SS-TTA: Test-Time Adaption for Self-Supervised Denoising Methods

Masud An-Nur Islam Fahim
University of Vaasa
Vaasa, Finland
masud.fahim@uwasa.fi

Jani Boutellier
University of Vaasa
Vaasa, Finland
jani.boutellier@uwasa.fi

Abstract

Even though image denoising has already been studied for decades, recent progress in deep learning has provided novel and considerably better results for this classical signal reconstruction problem. One of the most significant advances in recent years has been relaxing the requirement of having noise-free (clean) images in the training dataset. By leveraging self-supervised learning, recent methods already reach the reconstruction quality of classical and some supervised schemes. In this paper, we propose SS-TTA, a generic test-time adaptation policy that can be applied on top of various self-supervised denoising methods. Taking a pre-trained self-supervised denoising model and a test image as input, our SS-TTA algorithm improves the denoising performance through a proposed 'inference-guided regularization' process. Based on experiments with three synthetic and three real noise datasets, SS-TTA improves the denoising results of several state-of-the-art self-supervised methods, outperforms recent test-time adaptation approaches, and shows promising performance with supervised models. Finally, SS-TTA also generalizes to cases where the test-time noise distribution differs from the noise distribution of training images.

1. Introduction

Signal reconstruction from noisy observations is a research topic of continuous interest due to its numerous areas of application. One particularly important application area is image denoising, where input images are corrupted by noise, and a denoising algorithm attempts to recover the underlying clean image. Classic denoising studies [6, 9, 39] propose optimization-based solutions to recover clean images, and are successful to some extent. Optimization based approaches were later replaced by data driven approaches [31, 34, 42] that provided better denoising quality, but unfortunately require ground truth (clean images) – a problem that has been recently overcome by self-supervised denoising [20, 21, 26, 30].

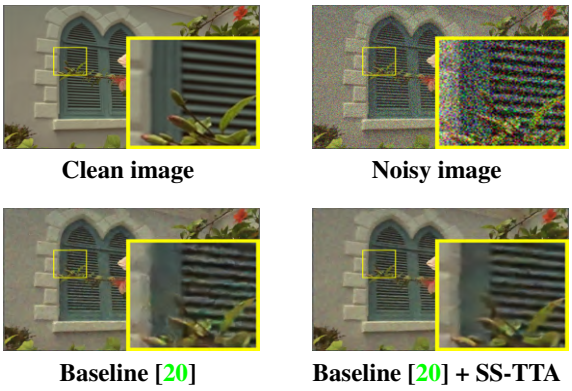

Figure 1. SS-TTA denoising example. The noisy image is the clean image corrupted by additive Gaussian noise with $\sigma^2 = 50$. Bottom row: denoising performance of Noise2Noise [20] with and without the proposed test-time adaptation approach SS-TTA.

Recent self-supervised denoising studies [3, 16, 18, 20, 21, 26, 30, 37] have achieved results that come close to the image denoising performance of fully supervised approaches [7, 31, 34, 40, 42]. Similarly, a few self-supervised studies have reported comparable performance by just training on single images [30], transfer learning [43], or real noise generation [28]. In practice, the studies mentioned above address the self-supervised learning problem by means of data augmentation [26, 38] or tailored regularization techniques [16, 37].

Even though the reported denoising performance of these self-supervised approaches is impressive, their best performance can only be observed in cases where the noise distribution of test images matches the noise characteristics of the training images (i.e. in-domain noise). In contrast, for test images with out-of-domain noise, their performance degrades [25], which implies shortcomings in their generalizability.

In general, self-supervised denoising methods [26, 29] rely on pairs of noisy images, instead of the noisy-clean image pairs used by fully supervised denoising methods.

In self-supervised training, the noise characteristics of the training set are a critical factor for obtaining good denoising performance, i.e., high noise intensity limits the resulting denoising performance, as first observed in [3]. In contrast, lower noise intensity in the training set can be expected to lead to better performance in self-supervised training. Following this rationale, we propose a denoising approach that synthesizes additional low-intensity noise training images *at test time* to adapt a pre-trained denoising model towards improved image restoration performance.

Based on this hypothesis, this work proposes a general test-time adaptation (TTA) method that can be applied on top of existing self-supervised denoising approaches. The proposed SS-TTA scheme takes as input an initial denoising model, and an image that has been denoised by the initial model. Consequently, our SS-TTA scheme creates two noise-added test-time training images from the denoised image, and adapts the initial denoising model towards improved performance considering the input image. The proposed SS-TTA method can be applied on top of a chosen neural network based denoising model (See Figure 1), and converges at test time within a few training epochs, improving the denoising performance of both synthetic and real noise images. In summary, the contributions of the proposed work are:

- To our knowledge, SS-TTA is the first study independent of additional trainable parameters or manual representation tuning within test-time adaption of denoising.
- SS-TTA is the first test-time adaptation scheme to report improved denoising performance for images with real-world, in- and out-of-domain synthetic noise.
- Compared to recent test-time adaptation approaches, SS-TTA requires a magnitude or two less iterations to convergence and outperforms existing approaches in denoising performance
- In the experimental section, results are presented for three synthetic and three real-world noise datasets, and for two different baseline neural architectures.

2. Related work

Most of the recent denoising studies [2, 31, 34, 42] rely on supervised learning, and present among others, cascaded network design [2, 31], feature ensembles [7], ensemble learning [40], custom loss functions [8, 24, 44], and application area specific data augmentation [23]. Although capable of providing superior performance, supervised methods require large quantities of paired training data — unfortunately, large-scale datasets with real-world noisy-clean image pairs are scarce. Self-supervised denoising methods [3, 16, 18, 20, 21, 26, 30, 37] relax the requirement of having clean, noise-free images, and leverage a form of noise augmentation [26, 29, 38, 43] or spatial regularization [3, 4, 16, 17, 30, 37] for training.

Unpaired denoising studies [11, 15, 36] generally consider two datasets for noise removal, where one dataset contains pairs (the guiding dataset), and the other one is unpaired (the denoising dataset). In order to provide good performance, both datasets should however exhibit similar noise characteristics. Unfortunately, unpaired denoising faces significant challenges [28] with datasets that consist of images corrupted by real-world noise.

There are also several studies that propose test time adaptation [10, 25, 33, 35, 43] for denoising. One of these previous works [33] presented a solution based on Stein's unbiased risk estimator (SURE) loss function, in order to adapt the model's parameter space to the test data. In contrast, [43] proposed noise resampling for retraining to improve the performance of the baseline model. However, [43] requires multiple phases of training to improve baseline performance. GainTuning [25] addresses the test time adaptation problem by injecting a gain parameter to each channel of the given network, and tuning the parameter for the given test image. Compared to previous works, GainTuning [25] achieves better performance and model generalizability. Finally, Lidia [35] has been built on a lightweight CNN, and relies on patch processing, non-local self-similarity and representation sparsity. Whereas Lidia presents high image restoration quality, the method relies on a specific neural architecture.

