Generalized Recorrupted-to-Recorrupted: Self-Supervised Learning Beyond Gaussian Noise (Supplemental Materials)

This supplementary material is divided into three parts: (i) proof of the propositions and theorem 1, (ii) additional information and derivation of the link between GR2R and SURE losses, and (iii) additional results.

1 Proofs

Proof of Proposition 1

Proof. The R2R loss can be re-expressed as

$$\mathbb{E}_{\boldsymbol{y}_1,\boldsymbol{y}_2|\boldsymbol{y}}\|f(\boldsymbol{y}_1)-\boldsymbol{y}_2\|_2^2=\mathbb{E}_{\boldsymbol{y}_1|\boldsymbol{y}}\|f(\boldsymbol{y}_1)\|_2^2+\mathbb{E}_{\boldsymbol{y}_2|\boldsymbol{y}}\|\boldsymbol{y}_2\|_2^2-2\sum_{i=1}^{n}\mathbb{E}_{\boldsymbol{y}_1,y_{2,i}|\boldsymbol{x}}\,y_{2,i}f_i(\boldsymbol{y}_1),$$

where $y_{2,i}\in\mathbb{R}$ denotes the ith entry of $\boldsymbol{y}_2$. If the following equality

$$\mathbb{E}_{\boldsymbol{y}_1,y_{2,i}|\boldsymbol{x}}\,y_{2,i}\,f_i(\boldsymbol{y}_1)=x_i\,\mathbb{E}_{\boldsymbol{y}_1|\boldsymbol{x}}\,f_i(\boldsymbol{y}_1), \tag{1}$$

holds (below is how to ensure this) for all $i=1,\dots,n$, then

$$\begin{aligned}\mathbb{E}_{\boldsymbol{y}_1,\boldsymbol{y}_2|\boldsymbol{x}}\|f(\boldsymbol{y}_1)-\boldsymbol{y}_2\|_2^2&=\mathbb{E}_{\boldsymbol{y}_1|\boldsymbol{x}}\|f(\boldsymbol{y}_1)\|_2^2+\mathbb{E}_{\boldsymbol{y}_2|\boldsymbol{x}}\|\boldsymbol{y}_2\|_2^2-2\sum_{i=1}^{n}x_i\mathbb{E}_{\boldsymbol{y}_1|\boldsymbol{x}}\,f_i(\boldsymbol{y}_1)\\&=\mathbb{E}_{\boldsymbol{y}_2|\boldsymbol{x}}\|f(\boldsymbol{y}_1)-\boldsymbol{x}\|_2^2-\|\boldsymbol{x}\|_2^2+\mathbb{E}_{\boldsymbol{y}_2|\boldsymbol{x}}\|\boldsymbol{y}_2\|_2^2\\&=\mathbb{E}_{\boldsymbol{y}_2|\boldsymbol{x}}\|f(\boldsymbol{y}_1)-\boldsymbol{x}\|_2^2+\text{const},\end{aligned} \tag{2}$$

where the second line comes from adding and subtracting $||\boldsymbol{x}||_2$.

A sufficient (but not necessary) condition for (1) to hold is that i) $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ are independent and ii) $\mathbb{E}_{\boldsymbol{y}_2|\boldsymbol{x}}\boldsymbol{y}_2=\boldsymbol{x}$. If this conditions hold, we trivially have $\mathbb{E}_{\boldsymbol{y}_1,\boldsymbol{y}_{2,i}|\boldsymbol{x}}y_{2,i}f_i(\boldsymbol{y}_1)=\left(\mathbb{E}_{\boldsymbol{y}_{2,i}|\boldsymbol{x}}y_{2,i}\right)\left(\mathbb{E}_{\boldsymbol{y}_1|\boldsymbol{x}}f_i(\boldsymbol{y}_1)\right)=x_i\mathbb{E}_{\boldsymbol{y}_1|\boldsymbol{x}}f_i(\boldsymbol{y}_1)$ for $i=1,\dots,n$. We will analyze the necessary condition (beyond independence) for the case of additive noise where $\boldsymbol{y}=\boldsymbol{x}+\boldsymbol{\epsilon}$ where ϵ is sampled from a symmetric noise distribution that is independent across pixel entries. We construct pairs $\boldsymbol{y}_1=\boldsymbol{y}+\boldsymbol{\omega}\tau$ and $\boldsymbol{y}_2=\boldsymbol{y}-\boldsymbol{\omega}/\tau$, with $\tau>0$ and $\boldsymbol{\omega}$ sampled from the same distribution as $\boldsymbol{\epsilon}$. Due to the independence across entries, we will drop the ith indices and define the scalar function $f_i(\cdot;\boldsymbol{y}_{1,-i}):\mathbb{R}\mapsto\mathbb{R}$, such that the left-hand side of (1) can be simplified to

$$\mathbb{E}_{\epsilon_i,\omega_i}(\underbrace{x_i+\epsilon_i-\omega_i/\tau}_{y_{2,i}})f_i(\underbrace{x_i+\epsilon_i+\tau\omega_i}_{y_{1,i}};\boldsymbol{y}_{1,-i})=x_i\,\mathbb{E}_{\epsilon_i,\omega_i}f_i(x_i+\epsilon_i+\tau\omega_i,\boldsymbol{y}_{-i})-\mathbb{E}_{\epsilon_i,\omega_i}\,(\epsilon_i-\frac{\omega_i}{\tau})f_i(x_i+\epsilon_i+\tau\omega_i;\boldsymbol{y}_{1,-i}), \tag{3}$$

where ω_i, x_i and ϵ_i refer to the ith entry, and are thus one-dimensional. In this additive case, showing (3) is equivalent to showing that

$$\mathbb{E}_{\epsilon_i,\omega_i}\,(\epsilon_i-\frac{\omega_i}{\tau})f_i(x_i+\epsilon_i+\tau\omega_i;\boldsymbol{y}_{1,-i})=0. \tag{4}$$

Assuming that f_i is analytic (that is, is infinitely differentiable and has a convergent Taylor expansion) and performing a Taylor expansion of f_i around x_i, we obtain

$$\mathbb{E}_{\boldsymbol{\epsilon}_{-i},\boldsymbol{\omega}_{-i}}\mathbb{E}_{\epsilon_i,\omega_i}\,(\epsilon_i-\frac{\omega_i}{\tau})f_i(x_i+\epsilon_i+\omega_i\tau;\boldsymbol{y}_{1,-i})=\mathbb{E}_{\boldsymbol{\epsilon}_{-i},\boldsymbol{\omega}_{-i}}\mathbb{E}_{\epsilon_i,\omega_i}\,(\epsilon_i-\frac{\omega_i}{\tau})\sum_{k\geq0}\frac{1}{k!}\frac{\partial^k f_i}{\partial {x_i}^k}(x_i;\boldsymbol{y}_{1,-i})\,(\epsilon_i+\tau\omega_i)^k \tag{5}$$

$$=\sum_{k\geq0}\frac{1}{k!}\mathbb{E}_{\boldsymbol{\epsilon}_{-i},\boldsymbol{\omega}_{-i}}\{\frac{\partial^k f_i}{\partial {x_i}^k}(x_i;\boldsymbol{y}_{1,-i})\}\,\mathbb{E}_{\epsilon_i,\omega_i}\,(\epsilon_i-\frac{\omega_i}{\tau})(\epsilon_i+\tau\omega_i)^k \tag{6}$$

where the case $k=0$ is removed from the last sum as $\mathbb{E}_{\epsilon_i,\omega_i}\{\epsilon_i-\frac{\omega_i}{\tau}\}=0$ if the two noises have zero mean.

□

Proof of Theorem 1

Proof. If the observation model belongs to the natural exponential family (NEF), we can write it as

$$p(\boldsymbol{y}|\boldsymbol{x}) = h(\boldsymbol{y}) \exp(\boldsymbol{y}^\top \eta(\boldsymbol{x}) - \phi(\boldsymbol{x})),$$

with $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, and $h : \mathbb{R} \mapsto \mathbb{R}$ $\eta : \mathbb{R} \mapsto \mathbb{R}$ and $\phi : \mathbb{R} \mapsto \mathbb{R}$ elementwise functions which change according to the distribution. NEF distributions verify the following properties [1]

1. η is an invertible function.
2. ϕ is strictly convex.
3. Given ϕ and η, h is given by the Laplace transform $h(\boldsymbol{y}) = \int \exp\left(-\boldsymbol{s}^\top \boldsymbol{y} + \phi(\eta^{-1}(\boldsymbol{s}))\right) d\boldsymbol{s}$.
4. The mean of each entry is given by

$$\mathbb{E}\{y_i|x_i\} = \frac{\partial \phi}{\partial x_i}(x_i) / \frac{\partial \eta}{\partial x_i}(x_i) = x_i, \tag{7}$$

   for $i = 1, \ldots, n$.

We look for the decomposition $\boldsymbol{y} = (1-\alpha)\boldsymbol{y}_1 + \alpha \boldsymbol{y}_2$ such that $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ also belong to the NEF, i.e.,

$$p_1(\boldsymbol{y}_1|\boldsymbol{x}) = h_1(\boldsymbol{y}_1) \exp\left(\boldsymbol{y}_1^\top \eta_1(\boldsymbol{x}) - \phi_1(\boldsymbol{x})\right), \tag{8}$$

and

$$p_2(\boldsymbol{y}_2|\boldsymbol{x}) = h_2(\boldsymbol{y}_2) \exp\left(\boldsymbol{y}_2^\top \eta_2(\boldsymbol{x}) - \phi_2(\boldsymbol{x})\right), \tag{9}$$

for some $\alpha \in (0,1)$. Hence, the element-by-element functions of $\boldsymbol{y}_1$, $\boldsymbol{y}_2$ are related to those of $\boldsymbol{y}$ as $\phi_1(\boldsymbol{x}) = (1-\alpha)\phi(\boldsymbol{x})$, $\phi_2(\boldsymbol{x}) = \alpha\phi(\boldsymbol{x})$, $\eta_1(\boldsymbol{x}) = (1-\alpha)\eta(\boldsymbol{x})$, $\eta_2(\boldsymbol{x}) = \alpha\eta(\boldsymbol{x})$, $h_1(\boldsymbol{y}_1) = \int \exp\left(-\boldsymbol{s}^\top \boldsymbol{y}_1 + (1-\alpha)\phi\left(\eta^{-1}(\frac{\boldsymbol{s}}{1-\alpha})\right)\right) d\boldsymbol{s}$ and $h_2(\boldsymbol{y}_2) = \int \exp\left(-\boldsymbol{s}^\top \boldsymbol{y}_2 + \alpha\phi\left(\eta^{-1}(\frac{\boldsymbol{s}}{\alpha})\right)\right) d\boldsymbol{s}$.
We first verify that this choice gives the right distribution for $\boldsymbol{y}$:

$$\begin{aligned} p(\boldsymbol{y}|\boldsymbol{x}) &= \int p_1(\boldsymbol{y}_1|\boldsymbol{x}) p_2(\frac{1}{\alpha}\boldsymbol{y} - \frac{1-\alpha}{\alpha}\boldsymbol{y}_1|\boldsymbol{x}) d\boldsymbol{y}_1 \\ &= \exp\left(\boldsymbol{y}^\top \eta(\boldsymbol{x}) - \phi(\boldsymbol{x})\right) \int h_1(\boldsymbol{y}_1) h_2(\frac{1}{\alpha}\boldsymbol{y} - \frac{1-\alpha}{\alpha}\boldsymbol{y}_1) d\boldsymbol{y}_1, \end{aligned}$$

where the second line uses the fact that

$$p_1(\boldsymbol{y}_1|\boldsymbol{x}) p_2(\frac{1}{\alpha}\boldsymbol{y} - \frac{1-\alpha}{\alpha}\boldsymbol{y}_1|\boldsymbol{x}) = h_1(\boldsymbol{y}_1) \exp\left((1-\alpha)\boldsymbol{y}_1^\top \eta(\boldsymbol{x}) - (1-\alpha)\phi(\boldsymbol{x})\right) h_2(\frac{1}{\alpha}\boldsymbol{y} - \frac{1-\alpha}{\alpha}\boldsymbol{y}_1) \times \tag{10}$$

$$\exp\left(\alpha\left(\frac{1}{\alpha}\boldsymbol{y} - \frac{1-\alpha}{\alpha}\boldsymbol{y}_1\right)^\top \eta(\boldsymbol{x}) - \alpha\phi(\boldsymbol{x})\right) \tag{11}$$

$$= \exp\left(\boldsymbol{y}^\top \eta(\boldsymbol{x}) - \phi(\boldsymbol{x})\right) h_1(\boldsymbol{y}_1) h_2(\frac{1}{\alpha}\boldsymbol{y} - \frac{1-\alpha}{\alpha}\boldsymbol{y}_1). \tag{12}$$

We can obtain the conditional distribution of $\boldsymbol{y}_1$ given $\boldsymbol{y}$ as

$$p(\boldsymbol{y}_1|\boldsymbol{y}, \boldsymbol{x}) = \frac{1}{p(\boldsymbol{y}|\boldsymbol{x})} p(\boldsymbol{y}|\boldsymbol{y}_1, \boldsymbol{x}) p_1(\boldsymbol{y}_1|\boldsymbol{x}),$$

due to Bayes theorem. Using the fact that $p(\boldsymbol{y}|\boldsymbol{y}_1, \boldsymbol{x}) = p_2(\frac{1}{\alpha}\boldsymbol{y} - \frac{1-\alpha}{\alpha}\boldsymbol{y}_1|\boldsymbol{x})$ we obtain

$$\begin{aligned} p(\boldsymbol{y}_1|\boldsymbol{y}, \boldsymbol{x}) &= \frac{1}{p(\boldsymbol{y}|\boldsymbol{x})} p_1(\boldsymbol{y}_1|\boldsymbol{x}) p_2(\frac{1}{\alpha}\boldsymbol{y} - \frac{1-\alpha}{\alpha}\boldsymbol{y}_1|\boldsymbol{x}) \\ &= \frac{h_1(\boldsymbol{y}_1) h_2(\boldsymbol{y} - \boldsymbol{y}_1)}{h(\boldsymbol{y})}, \end{aligned} \tag{13}$$

where we use again (12). Thus we have that $p(\boldsymbol{y}_1|\boldsymbol{y},\boldsymbol{x})$ does not depend on the unknown parameter $\boldsymbol{x}$, that is $p(\boldsymbol{y}_1|\boldsymbol{y},\boldsymbol{x}) = p(\boldsymbol{y}_1|\boldsymbol{y})$. Consequently, since $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ are independent conditional on $\boldsymbol{x}$ and $\mathbb{E}_{\boldsymbol{y}_2|\boldsymbol{x}}\{\boldsymbol{y}_2 - \boldsymbol{x}\} = \mathbb{E}\{\boldsymbol{y}_2|\boldsymbol{x}\} - \boldsymbol{x} = \boldsymbol{0}$, we have that

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{y}_1,\boldsymbol{y}_2|\boldsymbol{x}}\|f(\boldsymbol{y}_1)-\boldsymbol{y}_2\|_2^2 &= \mathbb{E}_{\boldsymbol{y}_1|\boldsymbol{x}}\|f(\boldsymbol{y}_1)-\boldsymbol{x}\|_2^2 + 2\mathbb{E}_{\boldsymbol{y}_1,\boldsymbol{y}_2|\boldsymbol{x}}\{(f(\boldsymbol{y}_1)-\boldsymbol{x})^\top(\boldsymbol{x}-\boldsymbol{y}_2)\} + \mathbb{E}_{\boldsymbol{y}_2|\boldsymbol{x}}\|\boldsymbol{x}-\boldsymbol{y}_2\|_2^2 \\
&= \mathbb{E}_{\boldsymbol{y}_1|\boldsymbol{x}}\|f(\boldsymbol{y}_1)-\boldsymbol{x}\|_2^2 + 2\mathbb{E}_{\boldsymbol{y}_1|x}\{f(\boldsymbol{y}_1)\}^\top\mathbb{E}_{\boldsymbol{y}_2|\boldsymbol{x}}\{\boldsymbol{y}_2-\boldsymbol{x}\} - \underbrace{2\mathbb{E}_{\boldsymbol{y}_2|\boldsymbol{x}}\{\boldsymbol{x}^\top(\boldsymbol{x}-\boldsymbol{y}_2)\} + \overbrace{\mathbb{E}_{\boldsymbol{y}_2|\boldsymbol{x}}\|\boldsymbol{x}-\boldsymbol{y}_2\|_2^2}^{n\mathbb{V}\{\boldsymbol{y}_2|\boldsymbol{x}\}}}_{\text{const}} \\
&= \mathbb{E}_{\boldsymbol{y}_1|\boldsymbol{x}}\|f(\boldsymbol{y}_1)-\boldsymbol{x}\|_2^2 + \text{const}.
\end{aligned}$$

□

Proof of Proposition 2

Proof. We can write the GR2R-MSE loss as

$$\begin{aligned}
\mathcal{L}^{\alpha}_{\text{GR2R-MSE}}(\boldsymbol{y};f) &= \mathbb{E}_{\boldsymbol{y}_2|\boldsymbol{y}}\,\|f(\frac{\boldsymbol{y}-\boldsymbol{y}_2\alpha}{1-\alpha}) - \boldsymbol{y}_2\|_2^2 && (14)\\
&= \mathbb{E}_{\boldsymbol{y}_2|\boldsymbol{y}}\,\|f(\frac{\boldsymbol{y}-\boldsymbol{y}_2\alpha}{1-\alpha}) - \boldsymbol{y} - (\boldsymbol{y}_2-\boldsymbol{y})\|_2^2 && (15)\\
&= \mathbb{E}_{\boldsymbol{y}_2|\boldsymbol{y}}\,\|f(\frac{\boldsymbol{y}-\boldsymbol{y}_2\alpha}{1-\alpha}) - \boldsymbol{y}\|_2^2 - \mathbb{E}_{\boldsymbol{y}_2|\boldsymbol{y}}\,2\sum_{i=1}^{n}(y_{2,i}-y_i)\,f_i(\frac{\boldsymbol{y}-\boldsymbol{y}_2\alpha}{1-\alpha}) + \text{const}. && (16)
\end{aligned}$$

Since by assumption f is analytic, we can apply a Taylor expansion to the second term, i.e., $f_i(\frac{\boldsymbol{y}-\boldsymbol{y}_2\alpha}{1-\alpha}) = \sum_{k\geq 0}\frac{1}{k!}\frac{\partial^k f_i}{\partial y_i^k}(\frac{\boldsymbol{y}-\boldsymbol{y}_{2,-i}\alpha}{1-\alpha})\frac{(-1)^k\alpha^k}{(1-\alpha)^k}y_{i,2}^k$, where $\boldsymbol{y}_{2,-i}\in\mathbb{R}^n$ has the ith entry equal to zero and the rest equal to $\boldsymbol{y}_2$. Thus we obtain:

$$\mathcal{L}^{\alpha}_{\text{GR2R-MSE}}(\boldsymbol{y};f) \propto \mathbb{E}_{\boldsymbol{y}_2|\boldsymbol{y}}\Big(\,\|f(\frac{\boldsymbol{y}-\boldsymbol{y}_2\alpha}{1-\alpha}) - \boldsymbol{y}\|_2^2 - 2\sum_{i=1}^{n}\sum_{k\geq 1}\frac{1}{k!}\frac{\partial^k f_i}{\partial y_i^k}(\frac{\boldsymbol{y}-\boldsymbol{y}_{2,-i}\alpha}{1-\alpha})\frac{(-1)^k}{(1-\alpha)^k}(y_{2,i}-y_i)(\alpha y_{2,i})^k\Big),$$

where we used the fact that for $k=0$ we have $\mathbb{E}\{y_{2,i}-y_i|y_i\}=0$. Taking the limit $\alpha\to 0$, we obtain[1]

$$\begin{aligned}
\lim_{\alpha\to 0}\mathcal{L}^{\alpha}_{\text{GR2R-MSE}}(\boldsymbol{y};f) &\propto \lim_{\alpha\to 0}\mathbb{E}_{\boldsymbol{y}_2|\boldsymbol{y}}\,\|f(\frac{\boldsymbol{y}-\boldsymbol{y}_2\alpha}{1-\alpha}) - \boldsymbol{y}\|_2^2 \\
&\quad - 2\sum_{i=1}^{n}\sum_{k\geq 1}\frac{1}{k!}\mathbb{E}_{\boldsymbol{y}_{2,-i}|\boldsymbol{y}}\left\{\frac{\partial^k f_i}{\partial y_i^k}(\frac{\boldsymbol{y}-\boldsymbol{y}_{2,-i}\alpha}{1-\alpha})\right\}\frac{(-1)^k}{(1-\alpha)^k}\mathbb{E}_{y_{2,i}|y_i}(y_{2,i}-y_i)(\alpha y_{2,i})^k \\
&\propto \|f(\boldsymbol{y})-\boldsymbol{y}\|_2^2 + 2\sum_{i=1}^{n}\sum_{k\geq 1}(-1)^{k+1}\frac{1}{k!}\frac{\partial^k f_i}{\partial y_i^k}(\boldsymbol{y})\lim_{\alpha\to 0}\mathbb{E}\{(y_{2,i}-y_i)(\alpha y_{2,i})^k|y_i,\alpha\}
\end{aligned}$$

where the last line uses the fact that $a_k(y_i) = \lim_{\alpha\to 0}\mathbb{E}_{y_{2,i}|y_i,\alpha}\{(y_{2,i}-y_i)(\alpha y_{2,i})^k\}$ converges for all positive integer k. Replacing the definition of a_k in the previous formula, we obtain the desired result:

$$\lim_{\alpha\to 0}\mathcal{L}^{\alpha}_{\text{GR2R-MSE}}(\boldsymbol{y};f) \propto \|f(\boldsymbol{y})-\boldsymbol{y}\|_2^2 + 2\sum_{i=1}^{n}\sum_{k\geq 1}(-1)^{k+1}a_k(y_i)\frac{1}{k!}\frac{\partial^k f_i}{\partial y_i^k}(\boldsymbol{y}).$$

