Patch-Craft Self-Supervised Training for Correlated Image Denoising (Supplementary Material)

Gregory Vaksman and Michael Elad
CS Department - The Technion
Haifa, Israel
grishav@campus.technion.ac.il, elad@cs.technion.ac.il

A. Proof of Lemma 1

Proof. We begin with rewriting the expression for $\nabla_\theta \tilde{l}$,

$$\begin{aligned}\nabla_\theta \tilde{l} &= \nabla_\theta^T f_\theta(\mathbf{y})\left(f_\theta(\mathbf{y}) - \tilde{\mathbf{x}}\right) = \\ &= \nabla_\theta^T f_\theta(\mathbf{y})\left(f_\theta(\mathbf{y}) - (\mathbf{x}+\mathbf{w})\right) = \\ &= \nabla_\theta^T f_\theta(\mathbf{y})\left(f_\theta(\mathbf{y}) - \mathbf{x}\right) - \nabla_\theta^T f_\theta(\mathbf{y})\,\mathbf{w} = \\ &= \nabla_\theta l - \nabla_\theta^T f_\theta(\mathbf{y})\,\mathbf{w}\ .\end{aligned} \tag{1}$$

We proceed to calculating the expectation of $\nabla_\theta \tilde{l}$,

$$\begin{aligned}\mathbb{E}\left[\nabla_\theta \tilde{l}\right] &= \mathbb{E}\left[\mathbb{E}\left[\nabla_\theta \tilde{l}\middle|\,\mathbf{x},\mathbf{z}\right]\right] = \\ &= \mathbb{E}\left[\mathbb{E}\left[\nabla_\theta l - \nabla_\theta^T f(\mathbf{y})\,\mathbf{w}\middle|\,\mathbf{x},\mathbf{z}\right]\right] = \\ &= \mathbb{E}\left[\mathbb{E}\left[\nabla_\theta l|\,\mathbf{x},\mathbf{z}\right] - \mathbb{E}\left[\nabla_\theta^T f(\mathbf{y})\,\mathbf{w}\middle|\,\mathbf{x},\mathbf{z}\right]\right] = \\ &= \mathbb{E}\left[\nabla_\theta l - \nabla_\theta^T f(\mathbf{y})\,\mathbb{E}\left[\mathbf{w}|\,\mathbf{x},\mathbf{z}\right]\right] = \\ &= \mathbb{E}\left[\nabla_\theta l - \nabla_\theta^T f(\mathbf{y})\,\mathbb{E}\left[\mathbf{w}\right]\right] = \\ &= \mathbb{E}\left[\nabla_\theta l - \nabla_\theta^T f(\mathbf{y})\cdot 0\right] = \\ &= \mathbb{E}\left[\nabla_\theta l\right] = \nabla_\theta L\ .\end{aligned} \tag{2}$$

The first equality, $\mathbb{E}\left[\nabla_\theta \tilde{l}\right] = \mathbb{E}\left[\mathbb{E}\left[\nabla_\theta \tilde{l}\middle|\,\mathbf{x},\mathbf{z}\right]\right]$, is correct by the law of total expectation, and $\mathbb{E}\left[\mathbf{w}|\,\mathbf{x},\mathbf{z}\right] = \mathbb{E}\left[\mathbf{w}\right]$ is true due to the independence of $\mathbf{w}$ in $\mathbf{x}$ and $\mathbf{z}$. Finally, we get

$$Bias\left[\nabla_\theta \tilde{l}\right] = \mathbb{E}\left[\nabla_\theta \tilde{l}\right] - \nabla_\theta L = 0\ , \tag{3}$$

which completes the proof. □

B. Dependency Reduction – Mathematical Analysis

In this section we bring proofs of statements used in Section 4.2, organized as follows. In Section B.1 we prove the convergence of $s_{\mathbf{y},\mathbf{r}}$ to a Gaussian distribution, and in Section B.2 we prove our statements regarding the shift of the mean in the case of dependencies of types (I) and (II).

B.1. Convergence to Gaussian

Our goal is to prove that $s_{\mathbf{y},\mathbf{r}}$ convergences in distribution to a Gaussian. We start with the assumption that the ground truth image is a sum of two statistically independent random variables, $\mathbf{x} = \bar{\mathbf{x}} + \mu_x$, where μ_x is an image mean (a scalar) and $\bar{\mathbf{x}}$ is a zero-mean vector. Similarly, we assume that $\mathbf{w} = \bar{\mathbf{w}} + \mu_w$. Our additional assumption is that $\bar{\mathbf{x}}$, $\mathbf{z}$, and $\bar{\mathbf{w}}$ are m*-dependent* in any dimension and follow zero-mean and *stationary* distributions.

Definition 1 (m-dependent sequences [1]).
*Let $X_1, X_2, \dots$ be a sequence of random variables. The sequence is **m-dependent** if $(X_1, \dots, X_i)$ and $(X_{i+k}, \dots, X_{i+k+l})$ are independent whenever $k > m$.*

By m-dependency in any dimension, we mean that $(X_{1,1}, \dots, X_{i,j})$ and $(X_{i+k,j+r}, \dots, X_{i+k+l,j+r+p})$ are independent if $\max\{k, r\} > m$. For natural images distribution, stationarity and m-dependency are common assumptions. Stationarity means translation invariance, and the m-dependency assumption implies that each pixel is dependent only on its local neighborhood of radius $m/2$. Indeed, m-dependency is the property that allows effective denoisers to have relatively small respective fields. To proceed with the proof, we need the following definition and theorem.

Definition 2 (Strongly mixing sequences [1]).
Let $X_1, X_2, \dots$ be a sequence of random variables, and let α_k be a number such that

$$\sup\left|P\left(A\cap B\right) - P\left(A\right)P\left(B\right)\right| \le \alpha_k$$

*for $A \in \sigma\left(X_1, \dots, X_i\right)$, $B \in \sigma\left(X_{i+k}, X_{i+k+1}, \dots\right)$, and $i \ge 1$, $k \ge 1$. The sequence is said to be **strongly mixing** if $\alpha_k \longrightarrow 0$ as $k \longrightarrow \infty$.*

Theorem 3 (Central limit theorem for strongly mixing sequences, Theorem 1.7 in [2]).

Let $X_1, X_2, \dots$ be a stationary and strongly mixing sequence with $\mathbb{E}[X_i] = 0$ such that for some $\delta > 0$

$$\mathbb{E}\left[|X_i|^{2+\delta}\right] < \infty \quad \text{and} \quad \sum_{i=1}^{\infty} (\alpha_i)^{\frac{\delta}{2+\delta}} < \infty .$$

Denote $S_n = X_1 + X_2 + \dots + X_n$, then

$$\lim_{n\to\infty} \frac{\mathbb{E}\left[S_n^2\right]}{n} = \sigma^2 = \mathbb{E}\left[X_1^2\right] + 2\sum_{i=1}^{\infty} \mathbb{E}\left[X_1 X_i\right] < \infty .$$

If $\sigma \neq 0$, then $\frac{S_n}{\sigma\sqrt{n}}$ converges in distribution to $\mathcal{N}(0,1)$ as n approaches infinity.

It follows from Theorem 3 follows that sum of m-dependent stationary sequence converges in distribution to a Gaussian, as stated in the following corollary.

Corollary 4. *Let $X_1, X_2, \dots$ be a stationary and m-dependent sequence with $\mathbb{E}[X_i] = 0$ such that for some $\delta > 0$, $\mathbb{E}\left[|X_i|^{2+\delta}\right] < \infty$.*
Denote $S_n = X_1 + X_2 + \dots + X_n$, then

$$\lim_{n\to\infty} \frac{\mathbb{E}\left[S_n^2\right]}{n} = \sigma^2 = \mathbb{E}\left[X_1^2\right] + 2\sum_{i=1}^{m} \mathbb{E}\left[X_1 X_i\right] < \infty .$$

If $\sigma \neq 0$, then $\frac{S_n}{\sigma\sqrt{n}}$ converges in distribution to $\mathcal{N}(0,1)$ as n approaches infinity.