In contrast to the aforementioned works, the proposed SS-TTA method can be applied on top of any self-supervised denoising approach, improving performance across synthetic and real noise dataset, and provides state-of-the-art results.

3. Proposed method

Before describing the proposed SS-TTA method, the theoretical background of self-supervised denoising is briefly presented. The necessary mathematical notation is explained in Table 1.

3.1. Self-supervised denoising

Let $\mathcal{F}_{\boldsymbol{\theta}}$ represent the baseline denoising model, $(\mathcal{Y}, \mathcal{X})$ the dataset with noisy-clean pairs, where a single pair $(\boldsymbol{y}, \boldsymbol{x}) \in (\mathcal{Y}, \mathcal{X})$, and $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{n}$, where additive noise is $\boldsymbol{n} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma^2})$. Following any standard training strategy and an appropriate set of hyperparameters, the supervised cost function $\|\mathcal{F}_\theta(\boldsymbol{y}) - \boldsymbol{x}\|_2^2$ for the denoising task can be minimized, and consequently the initial parameter space $\mathcal{F}_{\boldsymbol{\theta}}$ will converge to a local or global minima $\mathcal{F}_{\boldsymbol{\theta}^*}$.

If the same training procedure is applied to a self-supervised setting, the cost function will become $\|\mathcal{F}_\theta(\boldsymbol{y}) - \boldsymbol{y}\|_2^2$, where convergence to identity needs to be

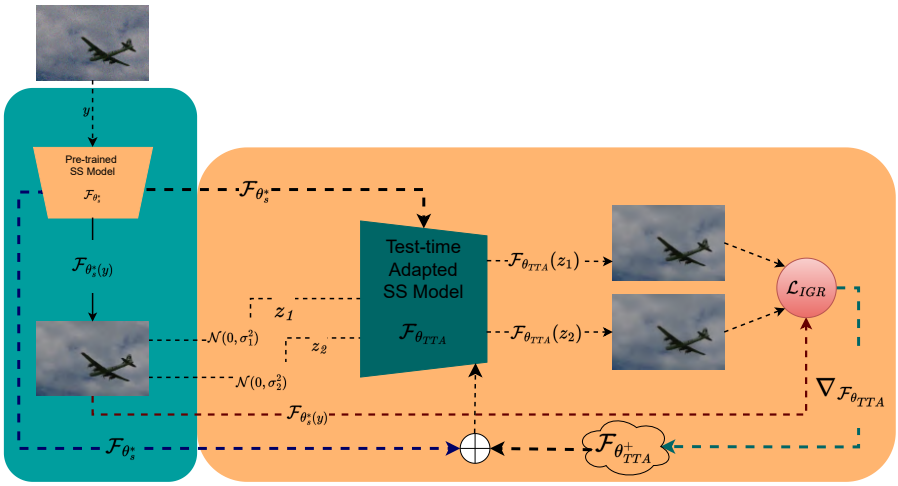

Figure 2. Flowchart of the proposed test-time adaptation scheme. Here, $\mathcal{F}_{\theta_s^*}$ is the pretrained self-supervised (SS) model, $\mathcal{F}_{\theta_{TTA}}$ is the model for test-time adaptation, $\mathcal{F}_{\theta^+_{TTA}}$ is the model after gradient update, and $\nabla_{\mathcal{F}_{\theta_{TTA}}}$ is the gradient w.r.t. $\mathcal{F}_{\theta_{TTA}}$.

Notation	Description
$\boldsymbol{x}$	Clean image
$\boldsymbol{y}$	Noisy image
$\boldsymbol{n} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma^2})$	Additive noise
$\boldsymbol{\sigma^2}$	Noise variance
$(\boldsymbol{y}, \boldsymbol{x})$	Noisy-Clean pair
$\mathbb{E}$	Expectation
$(\boldsymbol{y}, \boldsymbol{y'})$	Noisy-Noisy pair
$\mathcal{F}_{\boldsymbol{\theta}}(\boldsymbol{y})$	Output from the pre-trained model
$\mathcal{F}_{\theta_I}$	Ideal denoiser
$z = \mathcal{F}_{\boldsymbol{\theta}}(\boldsymbol{y}) + \boldsymbol{n}$	Noisy inferred image
$\boldsymbol{x} - \mathcal{F}_{\boldsymbol{\theta}}(\boldsymbol{x})$	Details of the clean image
$\mathcal{F}_{\boldsymbol{\theta}^*}$	Pre-trained supervised model
$\mathcal{F}_{\boldsymbol{\theta}_s^*}$	Pre-trained self-supervised model
$\mathcal{F}_{\boldsymbol{\theta}_{TTA}}$	Initial test-time adapted model
$\nabla_{\mathcal{F}_{\boldsymbol{\theta}_{TTA}}}$	Gradient w.r.t. $\mathcal{F}_{\boldsymbol{\theta}_{TTA}}$
$\mathcal{F}_{\boldsymbol{\theta}^+_{TTA}}$	Model after GD update
$\mathcal{F}_{\boldsymbol{\theta}^{avg}_{TTA}}$	Average of $\mathcal{F}_{\boldsymbol{\theta}^+_{TTA}}$ and $\mathcal{F}_{\boldsymbol{\theta}_s^*}$
$\mathcal{F}_{\boldsymbol{\theta}^*_{TTA}}$	Final test-time adapted model
$\mathcal{L}_{MSE}$	Mean-square loss function

Table 1. Descriptions of the necessary notations.

avoided for enabling minimization. For example, the well-known work Noise2Noise [20] avoids converging to identity by leveraging $\boldsymbol{y'} = \boldsymbol{x} + \boldsymbol{n'}$, which is a secondary noisy observation of the clean image (here, $\boldsymbol{n'} \sim \boldsymbol{\mathcal{N}(0, \sigma_1^2)}$). Hence, the minimization task becomes $\|\mathcal{F}_\theta(\boldsymbol{y}) - \boldsymbol{y'}\|_2^2$, and following the presentation in [29], it can be expanded as follows:

$$
\begin{aligned}
&\mathbb{E}_{\boldsymbol{n},\boldsymbol{n'}}\left\{\left\|\mathcal{F}_\theta(\boldsymbol{y}) - \boldsymbol{y'}\right\|_2^2\right\} \\
=&\mathbb{E}_{\boldsymbol{n},\boldsymbol{n'}}\left\{\left\|\mathcal{F}_\theta(\boldsymbol{y}) - \boldsymbol{x} - \boldsymbol{n'}\right\|_2^2\right\} \\
=&\mathbb{E}_{\boldsymbol{n},\boldsymbol{n'}}\left\{\|\mathcal{F}_\theta(\boldsymbol{y}) - \boldsymbol{x}\|_2^2 - 2\left(\boldsymbol{n'}\right)^\top\left(\mathcal{F}_{\boldsymbol{\theta}}(\boldsymbol{y}) - \boldsymbol{x}\right) + \left(\boldsymbol{n'}\right)^\top \boldsymbol{n'}\right\} \\
=&\mathbb{E}_{\boldsymbol{n},\boldsymbol{n'}}\left\{\|\mathcal{F}_\theta(\boldsymbol{y}) - \boldsymbol{x}\|_2^2\right\} + \text{ constant.}
\end{aligned}
$$