□

[1] We have that for $g:\mathbb{R}^n\mapsto\mathbb{R}$, the expectation $\mathbb{E}_{\boldsymbol{y}_2|\boldsymbol{y}}g(\alpha\boldsymbol{y}_2) = g(\boldsymbol{0})$ as $p(\alpha\boldsymbol{y}_2|\boldsymbol{y})\to\delta_{\boldsymbol{y}_2=\boldsymbol{0}}$ as $\alpha\to 0$.

Proof of Proposition 3

Proof. The $\mathcal{L}^{\alpha}_{\mathrm{GR2R}}(\boldsymbol{y}; f)$ is defined as

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{y}_1,\boldsymbol{y}_2|\boldsymbol{x}} - \log p_2\left(\boldsymbol{y}_2|\hat{\boldsymbol{x}} = f(\boldsymbol{y}_1)\right) &= \mathbb{E}_{\boldsymbol{y}_1,\boldsymbol{y}_2|\boldsymbol{x}} \left\{-\alpha \boldsymbol{y}_2^\top \eta\left(f(\boldsymbol{y}_1)\right) + \alpha\phi\left(f(\boldsymbol{y}_1)\right) - \log h_2(\boldsymbol{y}_2)\right\} && (17)\\
&= -\alpha\left(\mathbb{E}_{\boldsymbol{y}_2|\boldsymbol{x}}\, \boldsymbol{y}_2^\top\right) \mathbb{E}_{\boldsymbol{y}_1|\boldsymbol{x}}\, \eta\left(f(\boldsymbol{y}_1)\right) + \mathbb{E}_{\boldsymbol{y}_1|\boldsymbol{x}} \alpha\phi\left(f(\boldsymbol{y}_1)\right) - \mathbb{E}_{\boldsymbol{y}_2|\boldsymbol{x}} \log h_2(\boldsymbol{y}_2) && (18)\\
&= -\alpha\boldsymbol{x}^\top \mathbb{E}_{\boldsymbol{y}_1|\boldsymbol{x}}\, \eta\left(f(\boldsymbol{y}_1)\right) + \mathbb{E}_{\boldsymbol{y}_1|\boldsymbol{x}}\, \alpha\phi\left(f(\boldsymbol{y}_1)\right) - \mathbb{E}_{\boldsymbol{y}_2|\boldsymbol{x}} \log h_2(\boldsymbol{y}_2) && (19)\\
&= -\mathbb{E}_{\boldsymbol{y}_1|\boldsymbol{x}}\{\alpha\boldsymbol{x}^\top \eta\left(f(\boldsymbol{y}_1)\right) - \alpha\phi\left(f(\boldsymbol{y}_1)\right) + \log h_2(\boldsymbol{y}_2)\} + \text{const} && (20)\\
&= \mathbb{E}_{\boldsymbol{y}_1|\boldsymbol{x}} - \log p_2\left(\boldsymbol{x}|\hat{\boldsymbol{x}} = f(\boldsymbol{y}_1)\right) + \text{const}. && (21)
\end{aligned}$$

We now prove that $\mathbb{E}\{\boldsymbol{x}|\boldsymbol{y}_1\} = \arg\min_f \mathcal{L}^{\alpha}_{\mathrm{GR2R}}(\boldsymbol{y}; f)$. We can write this minimization as

$$\begin{aligned}
\min_f\ \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}_1}\{\eta\left(f(\boldsymbol{y}_1)\right)^\top \boldsymbol{x} - \phi\left(f(\boldsymbol{y}_1)\right)\} &= \min_f\ \mathbb{E}_{\boldsymbol{y}_1}\{\eta\left(f(\boldsymbol{y}_1)\right)^\top \mathbb{E}\{\boldsymbol{x}|\boldsymbol{y}_1\} - \phi\left(f(\boldsymbol{y}_1)\right)\} && (22)\\
&= \mathbb{E}_{\boldsymbol{y}_1}\{\min_f \eta\left(f(\boldsymbol{y}_1)\right)^\top \mathbb{E}\{\boldsymbol{x}|\boldsymbol{y}_1\} - \phi\left(f(\boldsymbol{y}_1)\right)\}, && (23)
\end{aligned}$$

where the last equality swaps the integration with the minimization since the minimizer exists for every fixed $\boldsymbol{y}_1$. Defining $\boldsymbol{z} := f(\boldsymbol{y}_1)$, we can minimize the term inside the expectation w.r.t. to

$$\arg\min_{\boldsymbol{z}}\ \mathbb{E}\{\boldsymbol{x}|\boldsymbol{y}_1\}\, \eta(\boldsymbol{z}) - \phi\left(\boldsymbol{z}\right). \qquad (24)$$

The problem is separable across entries, so it can be

$$\arg\min_{z_i}\ \mathbb{E}\{x_i|y_{1,i}\}\, \eta(z_i) - \phi\left(z_i\right), \qquad (25)$$

for $i = 1, \ldots, n$. Since the problem is strongly convex w.r.t. z_i, we can find the solution by setting its derivative to zero

$$\begin{aligned}
\mathbb{E}\{x_i|y_{1,i}\}\, \frac{\partial\eta}{\partial z_i}(\hat{z}_i) - \frac{\partial\phi}{\partial z_i}\left(\hat{z}_i\right) &= 0 && (26)\\
\frac{\partial\eta}{\partial z_i}(\hat{z}_i) / \frac{\partial\phi}{\partial z_i}\left(\hat{z}_i\right) &= \mathbb{E}\{x_i|y_{1,i}\} && (27)\\
\hat{z}_i &= \mathbb{E}\{x_i|y_{1,i}\}, && (28)
\end{aligned}$$

for $i = 1, \ldots, n$, where the second line uses property (7), and thus $\hat{f}(\boldsymbol{y}_1) = \mathbb{E}\{\boldsymbol{x}|\boldsymbol{y}_1\}$.

□

2 Additional information

Table 1 summarizes the NEF distributions $p(\boldsymbol{y}|\boldsymbol{x})$ used in the main document. This was used to create the recoruptions used in the main document. Specifically, the formulas to construct $\boldsymbol{y}_1$ in terms of $\boldsymbol{y}$ and the extra noise $\boldsymbol{\omega}$ can be derived from replacing $h(y), h_1(y_1)$ and $h_2(y_2)$ from Table 1 in Equation (13) for its respective NEF distribution; this is left as an exercise for the reader.

Model	$y \sim \mathcal{N}(x, \sigma^2)$	$y \sim \mathcal{P}(\frac{x}{\gamma})$	$y \sim \mathcal{G}(\ell, x/\ell)$	$y \sim \mathrm{Bin}(\ell, x)$
$\eta(x)$	x/σ^2	$\log(x)$	$-\ell/x$	$\log(x/(1-x))$
$\phi(x)$	$x^2/(2\sigma^2)$	x/γ	$\ell\log(x)$	$\ell\log(1-x)$
$h(y)$	$\sqrt{2\pi}\sigma\exp(y^2/(2\sigma^2))$	$(\gamma^y y!)^{-1}$	$\ell^\ell y^{\ell-1}/\Gamma(\ell)$	$\binom{\ell}{y}$
$h_1(y_1)$	$\sqrt{2\pi}\frac{\sigma}{\sqrt{1-\alpha}}\exp(y_1^2/(2\frac{\sigma^2}{1-\alpha}))$	$((1-\alpha)^{(1-\alpha)y_1+1}\gamma^{(1-\alpha)y_1}((1-\alpha)y_1)!)^{-1}$	$\frac{\ell^{(1-\alpha)\ell}((1-\alpha)y_1)^{(1-\alpha)\ell-1}}{(1-\alpha)\Gamma((1-\alpha)\ell)}$	$\frac{1}{1-\alpha}\binom{(1-\alpha)\ell}{(1-\alpha)y_1}$
$h_2(y_2)$	$\sqrt{2\pi}\frac{\sigma}{\sqrt{\alpha}}\exp(y_1^2/(2\frac{\sigma^2}{\alpha}))$	$(\alpha^{\alpha y_2+1}\gamma^{\alpha y_2}(\alpha y_2)!)^{-1}$	$\frac{\ell^{\alpha\ell}(\alpha y_2)^{\alpha\ell-1}}{\alpha\Gamma(\alpha\ell)}$	$\frac{1}{\alpha}\binom{\alpha\ell}{\alpha y_2}$

Table 1: Examples of one-dimensional natural exponential family distributions $p(y|x)$ and their respective decompositions. These can be extended to higher dimensions by considering separable distributions $p(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^n p(y_i|x_i)$, by $\eta(\boldsymbol{x}) = \sum_{i=1}^n \eta(x_i)$, $\phi(\boldsymbol{x}) = \sum_{i=1}^n \phi(x_i)$, $h(\boldsymbol{y}) = \prod_{i=1}^n h(y_i)$, $h_1(\boldsymbol{y}_1) = \prod_{i=1}^n h_1(y_{1,i})$ and $h_2(\boldsymbol{y}_2) = \prod_{i=1}^n h_2(y_{2,i})$.

