Supplementary Materials

I. Derivation of Equation (14)-(16).

Derivation from equation (14) to (15). Recall that we assumed an uniform prior distribution of $\boldsymbol{\mu}$ as follows

$$p(\boldsymbol{\mu}) = \begin{cases} 1/c & \text{if } \boldsymbol{\mu} \in \mathbb{U}, \\ 0 & \text{otherwise}, \end{cases} \tag{1}$$

where $\mathbb{U}$ is a sufficiently large bounded set that includes $\mathbf{0}$ and c is the measure of $\mathbb{U}$. For simplicity, we consider $\mathbb{U} = [-U, U]^J$, where J is the dimension of $\boldsymbol{\mu}$. On the other hand, $q(\boldsymbol{\mu}|\boldsymbol{\alpha}) = \prod_j q(\boldsymbol{\mu}_j|\boldsymbol{\alpha}_j)$, where $q(\boldsymbol{\mu}_j|\boldsymbol{\alpha}_j)$ is the Bernoulli distribution with probability q_j.

By definition, we have

$$\mathrm{KL}(q(\boldsymbol{\mu}|\boldsymbol{\alpha})||p(\boldsymbol{\mu})) = \int q(\boldsymbol{\mu}|\boldsymbol{\alpha}) \log q(\boldsymbol{\mu}|\boldsymbol{\alpha}) d\boldsymbol{\mu} - \int q(\boldsymbol{\mu}|\boldsymbol{\alpha}) \log p(\boldsymbol{\mu}) d\boldsymbol{\mu}. \tag{2}$$

Note that $p(\boldsymbol{\mu})$ is a continuous distribution while $q(\boldsymbol{\mu}|\boldsymbol{\alpha})$ is discrete. For the discrete Bernoulli distribution, $\int q(\boldsymbol{\mu}|\boldsymbol{\alpha}) \log q(\boldsymbol{\mu}|\boldsymbol{\alpha}) d\boldsymbol{\mu}$ is calculated by

$$\int q(\boldsymbol{\mu}|\boldsymbol{\alpha}) \log q(\boldsymbol{\mu}|\boldsymbol{\alpha}) d\boldsymbol{\mu} = \sum_j \big(q_j \log q_j + (1-q_j)\log(1-q_j)\big). \tag{3}$$

For the calculation of $\int q(\boldsymbol{\mu}|\boldsymbol{\alpha}) \log p(\boldsymbol{\mu}) d\boldsymbol{\mu}$, we use a continuous representation of the Bernoulli distribution as follows

$$q(\boldsymbol{\mu}_j|\boldsymbol{\alpha}_j) = (1-q_j) D_0(\boldsymbol{\mu}_j) + q_j D_{\boldsymbol{\alpha}_j}(\boldsymbol{\mu}_j). \tag{4}$$

where $D_a(x)$ is the Dirac delta function that satisfies

$$\int D_a(x)dx = 1, \ \int f(x) D_a(x) dx = f(a). \tag{5}$$

Next we consider two cases: $\boldsymbol{\alpha} \in \mathbb{U}$ and $\boldsymbol{\alpha} \notin \mathbb{U}$.

- $\boldsymbol{\alpha} \in \mathbb{U}$: It means that $\boldsymbol{\alpha}_j \in [-U, U], \ \forall j$. Then

  $$\int q(\boldsymbol{\mu}|\boldsymbol{\alpha}) \log p(\boldsymbol{\mu}) d\boldsymbol{\mu} = -\log c \prod_{q_j, \boldsymbol{\mu}_j \in \{0, \boldsymbol{\alpha}_j\}} \Big(q_j \mathbf{1}_{\boldsymbol{\alpha}_j}(\boldsymbol{\mu}_j) + (1-q_j)\mathbf{1}_0(\boldsymbol{\mu}_j)\Big), \tag{6}$$

  where $\mathbf{1}_a(x)$ is the indicator function:

  $$\mathbf{1}_a(x) = \begin{cases} 1 \text{ if } x = a, \\ 0 \text{ if } x \neq a. \end{cases} \tag{7}$$