Proof. To prove it, it suffices to show that from the m-dependency, it follows that for any $\delta > 0$,

$$\sum_{i=1}^{\infty} (\alpha_i)^{\frac{\delta}{2+\delta}} < \infty . \tag{4}$$

Note that for m-dependent sequences, $\alpha_k = 0$ for any $k > m$. Thus, the sum becomes finite and thereby bounded,

$$\sum_{k=1}^{\infty} (\alpha_k)^{\frac{\delta}{2+\delta}} = \sum_{k=1}^{m} (\alpha_k)^{\frac{\delta}{2+\delta}} < \infty . \tag{5}$$

□

Armed with Corollary 4, we return to the proof of $s_{\mathbf{y},\mathbf{r}}$ convergence. Assuming that μ_x and μ_w are known, $s_{\mathbf{y},\mathbf{r}}$ is defined as

$$s_{\mathbf{y},\mathbf{r}} = \frac{1}{n^2} \sum_{i,j=1}^{n} \left(y_{i,j} - \mu_x\right)\left(r_{i,j} - \mu_w\right) . \tag{6}$$

Note that in practice, μ_x and μ_w are unavailable, but they can be estimated empirically,

$$\mu_x \approx \frac{1}{n^2} \sum_{i,j=1}^{n} y_{i,j} , \quad \mu_w \approx \frac{1}{n^2} \sum_{i,j=1}^{n} r_{i,j} . \tag{7}$$

Proposition 5. *Let $s_{\mathbf{y},\mathbf{r}}$ be as defined in equation 6. Then, as $n \longrightarrow \infty$,*

$$\frac{n}{\sigma}\left(s_{\mathbf{y},\mathbf{r}} + \sigma_z^2\right) \longrightarrow \mathcal{N}(0,1) ,$$

where $\sigma_z^2 = \mathbb{E}\left[z_{i,j}^2\right]$ and

$$\sigma^2 = \lim_{n\to\infty} \mathbb{E}\left[s_{\mathbf{y},\mathbf{w}}^2\right] .$$

Proof. We begin by rewriting the expression for $\mathbf{r}$,

$$\mathbf{r} = \tilde{\mathbf{x}} - \mathbf{y} = \mathbf{x} + \mathbf{w} - \mathbf{x} - \mathbf{z} = \mathbf{w} - \mathbf{z} . \tag{8}$$

Substituting $\mathbf{r}$ into the expression for $s_{\mathbf{y},\mathbf{r}}$ we get

$$\begin{aligned} s_{\mathbf{y},\mathbf{r}} + \sigma_z^2 &= \frac{1}{n^2} \sum_{i,j=1}^{n} \left(y_{i,j} - \mu_x\right)\left(r_{i,j} - \mu_w\right) + \sigma_z^2 = \\ &= \frac{1}{n^2} \sum_{i,j=1}^{n} \left(\left(\bar{x}_{i,j} + z_{i,j}\right)\left(\bar{w}_{i,j} - z_{i,j}\right) + \sigma_z^2\right) = \\ &= \frac{1}{n^2} \sum_{i,j=1}^{n} \xi_{i,j} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{n} \sum_{j=1}^{n} \xi_{i,j}\right) = \frac{1}{n} \sum_{i=1}^{n} \zeta_i , \end{aligned} \tag{9}$$

where $\xi_{i,j} = \left(\bar{x}_{i,j} + z_{i,j}\right)\left(\bar{w}_{i,j} - z_{i,j}\right) + \sigma_z^2$ and $\zeta_i = \frac{1}{n}\sum_{j=1}^{n} \xi_{i,j}$. Clearly, sequence $(\xi_{1,1}, \dots, \xi_{n,n})$ is m-dependent in any dimension. Therefore $(\zeta_1, \dots, \zeta_n)$ is m-dependent. In addition, $\mathbb{E}[\zeta_i] = 0$ as $\bar{x}_{i,j}$, $z_{i,j}$, and $\bar{w}_{i,j}$ are zero-mean and independent. Also, $\mathbb{E}\left[|\zeta_i|^{2+\delta}\right] < \infty$ since $\bar{x}_{i,j}$, $z_{i,j}$, and $\bar{w}_{i,j}$ are bounded. Thus, according to Theorem 3

$$\frac{n}{\sigma}\left(s_{\mathbf{y},\mathbf{r}} + \sigma_z^2\right) \longrightarrow \mathcal{N}(0,1) .$$

□

B.2. Dependencies of Type (I) and (II)

We now turn to we prove our statements regarding the type (I) and type (II) dependencies. Through the section, we denote by $\sigma_{\alpha,\beta}$ the scalar covariances computed over pairs of images $\alpha, \beta \in \{\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \mathbf{w}, \mathbf{z}\}$, where σ_α^2 stands for the scalar variances ($\sigma_\alpha = \sqrt{\sigma_{\alpha,\alpha}}$). We start with the proof that type (I) dependency implies $\mathbb{E}[s_{\mathbf{y},\mathbf{r}}] > -\sigma_z^2$. Recall that type (I) dependency is characterized by a positive correlation between $\mathbf{z}$ and $\mathbf{w}$. In addition, assume that there is no dependency of type (II), i.e., $\mathbf{x}$ and $\mathbf{w}$ are independent.

Proposition 6. *Suppose that $\mathbf{w}$ is independent of $\mathbf{x}$, but there is a dependency between $\mathbf{w}$ and $\mathbf{z}$ such that $\sigma_{z,w} > 0$. Then $\mathbb{E}[s_{\mathbf{y},\mathbf{r}}] > -\sigma_z^2$.*

Proof. Substituting equations 8 into $\mathbb{E}[s_{\mathbf{y},\mathbf{r}}]$ we get

$$\mathbb{E}[s_{\mathbf{y},\mathbf{r}}] = \sigma_{y,r} = \sigma_{x+z,w-z} = \sigma_{z,w} - \sigma_z^2 > -\sigma_z^2 . \tag{10}$$

□

We proceed to a discussion of type (II) dependency, which occurs when $\tilde{\mathbf{x}}$ and $\mathbf{x}$ tend to be dissimilar. Through it, we assume that there is no dependency of type (I), i.e., $\mathbf{w}$ and $\mathbf{z}$ are independent. First, we show that this dependency is manifested in a negative correlation between $\mathbf{w}$ and $\mathbf{x}$, $\sigma_{x,w} < 0$. As mentioned in section 4.2,

$$\sigma_{x,\tilde{x}} = \sigma_x^2 + \sigma_{x,w} \ . \quad (11)$$

Recall that $\tilde{\mathbf{x}}$ is built of patches taken from noisy images. Thus, we can write

$$\tilde{\mathbf{x}} = \mathbf{x}_2 + \mathbf{z}_2 \ , \quad (12)$$

where $\mathbf{x}_2$ is a clean image, which may differ from $\mathbf{x}$, and $\mathbf{z}_2$ is an instantiation of input noise, which is independent of $\mathbf{x}$ and $\mathbf{x}_2$. Thus,

$$\sigma_{x,\tilde{x}} = \sigma_{x,x_2+z_2} = \sigma_{x,x_2} \leq \sigma_x \sigma_{x_2} = \sigma_x^2 \ . \quad (13)$$

A conclusion of equations 11 and 13 is that the dissimilarity between $\tilde{\mathbf{x}}$ and $\mathbf{x}$ reduces the value of $\sigma_{x,\tilde{x}}$, which means that $\sigma_{x,w}$ is necessarily negative. Finally, we prove that for the dependency of type (II), we get $\mathbb{E}\left[s_{\mathbf{y},\mathbf{r}}\right] < -\sigma_z^2$.

Proposition 7. *Suppose that* $\mathbf{w}$ *is independent on* $\mathbf{z}$*, but there is a dependency between* $\mathbf{w}$ *and* $\mathbf{x}$ *such that* $\sigma_{x,w} < 0$*. Then* $\mathbb{E}\left[s_{\mathbf{y},\mathbf{r}}\right] < -\sigma_z^2$*.*

Proof. Substituting equation 8 into $\mathbb{E}\left[s_{\mathbf{y},\mathbf{r}}\right]$ we have

$$\mathbb{E}\left[s_{\mathbf{y},\mathbf{r}}\right] = \sigma_{y,r} = \sigma_{x+z,w-z} = \sigma_{x,w} - \sigma_z^2 < -\sigma_z^2 \ . \quad (14)$$

□

C. Robustness of the Patch Matching

In section 4.2, we mention that we use large patches for patch matching. This section shows that patches must be large when dealing with correlated noise. The goal of patch matching is to find similar clean patches by checking the L_2 distance between their noisy versions. However, does the similarity between the noisy patches imply the similarity between their clean counterparts? Assuming that noise is independent of the image, the answer is yes, provided the patches are large enough. Intuitively, patch size should grow with the noise power, but do the noise correlations matter? In this section, we show that, in addition to the power of the noise, the patch size is heavily dependent on the noise correlation range. The distance between the noisy patches can be viewed as an estimator of the distance between their clean counterparts. We show that the variance of this estimator can increase dramatically with the noise correlation range. More specifically, we provide a lower bound for the estimator's variance, which depends on the patch size and the noise autocovariance.