Equivalence with SURE as $\alpha \to 0$

$$\lim_{\alpha\to 0} \mathcal{L}^{\alpha}_{\mathrm{GR2R-MSE}}(\boldsymbol{y}; f) = \|f(\boldsymbol{y}) - \boldsymbol{y}\|_2^2 + 2\sum_{i=1}^{n}\sum_{k\geq 1}(-1)^{k+1}a_k(y_i)\frac{1}{k!}\frac{\partial^k f_i}{\partial y_i^k}(\boldsymbol{y}) + \text{const.} \tag{29}$$

where

$$a_k(y_i) = \lim_{\alpha\to 0}\mathbb{E}_{y_{2,i}|y_i,\alpha}(y_{2,i} - y_i)(\alpha y_{2,i})^k. \tag{30}$$

Gaussian case. Based on the proposed re-corruption procedure for the Gaussian case, we have that the re-corruption of $\boldsymbol{y}_2$ in terms of $\boldsymbol{y}$ and the extra noise $\boldsymbol{\omega}$ as

$$\boldsymbol{y}_2 = \boldsymbol{y} - \sqrt{\frac{1-\alpha}{\alpha}}\boldsymbol{\omega} \tag{31}$$

Analyze for $k = 1$

$$a_1(y_i) = \lim_{\alpha\to 0}\mathbb{E}_{y_{2,i}|y_i,\alpha}\{(y_{2,i} - y_i)(\alpha y_{2,i})^1\} \tag{32}$$

for one element y_2, y

$$\begin{aligned}
\lim_{\alpha\to 0}\mathbb{E}_{y_2|y,\alpha}\{(y_2 - y)(\alpha y_2)\} &= \lim_{\alpha\to 0}\mathbb{E}_{\omega|y,\alpha}\{\alpha(y - \sqrt{\frac{1-\alpha}{\alpha}}\omega - y)(y - \sqrt{\frac{1-\alpha}{\alpha}}\omega)\} \\
&= \lim_{\alpha\to 0}\mathbb{E}_{\omega|y,\alpha}\{-\alpha\sqrt{\frac{1-\alpha}{\alpha}}\omega(y - \sqrt{\frac{1-\alpha}{\alpha}}\omega)\} \\
&= \lim_{\alpha\to 0}\mathbb{E}_{\omega|y,\alpha}\{-\alpha\sqrt{\frac{1-\alpha}{\alpha}}\omega y + \alpha\frac{1-\alpha}{\alpha}\omega^2)\} \\
&= \lim_{\alpha\to 0}\mathbb{E}_{\omega|y,\alpha}\{-\sqrt{\alpha(1-\alpha)}\omega y + (1-\alpha)\omega^2\} \\
&= \lim_{\alpha\to 0}(1-\alpha)\sigma^2 = \sigma^2
\end{aligned} \tag{33}$$

analyzing for $k > 1$ we have that $a_k(y) \to 0$ since the α^{k-1} term dominates in the expression

$$a_k(y) = \lim_{\alpha\to 0}\mathbb{E}_{\omega|y,\alpha}\Big\{(-\sqrt{\alpha(1-\alpha)}\omega y + (1-\alpha)\omega^2)(y - \sqrt{\frac{1-\alpha}{\alpha}}\omega)^{k-1}\alpha^{k-1}\Big\} \tag{34}$$

finally, substituting $a_k(y_i)$ in (29) for the Gaussian case we have that

$$\lim_{\alpha\to 0}\mathcal{L}^{\alpha}_{\mathrm{GR2R-MSE}}(\boldsymbol{y}; f) = \|f(\boldsymbol{y}) - \boldsymbol{y}\|_2^2 + 2\sigma^2\sum_{i=1}^{n}\frac{\partial^k f_i}{\partial y_i^k}(\boldsymbol{y}) + \text{const.} \tag{35}$$

Poisson case. Starting from $\mathcal{L}^{\alpha}_{\text{GR2R-MSE}}$ with $\boldsymbol{y}_1$ constructed in terms of $\boldsymbol{y}, \boldsymbol{y}_2$ and α as $\boldsymbol{y}_1 = (\boldsymbol{y} - \boldsymbol{y}_2\alpha)/(1-\alpha)$ we have that

$$\mathcal{L}^{\alpha}_{\mathrm{GR2R-MSE}}(\boldsymbol{y}; f) = \mathbb{E}_{\boldsymbol{y}_2|\boldsymbol{y}}\,\|f(\frac{\boldsymbol{y} - \boldsymbol{y}_2\alpha}{1-\alpha}) - \boldsymbol{y}\|_2^2 - \mathbb{E}_{\boldsymbol{y}_2|\boldsymbol{y}}\,2\sum_{i=1}^{n}(y_{2,i} - y_i)\,f_i(\frac{\boldsymbol{y} - \boldsymbol{y}_2\alpha}{1-\alpha}) + \text{const}, \tag{36}$$

evaluating $\lim_{\alpha\to 0}\mathcal{L}^{\alpha}_{\mathrm{GR2R-MSE}}(\boldsymbol{y}; f)$

$$\begin{aligned}
\lim_{\alpha\to 0}\mathcal{L}^{\alpha}_{\mathrm{GR2R-MSE}}(\boldsymbol{y}; f) &\propto \lim_{\alpha\to 0}\mathbb{E}_{\boldsymbol{y}_2|\boldsymbol{y}}\,\|f(\frac{\boldsymbol{y} - \boldsymbol{y}_2\alpha}{1-\alpha}) - \boldsymbol{y}\|_2^2 - \lim_{\alpha\to 0}\mathbb{E}_{\boldsymbol{y}_2|\boldsymbol{y}}\,2\sum_{i=1}^{n}(y_{2,i} - y_i)\,f_i(\frac{\boldsymbol{y} - \boldsymbol{y}_2\alpha}{1-\alpha}) \\
&\propto \|f(\boldsymbol{y}) - \boldsymbol{y}\|_2^2 + 2\lim_{\alpha\to 0}\mathbb{E}_{\boldsymbol{y}_2|\boldsymbol{y}}\left(\sum_{i=1}^{n} y_i f_i(\frac{\boldsymbol{y} - \boldsymbol{y}_2\alpha}{1-\alpha}) - y_{2,i} f_i(\frac{\boldsymbol{y} - \boldsymbol{y}_2\alpha}{1-\alpha})\right) \\
&\propto \|f(\boldsymbol{y}) - \boldsymbol{y}\|_2^2 + 2\sum_{i=1}^{n}\left(y_i f_i(\boldsymbol{y}) - \lim_{\alpha\to 0}\mathbb{E}_{\boldsymbol{y}_2|\boldsymbol{y}} y_{2,i} f_i(\frac{\boldsymbol{y} - \boldsymbol{y}_2\alpha}{1-\alpha})\right)
\end{aligned} \tag{37}$$

Recall that $\boldsymbol{y} = \gamma \boldsymbol{z}$ and $\boldsymbol{y}_2 = \gamma \boldsymbol{\omega}/\alpha$ with $\boldsymbol{\omega} \sim \mathrm{Bin}(\boldsymbol{z}, \alpha)$. Defining the function $g_{i,\alpha} : \omega_i \mapsto f_i(\frac{\boldsymbol{y}-\gamma\boldsymbol{\omega}}{1-\alpha})$, we have that the second term is

$$
\begin{aligned}
\lim_{\alpha\to 0} \mathbb{E}_{\boldsymbol{y}_2|\boldsymbol{y}} y_{2,i} f_i(\frac{\boldsymbol{y}-\boldsymbol{y}_2\alpha}{1-\alpha}) &= \lim_{\alpha\to 0} \mathbb{E}_{\boldsymbol{\omega}_{-i}|y_i} \mathbb{E}_{\omega_i|y_i} \gamma \frac{\omega_i}{\alpha} f_i(\frac{\boldsymbol{y}-\gamma\boldsymbol{\omega}}{1-\alpha}) && (38)\\
&= \lim_{\alpha\to 0} \sum_{k=1}^{z_i} \gamma \binom{z_i}{k} \alpha^{k-1}(1-\alpha)^{z_i-k} k\, \mathbb{E}_{\boldsymbol{\omega}_{-i}|y_i} g_{i,\alpha}(k) && (39)\\
&= \lim_{\alpha\to 0} \Big(\gamma z_i\,(1-\alpha)^{z_i-1} \mathbb{E}_{\boldsymbol{\omega}_{-i}|z_i} g_{i,\alpha}(1) + \sum_{k=2}^{z_i} \gamma \binom{z_i}{k} \alpha^{k-1}(1-\alpha)^{z_i-k} k\, \mathbb{E}_{\boldsymbol{\omega}_{-i}|y_i} g_{i,\alpha}(k) && (40)\\
&\propto \lim_{\alpha\to 0} \Big(\gamma z_i\,(1-\alpha)^{z_i-1} \mathbb{E}_{\boldsymbol{\omega}_{-i}|y_i} g_{i,\alpha}(1) + \mathcal{O}(\alpha)\Big) && (41)\\
&\propto \gamma z_i \lim_{\alpha\to 0} \mathbb{E}_{\boldsymbol{\omega}_{-i}|y_i} g_{i,0}(1) && (42)\\
&\propto y_i f_i(\boldsymbol{y} - \gamma \boldsymbol{e}_i), && (43)
\end{aligned}
$$

where $\boldsymbol{e}_i \in \mathbb{R}^n$ is the vector with i-th entry in 1 and with all others in 0. Thus, plugging in this result, we have

$$
\lim_{\alpha\to 0} \mathcal{L}^{\alpha}_{\mathrm{GR2R-MSE}}(\boldsymbol{y}; f) = \|f(\boldsymbol{y}) - \boldsymbol{y}\|_2^2 + 2\sum_{i=1}^{n} y_i \Big(f_i(\boldsymbol{y}) - f_i(\boldsymbol{y} - \gamma \boldsymbol{e}_i)\Big) + \text{const.} \tag{44}
$$

Gamma case. Based on the proposed re-corruption procedure for the Gamma case, we have that the re-corruption of $\boldsymbol{y}_2$ in terms of $\boldsymbol{y}$ and the extra noise $\boldsymbol{\omega} \sim \mathrm{Beta}(\ell\alpha, \ell(1-\alpha))$ as

$$
\boldsymbol{y}_2 = \frac{\boldsymbol{\omega}}{\alpha} \boldsymbol{y} \tag{45}
$$

then, replacing in the expression of $a_k(y_i)$ for one element y_2, y

$$
\begin{aligned}
a_k(y) &= \lim_{\alpha\to 0} \mathbb{E}_{y_2|y,\alpha}\{(y_2 - y)(\alpha y_2)^k\} = \lim_{\alpha\to 0} \mathbb{E}_{\omega|y,\alpha}\{(\frac{\omega}{\alpha} y - y)(\omega y)^k\}\\
&= \lim_{\alpha\to 0} \mathbb{E}_{\omega|\alpha}\{\frac{\omega^{k+1}}{\alpha} y^{k+1} - \omega^k y^{k+1}\}\\
&= y^{k+1} \lim_{\alpha\to 0} \Big(\frac{1}{\alpha} \mathbb{E}_{\omega|\alpha}\{\omega^{k+1}\} - \mathbb{E}_{\omega|\alpha}\{\omega^k\}\Big)
\end{aligned} \tag{46}
$$

