Self-supervised methods for low-level vision

Image recognition and computer vision

Victor Letzelter

January 17, 2022

Outline

The denoising problem

What is a noisy image ? *(Additive noise model)*

Decomposition in the form: $\boldsymbol{x} = \boldsymbol{s} + \boldsymbol{n}$ where :

- $\boldsymbol{s}$ is the signal, whose *closed* components are **not** statistically independent: $p(\boldsymbol{s}_i \mid \boldsymbol{s}_j) \neq p(\boldsymbol{s}_i)$
- $\boldsymbol{n}$ is the noise, conditionally pixel-wise independent given the signal $\boldsymbol{s}$: $p(\boldsymbol{n} \mid \boldsymbol{s}) = \prod_i p(\boldsymbol{n}_i \mid \boldsymbol{s}_i)$, with $\mathbb{E}[\boldsymbol{n}_i] = 0$.

The denoising problem

What is a noisy image ? *(Additive noise model)*

Decomposition in the form: $\boldsymbol{x} = \boldsymbol{s} + \boldsymbol{n}$ where :

- $\boldsymbol{s}$ is the signal, whose *closed* components are **not** statistically independent: $p(\boldsymbol{s}_i \mid \boldsymbol{s}_j) \neq p(\boldsymbol{s}_i)$
- $\boldsymbol{n}$ is the noise, conditionally pixel-wise independent given the signal $\boldsymbol{s}$: $p(\boldsymbol{n} \mid \boldsymbol{s}) = \prod_i p(\boldsymbol{n}_i \mid \boldsymbol{s}_i)$, with $\mathbb{E}[\boldsymbol{n}_i] = 0$.

Denoising performance: PSNR, in dB (The higher the best)

$$PSNR = 10 \log_{10}\left(\frac{MAX_I^2}{MSE}\right)$$

Where $MSE = \frac{1}{3mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\|I(i,j) - K(i,j)\|^2$ with I, K the clean and noisy images

MAX_I : maximum pixel value.

[Leh+18] Jaakko Lehtinen *et al.* "Noise2noise" (2018).

Framework

Data at disposal: $(\boldsymbol{s} + \boldsymbol{n}, \boldsymbol{s} + \boldsymbol{n}')$

☞ Mapping: Impossible task for the network

☞ The training will still converge to the correct solution (as $\mathbb{E}(\boldsymbol{n}) = 0$)

[Leh+18] Jaakko Lehtinen *et al.* "Noise2noise" (2018).

Framework

Data at disposal: $(\boldsymbol{s}+\boldsymbol{n}, \boldsymbol{s}+\boldsymbol{n}')$

☞ Mapping: Impossible task for the network

☞ The training will still converge to the correct solution (as $\mathbb{E}(\boldsymbol{n}) = 0$)

Remarks

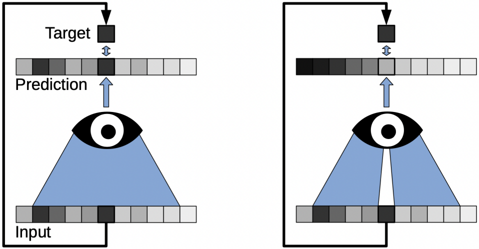
☞ Requires independent realizations of the corruption for each training image (only possible for static scenes)

[KBJ19] Alexander Krull *et al.* "Noise2void" (2019).

Strategy

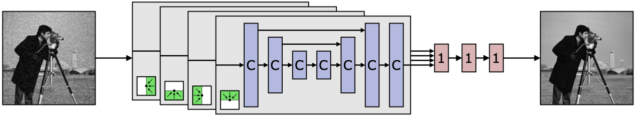
Using the CNN $f : (\boldsymbol{x}_{\mathrm{RF}(i)}; \boldsymbol{\theta}) \mapsto \hat{\boldsymbol{s}}_i$, minimize loss function

$$\arg\min_{\boldsymbol{\theta}} \sum_{i,j} L\left(f\left(\boldsymbol{x}^j_{\mathrm{RF}(i)}; \boldsymbol{\theta}\right) = \hat{\boldsymbol{s}}^j_i, \boldsymbol{s}^j_i\right)$$



Blind spot networks

[Lai+19] Samuli Laine *et al.* (2019).



New blind-spot network architecture

Bayesian Approach : $\underbrace{p\left(\boldsymbol{y} \mid \Omega_y\right)}_{\text{Training data}} = \int \underbrace{p(\boldsymbol{y} \mid \boldsymbol{x})}_{\text{Noise model}} \underbrace{p\left(\boldsymbol{x} \mid \Omega_y\right)}_{\text{Unobserved}} \mathrm{d}\boldsymbol{x}$

Approach

1. Train a NN to map $\Omega_y \mapsto (\mu_x, \Sigma_x)$, a Gaussian approximation to the prior $p(\boldsymbol{x}|\Omega_y)$.
2. At test time, first feed context Ω_y to neural network to yield μ_x and Σ_x; then compute posterior mean $\mathbb{E}_{\boldsymbol{x}}\left[p\left(\boldsymbol{x} \mid \boldsymbol{y}, \Omega_y\right)\right]$ by closed-form analytic integration.

