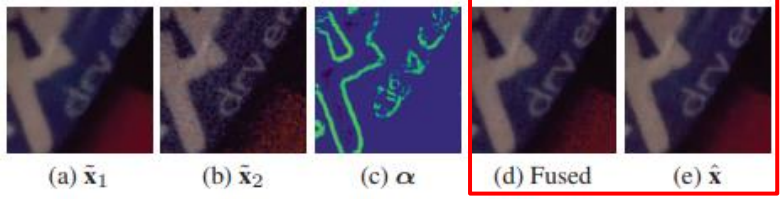
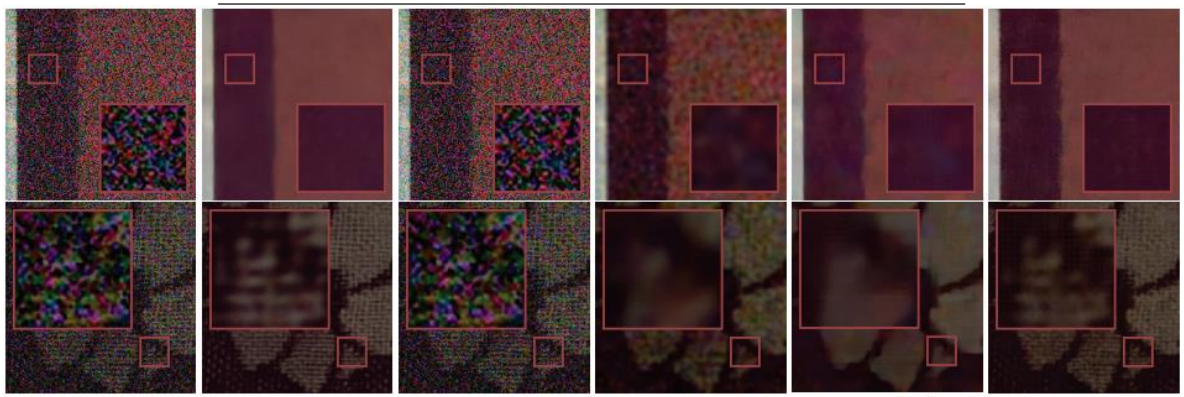


Figure 5. Visual comparison with image-level fusion strategy. (d) denotes the spatially adaptive fusion result $(1-\alpha)\cdot\tilde{\mathbf{x}}_1+\alpha\cdot\tilde{\mathbf{x}}_2$.

Paper Review

- Spatially Adaptive Self-Supervised Learning for Real-World Image Denoising
  - Experiments
    - SIDD의 train dataset에서 학습된 여러 모델을 SIDD의 validation/benchmark dataset과 DND benchmark dataset에서 성능을 비교하는 실험 진행

	Method	SIDD Validation PSNR↑ / SSIM↑ / LPIPS↓	SIDD Benchmark PSNR↑ / SSIM↑	DND Benchmark PSNR↑ / SSIM↑
Non-learning based	BM3D [11]	25.71 / 0.576 / 0.657	25.65 / 0.685	34.51 / 0.851
	WNNM [14]	26.05 / 0.592 / 0.635	25.78 / 0.809	34.67 / 0.865
Supervised (Synthetic pairs)	DnCNN [50]	26.21 / 0.604 / 0.712	26.25 / 0.599	32.43 / 0.790
	CBDNet [15]	33.07 / 0.863 / 0.288	33.28 / 0.868	38.05 / 0.942
	Zhou *et al.* [57]	33.96 / 0.899 / 0.258	34.00 / 0.898	38.40 / 0.945
Supervised (Real pairs)	DnCNN [50]	37.73 / 0.943 / 0.245	37.61 / 0.941	38.73 / 0.945
	Baseline, N2C [39]	38.98 / 0.954 / 0.201	38.92 / 0.953	39.37 / 0.954
	VDN [46]	39.29 / 0.956 / 0.208	39.26 / 0.955	39.38 / 0.952
	Restormer [48]	39.93 / 0.960 / 0.198	40.02 / 0.960	40.03 / 0.956
Unpaired	GCBD [8]	-	-	35.58 / 0.922
	UIDNet [16]	-	32.48 / 0.897	-
	C2N [18]	35.36 / 0.932 / 0.192	35.35 / **0.937**	37.28 / 0.924
	Wu *et al.* [44]	-	-	37.93 / 0.937
Self-Supervised	Noise2Void [24]	27.48 / 0.664 / 0.592	27.68 / 0.668	-
	Noise2Self [3]	29.94 / 0.782 / 0.556	29.56 / 0.808	-
	NAC [45]	-	-	36.20 / 0.925
	R2R [36]	-	34.78 / 0.898	-
	CVF-SID [35]	34.15 / 0.911 / 0.423	34.71 / 0.917	36.50 / 0.924
	AP-BSN+R^3 [27]	36.74 / **0.934** / 0.281	36.91 / 0.931	38.09 / 0.937
	Ours	**37.39** / **0.934** / **0.176**	**37.41** / 0.934	**38.18** / **0.938**



(a) Noisy Input (b) Baseline, N2C [39] (c) Noise2Void [24] (d) CVF-SID [35] (e) AP-BSN+R^3 [27] (f) Ours