The kth moment of ω can be expressed recursively as

$$
\mathbb{E}\{\omega^{k+1}\} = \frac{\ell\alpha + k - 1}{\ell + k - 1} \mathbb{E}\{\omega^k\} \tag{47}
$$

then

$$\lim_{\alpha\to 0}\Big(\frac{1}{\alpha}\mathbb{E}_{\omega|\alpha}\{\omega^{k+1}\} - \mathbb{E}_{\omega|\alpha}\{\omega^{k}\}\Big) = \lim_{\alpha\to 0}\mathbb{E}_{\omega|\alpha}\{\omega^{k}\}\Big(\frac{1}{\alpha}\frac{\ell\alpha+k-1}{\ell+k-1} - 1\Big) \tag{48}$$

$$= \lim_{\alpha\to 0}\mathbb{E}_{\omega|\alpha}\{\omega^{k}\}\frac{\ell\alpha+k-1-\alpha(\ell+k-1)}{\alpha(\ell+k-1)} \tag{49}$$

$$= \lim_{\alpha\to 0}\mathbb{E}_{\omega|\alpha}\{\omega^{k}\}\frac{(k-1)(1-\alpha)}{\alpha(\ell+k-1)} \tag{50}$$

$$= \lim_{\alpha\to 0}(\prod_{r=0}^{k-1}\frac{\alpha\ell+r}{\ell+r})\frac{(k-1)(1-\alpha)}{\alpha(\ell+k-1)} \tag{51}$$

$$= \lim_{\alpha\to 0}(\prod_{r=1}^{k-1}\frac{\alpha\ell+r}{\ell+r})\frac{\alpha\ell}{\ell}\frac{(k-1)(1-\alpha)}{\alpha(\ell+k-1)} \tag{52}$$

$$= \lim_{\alpha\to 0}(\prod_{r=1}^{k-1}\frac{\alpha\ell+r}{\ell+r})\frac{(k-1)(1-\alpha)}{(\ell+k-1)} \tag{53}$$

$$= (\prod_{r=1}^{k-1}\frac{r}{\ell+r})\frac{(k-1)}{(\ell+k-1)} \tag{54}$$

$$= \frac{(k-1)!\,\Gamma(\ell)}{\Gamma(\ell+k)}\frac{\ell(k-1)}{(\ell+k-1)} \tag{55}$$

$$\tag{56}$$

Finally, substituting $a_k(y_i)$ in (29) for the Gamma case we have that

$$\lim_{\alpha\to 0}\mathcal{L}^{\alpha}_{\mathrm{GR2R-MSE}}(\boldsymbol{y};f) = \|f(\boldsymbol{y})-\boldsymbol{y}\|_2^2 + 2\sum_{i=1}^{n}\sum_{k\geq 1}\frac{\ell(k-1)}{k(\ell+k-1)}\frac{(-y_i)^{k+1}\Gamma(\ell)}{\Gamma(\ell+k)}\frac{\partial^k f_i}{\partial y_i^k}(\boldsymbol{y}) + \text{const.} \tag{57}$$

3 Experimental details

The maximum-entropy sampling strategy, detailed below, is employed to generate noise that ensures the third moment is preserved in the experiment described in Section 4.1. Non-Gaussian Additive Noise, in the main paper.

3.1 Maximum-entropy sampling

Consider a random variable z with $\mu_i = \mathbb{E}\, z^i$ the desired moments of order $i = 1,\ldots,k$. We obtain maximum entropy samples verifying the desired moments up to order k by minimizing [2]

$$\arg\min_{\boldsymbol{z}}\sum_{i=0}^{k}\|\frac{1}{n}\sum_{j=1}z_j^i - \mu_i\|_2^2 \tag{58}$$

via gradient descent where we initialize $\boldsymbol{z}\sim\mathcal{N}(\mu_1\mathbf{1},\boldsymbol{I}(\mu_2-\mu_1^2))$. The optimization is stopped when the relative error is small, i.e.,

$$\frac{\frac{1}{n}|\sum_{j=1}z_j^i-\mu_i|}{|\mu_i|} < 0.1$$

for all $i = 1,\ldots,k$.

4 Additional Simulations and Results

4.1 Effect of the re-corruption hyper-parameter α.

We evaluate the performance of the proposed GR2R loss on the PSNR metric when examining the effect of the re-corruption parameter α on three noise distributions: Poisson, Gamma, and Gaussian. Specifically, the experimental setup consists of training the DnCNN model architecture by minimizing the proposed loss $\mathcal{L}^{\alpha}_{\text{GR2R-MSE}}$ for different values of α on the DIV2K dataset. All experiments share the same training configuration: Adam optimizer, with an initial learning rate of 1e-4 and 250 training epochs. For the noise model parameters, we set $\gamma = 0.5$ for the Poisson experiment, $\ell = 5$ for the Gamma experiment, and $\sigma = 0.1$ for the Gaussian experiment.

We test the GR2R loss for α values in the interval [0.1, 3.5] for Poisson and Gamma and in the interval [0.1, 0.9] for Gaussian. A scatter plot is shown in Figure 1 for all noise distributions tested, with the trends highlighted by polynomial fitting. A trade-off between the value of α and the PSNR score can be observed, where low values of α indicated less SNR in $\boldsymbol{y}_1$ and higher SNR in $\boldsymbol{y}_2$. For the Poisson and Gamma distributions, the optimal values of the re-corruption parameter α appear to be approximately $\alpha = 0.12$, while for the Gaussian distribution, the preferred value seems to be $\alpha = 0.3$. Furthermore, although the performance of the GR2R loss is sensitive to the choice of the re-corruption parameter α. The disparity between the highest and lowest PSNR scores is less than 0.2 dB for the Gamma and Gaussian distributions and less than 0.6 dB for Gaussian noise.

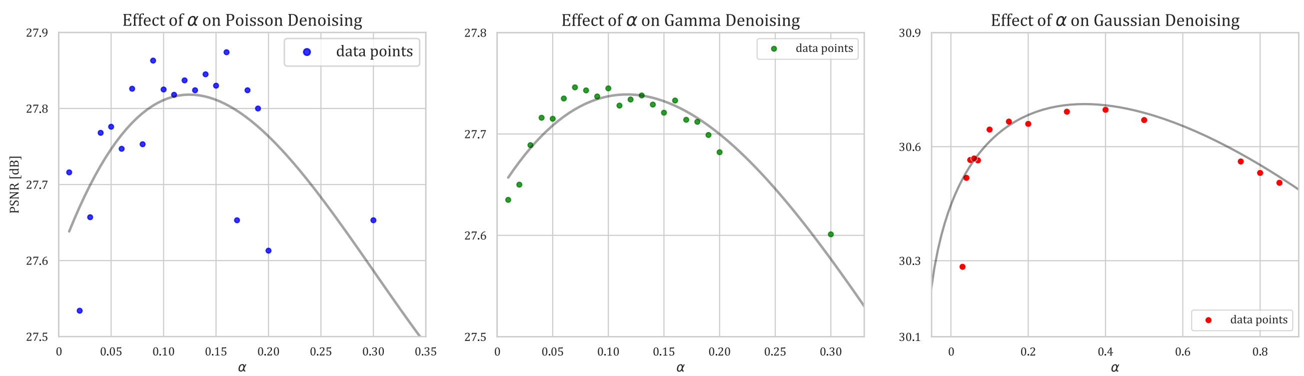

Figure 1: Effect of α parameter for different noise distributions. he results indicate that the optimal α parameter consistently lies within the range of 0.1 to 0.3 across all tested scenarios.

4.2 Log-Rayleigh Noise

In addition to the numerical comparisons presented in the main manuscript between R2R (matching second-order moment) and the proposed GR2R (matching third-order moment), presented in Section 4.1 in the main document, this section offers further elaboration on the experimental setups, as well as visual analyses of the noise estimation compared to the restored images. The training configuration consists of the DnCNN model along 100 epochs with a batch size of 15 with an initial learning rate of 5e-4 with the Adam optimizer in the DIV2K dataset. Figure 2 displays a histogram comparing the original Log-Rayleigh noise, which was utilized to corrupt the images, with the estimated additional noise provided by R2R and GR2R. It can be observed that extending the moment matching to the third moment significantly enhances the accuracy of the noise distribution estimation compared to matching only until the second moment. Restored images are presented in Figure 3, which demonstrate the effect of matching the third moment for the image denoising task.

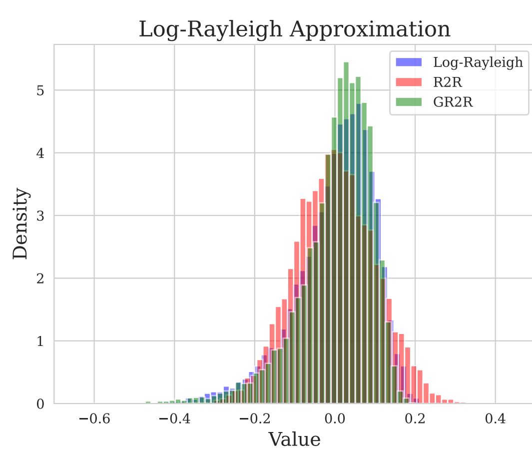

Figure 2: Histogram of noise estimations.