The $-2\mathbb{E}_{\boldsymbol{n},\boldsymbol{n'}}\left\{\left(\boldsymbol{n'}\right)^\top \mathcal{F}_{\boldsymbol{\theta}}(\boldsymbol{y})\right\}$ component is excluded from the last line as it will converge to zero due to the orthogonality of the noise components, and $\left(\boldsymbol{n'}\right)^\top \boldsymbol{n'}$ is a constant, which depends on the noise distribution of the input $\boldsymbol{y}$ and target images $\boldsymbol{y'}$. This constant term $\left(\boldsymbol{n'}\right)^\top \boldsymbol{n'}$ causes the self-supervised model and supervised model to reach different minima and potentially differing performance. There are several previous studies [3, 13, 19, 29] that point out this observation, and Appendix 1 shows a brief experimental analysis of the significance of constant term $\left(\boldsymbol{n'}\right)^\top \boldsymbol{n'}$.

This performance gap motivates proposing our solution for improving the denoising performance of $\mathcal{F}_{\boldsymbol{\theta}_s^*}$. In contrast to recent test-time adaption works [25, 35], the proposed approach adjusts all the parameters of $\mathcal{F}_{\boldsymbol{\theta}_s^*}$ to improve denoising quality. The following section explains the technical details of the SS-TTA algorithm.

3.2. SS-TTA

Figure 2 shows the generic procedure of the proposed test time adaptation algorithm. In the first step of test time adaptation, test images $\boldsymbol{Y}$ with random (zero mean or not, underlying noise variance unknown) additive noise are acquired. For a noisy test image $\boldsymbol{y} \in \boldsymbol{Y}$, and the given pre-trained self-supervised model $\mathcal{F}_{\boldsymbol{\theta}_s^*}$, $\mathcal{F}_{\boldsymbol{\theta}_s^*}(\boldsymbol{y})$ is the denoised approximation of the unknown clean image $\boldsymbol{x}$.

To enable SS-TTA test time adaptation, we sample additional noisy observations of $\mathcal{F}_{\boldsymbol{\theta}_s^*}(\boldsymbol{y})$ by adding synthetic noise to it. The resulting additional noisy images $\boldsymbol{z_1}, \boldsymbol{z_2}$ are acquired from $\mathcal{F}_{\boldsymbol{\theta}_s^*}(\boldsymbol{y})$ such that $\boldsymbol{z_1} = \mathcal{F}_{\boldsymbol{\theta}_s^*}(\boldsymbol{y}) + \boldsymbol{n_1}$, $\boldsymbol{z_2} = \mathcal{F}_{\boldsymbol{\theta}_s^*}(\boldsymbol{y}) + \boldsymbol{n_2}$, where $\boldsymbol{n_1} \sim \boldsymbol{\mathcal{N}(0, \sigma_1^2)}$, $\boldsymbol{n_2} \sim \boldsymbol{\mathcal{N}(0, \sigma_2^2)}$, and $\boldsymbol{\sigma_1^2 < \sigma_2^2}$. Initialized by the pre-trained self-supervised model $\mathcal{F}_{\boldsymbol{\theta}_s^*}$ and

$\boldsymbol{z_1}, \boldsymbol{z_2}$, the initial test-time adapted model $\mathcal{F}_{\theta_{TTA}}$ provides $\mathcal{F}_{\theta_{TTA}}(\boldsymbol{z_1})$, and $\mathcal{F}_{\theta_{TTA}}(\boldsymbol{z_2})$, and the test-time model updating can start. For this, two variants of the training loss are proposed:

SS-TTA$_1$. This general-purpose variant regularizes the pre-trained self-supervised model through a combination of mean-square error (MSE) loss terms. The first loss component is $\mathcal{L}_{s1} = \left\|\mathcal{F}_{\theta_s^*}(\boldsymbol{y}) - \mathcal{F}_{\theta_{TTA}}(\boldsymbol{z_1})\right\|_2^2$, which is a relaxed form of supervised minimization, as we consider $\mathcal{F}_{\theta_s^*}(\boldsymbol{y})$ the quasi-clean target. The second loss $\mathcal{L}_{s2}$ regulates the $\mathcal{F}_{\theta_{TTA}}$ by $\mathcal{L}_{s2} = \|\mathcal{F}_{\theta_{TTA}}(\boldsymbol{z_1}) - \mathcal{F}_{\theta_{TTA}}(\boldsymbol{z_2})\|_2^2$. Here, direct minimization with regard to the quasi-clean target $\mathcal{F}_{\theta_s^*}(\boldsymbol{y})$ would result in over-smoothed images, as empirically observed, and is hence replaced by $\mathcal{F}_{\theta_{TTA}}(\boldsymbol{z_1})$.

In order to properly explain the last loss component of SS-TTA, we briefly need to consider an *ideal denoiser* $\mathcal{F}_{\theta_I}$, which can exactly reconstruct clean images from noisy samples: if $\mathcal{F}_{\theta_I}$ is provided clean (noise-free) input, the observed output should be identical to the clean input, i.e., $\|\mathcal{F}_{\theta_I}(\boldsymbol{y}) - \mathcal{F}_{\theta_I}(\mathcal{F}_{\theta_I}(\boldsymbol{y}))\|_2^2 \equiv 0$. However, such identity mapping is missing from current state-of-the-art self-supervised architectures, and to mitigate the effects of missing identity mapping, the *identity mapping loss* $\mathcal{L}_3$ is added to the SS-TTA method: $\mathcal{L}_3 = \left\|\mathcal{F}_{\theta_s^*}(\boldsymbol{y}) - \mathcal{F}_{\theta_{TTA}}(\mathcal{F}_{\theta_s^*}(\boldsymbol{y}))\right\|_2^2$. In summary, the total loss of our first variant becomes

$$\mathcal{L}_{IGR1} = \mathcal{L}_{s1} + \mathcal{L}_{s2} + \mathcal{L}_3 \tag{1}$$

where IGR stands for *inference-guided loss*.

SS-TTA$_2$. It is well-known that besides reducing undesired noise, image denoising smoothenes the input image, removing small details. In order to address this undesired reduction of image sharpness, the detail-preserving SS-TTA$_2$ variant of our method is proposed; considering a clean image $\boldsymbol{x}$, the image details (high-frequency components) of $\boldsymbol{x}$ can be expressed as $\mathcal{I}_{HF} = \boldsymbol{x} - \mathcal{F}_{\theta_s^*}(\boldsymbol{x})$, where $\mathcal{F}_{\theta_s^*}$ can be understood to function like a low-pass filter. Similarly, in the case of self-supervised denoising, $\mathcal{I}_{HF} = \mathcal{F}_{\theta_s^*}(\boldsymbol{y}) - \mathcal{F}_{\theta_s^*}(\mathcal{F}_{\theta_s^*}(\boldsymbol{y}))$, where $\mathcal{F}_{\theta_s^*}(\boldsymbol{y})$ can be understood as a reasonable-quality reconstruction of $\boldsymbol{x}$, and $\mathcal{F}_{\theta_s^*}(\mathcal{F}_{\theta_s^*}(\boldsymbol{y}))$ as a further-smoothened version of $\mathcal{F}_{\theta_s^*}(\boldsymbol{y})$ (through re-applying $\mathcal{F}_{\theta_s^*}$). In the case of SS-TTA$_2$, the noisy input $\boldsymbol{y}$ is used for estimating $\mathcal{I}_{HF}$.