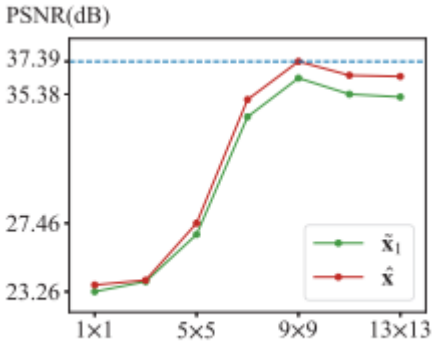
Paper Review

- Spatially Adaptive Self-Supervised Learning for Real-World Image Denoising
  - Experiments
    - Supervision component의 효과를 확인하기 위해 최종 denoising network를 학습 시키는 loss function에서 각 요소를 제거하며 denoising 성능을 비교

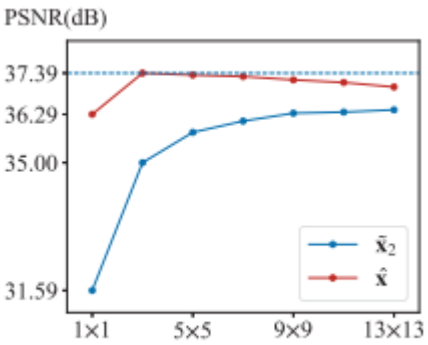
$$\mathcal{L}_D = (\mathbf{1} - \boldsymbol{\alpha}) \cdot \|sg(\tilde{\mathbf{x}}_1) - \hat{\mathbf{x}}\|_1 + \boldsymbol{\alpha} \cdot \|sg(\tilde{\mathbf{x}}_2) - \hat{\mathbf{x}}\|_1$$

Supervision of $\tilde{\mathbf{x}}_1$	✓		✓	✓
Supervision of $\tilde{\mathbf{x}}_2$		✓	✓	✓
Adaptive Coefficients $\boldsymbol{\alpha}$				✓
PSNR of $\hat{\mathbf{x}}$	36.84	35.95	37.30	37.39

    - BNN의 blind neighborhood size와 LAN의 receptive field size에 대한 ablation study 수행



(a) Blind-neighborhood size.

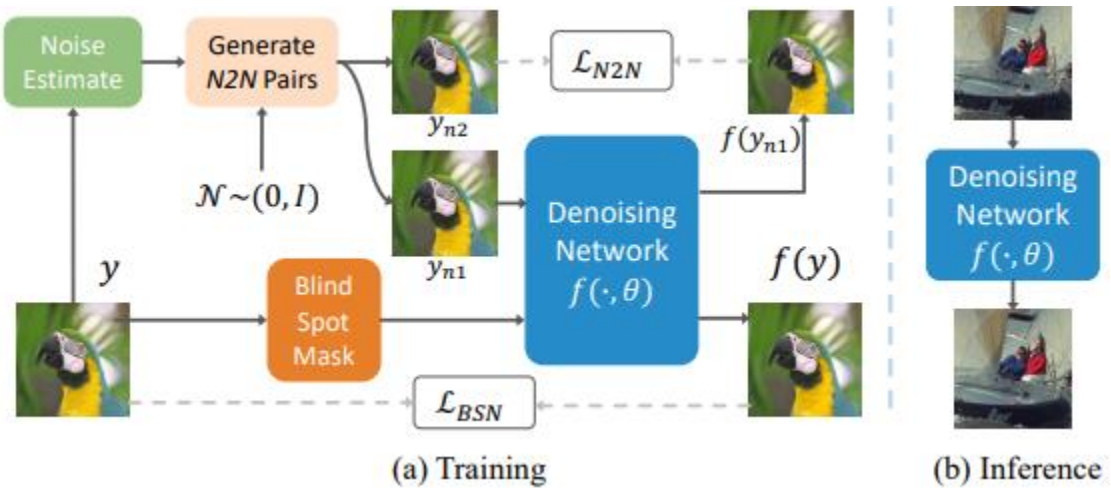


(b) Local receptive size.

Iterative Denoiser and Noise Estimator for Self-Supervised Image Denoising

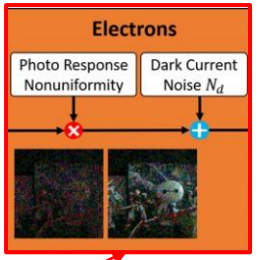
Paper Review

- Iterative Denoiser and Noise Estimator for Self-Supervised Image Denoising
  - Observations and Motivations
    - N2N은 충분한 training sample이 제공된다면 clean target을 사용한 경우와 동일한 성능을 얻을 수 있으며 self-supervised 방법보다 더 높은 성능을 달성할 수 있음
    - 기존에 N2N학습을 위한 data generation 방법은 clean image를 필요로 한다는 한계를 가짐
    - 본 논문은 noise estimator와 denoiser를 iterative하게 학습하는 방법을 통해 온전한 self-supervised 방식으로 N2N의 성능을 높이는 방법을 제안
      - Noise estimator는 noisy image의 noise level을 추정하여 N2N 학습을 위한 데이터를 생성하고, denoiser는 생성된 데이터를 사용하여 N2N 방식으로 학습



<제안 방법의 전체 개념도>

Paper Review



- Iterative Denoiser and Noise Estimator for Self-Supervised Image Denoising
  - Method
    - Previous works에 따르면 광학 센서에 발생하는 real noise는 heteroscedastic Gaussian model을 따르며, real noise n은 clean image x에 대해 다음과 같이 모델링 될 수 있음