  That is, $\int q(\boldsymbol{\mu}|\boldsymbol{\alpha}) \log p(\boldsymbol{\mu}) d\boldsymbol{\mu}$ is a constant regardless of the value of $\boldsymbol{\alpha}$ as long as $\boldsymbol{\alpha} \in \mathbb{U}$.
- $\boldsymbol{\alpha} \notin \mathbb{U}$: it is direct to get that

  $$\int q(\boldsymbol{\mu}|\boldsymbol{\alpha}) \log p(\boldsymbol{\mu}) d\boldsymbol{\mu} \leq \log p(\boldsymbol{\alpha}) = -\infty. \tag{8}$$

In summary, $\mathrm{KL}(q(\boldsymbol{\mu}|\boldsymbol{\alpha})||p(\boldsymbol{\mu}))$ is a constant if $\boldsymbol{\alpha} \in \mathbb{U}$ and $+\infty$ otherwise. Thus, we obtain

$$\begin{aligned} &\min_{\boldsymbol{\alpha}} \ \mathrm{KL}(q(\boldsymbol{\mu}|\boldsymbol{\alpha})||p(\boldsymbol{\mu})) - \mathbb{E}_{\boldsymbol{\mu}\sim q(\boldsymbol{\mu}|\boldsymbol{\alpha})} \log p(\boldsymbol{y}_0|\boldsymbol{\mu}) \\ \Longleftrightarrow &\max_{\boldsymbol{\alpha}} \ \mathbb{E}_{\boldsymbol{\mu}\sim q(\boldsymbol{\mu}|\boldsymbol{\alpha})} \log p(\boldsymbol{y}_0|\boldsymbol{\mu}) - \delta_{\mathbb{U}}(\boldsymbol{\alpha}) \\ \Longleftrightarrow &\max_{\boldsymbol{\alpha}\in\mathbb{U}} \ \mathbb{E}_{\boldsymbol{\mu}\sim p(\boldsymbol{\mu}|\boldsymbol{\alpha})} \log p(\boldsymbol{y}_0|\boldsymbol{\mu}). \end{aligned} \tag{9}$$

Derivation from equation (15) to (16). Since we assume $\boldsymbol{x}_0 = \mathcal{G}_{\boldsymbol{\mu}}(\boldsymbol{\epsilon}_0)$ and $\boldsymbol{y}_0 = |\boldsymbol{A}\boldsymbol{x}_0| + \boldsymbol{n}$, where $\boldsymbol{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{I})$, it can be obtained that $\boldsymbol{y}_0|\boldsymbol{\mu} \sim \mathcal{N}(|\boldsymbol{A}\mathcal{G}_{\boldsymbol{\mu}}(\boldsymbol{\epsilon}_0)|, \sigma^2\boldsymbol{I})$, *i.e.* $\log p(\boldsymbol{y}_0|\boldsymbol{\mu}) \propto -\|\boldsymbol{y}_0 - |\boldsymbol{A}\mathcal{G}_{\boldsymbol{\mu}}(\boldsymbol{\epsilon}_0)|\|_2^2$. Recall that $\boldsymbol{\mu} = \boldsymbol{\alpha} \odot \boldsymbol{d}$, then we have

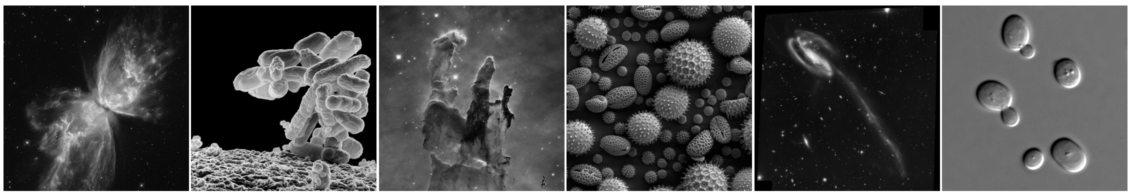
$$\begin{aligned} &\max_{\boldsymbol{\alpha}\in\mathbb{U}} \ \mathbb{E}_{\boldsymbol{\mu}\sim p(\boldsymbol{\mu}|\boldsymbol{\alpha})} \log p(\boldsymbol{y}_0|\boldsymbol{\mu}) \\ \Longleftrightarrow &\min_{\boldsymbol{\alpha}\in\mathbb{U}} \ \mathbb{E}_{\boldsymbol{\mu}\sim p(\boldsymbol{\mu}|\boldsymbol{\alpha})} \big\|\boldsymbol{y}_0 - |\boldsymbol{A}\mathcal{G}_{\boldsymbol{\mu}}(\boldsymbol{\epsilon}_0)|\big\|_2^2 \\ \Longleftrightarrow &\min_{\boldsymbol{\alpha}\in\mathbb{U}} \ \mathbb{E}_{\boldsymbol{d}} \big\|\boldsymbol{y}_0 - |\boldsymbol{A}\mathcal{G}_{\boldsymbol{\alpha}\odot\boldsymbol{d}}(\boldsymbol{\epsilon}_0)|\big\|_2^2. \end{aligned} \tag{10}$$