With this background, SS-TTA$_2$ computes adapted image details in every epoch as follows: $\mathcal{I}_{HF_a} = \mathcal{F}_{\theta_s^*}(\boldsymbol{y}) - \mathcal{F}_{\theta_{TTA}}(\mathcal{F}_{\theta_s^*}(\boldsymbol{y}))$. By $\mathcal{I}_{HF_a}$, the target variable $\mathcal{F}_{\theta_s^*}(\boldsymbol{y})$ is sharpened, and guided by image details, the loss terms become $\mathcal{L}_{D1} = \left\|\mathcal{F}_{\theta_s^*}(\boldsymbol{y}) + \mathcal{I}_{HF_a} - \mathcal{F}_{\theta_{TTA}}(\boldsymbol{z_1})\right\|_2^2$, and $\mathcal{L}_{D2} = \left\|\mathcal{F}_{\theta_s^*}(\boldsymbol{y}) + \mathcal{I}_{HF_a} - \mathcal{F}_{\theta_{TTA}}(\boldsymbol{z_2})\right\|_2^2$. While keeping the identity mapping term identical to SS-TTA$_1$, the total loss of SS-TTA$_2$ becomes:

$$\mathcal{L}_{IGR2} = \mathcal{L}_{D1} + \mathcal{L}_{D2} + \mathcal{L}_3 \tag{2}$$

SS-TTA model update. Using the total loss $\mathcal{L}_{IGR1}$ or $\mathcal{L}_{IGR2}$, the obtained gradient is labeled as $\nabla_{\mathcal{F}_{\theta_{TTA}}}$, and after gradient descent (GD), the updated model becomes $\mathcal{F}_{\theta_{TTA}^+}$. To perform the final GD, weight averaging is applied for improved model generalization [14, 22]: by averaging $\mathcal{F}_{\theta_{TTA}^+}$ and $\mathcal{F}_{\theta_s^*}$ we compute $\mathcal{F}_{\theta_{TTA}^{avg}}$ and update the current state of $\mathcal{F}_{\theta_{TTA}}$. The total overall procedure is repeated for several epochs to obtain the final test-time adapted model $\mathcal{F}_{\theta_{TTA}^*}$.

Rationale of SS-TTA. As discussed earlier, in contrast to supervised approaches, self-supervised methods tune the input model with regard to the noisy target. Additionally, depending upon the noisy input images, the performance of the self-supervised studies is limited by the constant $(\boldsymbol{n'})^\top \boldsymbol{n'}$ term that is not present in a supervised setup. *Hence, self-supervised denoising methods are tunable after the regular training phase* with a possibility of performance improvement. This does not fully generalize to the supervised case, as there is no theoretical performance limit apart from the network topology and training strategy.

In SS-TTA, we have an identical training setup (not more than five epochs, no annealing factor) for both SS-TTA$_1$ and SS-TTA$_2$, details described later.

4. Experiments and results

In this section, the proposed SS-TTA method is evaluated using several datasets, noise domains and baseline self-supervised denoising methods: Noise2Noise [20], Noisier2Noise [26] and Decay2Distill [4]. Furthermore, we compare the performance of SS-TTA to recent test-time adaptation approaches GainTuning [25] and Lidia [35]. In Tables 2, 3, 4, 5, 6 results from baseline methods [4, 20, 26] without test-time adaptation are shown in regular font, and baseline+SS-TTA results in boldface. Figure 4 shows visual examples for the approaches [4, 20, 26] with and without applying SS-TTA.

Datasets: SS-TTA is evaluated using synthetic and real noise datasets. For synthetic noise, BSD68 [42], Kodak24, and Urban100 [12] datasets are used with input noise variances of 25 and 50. To explore the performance limits of SS-TTA, we also experimented with noisy images with a high noise variance of 90. In the real noise domain, SIDD [1], PolyU [41], and CC [27] datasets were used. For SIDD, the validation set was used for our study (collected from the SIDD hosting website), whereas for CC and PolyU datasets, only the test sets were used.

Denoising methods: We have used pre-trained weights from the Noise2Noise [20], Noisier2Noise [26], and Decay2Distill [4] methods, built on the DnCNN [42] network.

Methods	BSD68 [42]	Kodak24	Urban100 [12]
N2N [20]	30.14	30.97	29.23
N2N+ SS-TTA$_1$	30.19	31.07	29.26
N2N+SS-TTA$_2$	**30.28**	**31.09**	**29.36**
Nr2N [26]	28.54	29.02	27.9
Nr2N+ SS-TTA$_1$	28.72	29.43	28.01
Nr2N+SS-TTA$_2$	**28.82**	**29.63**	**28.10**
D2D [4]	30.26	31.02	29.15
D2D+ SS-TTA$_1$	30.32	31.07	29.22
D2D+SS-TTA$_2$	**30.49**	**31.26**	**29.41**

Table 2. Average denoising performance on BSD68, Kodak24, and Urban100 datasets with synthetic in-domain $\sigma^2 = 25$ noise. Units are in PSNR and the model architecture is DnCNN [42].

The networks of [4, 20, 26] were trained [42] with images containing synthetic noise, both for real noise and synthetic noise datasets. Rotation and translation process was used along with cosine learning rate decay to train [42] with [4, 20, 26]. Synthetic train set was sampled from the DIV2K dataset. For images with real-world noise, self-supervised works rely on a variety of training strategies, such as single-image training [29], fine tuning [13], or regular training [28]. Unfortunately, the exact training procedures or pre-trained models for real noise cases are not publicly available.

Test-time adaptation details: For test time adaptation, the noisy test image, the pre-trained initial model $\mathcal{F}_{\theta_s^*}$, and adaption noise variance parameters σ_1^2, σ_2^2 are required. For synthetic noise removal experiments, it was observed that $\sigma_1^2 = 25, \sigma_2^2 = 30$ were appropriate values for successful denoising. These noise parameters were also used with the real-world noise SIDD dataset, however, for the CC dataset we used $\sigma_1^2 = 1$ and $\sigma_2^2 = 1$, and for the PolyU dataset $\sigma_1^2 = 5$ and $\sigma_2^2 = 6$.

The number of SS-TTA test time training epochs was kept at 5 across all experiments, using the Adam optimizer with a learning rate of 0.0001, without any scheduler. Increasing the epoch count resulted only in marginal denoising performance improvements. For test-time adaption parameter aggregation, we have used the convex sum policy and $\lambda_1 = 0.9, \lambda_2 = 0.1$ for $\mathcal{F}_{\theta_s^*}$ and $\mathcal{F}_{\theta_{TTA}^+}$, respectively.