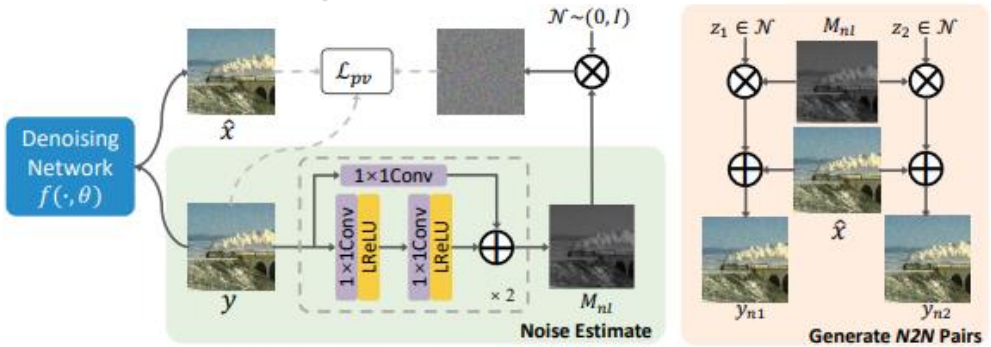
$$\boldsymbol{n} = \mathcal{N} \sim (0, \sigma_d^2 \boldsymbol{x} + \sigma_i^2)$$

→Noise level은 image content의 영향을 받음을 의미

    - 즉, signal-dependent noise σ_d는 image의 밝은 영역에서는 noise variance가 높아지는 것을 의미하고 signal-independent noise σ_i는 image에 상관없이 모든 픽셀에 추가되는 기본 noise level을 의미
    - 먼저 noisy image y를 denoiser network f에 통과시켜 pre-denoised image $\hat{x}$를 얻음
    - 이후 pre-denoised image와 noisy image를 estimator에 통과시켜 pixel-wise noise level map M_{nl} 추정
    - 계산된 M_{nl}을 사용하여 N2N pair를 생성하는데, 여기서 reparameterization trick 사용
    - Noise estimator를 constrain하기 위해 patch variance loss를 적용

$$\mathcal{L}_{pv} = \sum_i \| Var(\mathcal{P}(\boldsymbol{y}_{ni}, p)) - Var(\mathcal{P}(\boldsymbol{y}, p)) \|$$

→ $p \times p$ size patch에서 real noisy image y와 생성된 noisy image의 variance가 같아지도록 유도



(a) Noise level estimation (b) N2N pair generation

<제안한 noise estimation 과정>

서강대학교 SOGANG UNIVERSITY
VDS LAB

Paper Review

- Iterative Denoiser and Noise Estimator for Self-Supervised Image Denoising
  - Method
    - Noise estimator를 통해 생성된 N2N pair를 사용하여 denoiser를 학습

$$\mathcal{L}_{N2N} = \|f(\boldsymbol{y}_{n1}) - \boldsymbol{y}_{n2}\|_2^2$$

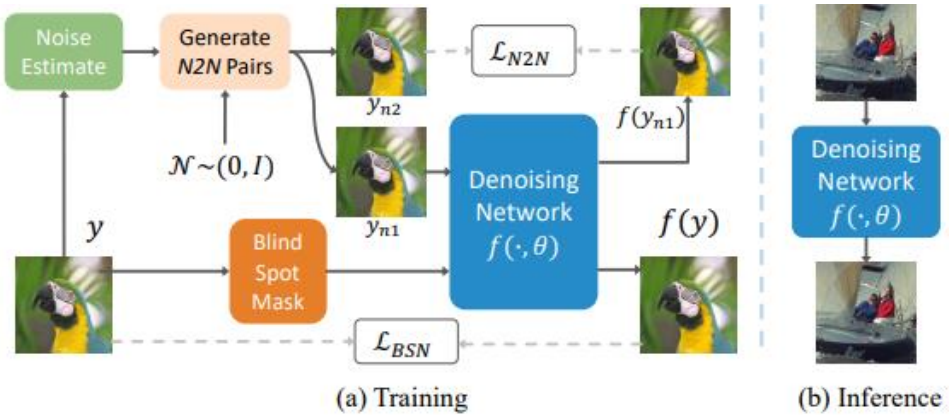
    - 학습 초기에 denoiser와 noise estimator의 성능이 저조해서 의도한 동작을 수행하지 못하고 이에 따라 학습이 불안정해지는 현상 발생
    - 이를 해결하기 위해 학습 초기에는 BSN을 사용하여 coarse denoised image를 얻고 이를 사용하여 denoiser를 학습하는 방법을 적용

$$\mathcal{L}_{BSN} = \|f(\boldsymbol{y}_{RF};\theta) - \boldsymbol{y}\|_2^2$$

    - 최종 loss function은 다음과 같음

$$\mathcal{L} = \mu\mathcal{L}_{pu} + \gamma\mathcal{L}_{N2N} + \lambda\mathcal{L}_{BSN}$$