In this section, we use the following notations and assumptions. Let $\mathbf{y}^{(1)}, \mathbf{y}^{(2)} \in \mathbb{R}^{n \times n}$ be two noisy patches, and $\mathbf{x}^{(1)}, \mathbf{x}^{(2)} \in \mathbb{R}^{n \times n}$ be their clean versions, such that

$$\begin{aligned} \mathbf{y}^{(1)} &= \mathbf{x}^{(1)} + \mathbf{z}^{(1)} \\ \mathbf{y}^{(2)} &= \mathbf{x}^{(2)} + \mathbf{z}^{(2)} \ . \end{aligned} \quad (15)$$

We assume that $\mathbf{z}^{(1)}$ and $\mathbf{z}^{(2)}$ are independent realizations of 2D zero-mean random process $\{Z_{i,j}\}$ with autocovariance $R_{ZZ}(\tau_1, \tau_2)$ and denote $\sigma_z^2 = R_{ZZ}(0,0)$. For this section only, we assume that $\{Z_{i,j}\}$ is a Gaussian process. We denote by δ_x the normalized squared L_2 distance between $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$. Similarly, δ_y stands for the normalized squared L_2 distance between $\mathbf{y}^{(1)}$ and $\mathbf{y}^{(2)}$,

$$\begin{aligned} \delta_x &= \frac{1}{n^2} \sum_{i,j=1}^{n} \left(x_{i,j}^{(2)} - x_{i,j}^{(1)} \right)^2 \\ \delta_y &= \frac{1}{n^2} \sum_{i,j=1}^{n} \left(y_{i,j}^{(2)} - y_{i,j}^{(1)} \right)^2 \ . \end{aligned} \quad (16)$$

Theorem 8. *Let* δ_y *be an estimator of* δ_x*, then* δ_y *has a constant bias* $2\sigma_z^2$*, and its variance is bounded from below by*

$$var\left[\delta_y\right] \geq \frac{8}{n^2} \sigma_z^4 \rho \ ,$$

where

$$\rho = \frac{1}{n^2} \sum_{i_1,j_1,i_2,j_2=1}^{n} \left(\frac{R_{ZZ}(i_1 - j_1, i_2 - j_2)}{\sigma_z^2} \right)^2 \ .$$

The bound is sharp since equality holds for $\delta_x = 0$*.*

The proof of the theorem is given in appendix C.1. The theorem shows that the bound is proportional to ρ, whereρ can grow fast with the noise correlation range. We illustrate this by the example of R_{ZZ} with bilinear decay,

$$\begin{aligned} R_{ZZ}(\tau_1, \tau_2) &= g(\tau_1)\, g(\tau_2) \\ g(\tau) &= \sigma_z \cdot \max\left\{1 - \frac{|\tau|}{\theta}, 0\right\} \ , \end{aligned} \quad (17)$$

where $\frac{1}{\theta}$ is the decay incline.

Proposition 9. *Suppose* R_{ZZ} *as defined in equation 17. Then*

$$\rho \geq \frac{1}{4}\left(r + \frac{1}{r}\right)^2 \ , \quad r = \min\{n, \lfloor\theta\rfloor\} \ .$$

The bound is sharp, the equality holds for $\theta = n$ *and* $\theta = 1$*.*

The proof is given in appendix C.2. Proposition 10 shows that for R_{ZZ} with bilinear decay, $var\left[\delta_y\right]$ exhibits quadratic growth with respect to the correlation range (for $\theta \leq n$).

C.1. Proof of Theorem 8

Proof. We begin with introducing notations. Denote three difference images by $\mathbf{d}^{(x)}$, $\mathbf{d}^{(y)}$, and $\mathbf{d}^{(z)}$, where

$$\mathbf{d}^{(x)} = \mathbf{x}^{(2)} - \mathbf{x}^{(1)}, \mathbf{d}^{(y)} = \mathbf{y}^{(2)} - \mathbf{y}^{(1)}, \mathbf{d}^{(z)} = \mathbf{z}^{(2)} - \mathbf{z}^{(1)}.$$

In addition, we define δ_x, δ_y, δ_z, and δ_{xz} as follows

$$\begin{aligned} \delta_x &= \frac{1}{n^2} \sum_{i,j=1}^{n} \left(d_{i,j}^{(x)}\right)^2, \quad \delta_y = \frac{1}{n^2} \sum_{i,j=1}^{n} \left(d_{i,j}^{(y)}\right)^2, \\ \delta_z &= \frac{1}{n^2} \sum_{i,j=1}^{n} \left(d_{i,j}^{(z)}\right)^2, \quad \delta_{x,z} = \frac{1}{n^2} \sum_{i,j=1}^{n} d_{i,j}^{(x)} d_{i,j}^{(z)}, \end{aligned} \tag{18}$$

where δ_x, δ_y, and δ_z are the normalized L_2 norms of $\mathbf{d}^{(x)}$, $\mathbf{d}^{(y)}$, and $\mathbf{d}^{(z)}$, respectively, and δ_{xz} stands for a mixed expression. Then

$$\mathbf{d}^{(y)} = \mathbf{y}^{(2)} - \mathbf{y}^{(1)} = \mathbf{x}^{(2)} - \mathbf{x}^{(1)} + \mathbf{z}^{(2)} - \mathbf{z}^{(1)} = \mathbf{d}^{(x)} + \mathbf{d}^{(z)}.$$

Rewriting δ_y, we get

$$\begin{aligned} \delta_y =& \frac{1}{n^2} \sum_{i,j=1}^{n} \left(d_{i,j}^{(y)}\right)^2 = \frac{1}{n^2} \sum_{i,j=1}^{n} \left(d_{i,j}^{(x)} + d_{i,j}^{(z)}\right)^2 = \\ =& \frac{1}{n^2} \sum_{i,j=1}^{n} \left(d_{i,j}^{(x)}\right)^2 + \frac{2}{n^2} \sum_{i,j=1}^{n} d_{i,j}^{(x)} d_{i,j}^{(z)} + \\ & \frac{1}{n^2} \sum_{i,j=1}^{n} \left(d_{i,j}^{(z)}\right)^2 = \delta_x + 2\delta_{x,z} + \delta_z . \end{aligned} \tag{19}$$

Therefore,

$$\begin{aligned} Bias\left[\delta_y\right] &= \mathbb{E}\left[\delta_y\right] - \delta_x = \\ &= \mathbb{E}\left[(\delta_x + 2\delta_{x,z} + \delta_z)\right] - \delta_x = \\ &= 2\,\mathbb{E}\left[\delta_{x,z}\right] + \mathbb{E}\left[\delta_z\right] , \end{aligned} \tag{20}$$

where

$$\begin{aligned} 2\,\mathbb{E}\left[\delta_{x,z}\right] &= \frac{2}{n^2} \sum_{i,j=1}^{n} \mathbb{E}\left[d_{i,j}^{(x)} d_{i,j}^{(z)}\right] = \\ &= \frac{2}{n^2} \sum_{i,j=1}^{n} d_{i,j}^{(x)}\, \mathbb{E}\left[d_{i,j}^{(z)}\right] = \\ &= \frac{2}{n^2} \sum_{i,j=1}^{n} d_{i,j}^{(x)}\, \mathbb{E}\left[z_{i,j}^{(2)} - z_{i,j}^{(1)}\right] = \\ &= \frac{2}{n^2} \sum_{i,j=1}^{n} d_{i,j}^{(x)}\, (0 - 0) = 0 . \end{aligned} \tag{21}$$

Since $\mathbf{z}^{(1)}$ and $\mathbf{z}^{(2)}$ are independent,

$$\mathbb{E}\left[\left(d_{i,j}^{(z)}\right)^2\right] = \mathbb{E}\left[\left(z_{i,j}^{(2)} - z_{i,j}^{(1)}\right)^2\right] = 2\sigma_z^2 . \tag{22}$$

Then,

$$\mathbb{E}\left[\delta_z\right] = \frac{1}{n^2} \sum_{i,j=1}^{n} \mathbb{E}\left[\left(d_{i,j}^{(z)}\right)^2\right] = \frac{1}{n^2} \sum_{i,j=1}^{n} 2\sigma_z^2 = 2\sigma_z^2 . \tag{23}$$

Substituting equations 21 and 22 into 20, we have

$$Bias\left[\delta_y\right] = 2\,\mathbb{E}\left[\delta_{x,z}\right] + \mathbb{E}\left[\delta_z\right] = 0 + 2\sigma_z^2 = 2\sigma_z^2 . \tag{24}$$

We proceed to calculate the variance.

$$\begin{aligned} var\left[\delta_y\right] &= var\left[(\delta_y - \delta_x)\right] = \\ &= \mathbb{E}\left[(\delta_y - \delta_x)^2\right] - \left(\mathbb{E}\left[(\delta_y - \delta_x)\right]\right)^2 = \\ &= \mathbb{E}\left[(\delta_y - \delta_x)^2\right] - \left(Bias\left[\delta_y\right]\right)^2 = \\ &= \mathbb{E}\left[(\delta_y - \delta_x)^2\right] - 4\sigma_z^4 . \end{aligned} \tag{25}$$

Substituting equation 19 into 25, we get

$$\begin{aligned} var\left[\delta_y\right] &= \mathbb{E}\left[(\delta_y - \delta_x)^2\right] - 4\sigma_z^4 = \\ &= \mathbb{E}\left[(2\delta_{x,z} + \delta_z)^2\right] - 4\sigma_z^4 = \\ &= 4\,\mathbb{E}\left[\delta_{x,z}^2\right] + 4\,\mathbb{E}\left[\delta_{x,z}\delta_z\right] + \mathbb{E}\left[\delta_z^2\right] - 4\sigma_z^4 . \end{aligned} \tag{26}$$

Since $\mathbb{E}\left[\delta_{x,z}^2\right] \geq 0$, the variance can be bounded from below using the following inequality,

$$var\left[\delta_y\right] \geq 4\,\mathbb{E}\left[\delta_{x,z}\right] + \mathbb{E}\left[\delta_z^2\right] - 4\sigma_z^4 , \tag{27}$$

where the bound is strict since provided $\mathbf{d}^{(x)} = \mathbf{0}$, we have

$$\begin{aligned} \mathbb{E}\left[\delta_{x,z}^2\right] &= \mathbb{E}\left[\left(\frac{1}{n^2} \sum_{i,j=1}^{n} d_{i,j}^{(x)} d_{i,j}^{(z)}\right)^2\right] = \\ &= \mathbb{E}\left[\left(\frac{1}{n^2} \sum_{i,j=1}^{n} 0 \cdot d_{i,j}^{(z)}\right)^2\right] = \mathbb{E}\left[0\right] = 0 . \end{aligned} \tag{28}$$