4.1. SS-TTA denoising performance

Table 2 shows the effect of SS-TTA test time adaptation when applied on top of three recent self-supervised denoising methods in the case of synthetic noise with $\sigma^2 = 25$: SS-TTA surpasses the improves the performance of all three baseline methods, and SS-TTA$_2$ outperforms SS-TTA$_1$ in each case. Similarly, Table 3 shows the results for synthetic noise with $\sigma^2 = 50$, which reveals that in the presence of stronger noise, the noise removal performance SS-TTA is likewise higher than in the $\sigma^2 = 25$ case.

Methods	BSD68 [42]	Kodak24	Urban100 [12]
N2N [20]	26.63	27.51	25.55
N2N+ SS-TTA$_1$	26.72	27.69	25.63
N2N+SS-TTA$_2$	**26.84**	**27.80**	**25.71**
Nr2N [26]	25.60	26.35	24.56
Nr2N+ SS-TTA$_1$	25.85	26.80	24.76
Nr2N+SS-TTA$_2$	**26.04**	**27.02**	**24.91**
D2D [4]	26.53	27.48	25.49
D2D+ SS-TTA$_1$	26.69	27.72	25.64
D2D+SS-TTA$_2$	**26.90**	**27.94**	**25.84**

Table 3. Average denoising performance on BSD68, Kodak24, and Urban100 datasets with synthetic in-domain $\sigma^2 = 50$ noise. Units are in PSNR and the model architecture is DnCNN [42].

Methods	BSD68 [42]	Kodak24	Urban100 [12]
N2N [20]	21.33	21.91	20.56
N2N+ SS-TTA$_1$	21.94	22.73	20.88
N2N+SS-TTA$_2$	**22.28**	**23.18**	**21.07**
Nr2N [26]	20.30	20.75	19.55
Nr2N+ SS-TTA$_1$	20.82	21.55	19.85
Nr2N+SS-TTA$_2$	**21.17**	**22.02**	**20.06**
D2D [4]	19.17	19.33	19.01
D2D+ SS-TTA$_1$	20.25	21.02	19.74
D2D+SS-TTA$_2$	**21.02**	**22.06**	**20.27**

Table 4. Average denoising performance on BSD68, Kodak24, and Urban100 datasets, synthetic out-of-domain $\sigma^2 = 90$ noise. Units are in PSNR and the model architecture is DnCNN [42].

To assess the generalizability of the SS-TTA method, the performance of the model with $\sigma_1^2 = 25, \sigma_2^2 = 30$ was also tried in denoising of $\sigma^2 = 90$ noisy input images. The results of Table 4 show that the baseline methods perform denoising much worse than with $\sigma^2 = 25$ or $\sigma^2 = 50$ noise, but SS-TTA improves the PSNR performance of each denoiser by around 1 dB.

Methods	CC	PolyU [41]	SIDD [1]
N2N [20]	34.39	35.29	27.95
N2N+ SS-TTA$_1$	35.15	36.54	30.01
N2N+SS-TTA$_2$	**35.17**	**36.85**	**30.22**
Nr2N [26]	34.70	35.23	28.53
Nr2N+ SS-TTA$_1$	35.52	36.10	29.49
Nr2N+SS-TTA$_2$	**35.68**	**36.56**	**29.84**
D2D [4]	33.98	36.27	28.41
D2D+ SS-TTA$_1$	34.32	36.50	29.02
D2D+SS-TTA$_2$	**34.55**	**36.33**	**29.15**

Table 5. Average denoising performance on SIDD, PolyU, and CC datasets, with real-world noise. Units are in PSNR and the model architecture is DnCNN [42].

For real-world noisy images, three datasets: CC, PolyU, and SIDD, were considered. The same baseline methods [4, 20, 26] as with the synthetic noise cases were used for real-world noise as well. From Table 5, the performance

Methods	BSD68
N2N [20]	29.48
N2N+SS-TTA$_1$	29.58
N2N+SS-TTA$_2$	**29.66**
Nr2N [26]	27.60
Nr2N+SS-TTA$_1$	27.89
Nr2N+SS-TTA$_2$	**28.01**
D2D [4]	29.44
D2D+SS-TTA$_1$	29.47
D2D+SS-TTA$_2$	**28.01**

Table 6. Performance increase provided by SS-TTA for $\lambda \in [5, 50]$ Poisson distribution noise for the BSD68 dataset.

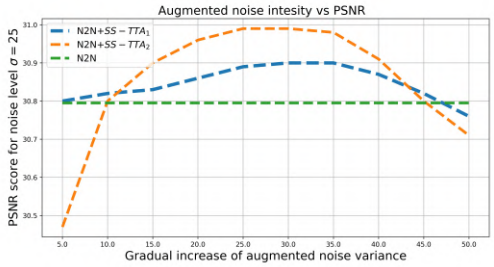

Figure 3. The effect of σ_1^2, σ_2^2 on test-time adaptation. The x-axis shows the value of the SS-TTA σ_1^2, σ_2^2 for the parameters, whereas the y-axis shows the obtained PSNR score. The noise variance for the input image is $\sigma^2 = 25$. The baseline model is Noise2Noise [20]. Note that, $\sigma_2^2 > \sigma_1^2$ and $\sigma_2^2 = \sigma_1^2 + 5$.

increase over the baseline studies [4, 20, 26] by SS-TTA is shown, and it is evident that SS-TTA works equally well across synthetic and real-world noise.

A short experiment on Poisson distribution noise removal was also performed. As shown in Table 6, our test-time adaptation algorithm generalizes well for Poisson noise when applied on top of several baseline algorithms [4, 20, 26].

4.2. Comparison with other TTA methods

For comparing the proposed SS-TTA method with other test-time adaptation methods for denoising, we are required to rely upon the comparison table provided in [25], since the pre-trained models of related methods GainTuning [25] and Lidia [35] are not publicly available. Table 7 shows the results for the BSD68 dataset with $\sigma^2 = 30, 40, 50$, and shows a comparison between SS-TTA, GainTuning [25] and Lidia [35]. As shown in the table, SS-TTA outperforms [25] and [35] in each case.

4.3. The effect of noise variance

In the previous denoising experiments, the σ_1^2 and σ_2^2 parameters were kept at fixed values, as given early in Section 4. Here, we show the denoising effect of adjusting

Method	Inference	$\sigma^2 = 30$	$\sigma^2 = 40$	$\sigma^2 = 50$
GainTuning [25]	Pre-trained	28.39	27.16	26.27
	Adapted	**28.47**	**27.23**	**26.33**
Lidia [35]	Pre-trained	28.24	26.91	25.74
	Adapted	**28.23**	**26.97**	**26.02**
SS-TTA$_1$+ [20]+ [42]	Pre-trained	29.30	27.87	26.63
	Adapted	**29.42**	**28.00**	**26.84**
SS-TTA$_1$+ [20]+ [32]	Pre-trained	30.64	28.46	28.22
	Adapted	**30.91**	**28.78**	**28.44**

Table 7. SS-TTA performance compared with other test-time adaptation methods in the synthetic noise domain for the BSD68 dataset. Here, SS-TTA is applied on pre-trained DnCNN [42] and UNet [32] using Noise2Noise [20]. The input noise variances are $\sigma^2 = 30, 40, 50$. The test-time adapted versions are in boldface for all methods.