→ μ, γ, λ를 학습을 시작할 때는 1, 0, 1로 설정하고 epoch이 진행됨에 따라 0, 1, 0이 되도록 설정



<제안 방법의 전체 개념도>

서강대학교 SOGANG UNIVERSITY
VDS LAB

Paper Review

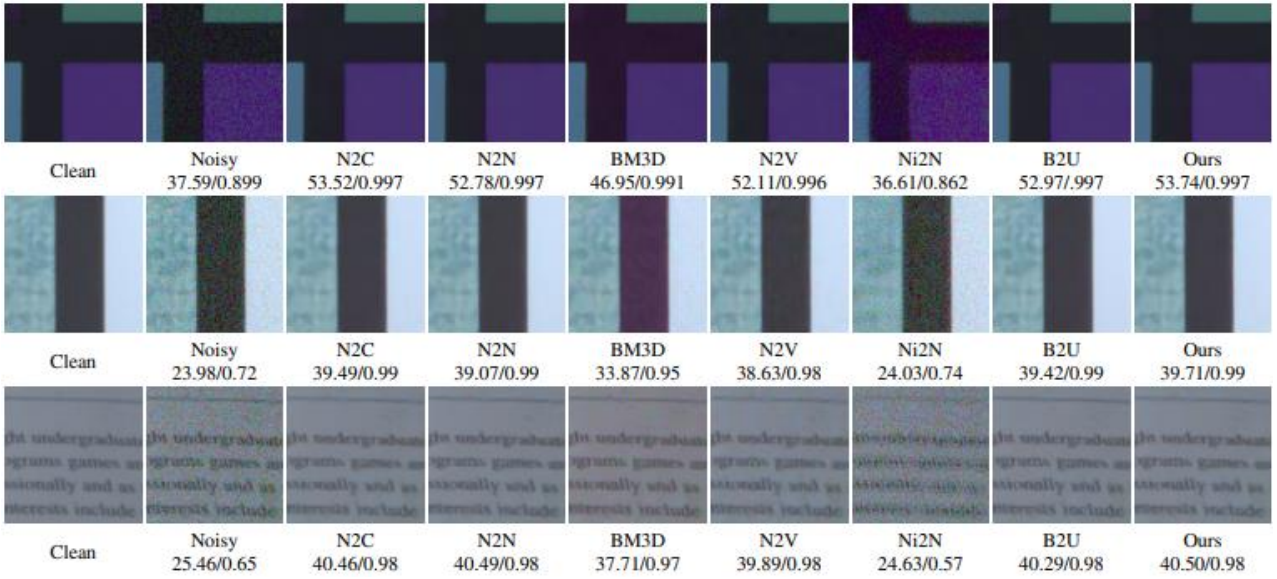
- Iterative Denoiser and Noise Estimator for Self-Supervised Image Denoising
  - Experiments
    - Synthetic sRGB noisy image에서 기존 방법들과 denoising 성능을 비교
      - 제안 방법이 가장 N2N 또는 N2C (Noise2Clean, supervised 방식을 의미)에 근접한 성능을 달성

Noise Type	Method	KODAK	BSD300	SET14
Gaussian $\sigma = 25$	N2C	32.46/0.884	31.20/0.881	31.43/0.868
	N2N [21]	32.48/0.885	31.22/0.882	31.45/0.869
	BM3D [7]	29.97/0.808	28.48/0.788	29.63/0.818
	N2V [18]	31.81/0.875	30.52/0.870	30.53/0.853
	SSDN [19]	30.62/0.840	28.62/0.803	29.93/0.830
	R2R [30]	32.25/0.880	30.91/0.872	31.32/**0.865**
	NAC [42]	25.69/0.521	25.51/0.583	25.67/0.586
	Ni2N [29]	30.45/0.811	29.34/0.803	29.75/0.815
	NBR2NBR [16]	32.08/0.879	30.79/0.873	31.09/0.864
	B2U [38]	**32.27**/0.880	30.87/0.872	31.27/0.864
	Ours	**32.27/0.881**	**31.01/0.876**	**31.29**/ 0.862
Gaussian $\sigma \in [5, 50]$	N2C	32.58/0.876	31.27/0.870	31.50/0.864
	N2N [21]	32.57/0.876	31.26/0.870	31.46/0.863
	BM3D [7]	29.38/0.781	28.83/0.795	30.74/0.834
	N2V [18]	31.72/0.863	30.39/0.855	30.24/0.843
	SSDN [19]	30.52/0.833	28.43/0.794	29.71/0.822
	R2R [30]	31.50/0.850	30.56/0.855	30.84/0.850
	NAC [42]	25.40/0.516	24.98/0.560	25.44/0.575
	Ni2N [29]	32.17/0.868	30.93/0.862	30.87/0.852
	NBR2NBR [16]	32.10/0.870	30.73/0.861	31.05/**0.858**
	B2U [38]	32.34/**0.872**	30.86/0.861	**31.14**/0.857
	Ours	**32.35/0.872**	**31.09/0.866**	31.09/0.855
Poisson $\lambda = 30$	N2C	31.84/0.877	30.54/0.872	30.63/0.859
	N2N [21]	31.84/0.877	30.54/0.872	30.63/0.858
	BM3D [7]	27.89/0.738	26.58/0.717	27.11/0.744
	N2V [18]	31.18/0.864	29.88/0.858	29.79/0.841
	SSDN [19]	30.19/0.833	28.25/0.794	29.35/0.820
	R2R [30]	30.50/0.801	29.47/0.811	29.53/0.801
	NAC [42]	24.36/0.486	24.33/0.559	23.93/0.541
	Ni2N [29]	29.43/0.775	28.29/0.764	28.63/0.778
	NBR2NBR [16]	31.44/0.870	30.10/0.863	30.29/0.853
	B2U [38]	**31.64/0.871**	**30.25**/0.862	**30.46**/0.850
	Ours	31.60/0.870	30.22/**0.865**	30.41/**0.855**
Poisson $\lambda \in [5, 50]$	N2C	31.25/0.862	30.17/0.859	30.28/0.848
	N2N [21]	31.17/0.861	30.10/0.859	30.19/0.847
	BM3D [7]	27.08/0.702	25.85/0.688	26.44/0.724
	N2V [18]	30.55/0.844	29.46/0.844	29.44/0.831
	SSDN [19]	29.76/0.820	27.89/0.778	28.94/0.808
	R2R [30]	29.14/0.732	28.68/0.771	28.77/0.765
	NAC [42]	23.12/0.447	23.47/0.534	23.14/0.516
	Ni2N [29]	30.31/0.812	29.45/0.821	29.40/0.812
	NBR2NBR [16]	30.86/0.855	29.54/0.843	29.79/0.838
	B2U [38]	**31.07/0.857**	29.92/0.852	**30.10/0.844**
	Ours	31.00/**0.857**	**29.99/0.855**	29.99/0.843