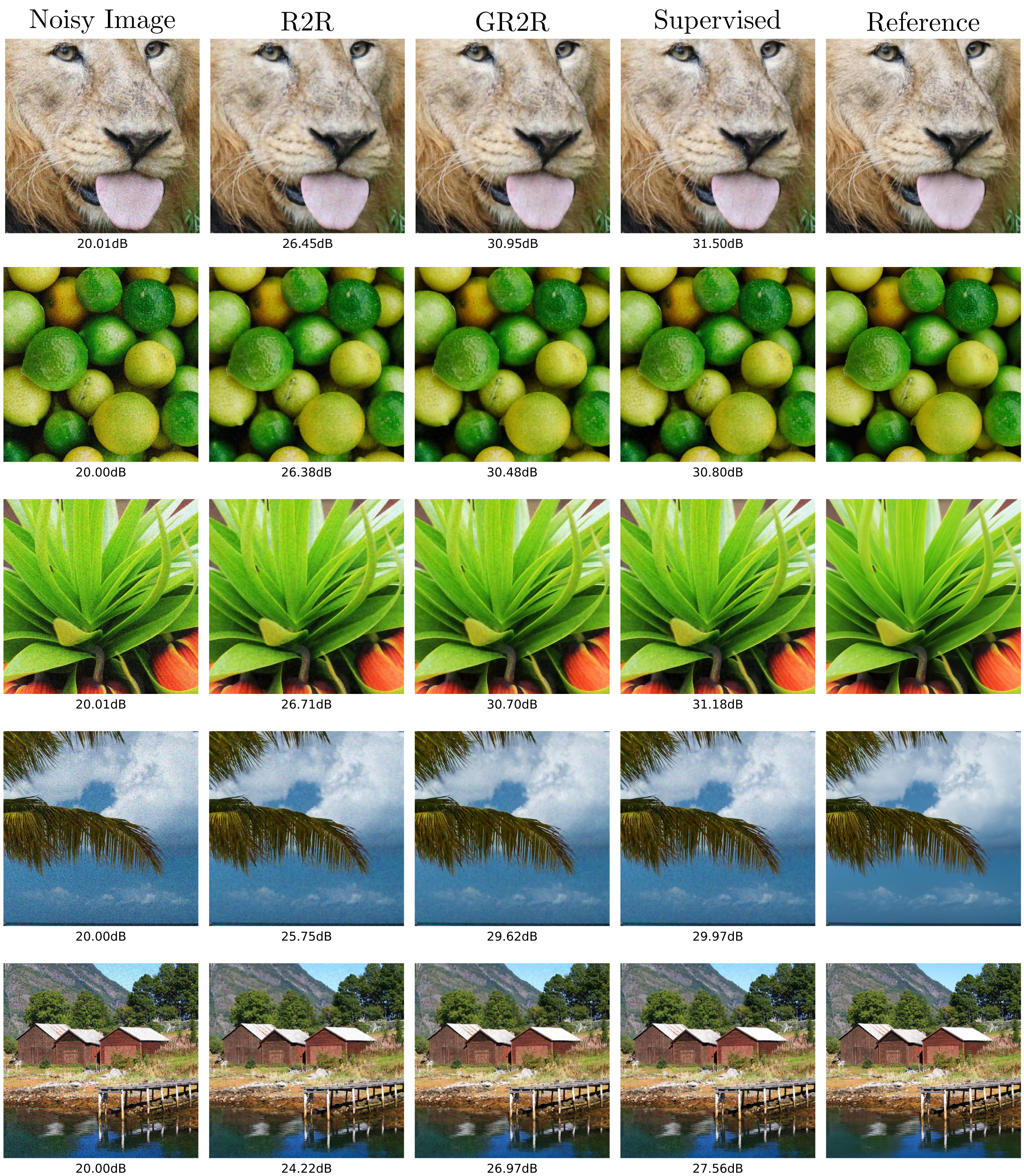

Figure 3: Visual Results for a Log-Rayleigh Noise with a standard deviation of $\sigma = 0.1$.

4.3 Additional Results

The following subsections present results of the PSNR mean and standard deviation obtained for the different methods for Poisson, Gamma, and Gaussian distributions. Each subsection also shows additional visual results.

4.3.1 Poission Noise

Table 2: PSNR results on Poisson noise. GR2R-NLL stands for the proposed GR2R with Negative Log-Likelihood.

Poisson Noise	Methods				
Noise Level (γ)	PURE [3]	Neigh2Neigh [4]	GR2R-NLL (ours)	GR2R-MSE (ours)	Supervised-MSE
0.01	32.69±2.13	33.37±2.20	33.90±2.26	33.92±2.20	**33.96±2.23**
0.1	24.37±1.89	28.27±2.60	28.30±2.65	28.35±2.64	**28.39±2.65**
0.5	22.98±1.53	24.90±2.68	25.07±2.71	24.69±2.74	**25.32±2.75**
1.0	17.94±1.13	23.56±2.67	23.69±2.70	23.49±2.71	**23.85±2.72**

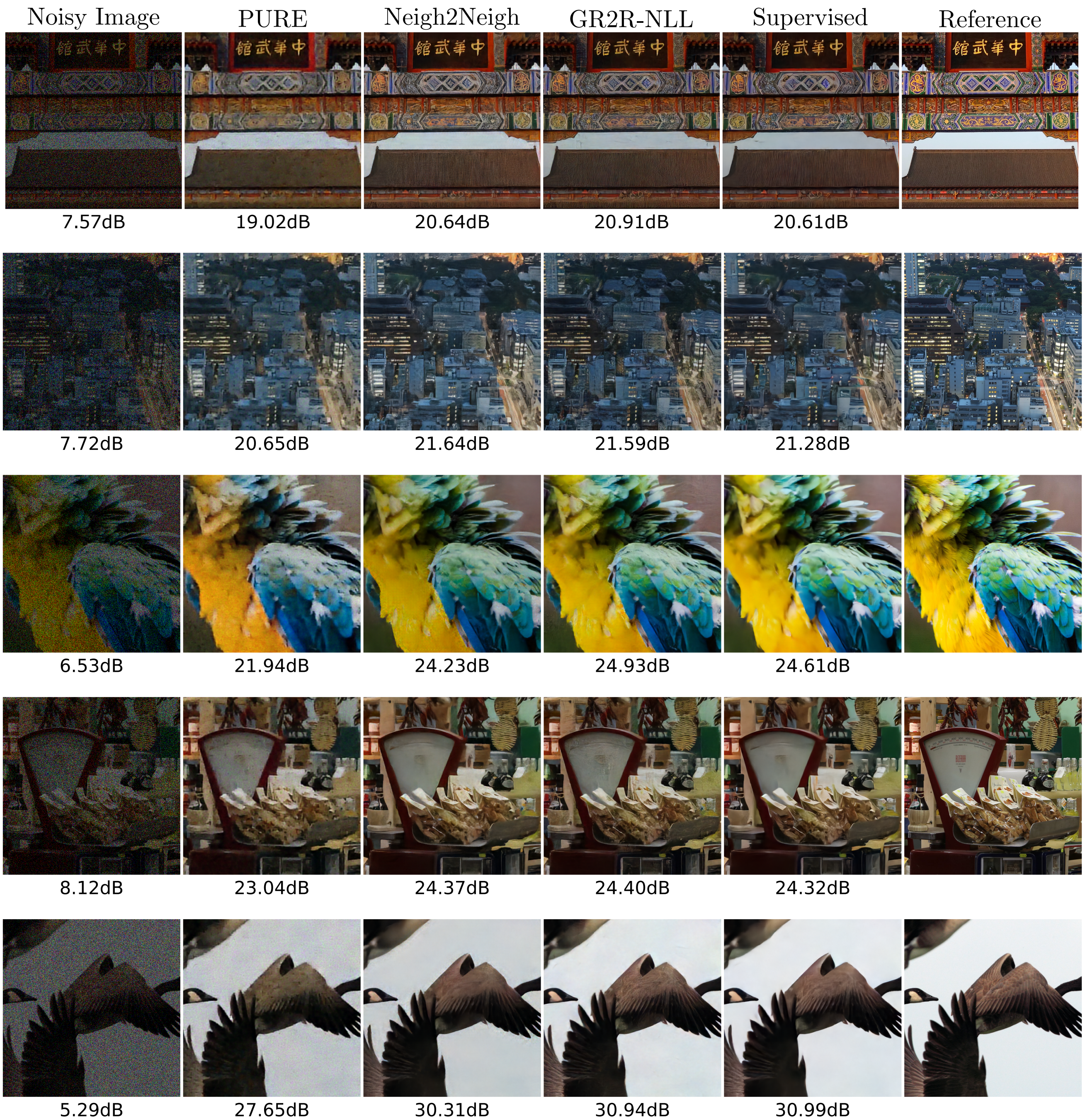

Figure 4: Poisson Denoising in DIV2K Dataset.

4.3.2 Gamma Noise

Table 3: PSNR results on Gamma noise. GR2R-NLL stands for the proposed GR2R with Negative Log-Likelihood.

Gaussian Noise	Methods			
Number of looks (ℓ)	Neigh2Neigh [4]	GR2R-NLL (ours)	GR2R-MSE (ours)	Supervised-MSE
30	30.34±1.60	30.43±1.61	<u>31.58±1.72</u>	**31.86±1.73**
15	28.56±1.58	28.71±1.59	<u>29.55±1.68</u>	**29.76±1.70**
5	25.71±1.53	25.79±1.49	<u>26.35±1.57</u>	**26.72±1.62**
1	22.19±1.40	22.19±1.34	<u>22.38±1.40</u>	**22.56±1.44**