Next, we show that $\mathbb{E}\left[\delta_{x,z}\delta_z\right] = 0$.

$$\begin{aligned}\mathbb{E}\left[\delta_{x,z}\delta_z\right] &= \frac{1}{n^4}\,\mathbb{E}\left[\left(\sum_{i,j}^{n} d^{(x)}_{i,j} d^{(z)}_{i,j}\right)\left(\sum_{i,j}^{n}\left(d^{(z)}_{i,j}\right)^2\right)\right] = \\ &= \frac{1}{n^4}\,\mathbb{E}\left[\sum_{i_1,j_1,i_2,j_2}^{n} d^{(x)}_{i_1,j_1} d^{(z)}_{i_1,j_1}\left(d^{(z)}_{i_2,j_2}\right)^2\right] = \\ &= \frac{1}{n^4}\sum_{i_1,j_1,i_2,j_2}^{n}\mathbb{E}\left[d^{(x)}_{i_1,j_1} d^{(z)}_{i_1,j_1}\left(d^{(z)}_{i_2,j_2}\right)^2\right] = \\ &= \frac{1}{n^4}\sum_{i_1,j_1,i_2,j_2}^{n} d^{(x)}_{i_1,j_1}\,\mathbb{E}\left[d^{(z)}_{i_1,j_1}\left(d^{(z)}_{i_2,j_2}\right)^2\right] = \\ &= \frac{1}{n^4}\sum_{i_1,j_1,i_2,j_2}^{n} d^{(x)}_{i_1,j_1}\cdot 0 = 0\,,\end{aligned} \tag{29}$$

where $\mathbb{E}\left[d^{(z)}_{i_1,j_1}\left(d^{(z)}_{i_2,j_2}\right)^2\right] = 0$ as it is the third moment of multivariate Gaussian distribution. Substituting equation 29 into 27, we get

$$var\left[\hat{\delta}_x\right] \geq \mathbb{E}\left[\delta_z^2\right] - 4\sigma_z^4\,. \tag{30}$$

It remains to calculate the value of $\mathbb{E}\left[\delta_z^2\right]$.

$$\begin{aligned}\mathbb{E}\left[\delta_z^2\right] &= \mathbb{E}\left[\delta_z\delta_z\right] = \\ &= \frac{1}{n^4}\,\mathbb{E}\left[\left(\sum_{i,j=1}^{n}\left(d^{(z)}_{i,j}\right)^2\right)\left(\sum_{i,j=1}^{n}\left(d^{(z)}_{i,j}\right)^2\right)\right] = \\ &= \frac{1}{n^4}\,\mathbb{E}\left[\sum_{i_1,j_1,i_2,j_2=1}^{n}\left(d^{(z)}_{i_1,j_1}\right)^2\left(d^{(z)}_{i_2,j_2}\right)^2\right] = \\ &= \frac{1}{n^4}\sum_{i_1,j_1,i_2,j_2=1}^{n} E\left[\left(d^{(z)}_{i_1,j_1}\right)^2\left(d^{(z)}_{i_2,j_2}\right)^2\right]\,.\end{aligned} \tag{31}$$

$$\begin{aligned}&\mathbb{E}\left[\left(d^{(z)}_{i_1,j_1}\right)^2\left(d^{(z)}_{i_2,j_2}\right)^2\right] = \\ &= \mathbb{E}\left[\left(z^{(1)}_{i_1,j_1} - z^{(2)}_{i_i,j_i}\right)^2\left(z^{(1)}_{i_2,j_2} - z^{(2)}_{i_2,j_2}\right)^2\right] = \\ &= \mathbb{E}\Bigg[\left(\left(z^{(1)}_{i_1,j_1}\right)^2 - 2z^{(1)}_{i_1,j_1}z^{(2)}_{i_1,j_1} + \left(z^{(2)}_{i_1,j_1}\right)^2\right)\times \\ &\qquad\left(\left(z^{(1)}_{i_2,j_2}\right)^2 - 2z^{(1)}_{i_2,j_2}z^{(2)}_{i_2,j_2} + \left(z^{(2)}_{i_2,j_2}\right)^2\right)\Bigg] = \\ &= \mathbb{E}\left[\left(z^{(1)}_{i_1,j_1}\right)^2\left(z^{(1)}_{i_2,j_2}\right)^2\right] - \\ &\qquad 2\,\mathbb{E}\left[\left(z^{(1)}_{i_1,j_1}\right)^2 z^{(1)}_{i_2,j_2}z^{(2)}_{i_2,j_2}\right] + \\ &\qquad \mathbb{E}\left[\left(z^{(1)}_{i_1,j_1}\right)^2\left(z^{(2)}_{i_2,j_2}\right)^2\right] - \\ &\qquad 2\,\mathbb{E}\left[z^{(1)}_{i_1,j_1}z^{(2)}_{i_1,j_1}\left(z^{(1)}_{i_2,j_2}\right)^2\right] + \\ &\qquad 4\,\mathbb{E}\left[z^{(1)}_{i_1,j_1}z^{(2)}_{i_1,j_1}z^{(1)}_{i_2,j_2}z^{(2)}_{i_2,j_2}\right] - \\ &\qquad 2\,\mathbb{E}\left[z^{(1)}_{i_1,j_1}z^{(2)}_{i_1,j_1}\left(z^{(2)}_{i_2,j_2}\right)^2\right] + \\ &\qquad \mathbb{E}\left[\left(z^{(2)}_{i_1,j_1}\right)^2\left(z^{(1)}_{i_2,j_2}\right)^2\right] - \\ &\qquad 2\,\mathbb{E}\left[\left(z^{(2)}_{i_1,j_1}\right)^2 z^{(1)}_{i_2,j_2}z^{(2)}_{i_2,j_2}\right] + \\ &\qquad \mathbb{E}\left[\left(z^{(2)}_{i_1,j_1}\right)^2\left(z^{(2)}_{i_2,j_2}\right)^2\right]\,.\end{aligned} \tag{32}$$

All summands in equation 32 are the fourth moments of multivariate Gaussian distribution. Using formula for the Gaussian moment, we get

$$\begin{aligned}&\mathbb{E}\left[\left(z^{(1)}_{i_1,j_1}\right)^2\left(z^{(1)}_{i_2,j_2}\right)^2\right] = \\ &\qquad = \sigma_z^4 + 2\left(R_{ZZ}(i_1-j_1, i_2-j_2)\right)^2\end{aligned} \tag{33}$$

$$\begin{aligned}&2\,\mathbb{E}\left[\left(z^{(1)}_{i_1,j_1}\right)^2 z^{(1)}_{i_2,j_2}z^{(2)}_{i_2,j_2}\right] = \\ &\qquad = 2\,\mathbb{E}\left[\left(z^{(1)}_{i_1,j_1}\right)^2 z^{(1)}_{i_2,j_2}\right]\mathbb{E}\left[z^{(2)}_{i_2,j_2}\right] = 0\end{aligned} \tag{34}$$

$$\begin{aligned}&\mathbb{E}\left[\left(z^{(1)}_{i_1,j_1}\right)^2\left(z^{(2)}_{i_2,j_2}\right)^2\right] = \\ &\qquad = \mathbb{E}\left[\left(z^{(1)}_{i_1,j_1}\right)^2\right]\mathbb{E}\left[\left(z^{(2)}_{i_2,j_2}\right)^2\right] = \sigma_z^4\end{aligned} \tag{35}$$

$$2\,\mathbb{E}\left[z^{(1)}_{i_1,j_1} z^{(2)}_{i_1,j_1}\left(z^{(1)}_{i_2,j_2}\right)^2\right] = = 2\,\mathbb{E}\left[z^{(1)}_{i_1,j_1}\left(z^{(1)}_{i_2,j_2}\right)^2\right]\mathbb{E}\left[z^{(2)}_{i_1,j_1}\right] = 0 \tag{36}$$

$$4\,\mathbb{E}\left[z^{(1)}_{i_1,j_1} z^{(2)}_{i_1,j_1} z^{(1)}_{i_2,j_2} z^{(2)}_{i_2,j_2}\right] = = 4\,\mathbb{E}\left[z^{(1)}_{i_1,j_1} z^{(1)}_{i_2,j_2}\right]\mathbb{E}\left[z^{(2)}_{i_1,j_1} z^{(2)}_{i_2,j_2}\right] = = 4\left(R_{ZZ}(i_1-j_1,i_2-j_2)\right)^2 \tag{37}$$

$$2\,\mathbb{E}\left[z^{(1)}_{i_1,j_1} z^{(2)}_{i_1,j_1}\left(z^{(2)}_{i_2,j_2}\right)^2\right] = = 2\,\mathbb{E}\left[z^{(1)}_{i_1,j_1}\right]\mathbb{E}\left[z^{(2)}_{i_1,j_1}\left(z^{(2)}_{i_2,j_2}\right)^2\right] = 0 \tag{38}$$

$$\mathbb{E}\left[\left(z^{(2)}_{i_1,j_1}\right)^2\left(z^{(1)}_{i_2,j_2}\right)^2\right] = = \mathbb{E}\left[\left(z^{(2)}_{i_1,j_1}\right)^2\right]\mathbb{E}\left[\left(z^{(1)}_{i_2,j_2}\right)^2\right] = \sigma_z^4 \tag{39}$$

$$2\,\mathbb{E}\left[\left(z^{(2)}_{i_1,j_1}\right)^2 z^{(1)}_{i_2,j_2} z^{(2)}_{i_2,j_2}\right] = = 2\,\mathbb{E}\left[\left(z^{(2)}_{i_1,j_1}\right)^2 z^{(2)}_{i_2,j_2}\right]\mathbb{E}\left[z^{(1)}_{i_2,j_2}\right] = 0 \tag{40}$$

$$\mathbb{E}\left[\left(z^{(2)}_{i_1,j_1}\right)^2\left(z^{(2)}_{i_2,j_2}\right)^2\right] = = \sigma_z^4 + 2\left(R_{ZZ}(i_1-j_1,i_2-j_2)\right)^2 . \tag{41}$$