Paper Review

- Iterative Denoiser and Noise Estimator for Self-Supervised Image Denoising
  - Experiments
    - Real raw noisy image에서 기존 방법들과 denoising 성능 비교
      - 제안 방법이 가장 높은 성능을 달성
      - 일반적인 supervised 방식의 train cost를 “×1”라고 할 때, B2U는 ×17의 cost를 가지고, 제안 방법은 ×3의 cost를 가지므로 더 효율적이라고 볼 수 있음

Methods	Network	Train Cost	PSNR	SSIM
N2C	U-Net	×1	51.27	0.983
N2N [21]	U-Net	×1	51.29	0.991
BM3D [7]	-	-	48.13	0.983
N2V [18]	U-Net	×1	50.46	0.990
SSDN [19] (Gaussian)	U-Net	×4	50.44	0.990
SSDN [19] (Poisson)	U-Net	×4	50.89	0.990
R2R [30]	U-Net	×1	47.20	0.980
NAC [42]	U-Net	×1	43.24	0.961
Ni2N [29]	U-Net	×1	33.74	0.752
NBR2NBR [16]	U-Net	×2	51.06	0.991
B2U [38]	U-Net	×17	51.36	**0.992**
Ours	U-Net	×3	**51.40**	**0.992**



Paper Review

- Iterative Denoiser and Noise Estimator for Self-Supervised Image Denoising
  - Experiments
    - Patch variance loss의 유효성을 검증하는 ablation study 진행

$$\mathcal{L} = \mu\mathcal{L}_{pv} + \gamma\mathcal{L}_{N2N} + \lambda\mathcal{L}_{BSN}$$

	$\mu = 0$	$\mu = 1$	$\mu = 1 \rightarrow 0$
PSNR	24.21	31.86	**32.27**
SSIM	0.675	0.872	**0.881**

$$\mathcal{L}_{pv} = \sum_i \|Var(\mathcal{P}(\boldsymbol{y}_{ni}, p)) - Var(\mathcal{P}(\boldsymbol{y}, p))\|$$

patch size	4	8	32	64	Global
PSNR	32.21	**32.27**	32.23	32.24	32.18
SSIM	0.877	**0.881**	0.880	0.879	0.875

- N2N loss와 BSN loss의 유성을 검증하는 ablation study 진행

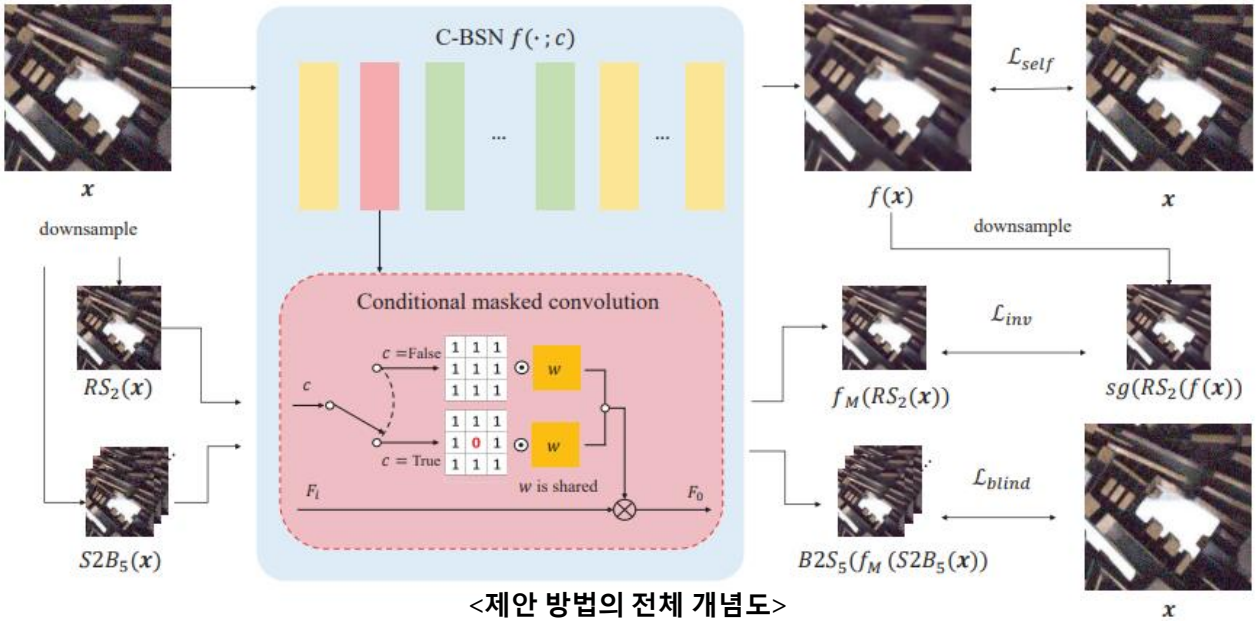
	$\gamma = 0$	$\gamma = 1$	$\gamma = 0 \rightarrow 1$
PSNR	31.81	32.14	**32.27**
SSIM	0.875	0.873	**0.881**

	$\lambda = 0$	$\lambda = 1$	$\lambda = 1 \rightarrow 0$
PSNR	14.46	31.75	**32.27**
SSIM	0.5312	0.875	**0.881**

Self-supervised Image Denoising with Downsampled Invariance Loss and Conditional Blind-Spot Network