The constraint $\boldsymbol{\alpha} \in \mathbb{U}$ is omitted in (16) as the feasible set $\mathbb{U}$ is sufficiently large.

II. TEST SETS

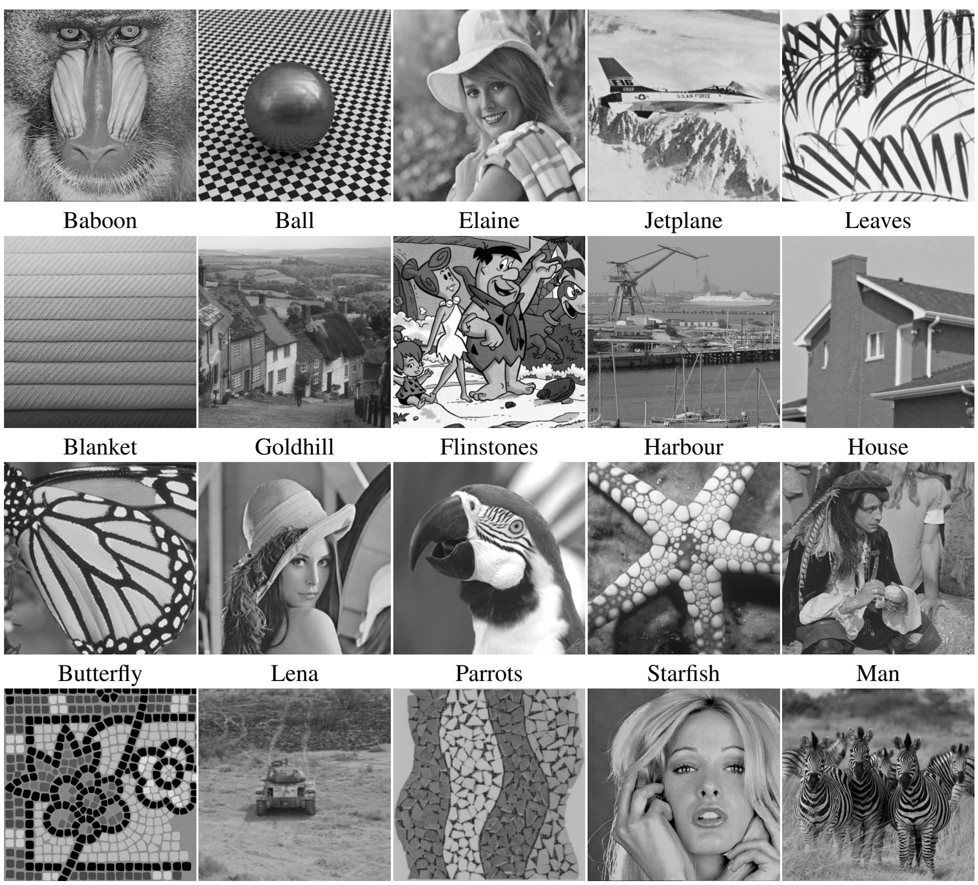
Barbara Boat Cameraman Couple Peppers Bridge

Fig. 1: Images of Natural-6 test set.



Butterfly Nebula E.Coli Pillars of Creation Pollen Tadpole Galaxy Yeast

Fig. 2: Images of Unnatural-6 test set.



Baboon Ball Elaine Jetplane Leaves

Blanket Goldhill Flinstones Harbour House

Butterfly Lena Parrots Starfish Man

Masaike Tank Mosaic Woman blonde Zebra

Fig. 3: Images of Set-20 test set.