Summarizing the expressions in equations 33-41, we get

$$\mathbb{E}\left[\left(d^{(z)}_{i_1,j_1}\right)^2\left(d^{(z)}_{i_2,j_2}\right)^2\right] = \sigma_z^4 + 2\left(R_{ZZ}(i_1-j_1,i_2-j_2)\right)^2 + \sigma_z^4 + 4\left(R_{ZZ}(i_1-j_1,i_2-j_2)\right)^2 + \sigma_z^4 + \sigma_z^4 + 2\left(R_{ZZ}(i_1-j_1,i_2-j_2)\right)^2 = = 4\sigma_z^4 + 8\left(R_{ZZ}(i_1-j_1,i_2-j_2)\right)^2 \ . \tag{42}$$

Substituting equation 42 into 31, we get

$$\mathbb{E}\left[\delta_z^2\right] = \frac{1}{n^4}\sum_{i_1,j_1,i_2,j_2=1}^{n} E\left[\left(d^{(z)}_{i_1,j_1}\right)^2\left(d^{(z)}_{i_2,j_2}\right)^2\right] = = \frac{1}{n^4}\sum_{i_1,j_1,i_2,j_2=1}^{n}\left(4\sigma_z^4 + 8\left(R_{ZZ}(i_1-j_1,i_2-j_2)\right)^2\right) = = 4\sigma_z^4 + \frac{8}{n^4}\sum_{i_1,j_1,i_2,j_2=1}^{n}\left(R_{ZZ}(i_1-j_1,i_2-j_2)\right)^2 \tag{43}$$

Substituting equation 43 into 30, we have

$$var\left[\delta_y\right] \geq \mathbb{E}\left[\delta_z^2\right] - 4\sigma_z^4 = = \frac{8}{n^4}\sum_{i_1,j_1,i_2,j_2=1}^{n}\left(R_{ZZ}\left(i_1-j_1,i_2-j_2\right)\right)^2 = = \frac{8}{n^2}\sigma_z^4\left(\frac{1}{n^2}\sum_{i_1,j_1,i_2,j_2=1}^{n}\left(\frac{R_{ZZ}\left(i_1-j_1,i_2-j_2\right)}{\sigma_z^2}\right)^2\right) . \tag{44}$$

Finally, we get

$$var\left[\delta_y\right] \geq \frac{8}{n^2}\sigma_z^4\rho \ , \tag{45}$$

where

$$\rho = \frac{1}{n^2}\sum_{i_1,j_1,i_2,j_2=1}^{n}\left(\frac{R_{ZZ}\left(i_1-j_1,i_2-j_2\right)}{\sigma_z^2}\right)^2 .$$

□

C.2. Proof of Proposition 9

Proof. Substituting equation 17 into ρ in Theorem 8, we get

$$\rho = \frac{1}{n^2}\sum_{i_1,j_1,i_2,j_2=1}^{n}\left(\frac{R_{ZZ}\left(i_1-j_1,i_2-j_2\right)}{\sigma_z^2}\right)^2 = = \frac{1}{n^2\sigma_z^4}\sum_{i_1,j_1,i_2,j_2=1}^{n} g^2\left(i_1-j_1\right)g^2\left(i_2-j_2\right) = = \frac{1}{n^2\sigma_z^4}\sum_{i_1,j_1=1}^{n} g^2\left(i_1-j_1\right)\sum_{i_2,j_2=1}^{n} g^2\left(i_2-j_2\right) = = \frac{1}{n^2\sigma_z^4}\left(\sum_{i,j=1}^{n} g^2\left(i-j\right)\right)^2 . \tag{46}$$

It is easy to see that if $a \geq b > 0$, then for any τ,

$$1 - \frac{|\tau|}{a} \geq 1 - \frac{|\tau|}{b} \tag{47}$$

and

$$\max\left\{1 - \frac{|\tau|}{a}, 0\right\} \geq \max\left\{1 - \frac{|\tau|}{b}, 0\right\} \ . \tag{48}$$

We denote $r = \min\{\lfloor\theta\rfloor, n\}$, where $\lfloor\cdot\rfloor$ stands for the floor function. Applying Lemma 10 on equation 46, substituting the expression for $g\left(\tau\right)$ in equation 17, and using inequalities 47 and 48, we get

$$
\begin{aligned}
\rho &= \frac{1}{n^2\sigma_z^4}\left(\sum_{i,j=1}^{n} g^2\left(i-j\right)\right)^2 = \\
&= \frac{1}{n^2\sigma_z^4}\left(n\sum_{\tau=-(n-1)}^{n-1}\left(1-\frac{|\tau|}{n}\right)g^2\left(\tau\right)\right)^2 = \\
&= \frac{1}{\sigma_z^4}\left(\sum_{\tau=-(n-1)}^{n-1}\left(1-\frac{|\tau|}{n}\right)\times\right. \\
&\qquad\qquad \left.\left(\sigma_z\cdot\max\left\{1-\frac{|\tau|}{\theta},0\right\}\right)^2\right)^2 \geq \\
&\geq \left(\sum_{\tau=-(n-1)}^{n-1}\left(1-\frac{|\tau|}{n}\right)\times\right. \\
&\qquad\qquad \left.\left(\max\left\{1-\frac{|\tau|}{\lfloor\theta\rfloor},0\right\}\right)^2\right)^2 = \\
&= \left(\sum_{\tau=-(r-1)}^{r-1}\left(1-\frac{|\tau|}{n}\right)\left(1-\frac{|\tau|}{\lfloor\theta\rfloor}\right)^2\right)^2 \geq \\
&\geq \left(\sum_{\tau=-(r-1)}^{r-1}\left(1-\frac{|\tau|}{r}\right)\left(1-\frac{|\tau|}{r}\right)^2\right)^2 = \\
&= \left(\sum_{\tau=-(r-1)}^{r-1}\left(1-\frac{|\tau|}{r}\right)^3\right)^2 = \\
&= \left(1+2\sum_{\tau=1}^{r-1}\left(1-\frac{\tau}{r}\right)^3\right)^2 = \\
&= \left(1+\frac{2}{r^3}\sum_{\tau=1}^{r-1}\left(r-\tau\right)^3\right)^2 .
\end{aligned} \tag{49}
$$

Applying variable change $\tau = r - \tau$ and using the formula for the sum of cubes, we get

$$
\begin{aligned}
\rho &\geq \left(1+\frac{2}{r^3}\sum_{\tau=1}^{r-1}\left(r-\tau\right)^3\right)^2 = \\
&= \left(1+\frac{2}{r^3}\sum_{\tau=1}^{r-1}\tau^3\right)^2 = \\
&= \left(1+\frac{2}{r^3}\frac{\left(r-1\right)^2 r^2}{4}\right)^2 = \\
&= \left(1+\frac{\left(r-1\right)^2}{2r}\right)^2 = \\
&= \left(\frac{2r+r^2-2r+1}{2r}\right)^2 = \\
&= \frac{1}{4}\left(r+\frac{1}{r}\right)^2 .
\end{aligned} \tag{50}
$$

□

Lemma 10.

$$
\sum_{i,j=1}^{n} f\left(i-j\right) = n\sum_{\tau=-(n-1)}^{n-1}\left(1-\frac{|\tau|}{n}\right)f\left(\tau\right) .
$$

Proof. We start by splitting the inner sum into two ranges, $1 \leq j \leq i$ and $i \leq j \leq n$

$$
\begin{aligned}
&\sum_{i,j=1}^{n} f\left(i-j\right) = \\
&= \sum_{i=1}^{n}\left(\sum_{j=1}^{i} f\left(i-j\right)+\sum_{j=i}^{n} f\left(i-j\right)-f\left(i-i\right)\right) = \\
&= \sum_{i=1}^{n}\sum_{j=1}^{i} f\left(i-j\right)+\sum_{i=1}^{n}\sum_{j=i}^{n} f\left(i-j\right)-nf\left(0\right) .
\end{aligned} \tag{51}
$$