Paper Review

- Self-Supervised Image Denoising with Downsampled Invariance Loss and Conditional Blind-Spot Network
  - Abstract
    - BSN은 center pixel 정보를 사용할 수 없다는 한계를 가짐
    - 본 논문은 supervised loss의 이론적 upper bound를 유도하고, 그 값을 single noisy image만을 사용하여 계산할 수 있도록 변형한 downsampled invariance loss를 제안
    - Downsampled invariance loss를 구현하기 위해 center kernel의 blind / non-blind를 선택적으로 설정할 수 있는 conditional blind spot network (CBSN)을 제안
    - 또한, 기존에 noise correlation을 완화하기 위해 사용되던 pixel down-shuffle의 문제점인 check board artifact를 완화할 수 있는 random subsampler (RS) 방법을 제안



<제안 방법의 전체 개념도>

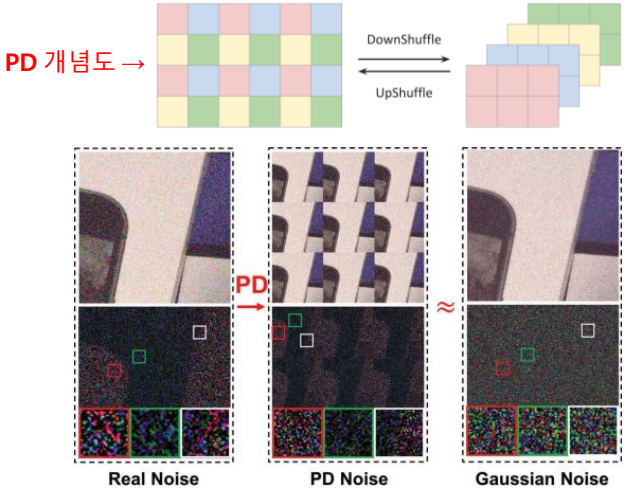
Paper Review

- Self-Supervised Image Denoising with Downsampled Invariance Loss and Conditional Blind-Spot Network
  - Method

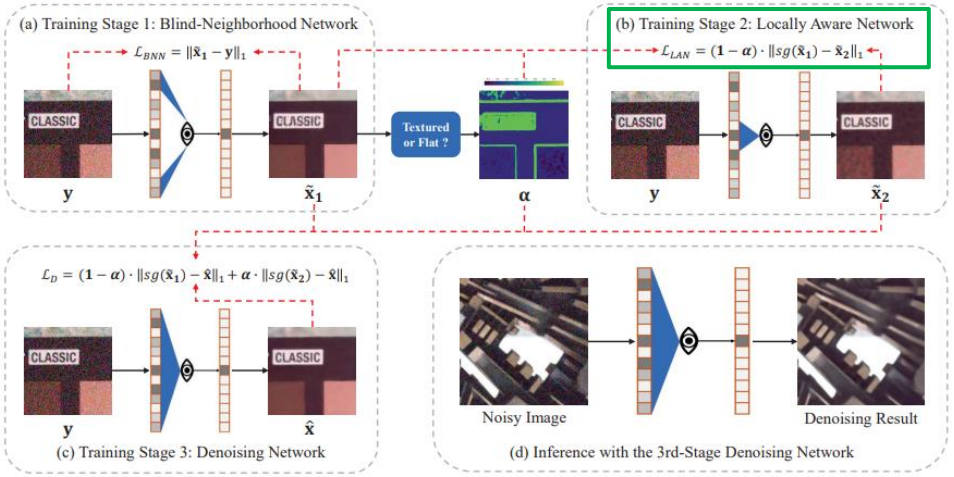
Learnable term　　Objective term

$$\mathcal{L}_{inv} = \sqrt{\frac{s^2}{m}} \left\| d_s(f(\boldsymbol{x})) - sg(f_M(d_s(\boldsymbol{x}))) \right\|_2$$

- d_s: s-fold down sampler → Noise correlation을 약화
- f: non-blind denoiser → Image는 잘 복원하지만 supervised 학습 필요 (첫 번째 논문의 LAN과 유사)
- f_M: blind denoiser → pixel-wise independent한 noise를 잘 복원함
- Objective term은 denoising이 잘 된 image, 반면 learnable term은 center pixel도 활용할 수 있지만 supervised 학습 필요 → objective term을 clean image처럼 활용하여 non-blind network 학습
- 첫 번째 논문에서 LAN이 BNN의 출력을 보고 학습했던 것과 유사한 방법이라고 생각됨



<Down shuffle과 real noise의 관계>



<첫 번째 논문의 전체 개념도>

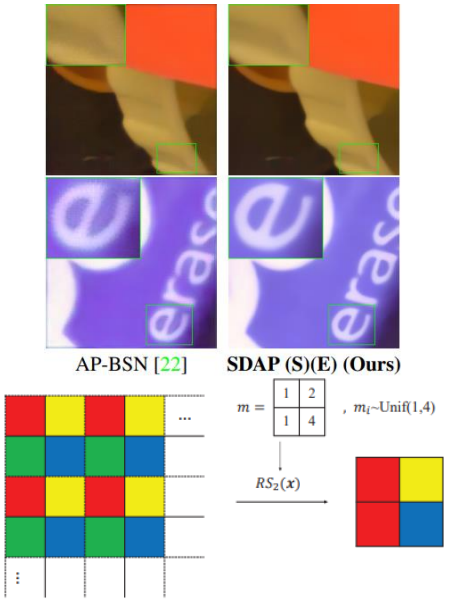
Paper Review

- Self-Supervised Image Denoising with Downsampled Invariance Loss and Conditional Blind-Spot Network
  - Method