III. Additional Visual Results

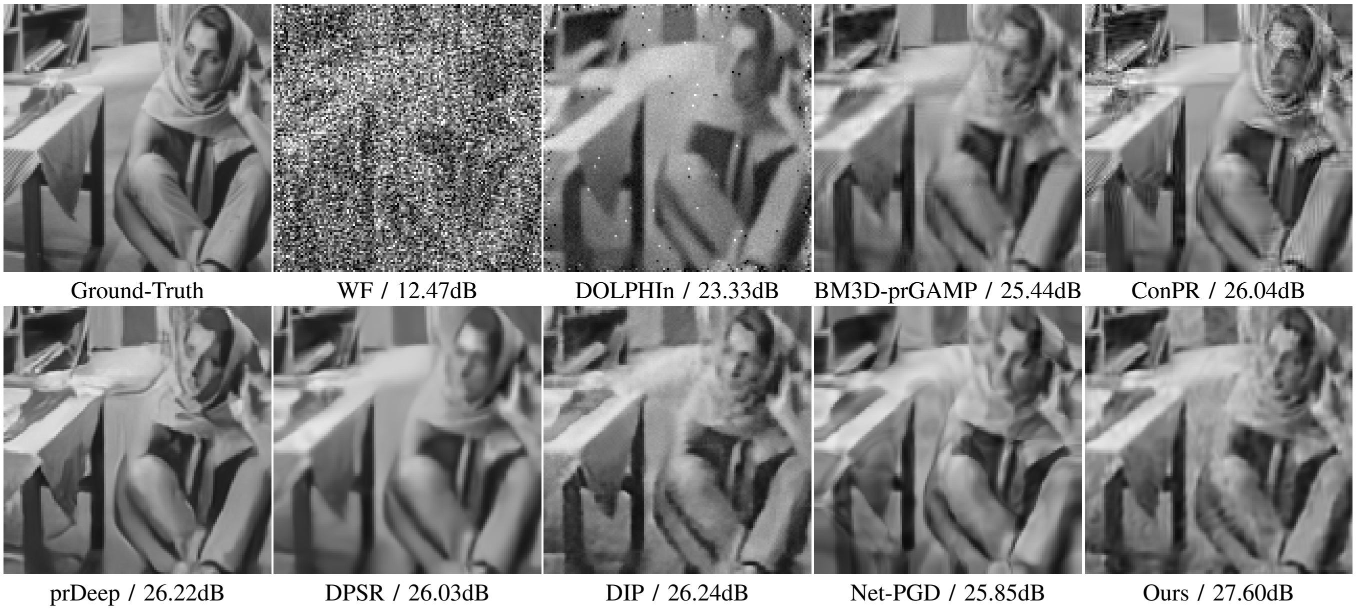


Fig. 4: Reconstructions on image "Barbara" in the presence of AWGN (SNR=15dB) with $R = 1$ bipolar mask.