We calculate the first double sum. Substituting $\tau = i-j+1$

and changing the order of the double summation, we have

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{i} f(i-j) &= \sum_{i=1}^{n}\sum_{\tau=1}^{i} f(\tau-1) = \\ &= \sum_{1\leq\tau\leq i\leq n} f(\tau-1) = \\ &= \sum_{\tau=1}^{n}\sum_{i=\tau}^{n} f(\tau-1) = \\ &= \sum_{\tau=1}^{n} f(\tau-1)\sum_{i=\tau}^{n} 1 = \\ &= \sum_{\tau=1}^{n} f(\tau-1)(n-(\tau-1)) = \\ &= \sum_{\tau=0}^{n-1} (n-\tau) f(\tau) = \\ &= \sum_{\tau=0}^{n-1} (n-|\tau|) f(\tau) \ .\end{aligned} \tag{52}$$

Substituting variable change $t = n - j + 1$, $k = n - i + 1$ into the second double summation in equation 51, we get

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=i}^{n} f(i-j) &= \sum_{k=1}^{n}\sum_{t=1}^{k} f(t-k) = \\ &= \sum_{k=1}^{n}\sum_{t=1}^{k} g(k-t) \ ,\end{aligned} \tag{53}$$

where $g(\tau) = f(-\tau)$. Substituting equation 52 into 53, we have

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=i}^{n} f(i-j) &= \sum_{\tau=0}^{n-1} (n-|\tau|) g(\tau) = \\ &= \sum_{\tau=-(n-1)}^{0} (n-|\tau|) f(\tau) \ .\end{aligned} \tag{54}$$

Substituting equations 52 and 54 into equation 51 finalizes the proof

$$\begin{aligned}&\sum_{i,j=1}^{n} f(i-j) = \\ &= \sum_{i=1}^{n}\sum_{j=1}^{i} f(i-j) + \sum_{i=1}^{n}\sum_{j=i}^{n} f(i-j) - nf(0) \\ &= \sum_{\tau=0}^{n-1} (n-|\tau|) f(\tau) + \\ &\qquad \sum_{\tau=-(n-1)}^{0} (n-|\tau|) f(\tau) - nf(0) = \\ &= \sum_{\tau=-(n-1)}^{n-1} (n-|\tau|) f(\tau) \ .\end{aligned} \tag{55}$$

□

	$\sigma = 5$	$\sigma = 10$	$\sigma = 15$	$\sigma = 20$
$k = 2$	19	27	31	33
$k = 3$	19	37	41	43
$k = 4$	25	43	43	45

Table 1. Patch size n for correlated Gaussian experiments.

ISO	1600	3200	6400	12800	25600
n	15	25	27	35	37

Table 2. Patch size n for experiments with real-world noise.

D. Training Details

In all experiments, we train the denoisers for 30 epochs using the Adam [3] optimizer while decreasing the learning rate by 0.5 every 5 epochs. The initial learning rate is 0.001 for the correlated Gaussian experiment and 0.00001 for the experinet with real-world noise. During the training, we extract random patches of size 50×50 from the training data and use a batch size of 32 for correlated Gaussian denoising and 128 for the experiment with real-world noise. The values of n used in the patch search are summarized in tables 1 and 2. Examples of CRVD sequences are presented in figure 1.

E. Additional Results

This section provides additional visual comparisons of our framework versus leading competitors. Figures 2, 3, and 4 show denoising examples from the correlated Gaussian experiments, while figures 5, 6, and 7 present real-world denoising examples.

(a) Frame 1 (b) Frame 2 (c) Frame 3
(d) Frame 4 (e) Frame 5 (f) Frame 6
(g) Frame 7 (h) Frame 1 (i) Frame 2
(j) Frame 3 (k) Frame 4 (l) Frame 5
(m) Frame 6 (n) Frame 7

Figure 1. Examples of CRVD sequences. Figures 1a to 1g present sequence 5, while figures 1h to 1n shown sequence 7.

References

[1] Patrick Billingsley. *Probability and measure*. John Wiley & Sons, 2008. 1

[2] Ildar A Ibragimov. Some limit theorems for stationary processes. *Theory of Probability & Its Applications*, 7(4):349–382, 1962. 1

[3] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2015. 8

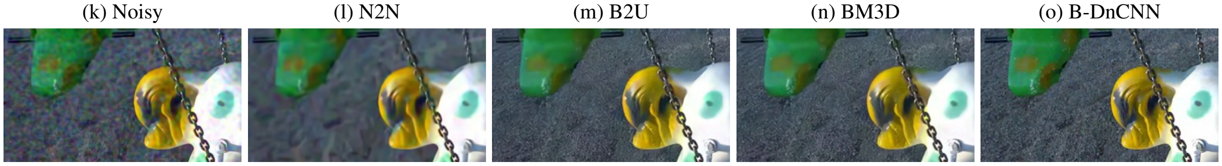
(a) Noisy 28.12 / 0.522
(b) N2N 28.70 / 0.555
(c) B2U 24.27 / 0.335
(d) BM3D 31.41 / 0.726
(e) B-DnCNN 30.55 / 0.677
(f) R2R 34.05 / 0.836
(g) BM3D-O 36.37 / 0.939
(h) PC-UNet 38.65 / 0.965
(i) PC-DnCNN 39.02 / 0.968
(j) Clean
(k) Noisy
(l) N2N
(m) B2U
(n) BM3D
(o) B-DnCNN
(p) R2R
(q) BM3D-O
(r) PC-UNet
(s) PC-DnCNN
(t) Clean
(u) Noisy 28.11 / 0.596
(v) N2N 28.52 / 0.620
(w) B2U 24.09 / 0.363
(x) BM3D 30.26 / 0.725
(y) B-DnCNN 29.48 / 0.686
(z) R2R 31.83 / 0.781
(aa) BM3D-O 33.84 / 0.876
(ab) PC-UNet 36.26 / 0.934
(ac) PC-DnCNN 36.42 / 0.936
(ad) Clean
(ae) Noisy
(af) N2N
(ag) B2U
(ah) BM3D
(ai) B-DnCNN
(aj) R2R
(ak) BM3D-O
(al) PC-UNet
(am) PC-DnCNN
(an) Clean

Figure 2. Denoising examples with correlated Gaussian noise. The first four rows show frame 13 of the sequence *planes-crossing* with $\sigma = 10$ and $k = 3$. The last four rows present frame 5 of the sequence *helicopter* with $\sigma = 10$ and $k = 4$. As can be seen, oracle BM3D leaves a substantial amount of low-frequency noise unfiltered, while other algorithms, except ours (PC-UNet and PC-DnCNN), do not succeed in removing the noise.

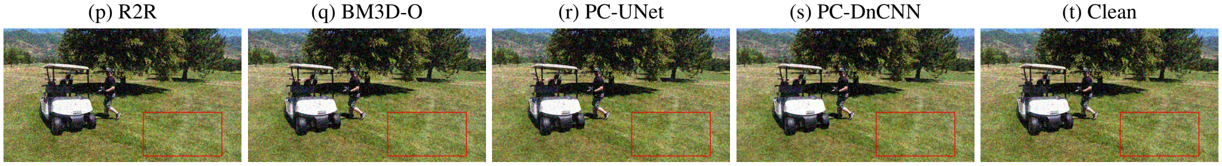
(a) Noisy 24.60 / 0.710 | (b) N2N 25.00 / 0.725 | (c) B2U 21.59 / 0.531 | (d) BM3D 26.07 / 0.757 | (e) B-DnCNN 25.73 / 0.762

(f) R2R 26.69 / 0.756 | (g) BM3D-O 25.85 / 0.703 | (h) PC-UNet 29.30 / 0.877 | (i) PC-DnCNN 29.29 / 0.881 | (j) Clean

(k) Noisy | (l) N2N | (m) B2U | (n) BM3D | (o) B-DnCNN

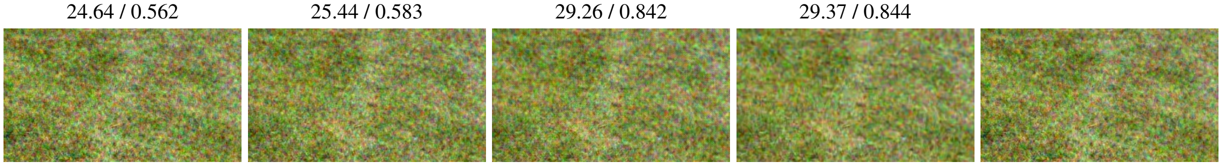


(p) R2R | (q) BM3D-O | (r) PC-UNet | (s) PC-DnCNN | (t) Clean

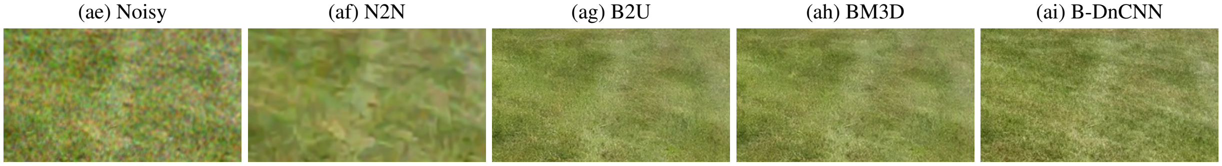


(u) Noisy 22.10 / 0.543 | (v) N2N 22.73 / 0.569 | (w) B2U 21.98 / 0.482 | (x) BM3D 23.82 / 0.582 | (y) B-DnCNN 23.36 / 0.608