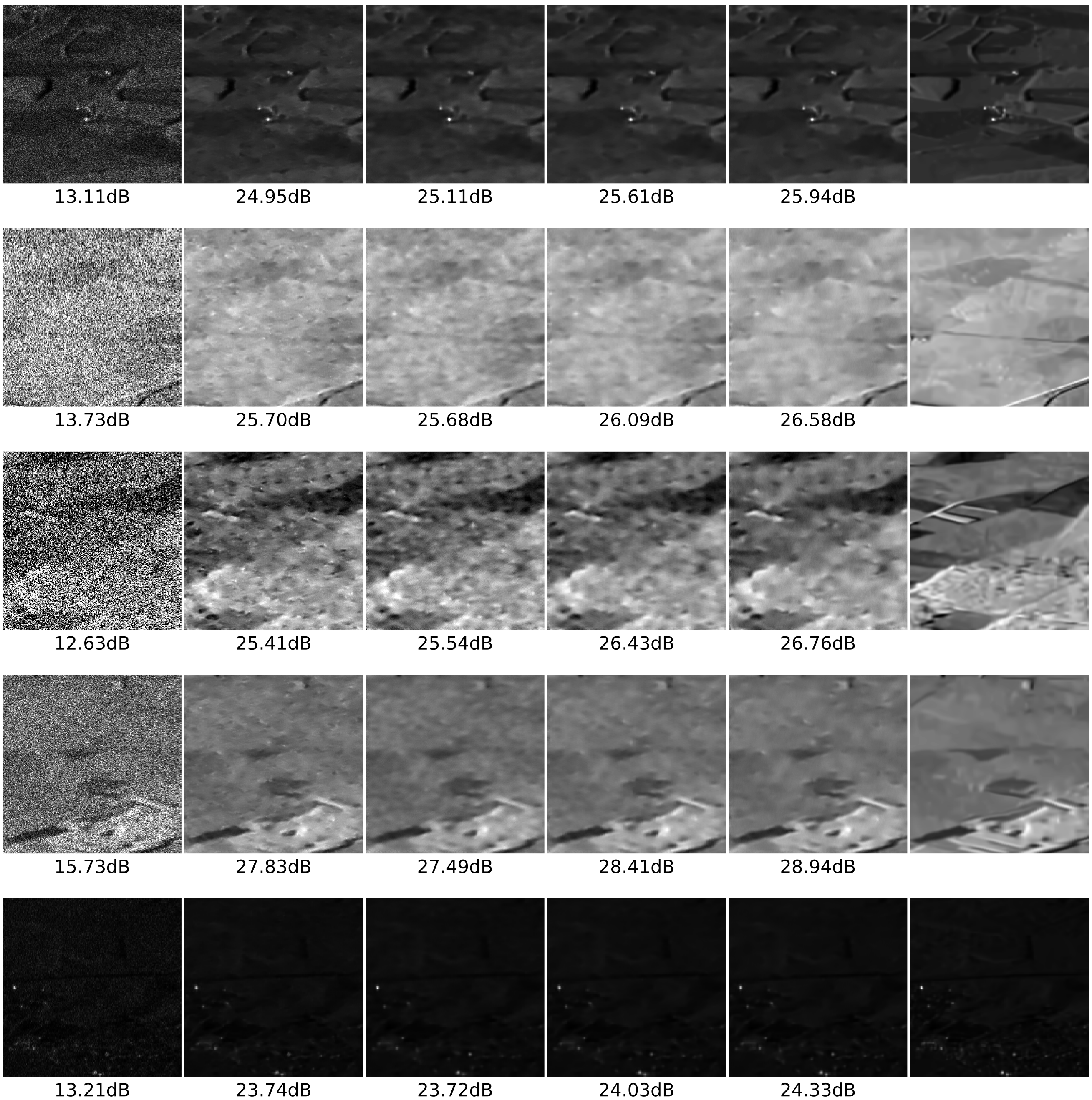

Figure 5: Gamma Denoising in SARDataset.

4.3.3 Gaussian Noise

Table 4: PSNR results for Gaussian noise. For this case, the MSE and NLL variants of GR2R are the same.

Gaussian noise	Methods				
Noise Level (σ)	Noise2Score [5]	SURE [6]	Neigh2Neigh [4]	GR2R (ours)	Supervised-MSE
0.05	34.42±1.16	35.31±1.43	35.07±1.41	35.38±1.47	**35.41±1.47**
0.1	31.02±0.74	32.76±1.22	32.57±1.22	33.03±1.29	**33.14±1.28**
0.2	29.34±0.62	29.77±1.02	29.73±1.05	30.24±1.05	**30.38±1.05**
0.5	22.94±0.65	25.52±1.02	25.61±0.99	25.81±0.97	**25.93±0.94**

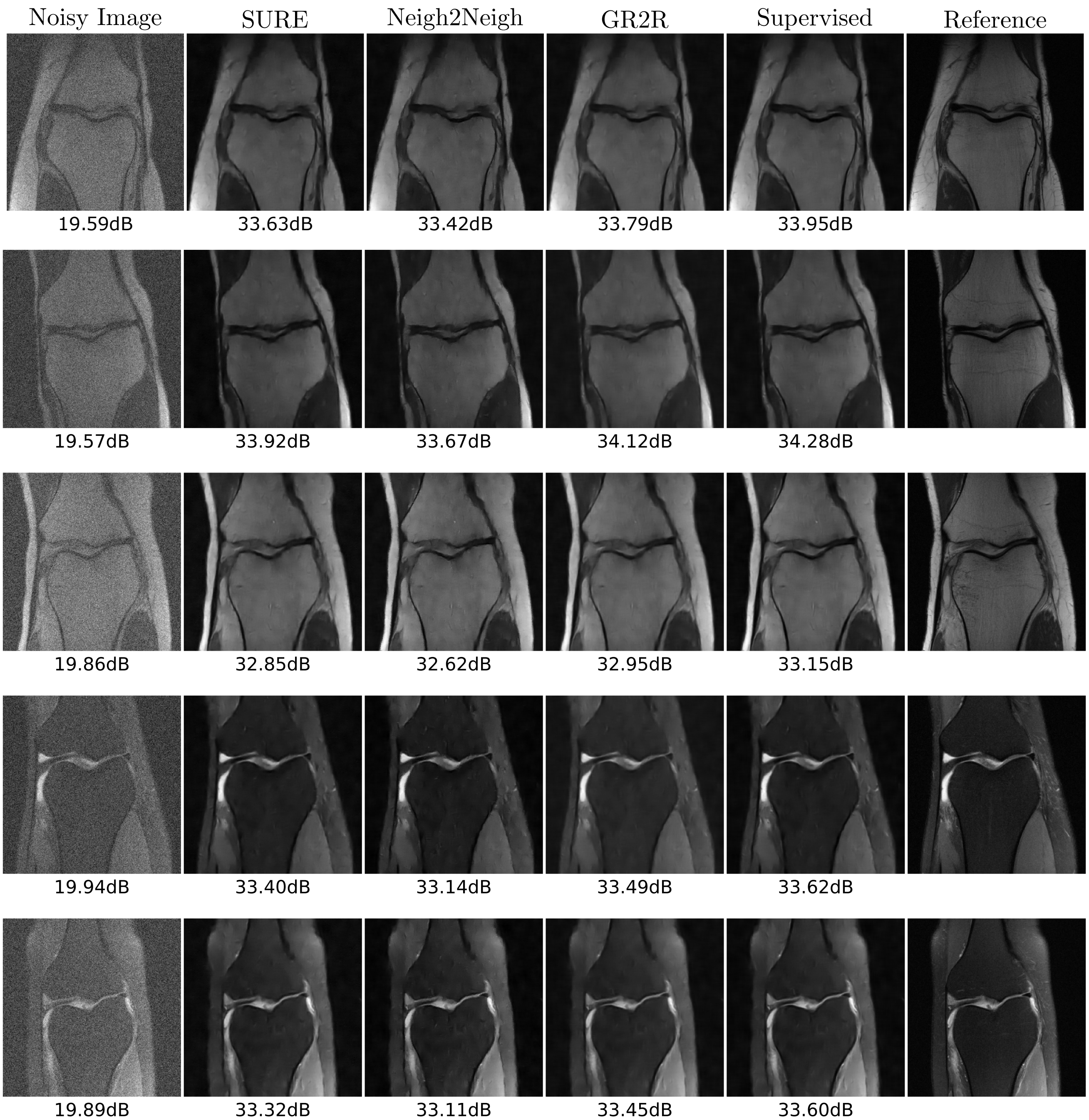

Figure 6: Gaussian Denoising in MRI Dataset.

5 General Inverse Problems

This section extends the results of the self-supervised inpainting (Section 5 of the main paper) for Poisson, Gamma, and Gaussian using DIV2K Dataset.

Table 5: PSNR/SSIM results for different noise models on inpainting in DIV2K dataset.

Noise Model	EI [7]	REI [8]	GR2R (ours)	Supervised-MSE
Poisson $\gamma = 0.5$	22.53/0.627	27.05/0.777	27.41/0.791	**28.42/0.832**
Gamma $\ell = 5$	17.06/0.467	-	26.81/0.784	**27.12/0.802**
Gaussian $\sigma = 0.1$	23.68/0.671	29.53/0.853	29.58/0.854	**29.93/0.866**

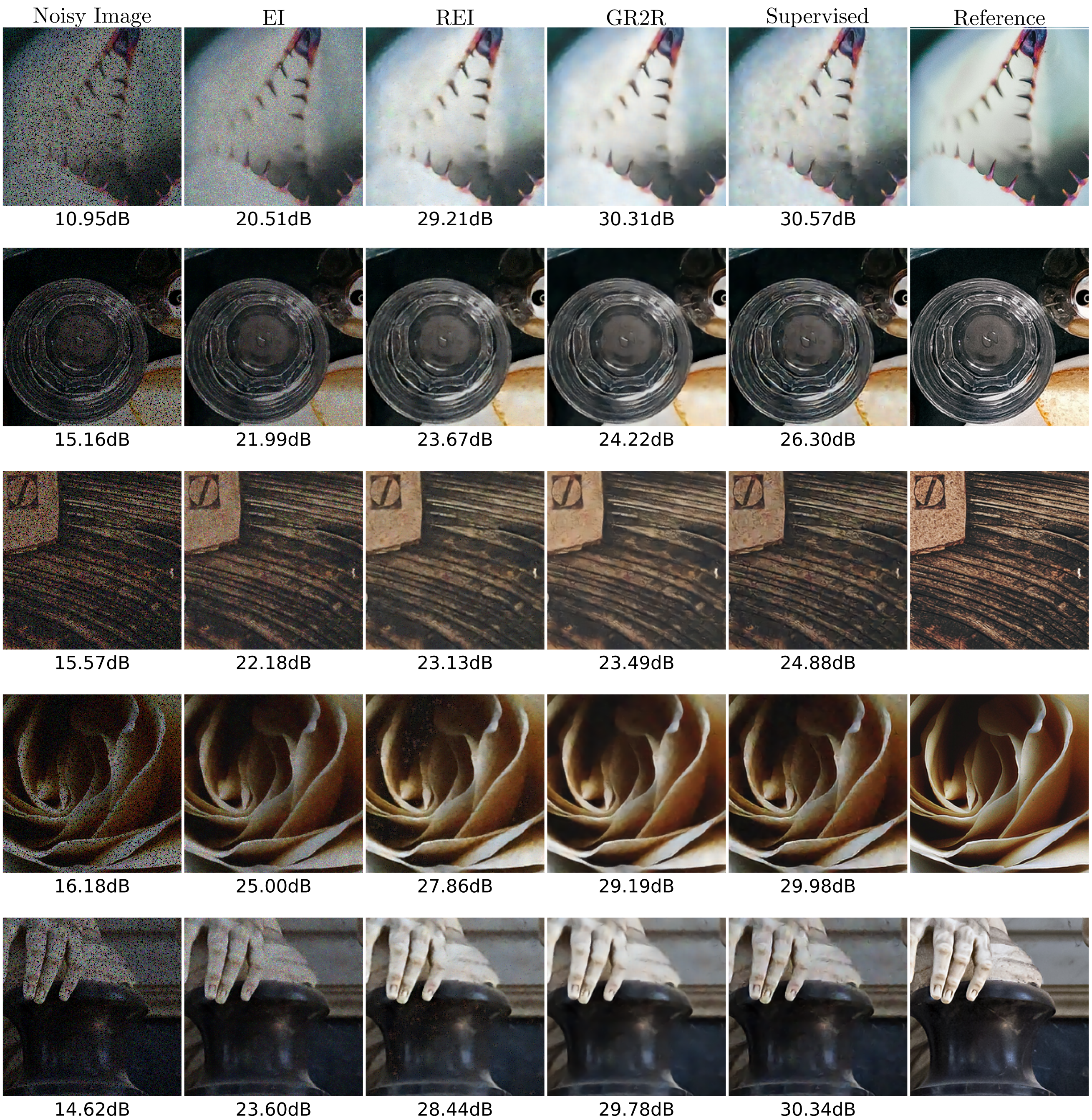

Figure 7: Inpaiting with Poisson noise in DIV2K Dataset.