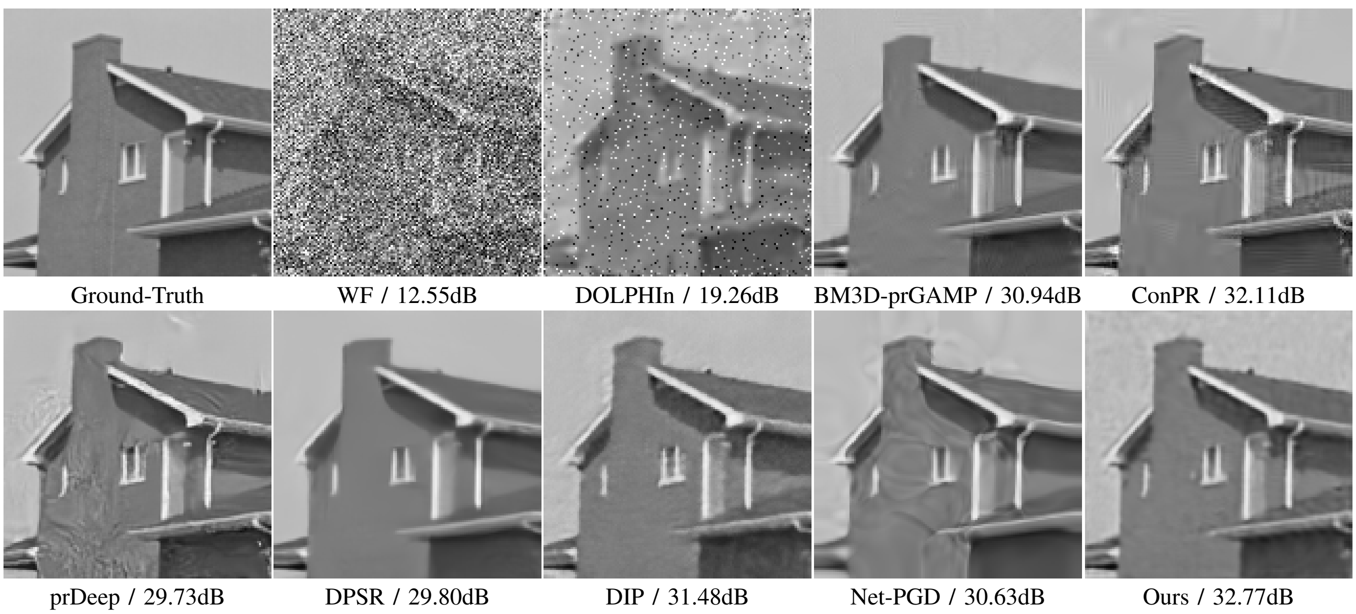


Fig. 5: Reconstructions on image "House" in the presence of AWGN (SNR=20dB) with $R = 1$ bipolar mask.