(z) R2R 24.64 / 0.562 | (aa) BM3D-O 25.44 / 0.583 | (ab) PC-UNet 29.26 / 0.842 | (ac) PC-DnCNN 29.37 / 0.844 | (ad) Clean

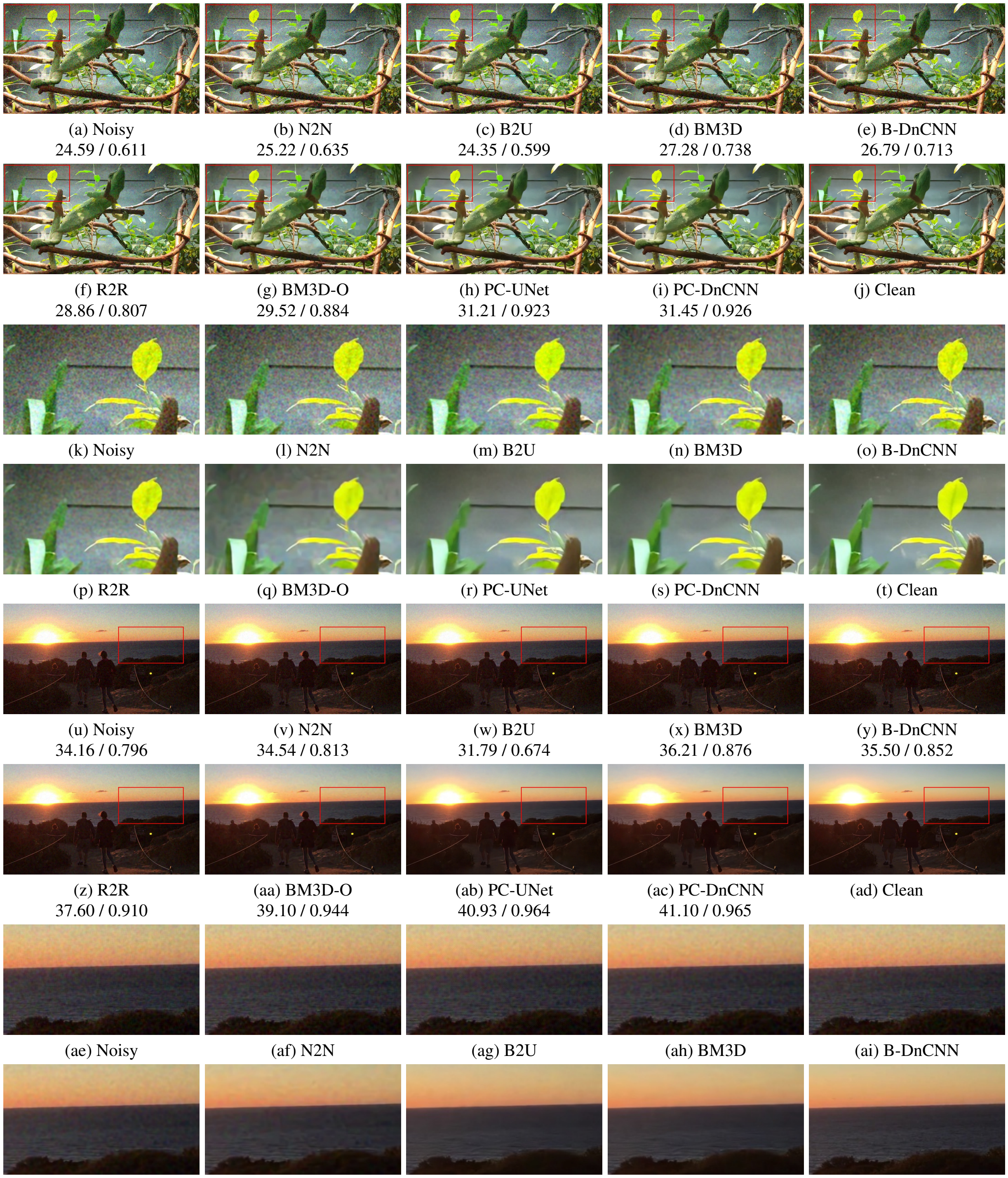


(ae) Noisy | (af) N2N | (ag) B2U | (ah) BM3D | (ai) B-DnCNN



(aj) R2R | (ak) BM3D-O | (al) PC-UNet | (am) PC-DnCNN | (an) Clean

Figure 3. Denoising examples with correlated Gaussian noise. The first four rows show frame 5 of the sequence *carousel* with $\sigma = 15$ and $k = 4$. The last four rows present frame 13 of the sequence *golf* with $\sigma = 20$ and $k = 4$. As can be seen, oracle BM3D produces blurred images while leaving a substantial amount of low-frequency noise unfiltered. Other algorithms, except ours (PC-UNet and PC-DnCNN), do not succeed in removing the noise.



(a) Noisy 24.59 / 0.611 (b) N2N 25.22 / 0.635 (c) B2U 24.35 / 0.599 (d) BM3D 27.28 / 0.738 (e) B-DnCNN 26.79 / 0.713

(f) R2R 28.86 / 0.807 (g) BM3D-O 29.52 / 0.884 (h) PC-UNet 31.21 / 0.923 (i) PC-DnCNN 31.45 / 0.926 (j) Clean

(k) Noisy (l) N2N (m) B2U (n) BM3D (o) B-DnCNN

(p) R2R (q) BM3D-O (r) PC-UNet (s) PC-DnCNN (t) Clean

(u) Noisy 34.16 / 0.796 (v) N2N 34.54 / 0.813 (w) B2U 31.79 / 0.674 (x) BM3D 36.21 / 0.876 (y) B-DnCNN 35.50 / 0.852

(z) R2R 37.60 / 0.910 (aa) BM3D-O 39.10 / 0.944 (ab) PC-UNet 40.93 / 0.964 (ac) PC-DnCNN 41.10 / 0.965 (ad) Clean

(ae) Noisy (af) N2N (ag) B2U (ah) BM3D (ai) B-DnCNN

(aj) R2R (ak) BM3D-O (al) PC-UNet (am) PC-DnCNN (an) Clean

Figure 4. Denoising examples with correlated Gaussian noise. The first four rows show frame 24 of the sequence *chameleon* with $\sigma = 15$ and $k = 3$. The last four rows present frame 5 of the sequence *people-sunset* with $\sigma = 5$ and $k = 4$. As can be seen, oracle BM3D leaves a substantial amount of low-frequency noise unfiltered, while other algorithms, except ours (PC-UNet and PC-DnCNN), do not succeed in removing the noise.