Technical configuration

- Training dataset: 44328 images from ImageNet. Between 256x256 and 512x512 pixels in size (6.81 Go).

Technical configuration

- Training dataset: 44328 images from ImageNet. Between 256x256 and 512x512 pixels in size (6.81 Go).
- Pretrained networks from [Lai+19] used.

Technical configuration

- Training dataset: 44328 images from ImageNet. Between 256x256 and 512x512 pixels in size (6.81 Go).
- Pretrained networks from [Lai+19] used.
- Technical configuration: Code executed on Google Colab Pro with GPU T4.

Experiments: With Gaussian Noise, $\boldsymbol{n} \sim \mathcal{N}(0, \sigma^2 = 25^2)$

Noisy input: 20.37 dB

N2C: 35.242 dB

N2N: 35.20 dB

N2V (σ known): 35.237 dB

Experiments: With Poisson Noise, $\boldsymbol{x} \sim \frac{\mathcal{P}(\lambda \boldsymbol{s})}{\lambda}$ ($\lambda = 30$)

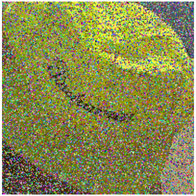
Noisy input: 19.81 dB

N2C: 30.36 dB

N2N: 30.34 dB

N2V: 30.27 dB

With Impulse Noise: $\forall i,\ x[i] \sim \mathcal{U}[0,1]^3$ w.p $\alpha = 0.5$



Noisy input: 11.85 dB

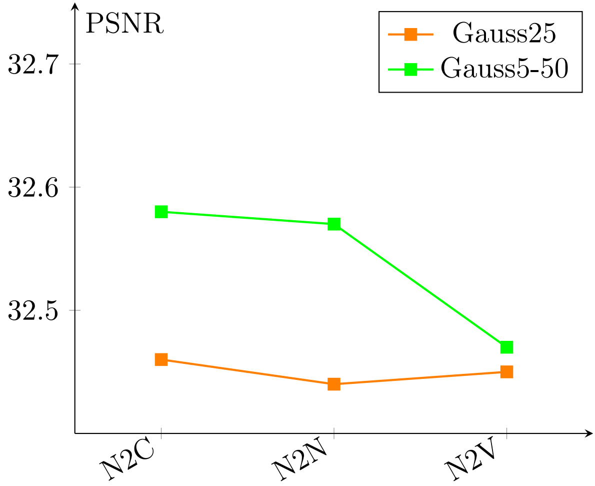
N2C: 37.87 dB

N2N: 37.55 dB

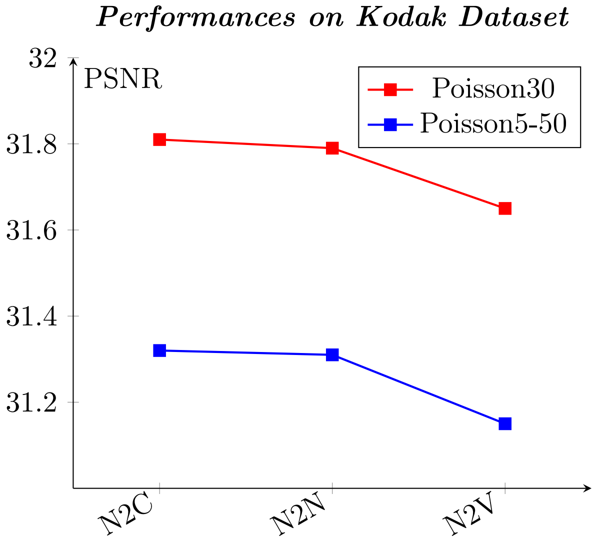
N2V: 37.90 dB

Experiments: With Gaussian Noise, $\boldsymbol{n} \sim \mathcal{N}(0, \sigma^2)$

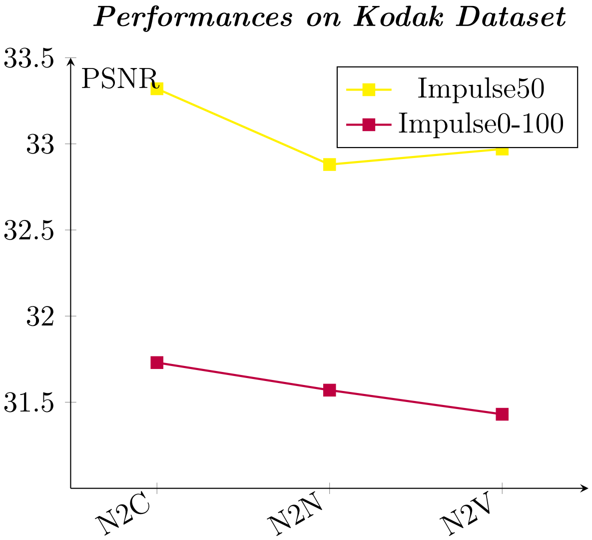
Performances on Kodak Dataset ($\simeq 2$ min ***per point***)



Experiments: With Poisson Noise, $\boldsymbol{x} \sim \frac{\mathcal{P}(\lambda \boldsymbol{s})}{\lambda}$



With Impulse color Noise $x[i] \sim \mathcal{U}[0,1]^3$ w.p α



Conclusion

- N2C produces the best results on average.