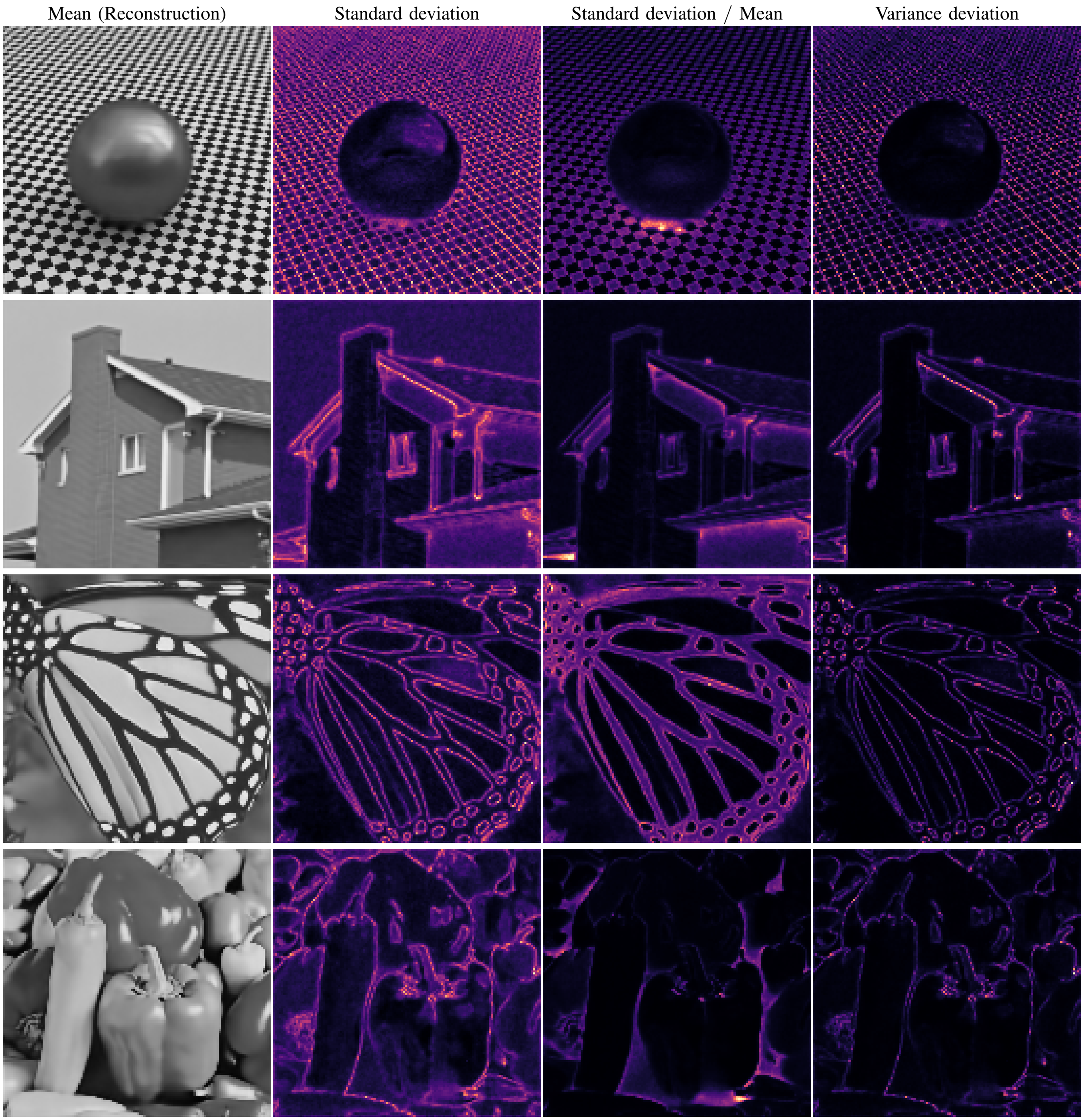


Fig. 6: Pixel-wise statistics over 100 predictions on image “Ball”, “House”,“Butterfly” , and “Peppers”. Four bipolar masks are used in these experiments.

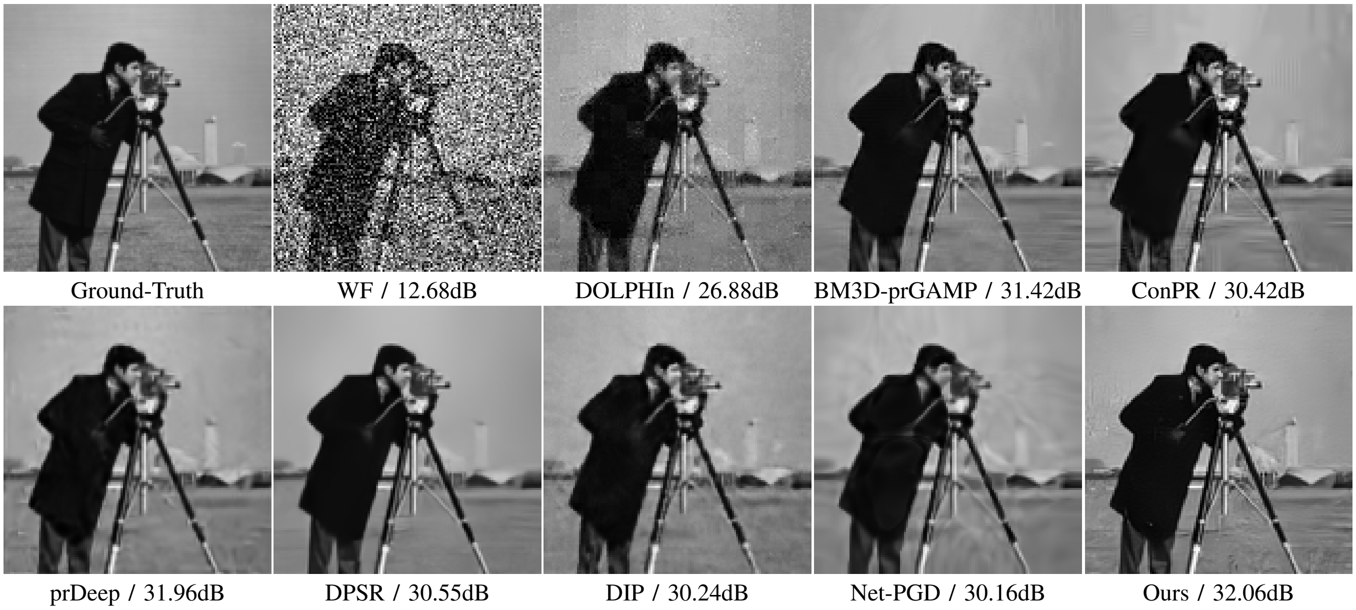


Fig. 7: Reconstructions on image “Cameraman” in the presence of Poisson noise ($\alpha = 27$) with $R = 2$ bipolar masks.