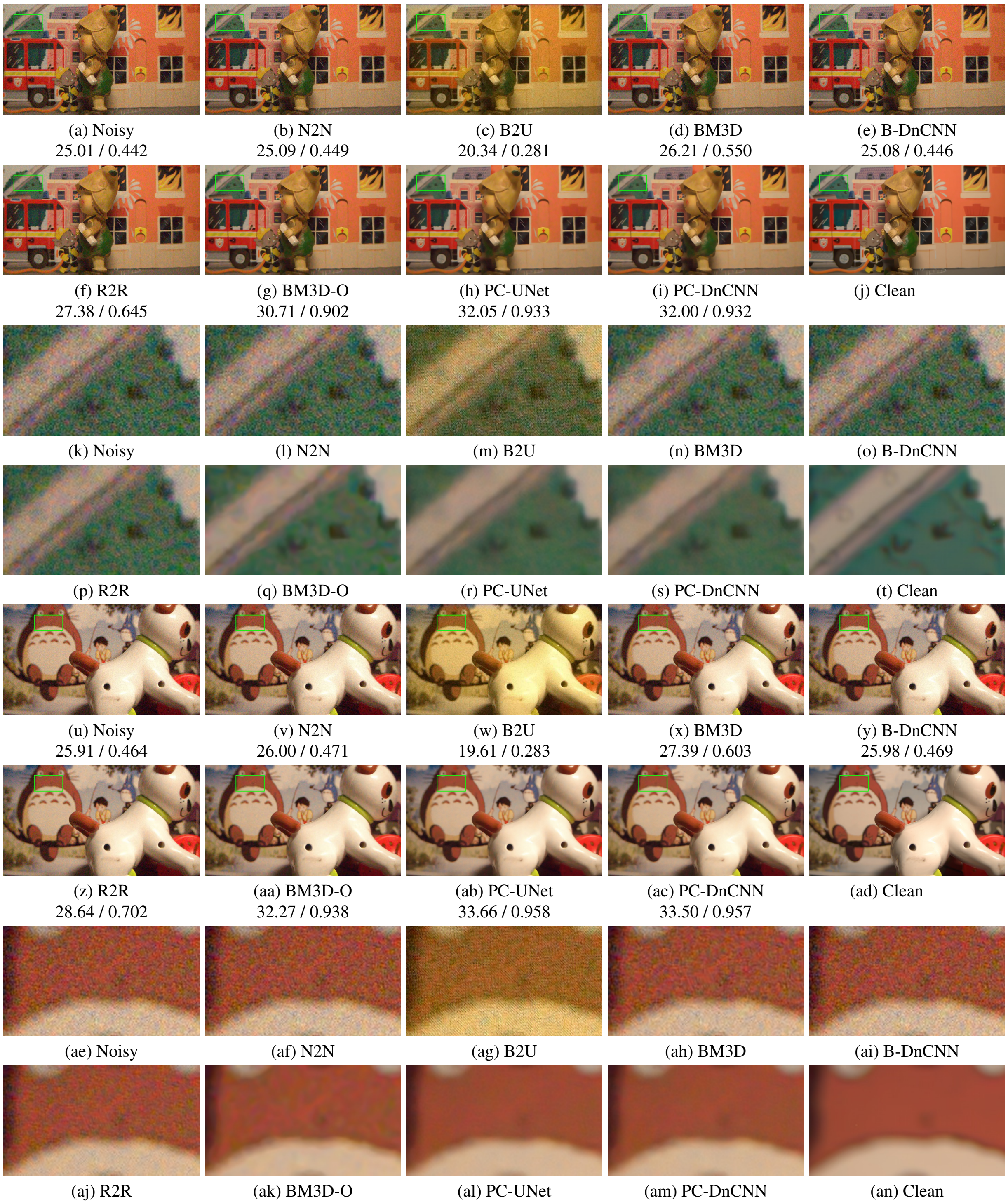
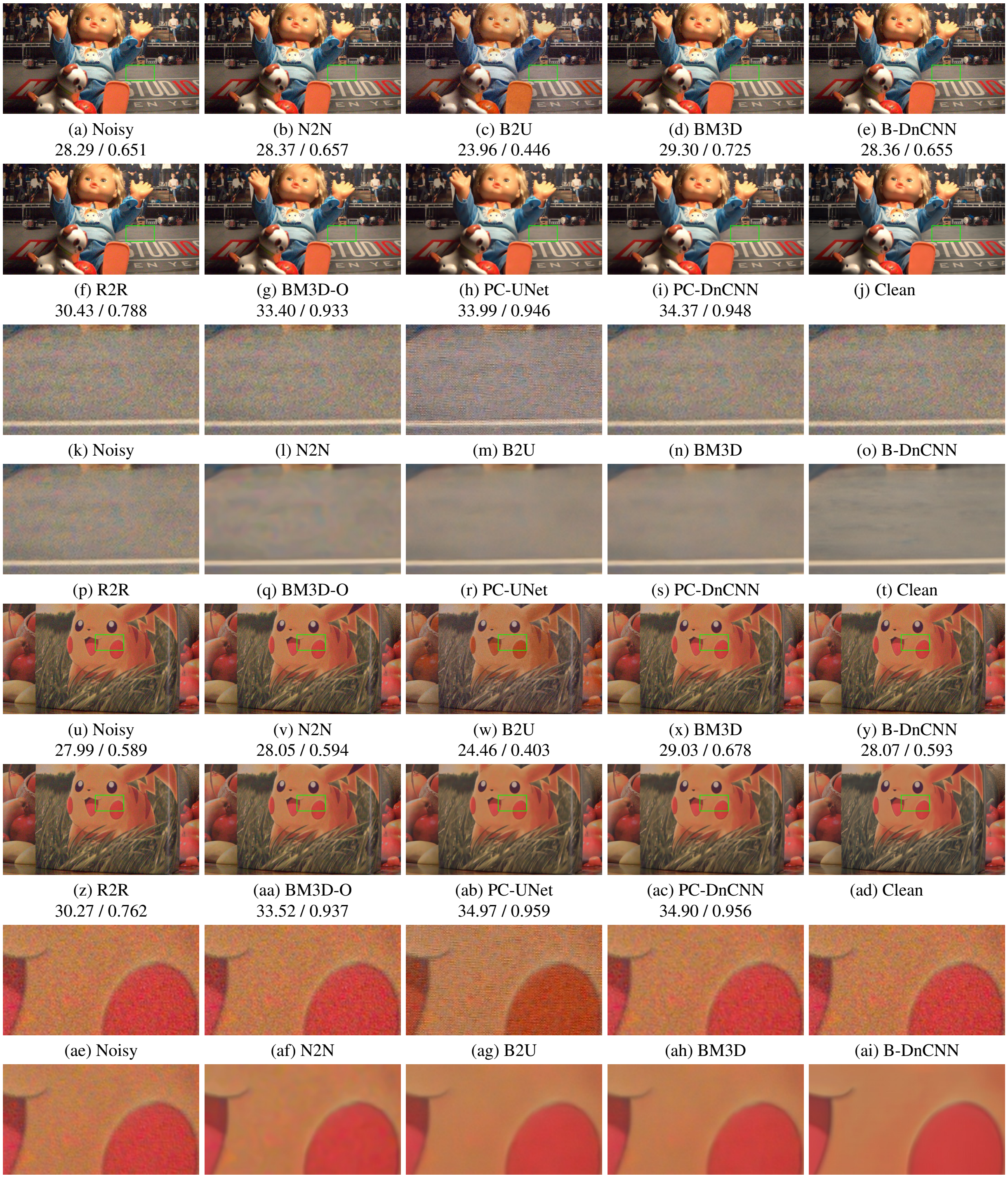


Figure 5. Denoising examples with real-world noise. The first four rows show frame 6 of scene 9. The last four rows present frame 4 of scene 2. Both images are captured with ISO 25600. As can be seen, oracle BM3D leaves a substantial amount of low-frequency noise unfiltered, while other algorithms, except ours (PC-UNet and PC-DnCNN), do not succeed in removing the noise.



(a) Noisy 28.29 / 0.651
(b) N2N 28.37 / 0.657
(c) B2U 23.96 / 0.446
(d) BM3D 29.30 / 0.725
(e) B-DnCNN 28.36 / 0.655
(f) R2R 30.43 / 0.788
(g) BM3D-O 33.40 / 0.933
(h) PC-UNet 33.99 / 0.946
(i) PC-DnCNN 34.37 / 0.948
(j) Clean
(k) Noisy
(l) N2N
(m) B2U
(n) BM3D
(o) B-DnCNN
(p) R2R
(q) BM3D-O
(r) PC-UNet
(s) PC-DnCNN
(t) Clean
(u) Noisy 27.99 / 0.589
(v) N2N 28.05 / 0.594
(w) B2U 24.46 / 0.403
(x) BM3D 29.03 / 0.678
(y) B-DnCNN 28.07 / 0.593
(z) R2R 30.27 / 0.762
(aa) BM3D-O 33.52 / 0.937
(ab) PC-UNet 34.97 / 0.959
(ac) PC-DnCNN 34.90 / 0.956
(ad) Clean
(ae) Noisy
(af) N2N
(ag) B2U
(ah) BM3D
(ai) B-DnCNN
(aj) R2R
(ak) BM3D-O
(al) PC-UNet
(am) PC-DnCNN
(an) Clean

Figure 6. Denoising examples with real-world noise. The first four rows show frame 3 of scene 6. The last four rows present frame 6 of scene 11. Both images are captured with ISO 12800. As can be seen, oracle BM3D leaves a substantial amount of low-frequency noise unfiltered, while other algorithms, except ours (PC-UNet and PC-DnCNN), do not succeed in removing the noise.

(a) Noisy 32.88 / 0.864
(b) N2N 32.96 / 0.866
(c) B2U 4.36 / -0.001
(d) BM3D 33.68 / 0.894
(e) B-DnCNN 32.95 / 0.866
(f) R2R 34.65 / 0.921
(g) BM3D-O 34.59 / 0.925
(h) PC-UNet 35.82 / 0.948
(i) PC-DnCNN 35.80 / 0.947
(j) Clean
(k) Noisy
(l) N2N
(m) B2U
(n) BM3D
(o) B-DnCNN
(p) R2R
(q) BM3D-O
(r) PC-UNet
(s) PC-DnCNN
(t) Clean
(u) Noisy 35.56 / 0.886
(v) N2N 35.54 / 0.886
(w) B2U 4.66 / 0.012
(x) BM3D 36.57 / 0.916
(y) B-DnCNN 35.67 / 0.888
(z) R2R 37.65 / 0.941
(aa) BM3D-O 39.17 / 0.966
(ab) PC-UNet 40.14 / 0.976
(ac) PC-DnCNN 40.16 / 0.976
(ad) Clean
(ae) Noisy
(af) N2N
(ag) B2U
(ah) BM3D
(ai) B-DnCNN
(aj) R2R
(ak) BM3D-O
(al) PC-UNet
(am) PC-DnCNN
(an) Clean

Figure 7. Denoising examples with real-world noise. The first four rows show frame 6 of scene 2 captured with ISO 6400. The last four rows present frame 5 of scene 1 taken with ISO 3200. As can be seen, oracle BM3D leaves a noticeable amount of low-frequency noise unfiltered, while other algorithms, except ours (PC-Unet and PC-DnCNN), do not succeed in removing the noise.