Method	Inference	$\sigma^2 = 25$	$\sigma^2 = 50$	$\sigma^2 = 90$
Nr2N [26] + SS-TTA$_1$	with $\mathcal{L}_3$	**30.85**	**27.30**	**21.57**
	without $\mathcal{L}_3$	30.78	27.24	21.52

Table 8. SS-TTA$_1$ ablation with and without the $\mathcal{L}_3$ loss for the 5set dataset. Here, pre-trained model is [26] and input image noise variances are ($\sigma^2 = 25, 50, 90$).

σ_1^2 and σ_2^2 for $\mathcal{F}_{\theta_s^*}$, which is crucial to the SS-TTA performance. Figure 3 shows the noise variance $\sigma_1^2 = 25$ versus PSNR performance for SS-TTA$_1$ applied on top of the Noise2Noise method. Here, the noisy input images have noise variance $\sigma^2 = 25$, and evidently, the best PSNR score is achieved when augmented noise variance is also $\sigma_1^2 = 25$. Similarly, Figure 5 presents an experiment for input image noise variance $\sigma^2 = 50$ for SS-TTA with the underlying baseline method Decay2Distill [4]. For this case, SS-TTA denoising performance increases when augmented noise variance surpasses $\sigma_1^2 = 20$. Although Figure 3 and Figure 5 do not reflect general behavior, the graphs indicate that performance improvement starts or peaks around $\sigma_1^2 = 25$, which was also observed in other experiments. The baseline denoiser [4, 20, 21, 26] models have generally been trained with $\sigma^2 = 25$, which might affect the observed behavior, but the topic requires more investigation. Similarly, the current approach that σ_1^2 and σ_2^2 are manual hyperparameters of our method, is also a weakness that we aim to address in the future.

4.4. Technical remarks

This section discusses several technical aspects of the proposed test-time adaptation method.

Number of σ^2 components. Instead of using only σ_1^2 and σ_2^2 as suggested earlier in the paper, SS-TTA performance was also evaluated with higher orders of σ^2 and it was observed that increasing noisy inputs improves the average restoration quality up to 0.06 dB. However, this also increases inference complexity dramatically. Consequently, two noisy inputs was determined as the best quality

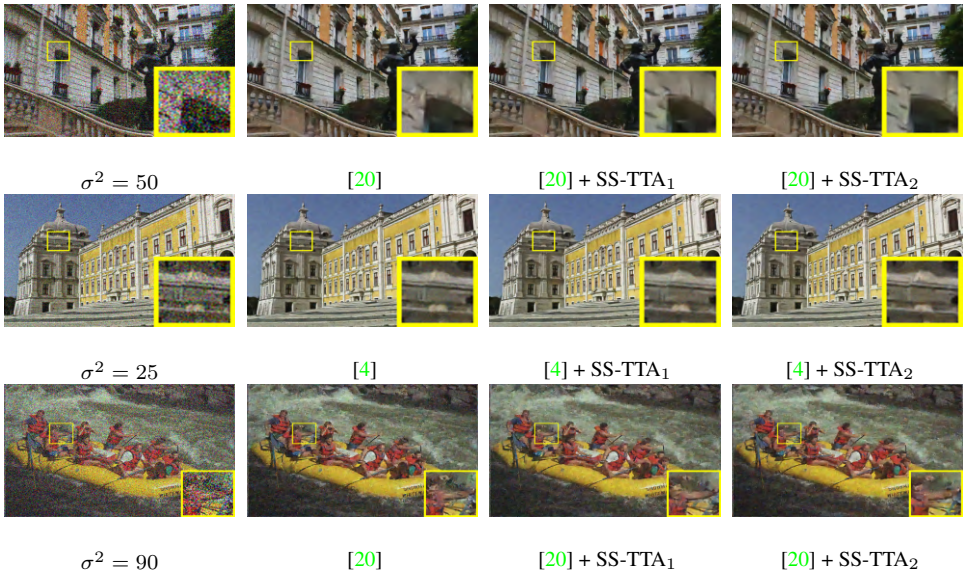

Figure 4. Visual comparison of denoising sRGB images in the setting of $\sigma^2 = [25, 50, 90]$.

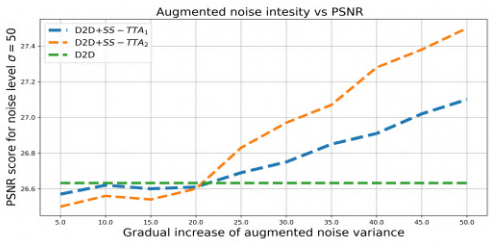

Figure 5. Performance of Decay2Distill [4] for noisy images with $\sigma^2 = 50$. The denoising performance of SS-TTA starts taking effect when the σ_1^2 parameter is increased beyond the value 20. However, this trend is not generic to every experiment/dataset.

vs. computation time trade-off.

Applicability. SS-TTA has its highest potential in the pre-processing steps of scientific imaging such as biomedical, microscopic [25], and astronomical imaging [35], as well as with regular imaging where quality is prioritized over the processing time.

Estimating σ^2 from test images. For estimating the σ values, the approach proposed in [5] was followed. Suitable σ values for synthetic images are discoverable using [5], but for real-world noise, our approach currently relies on empirical evidence.

Avoiding over-smoothing. To address potential over-smoothing/enhancement during adaptation, $\mathcal{F}_{\theta_{TTA}}$ is gradually updated through the convex sum policy with $\mathcal{F}_{\theta_s^*}$. Consequently, our final adapted model $\mathcal{F}_{\theta^*_{TTA}}$ for a given sample is always within nearby region of $\mathcal{F}_{\theta_s^*}$, preventing over-enhancement.

Effect of loss $\mathcal{L}_3$ in SS-TTA. Both variants SS-TTA$_1$ and SS-TTA$_2$ utilize *identity mapping loss* $\mathcal{L}_3$ for smoother output. Table 8 illustrates the filtering result of SS-TTA with and without $\mathcal{L}_3$.

Batch denoising. Batch processing of images using SS-TTA was experimented with batch sizes between 2 and 16, which caused PSNR decline between 0.05-0.20 dB. Since the adaptation parameters are instance-dependent, this performance loss is expected.

4.5. SS-TTA with supervised models

Even though our test-time adaption algorithm is primarily designed for self-supervised methods and shows considerable improvement, we also briefly extend our experiments to fully supervised models. Table 9 shows the PSNR improvements provided by applying SS-TTA to UNet and DnCNN models that have been pre-trained in a supervised fashion. Although the performance improvements are modest compared to the main application area of SS-TTA, self-supervised methods, especially with strong out-of-domain noise ($\sigma^2 = 90$), SS-TTA delivers a significant increase in denoising effect.