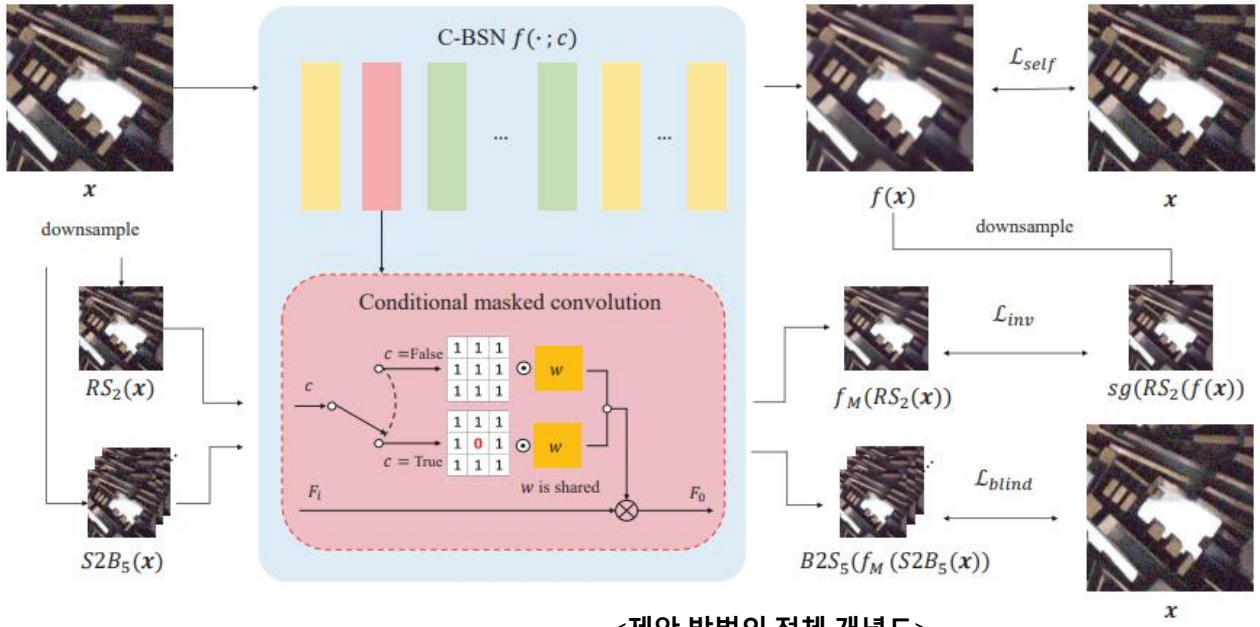
Weight는 같은데 masking만 선택적으로 되어야 함

$$\mathcal{L}_{inv} = \sqrt{\frac{s^2}{m}} \left\| d_s(f(\boldsymbol{x})) - sg(f_M(d_s(\boldsymbol{x}))) \right\|_2$$

- Downsampled Invariance loss는 blind / non-blind network가 동일한 weight를 사용해야 함
- 이를 구현하기 위해 conditional blind spot network (CBSN)을 제안
- Network의 새로운 입력 condition c를 받아 선택적으로 weight의 center를 masking할 수 있도록 설계
- 또한 down sampler로 흔하게 사용되는 space2batch (B2S)는 checkboard artifact를 만드므로, randomness를 부여하여 artifact를 완화하는 random subsampler를 제안



<제안한 random sampler>



<제안 방법의 전체 개념도>

Paper Review

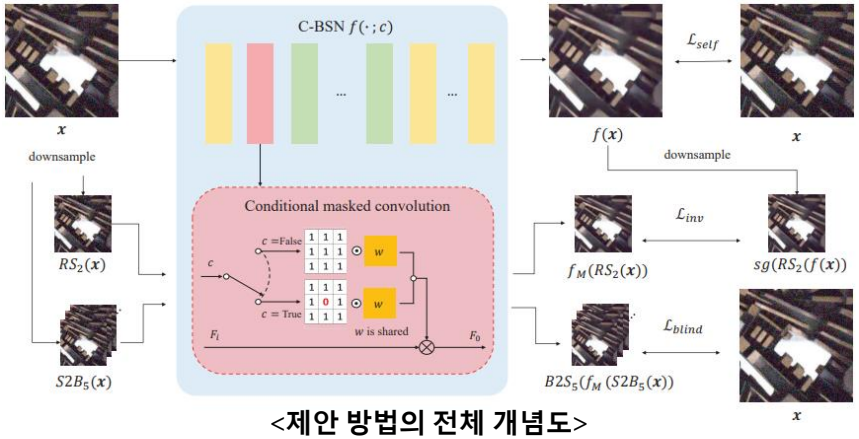
- Self-Supervised Image Denoising with Downsampled Invariance Loss and Conditional Blind-Spot Network
  - Method
    - 추가적으로 학습 초기에 안정성을 높이기 위해 blind loss를 사용하여 f_M를 학습
      - 초기에 f_M을 학습하지 않으면, objective term이 noisy해서 학습이 불안정해짐

$$\mathcal{L}_{blind} = \left\| B2S_5(f_M(S2B_5(\boldsymbol{x}))) - \boldsymbol{x} \right\|$$

    - 최종 loss function은 다음과 같음

$$\mathcal{L}_{self} = \left\| f(x) - x \right\|$$
$$\mathcal{L}_{invRS} = \left\| RS_2(f(\boldsymbol{x})) - sg(f_M(RS_2(\boldsymbol{x})) \right\|$$
$$\mathcal{L}_{CBSN} = \mathcal{L}_{self} + \lambda_{inv} \cdot \mathcal{L}_{invRS}$$
$$\mathcal{L}_{total} = \mathcal{L}_{blind} + \lambda_{sch} \cdot \mathcal{L}_{CBSN}$$