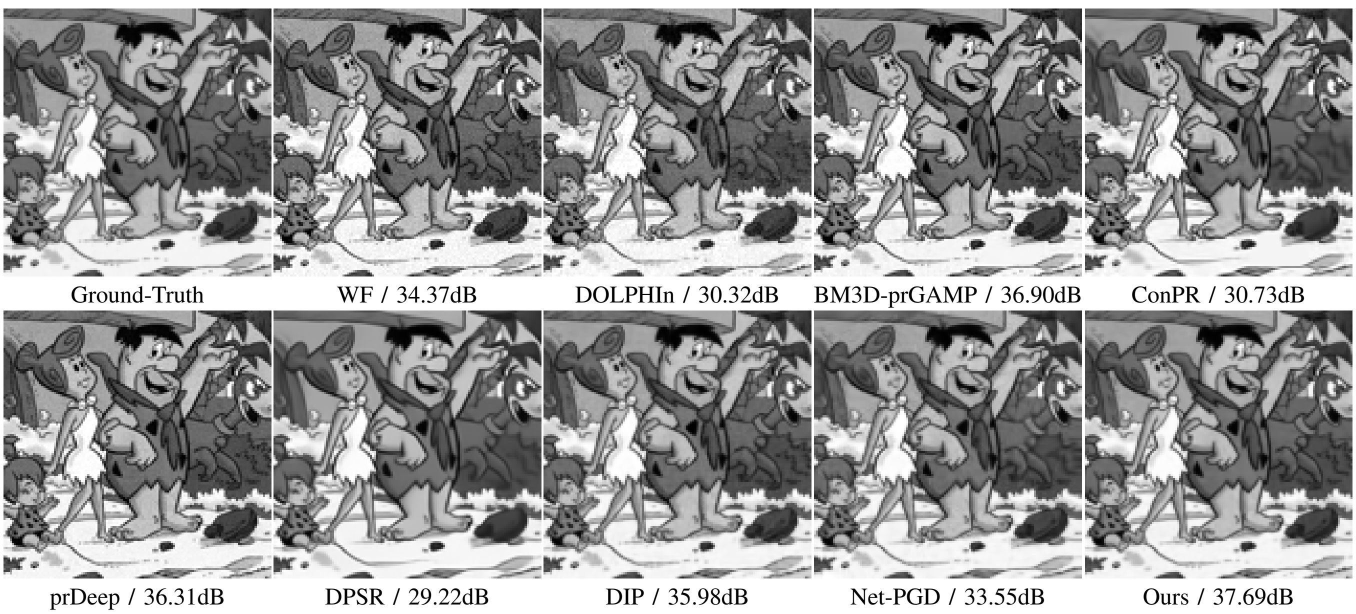


Fig. 8: Reconstructions on image “Flinstones” in the presence of Poisson noise ($\alpha = 9$) with $R = 4$ bipolar masks.