4.6. Computational complexity

Finally, the SS-TTA algorithm is computationally much lighter than the related test-time adaptation works GainTuning [25] and Lidia [35]. For every presented experiment, the SS-TTA test time adaptation was performed in five epochs for a given noisy image (independent of the variant), while GainTuning [25] requires some hundreds of epochs. Additionally, GainTuning [25] also segments the input image into a large number of random patches to complete the tuning process. Even though Lidia [35] allows enable

Method	Inference	$\sigma^2 = 25$	$\sigma^2 = 50$	$\sigma^2 = 90$
UNet [32]	Baseline	33.70	30.03	21.40
	Baseline + SS-TTA$_1$	33.71	30.04	21.51
	Baseline + SS-TTA$_2$	**33.72**	**30.06**	**21.61**
DnCNN [42]	Baseline	30.21	27.16	18.64
	Baseline + SS-TTA$_1$	30.30	27.16	19.47
	Baseline + SS-TTA$_2$	**30.32**	**27.23**	**20.12**

Table 9. SS-TTA performance with supervised models in the synthetic noise domain for the BSD68 dataset. Here, pre-trained weights are from [32, 42] and input image noise variances are ($\sigma^2 = 25, 50, 90$).

execution after five epochs, the method's internal mechanism relies on patch processing, leveraging non-local self-similarity, and exploiting representation sparsity, which implies a significant computation burden. On the contrary, the proposed SS-TTA method does not require patch conversion or other similar operations. SS-TTA requires, on average, $0.80s$ to process each image instance (w/o SS-TTA $0.09s$) with DnCNN [42] running on Intel Core i7-10700K CPU and an NVidia RTX3090 GPU, which is reasonable considering the potential application domains.

4.7. Discussion

Our current choice of noise injection for the proposed SS-TTA algorithm is not entirely adaptive, as observed with real-world noisy mitigation. In future studies, we would like to address solutions that are more adaptive than the present design. Additionally, we aim to present results for more diverse noise domains for both image and video samples. Finally, readers might get the impression that our SS-TTA$_2$ outperforms SS-TTA$_1$ in almost every case. However, for some applications from biometric or clinical domains, SS-TTA$_2$ keeps the potential to introduce false details despite the higher metric performance. Our future study will aim to address the above scenarios with proper alternatives.

Apart from that, from Figure 3 and Figure 5, it is clear that SS-TTA performs better when the augmented noise is higher. From Figure 5, noisy input images are corrupted by noise of high intensity $\sigma^2 = 50$, which implies noise constraints are higher for such samples, and larger σ_1^2,σ_2^2 boost SS-TTA performance. However, this trend is not generalizable for lower input noise intensities such as $\sigma^2 = [15, 25]$. From Figure 3, for input noise variance of $\sigma^2 = 25$, SS-TTA with larger σ_1^2,σ_2^2 leads to better performance up to some extent, but then decreases at $\sigma_1^2 = 40$. Generally, self-supervised models are trained for images with $\sigma^2 = 25$, and for input images with noise levels $\sigma^2 = 25$ or 50, SS-TTA works well with $\sigma_1^2 = 25$, $\sigma_2^2 = 30$, as empirically observed. Hence, the noise parameters have been fixed accordingly. However, we have changed $\sigma_1^2 = 1$,$\sigma_2^2 = 2$ for the supervised case (Table 9) with $\sigma^2 = [25, 50]$, as the noise constraint is absent in the supervised setup.

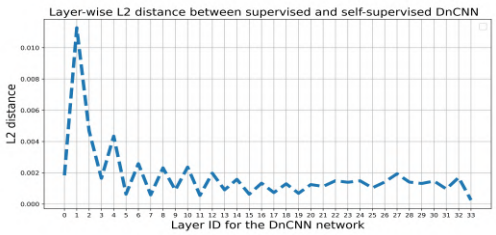

Figure 6. Layer-wise parameter distance between supervised [42] and self-supervised [26] DnCNN models.

5. Conclusion

This paper presents SS-TTA, a test-time adaptation approach for improving the performance of self-supervised denoising methods. SS-TTA can be applied on top of a variety of self-supervised denoising methods, as well as supervised ones. Our evaluation shows that SS-TTA is model-independent and outperforms recent test-time adaptation works across different noise categories: in/out-domain, synthetic, and real noise scenarios. Internally, SS-TTA relies on the controlled addition of synthetic noise and *inference-guided regularization*, a novel class of loss functions. Regarding computational efficiency, SS-TTA is considerably simple compared to other similar works. For future work, we intend to concentrate on automating the setting of two internal noise parameters that currently require manual adjustment. However, our extensive experiments have shown that even by default values of noise parameters, SS-TTA delivers state-of-the-art performance in test-time adapted denoising.

Appendix 1: Parameter distance between supervised and self-supervised models

One might assume that the constant term $(\boldsymbol{n'})^\top \boldsymbol{n'}$ only offsets the loss making the minima of both $\boldsymbol{\mathcal{F}_{\theta^*}}$ and $\boldsymbol{\mathcal{F}_{\theta_s^*}}$ the same. As an illustration that this is not the case, an identical training setup (epoch count, batch size, learning rate, stochastic order, and initialization) was established and used to train supervised and self-supervised DnCNN [42] models. Comparison of layer-wise global norms between each layer (Figure 6) shows that the global norm difference for each layer is sufficiently large to indicate that supervised and self-supervised local minima are not identical nor equivalent. Hence, it can be concluded that the constant term $(\boldsymbol{n'})^\top \boldsymbol{n'}$ causes the self-supervised and supervised models to have different minima.

Acknowledgements

This work was partially funded by the Academy of Finland project 327912 REPEAT.

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. 4, 5

[2] Saeed Anwar and Nick Barnes. Real image denoising with feature attention. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3155–3164, 2019. 2

[3] Joshua Batson and Loic Royer. Noise2self: Blind denoising by self-supervision. In *International Conference on Machine Learning*, pages 524–533. PMLR, 2019. 1, 2, 3

[4] Manisha Das Chaity and Masud An Nur Islam Fahim. Decay2distill: Leveraging spatial perturbation and regularization for self-supervised image denoisin. *arXiv preprint arXiv:2208.01948*, 2022. 2, 4, 5, 6, 7

[5] Guangyong Chen, Fengyuan Zhu, and Pheng Ann Heng. An efficient statistical method for image noise level estimation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 477–485, 2015. 7

[6] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095, 2007. 1

[7] Masud An Nur Islam Fahim, Nazmus Saqib, Shafkat Khan Siam, and Ho Yub Jung. Denoising single images by feature ensemble revisited. *Sensors*, 22(18):7080, 2022. 1, 2

[8] Maryam Gholizadeh-Ansari, Javad Alirezaie, and Paul Babyn. Deep learning for low-dose ct denoising using perceptual loss and edge detection layer. *Journal of digital imaging*, 33(2):504–515, 2020. 2