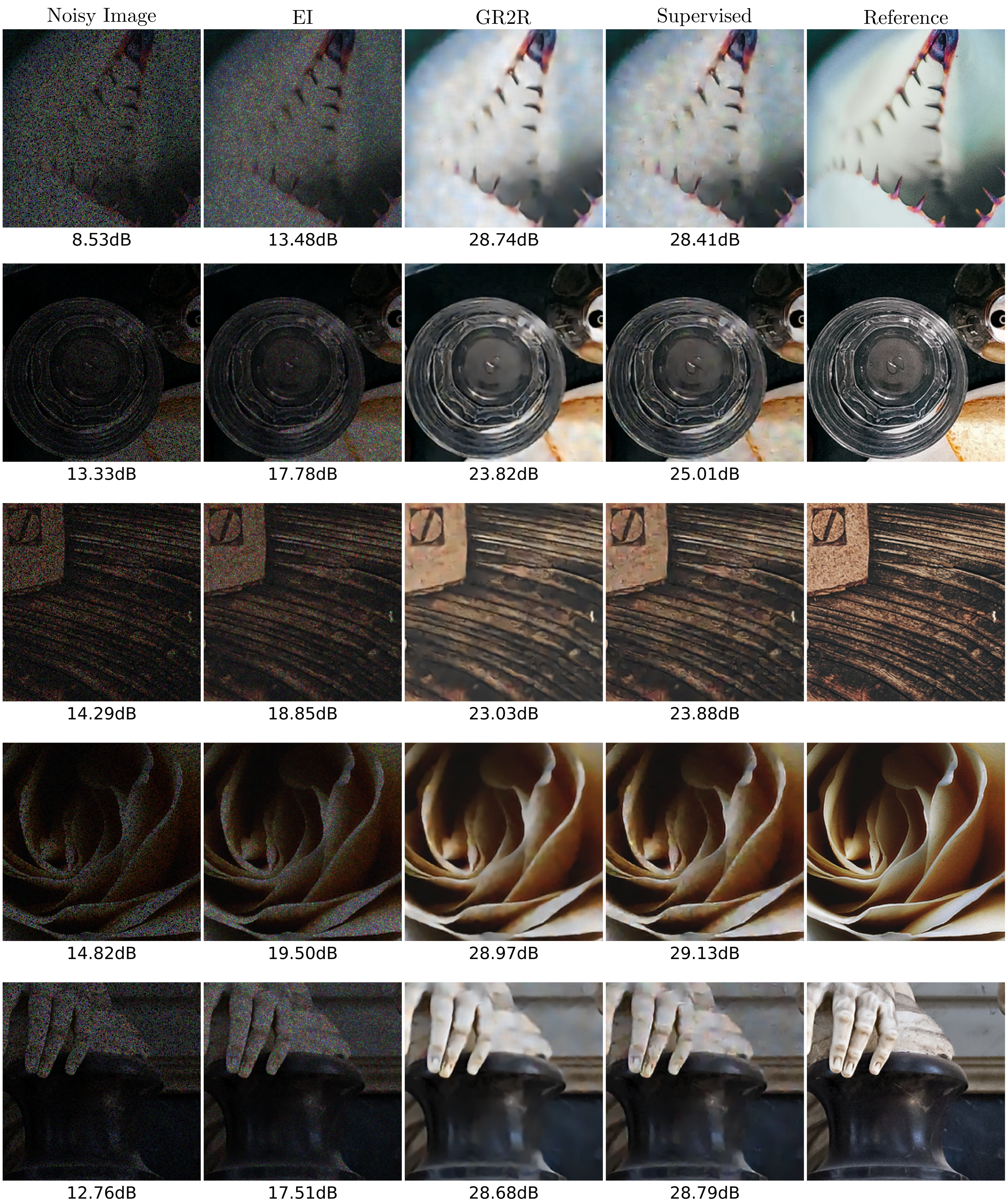

Figure 8: Inpaiting with Gamma noise in DIV2K Dataset.

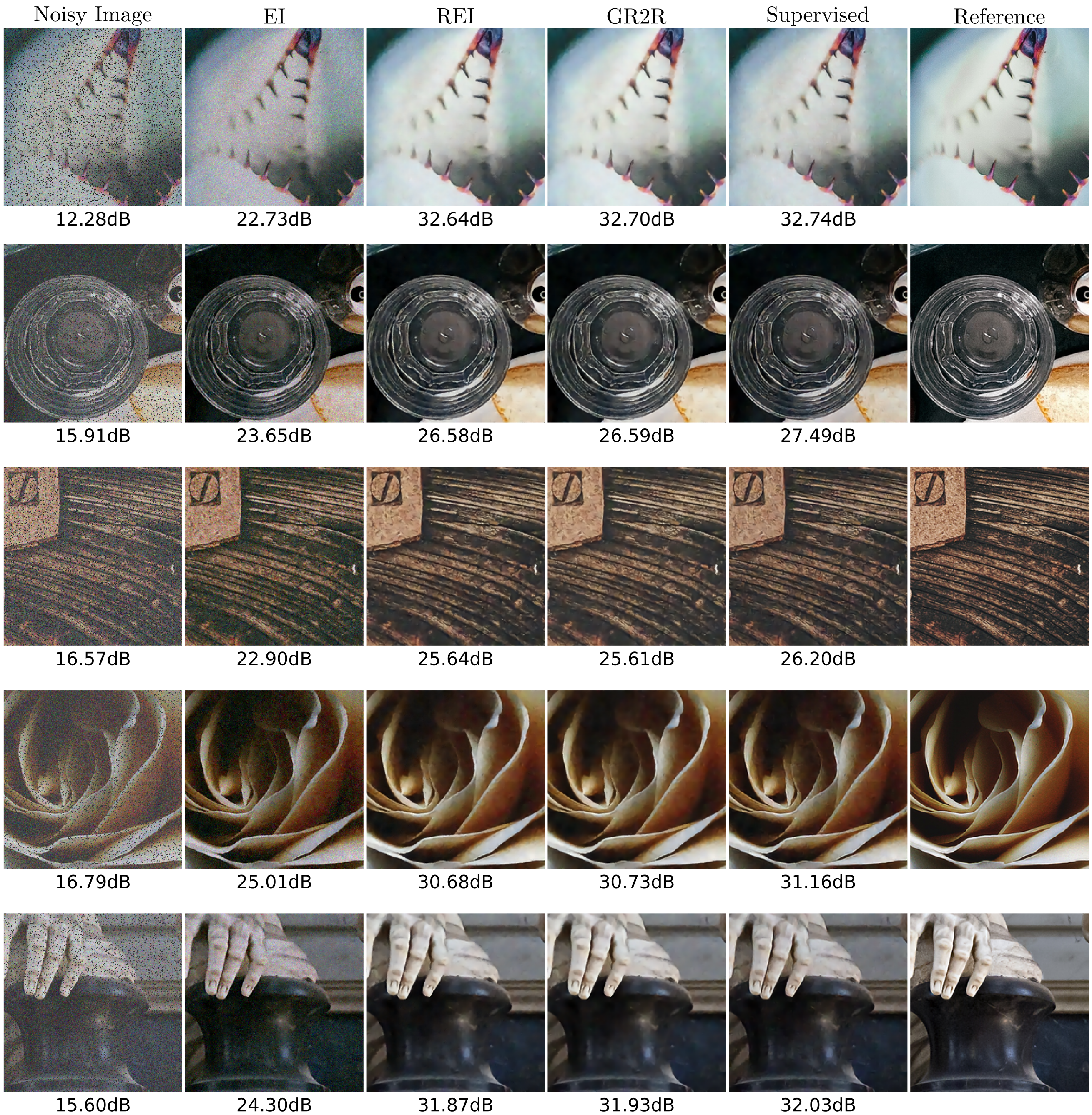

Figure 9: Inpaiting with Gaussian noise in DIV2K Dataset.

References

[1] B. Efron, *Exponential families in theory and practice.* Cambridge University Press, 2022.

[2] J. Bruna and S. Mallat, "Audio texture synthesis with scattering moments," *arXiv preprint arXiv:1311.0407*, 2013.

[3] Y. Le Montagner, E. D. Angelini, and J.-C. Olivo-Marin, "An unbiased risk estimator for image denoising in the presence of mixed poisson–gaussian noise," *IEEE Transactions on Image processing*, vol. 23, no. 3, pp. 1255–1268, 2014.

[4] T. Huang, S. Li, X. Jia, H. Lu, and J. Liu, "Neighbor2neighbor: Self-supervised denoising from single noisy images," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 14 781–14 790.

[5] K. Kim and J. C. Ye, "Noise2score: tweedie's approach to self-supervised image denoising without clean images," *Advances in Neural Information Processing Systems*, vol. 34, pp. 864–874, 2021.

[6] S. Ramani, T. Blu, and M. Unser, "Monte-carlo sure: A black-box optimization of regularization parameters for general denoising algorithms," *IEEE Transactions on image processing*, vol. 17, no. 9, pp. 1540–1554, 2008.

[7] D. Chen, J. Tachella, and M. E. Davies, "Equivariant imaging: Learning beyond the range space," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 4379–4388.

[8] ——, "Robust equivariant imaging: a fully unsupervised framework for learning to image from noisy and partial measurements," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5647–5656.