Conclusion

- N2C produces the best results on average.
- Similar quality in the self-supervised way but N2V more sensible to parameters variation.

Conclusion

- N2C produces the best results on average.
- Similar quality in the self-supervised way but N2V more sensible to parameters variation.
- N2V outperforms N2C with small training sets (See Annex).

Conclusion

- N2C produces the best results on average.
- Similar quality in the self-supervised way but N2V more sensible to parameters variation.
- N2V outperforms N2C with small training sets (See Annex).

Thanks for your attention!

Experiments: With Gaussian Noise, $\boldsymbol{n} \sim \mathcal{N}(0, \sigma^2)$

$\sigma = 25$ (8-bits unit)

Means of the PSNR				
Method	σ known ?	Kodak	BSD300	Set14
N2C	Yes	32.466	31.08	31.259
N2N	Yes	32.448	31.074	31.231
N2V	Yes	32.451	31.027	31.247

$\sigma = 50$ (8-bits unit)

Means of the PSNR				
Method	σ known ?	Kodak	BSD300	Set14
N2C	Yes	32.581	31.250	31.251
N2N	Yes	32.572	31.246	31.241
N2V	Yes	32.474	31.156	30.541

Experiments: With Poisson Noise, $\boldsymbol{x} \sim \frac{\mathcal{P}(\lambda \boldsymbol{s})}{\lambda}$

$$\lambda = 30$$

Means of the PSNR				
Method	σ known ?	Kodak	BSD300	Set14
N2C	Yes	31.806	30.401	30.451
N2N	Yes	31.795	30.392	30.442
N2V	Yes	31.653	30.249	30.290

$$5 \leq \lambda \leq 50$$

Means of the PSNR				
Method	σ known ?	Kodak	BSD300	Set14
N2C	Yes	31.322	29.897	29.966
N2N	Yes	31.312	29.891	29.966
N2V	Yes	31.152	29.742	29.824

With Impulse color Noise $x[i] \sim \mathcal{U}[0,1]^3$ w.p α

$$\alpha = 0.5$$

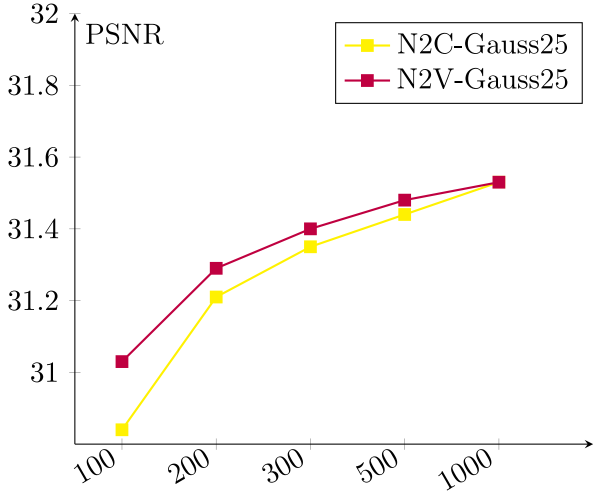
Means of the PSNR				
Method	σ known ?	Kodak	BSD300	Set14
N2C	Yes	33.318	31.194	31.447
N2N	Yes	32.877	30.847	30.961
N2V	Yes	32.978	30.772	31.070

$$0 \leq \alpha \leq 1$$

Means of the PSNR				
Method	σ known ?	Kodak	BSD300	Set14
N2C	Yes	31.725	30.388	29.645
N2N	Yes	31.566	30.262	29.368
N2V	Yes	31.431	30.156	29.335

On small training datasets: with $\boldsymbol{n} \sim \mathcal{N}(0, \sigma^2 = 25^2)$

Avg. PSNR against # of training images (from [Lai+19])



Bilbiography

[Leh+18] Jaakko Lehtinen et al. "Noise2noise: Learning image restoration without clean data". In: *arXiv preprint arXiv:1803.04189* (2018).

[KBJ19] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. "Noise2void-learning denoising from single noisy images". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 2129–2137.

[Lai+19] Samuli Laine et al. "High-quality self-supervised deep image denoising". In: *Advances in Neural Information Processing Systems* 32 (2019), pp. 6970–6980.