[9] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. Weighted nuclear norm minimization with application to image denoising. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2862–2869, 2014. 1

[10] Agus Gunawan, Muhammad Adi Nugroho, and Se Jin Park. Test-time adaptation for real image denoising via meta-transfer learning. *arXiv preprint arXiv:2207.02066*, 2022. 2

[11] Zhiwei Hong, Xiaocheng Fan, Tao Jiang, and Jianxing Feng. End-to-end unpaired image denoising with conditional adversarial networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 4140–4149, 2020. 2

[12] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5197–5206, 2015. 4, 5

[13] Tao Huang, Songjiang Li, Xu Jia, Huchuan Lu, and Jianzhuang Liu. Neighbor2neighbor: Self-supervised denoising from single noisy images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14781–14790, 2021. 3, 5

[14] Pavel Izmailov, Dmitrii Podoprikhin, Timur Garipov, Dmitry Vetrov, and Andrew Gordon Wilson. Averaging weights leads to wider optima and better generalization. *arXiv preprint arXiv:1803.05407*, 2018. 4

[15] Geonwoon Jang, Wooseok Lee, Sanghyun Son, and Kyoung Mu Lee. C2n: Practical generative noise modeling for real-world denoising. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2350–2359, 2021. 2

[16] 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*, pages 2129–2137, 2019. 1, 2

[17] Samuli Laine, Tero Karras, Jaakko Lehtinen, and Timo Aila. High-quality self-supervised deep image denoising. *Advances in Neural Information Processing Systems*, 32, 2019. 2

[18] Kanggeun Lee and Won-Ki Jeong. Noise2kernel: Adaptive self-supervised blind denoising using a dilated convolutional kernel architecture. *Sensors*, 22(11):4255, 2022. 1, 2

[19] Wooseok Lee, Sanghyun Son, and Kyoung Mu Lee. Ap-bsn: Self-supervised denoising for real-world images via asymmetric pd and blind-spot network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17725–17734, 2022. 3

[20] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189*, 2018. 1, 2, 3, 4, 5, 6, 7

[21] Jason Lequyer, Reuben Philip, Amit Sharma, and Laurence Pelletier. Noise2fast: Fast self-supervised single image blind denoising. *arXiv preprint arXiv:2108.10209*, 2021. 1, 2, 6

[22] Tao Li, Zhehao Huang, Qinghua Tao, Yingwen Wu, and Xiaolin Huang. Trainable weight averaging for fast convergence and better generalization. *arXiv preprint arXiv:2205.13104*, 2022. 4

[23] Jiaming Liu, Chi-Hao Wu, Yuzhi Wang, Qin Xu, Yuqian Zhou, Haibin Huang, Chuan Wang, Shaofan Cai, Yifan Ding, Haoqiang Fan, et al. Learning raw image denoising with bayer pattern unification and bayer preserving augmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. 2

[24] Yinjin Ma, Biao Wei, Peng Feng, Peng He, Xiaodong Guo, and Ge Wang. Low-dose ct image denoising using a generative adversarial network with a hybrid loss function for noise learning. *IEEE Access*, 8:67519–67529, 2020. 2

[25] Sreyas Mohan, Joshua L Vincent, Ramon Manzorro, Peter Crozier, Carlos Fernandez-Granda, and Eero Simoncelli. Adaptive denoising via gaintuning. *Advances in Neural Information Processing Systems*, 34:23727–23740, 2021. 1, 2, 3, 4, 6, 7

[26] Nick Moran, Dan Schmidt, Yu Zhong, and Patrick Coady. Noisier2noise: Learning to denoise from unpaired noisy data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12064–12072, 2020. 1, 2, 4, 5, 6, 8

[27] Seonghyeon Nam, Youngbae Hwang, Yasuyuki Matsushita, and Seon Joo Kim. A holistic approach to cross-channel image noise modeling and its application to image denoising.

In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1683–1691, 2016. 4

[28] Reyhaneh Neshatavar, Mohsen Yavartanoo, Sanghyun Son, and Kyoung Mu Lee. Cvf-sid: Cyclic multi-variate function for self-supervised image denoising by disentangling noise from image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17583–17591, 2022. 1, 2, 5

[29] Tongyao Pang, Huan Zheng, Yuhui Quan, and Hui Ji. Recorrupted-to-recorrupted: unsupervised deep learning for image denoising. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2043–2052, 2021. 1, 2, 3, 5

[30] Yuhui Quan, Mingqin Chen, Tongyao Pang, and Hui Ji. Self2self with dropout: Learning self-supervised denoising from single image. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1890–1898, 2020. 1, 2

[31] Chao Ren, Xiaohai He, Chuncheng Wang, and Zhibo Zhao. Adaptive consistency prior based deep network for image denoising. In *proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8596–8606, 2021. 1, 2

[32] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 6, 8

[33] Shakarim Soltanayev and Se Young Chun. Training deep learning based denoisers without ground truth data. *Advances in neural information processing systems*, 31, 2018. 2

[34] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9446–9454, 2018. 1, 2

[35] Gregory Vaksman, Michael Elad, and Peyman Milanfar. Lidia: Lightweight learned image denoising with instance adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 524–525, 2020. 2, 3, 4, 6, 7

[36] Xiaohe Wu, Ming Liu, Yue Cao, Dongwei Ren, and Wangmeng Zuo. Unpaired learning of deep image denoising. In *European conference on computer vision*, pages 352–368. Springer, 2020. 2

[37] Yaochen Xie, Zhengyang Wang, and Shuiwang Ji. Noise2same: Optimizing a self-supervised bound for image denoising. *Advances in Neural Information Processing Systems*, 33:20320–20330, 2020. 1, 2

[38] Jun Xu, Yuan Huang, Ming-Ming Cheng, Li Liu, Fan Zhu, Zhou Xu, and Ling Shao. Noisy-as-clean: Learning self-supervised denoising from corrupted image. *IEEE Transactions on Image Processing*, 29:9316–9329, 2020. 1, 2

[39] Li Xu, Cewu Lu, Yi Xu, and Jiaya Jia. Image smoothing via l 0 gradient minimization. In *Proceedings of the 2011 SIGGRAPH Asia conference*, pages 1–12, 2011. 1

[40] Xuhui Yang, Yong Xu, Yuhui Quan, and Hui Ji. Image denoising via sequential ensemble learning. *IEEE Transactions on Image Processing*, 29:5038–5049, 2020. 1, 2

[41] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for real image restoration and enhancement. In *European Conference on Computer Vision*, pages 492–511. Springer, 2020. 4, 5

[42] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7):3142–3155, 2017. 1, 2, 4, 5, 6, 8

[43] Yi Zhang, Dasong Li, Ka Lung Law, Xiaogang Wang, Hongwei Qin, and Hongsheng Li. Idr: Self-supervised image denoising via iterative data refinement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2098–2107, 2022. 1, 2

[44] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz. Loss functions for image restoration with neural networks. *IEEE Transactions on computational imaging*, 3(1):47–57, 2016. 2