    - λ_{sch}는 0으로 시작하고 200,000 iteration 동안 점진적으로 1로 증가

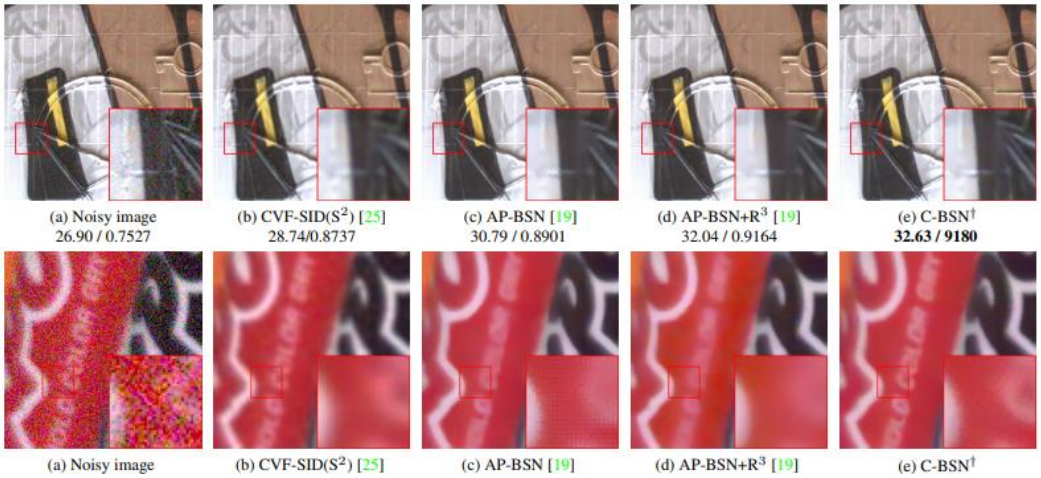


<제안 방법의 전체 개념도>

서강대학교 SOGANG UNIVERSITY
VDS LAB

Paper Review

- Self-Supervised Image Denoising with Downsampled Invariance Loss and Conditional Blind-Spot Network
  - Experiments
    - 기존 방법을 supervised / generation-based / self-supervised 방식으로 분류하여 이들과 성능 비교



Supervision	Method	SIDD		DND	
		PSNR(dB)	SSIM	PSNR(dB)	SSIM
Model-based	BM3D [7]	25.65	0.685	34.51	0.851
	WNNM [9]	25.78	0.809	34.67	0.865
Supervised	DNCNN [43]	35.13	0.896	37.89	0.932
	CBDNet [10]	33.28	0.868	38.05	0.942
	RIDNet [3]	38.70	0.950	39.24	0.952
	AINDNet (R)* [14]	38.84	0.951	39.34	0.952
	VDN [39]	39.26	0.955	39.38	0.952
	MIRNet [41]	39.72	0.959	39.88	0.956
	MAXIM-3S [33]	39.96	0.960	39.84	0.957
Generation-based	GCBD [37]	-	-	35.58	0.922
	C2N* [12] + DIDN [38]	35.35	0.937	36.38	0.887
Self-supervised	NAC [37]	-	-	36.20	0.925
	R2R [26]	34.78	0.898	-	-
	CVF-SID(T) [25]	34.43	0.912	36.31	0.923
	CVF-SID(S^2)† [25]	34.71	0.917	36.50	0.924
	AP-BSN [19]	34.90	0.900	37.46	0.924
	AP-BSN + R^3 [19]	35.97	0.925	38.09	0.937
	C-BSN	36.82	**0.934**	38.45	0.939
	C-BSN†	**36.84**	0.933	**38.60**	**0.941**

Paper Review

- Self-Supervised Image Denoising with Downsampled Invariance Loss and Conditional Blind-Spot Network
  - Experiments
    - Loss function에 대한 ablation study 진행

$$\mathcal{L}_{invRS} = \left\| RS_2(f(\boldsymbol{x})) - sg(f_M(RS_2(\boldsymbol{x})) \right\|$$

Loss function	PSNR(dB)	SSIM
$\mathcal{L}_{N2Same}$	25.58	0.807
$\mathcal{L}_{total}$ with blind-spot	35.86	0.931
$\mathcal{L}_{inv}$ with RMS	35.63	0.920
$\mathcal{L}_{total}$	**36.22**	**0.935**

$\mathcal{L}_{N2Same}$: condition c를 항상 False로 설정해서 non-blind network만 사용
$\mathcal{L}_{total}$ with blind-spot: condition c를 항상 True로 설정해서 blind network만 사용
$\mathcal{L}_{inv}$ with RMS : $\mathcal{L}_{inv}$를 기존 연구 (noise2same)를 따라서 L2 loss를 사용

    - Down sampler에 대한 ablation study 진행

downsampler	stride	PSNR(dB)	SSIM
PD	5	34.71	0.905
	2	35.32	0.914
$S2B$	5	35.62	0.924
	2	36.02	0.922
RS	5	35.24	0.922
	2	**36.22**	**0.935**

    - Blind loss 사용에 대한 ablation study 진행

$$\mathcal{L}_{total} = \mathcal{L}_{blind} + \lambda_{sch} \cdot \mathcal{L}_{CBSN}$$

λ_{sch}	PSNR(dB)	SSIM
∞	25.92	0.810
0	29.59	0.757
1	35.65	0.926
warm-up	**36.22**	**0.935**

감사합니다.