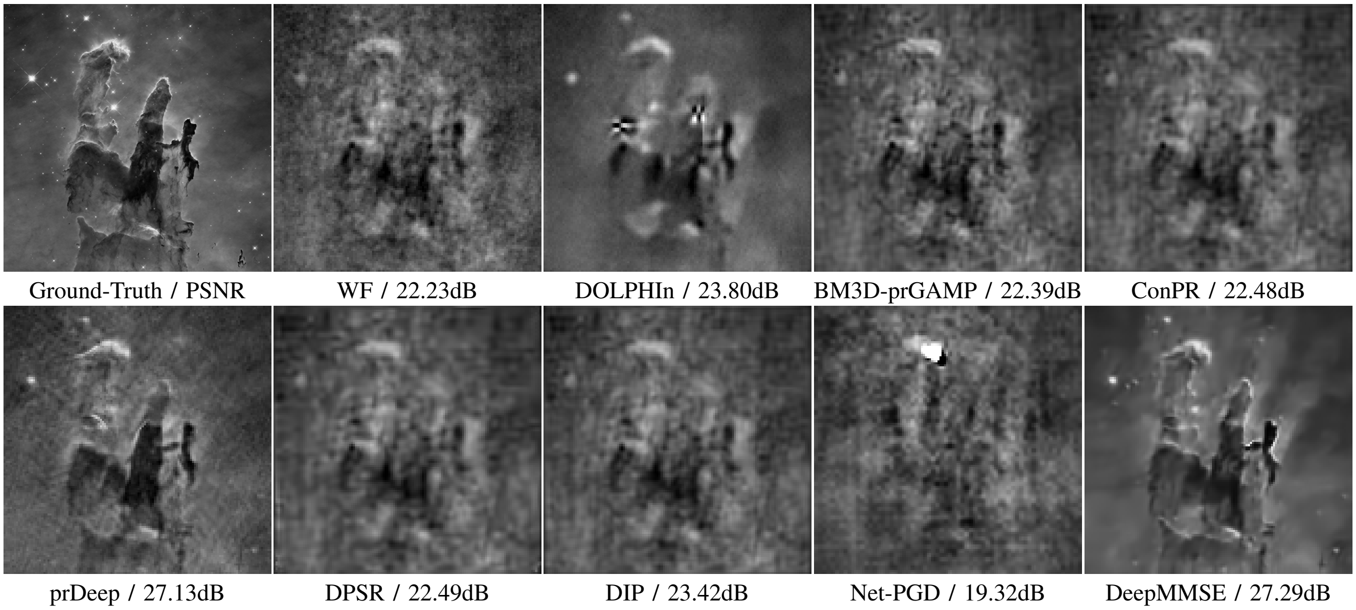


Fig. 9: Reconstructions on image “Pillars of Creation” from $4\times$ oversampled Fourier magnitude measurements with Poisson noise $(\gamma = 2)$.

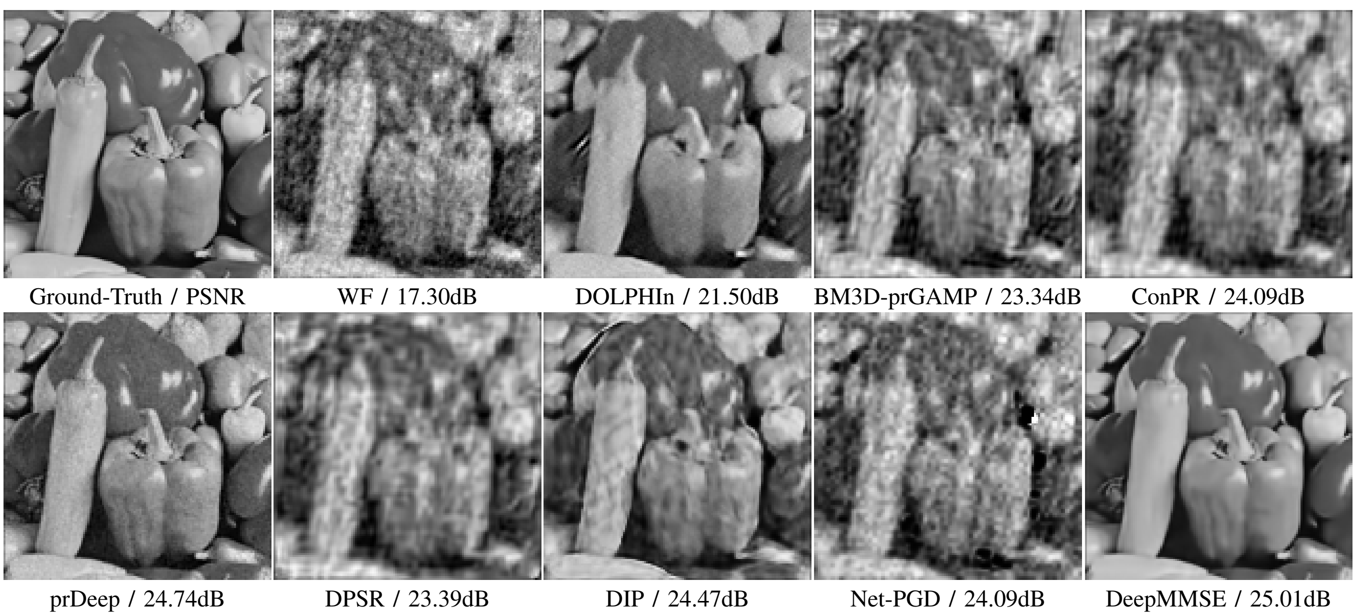


Fig. 10: Reconstructions on image “Peppers” from $4\times$ oversampled Fourier magnitude measurements with Poisson noise $(\gamma = 4)$.

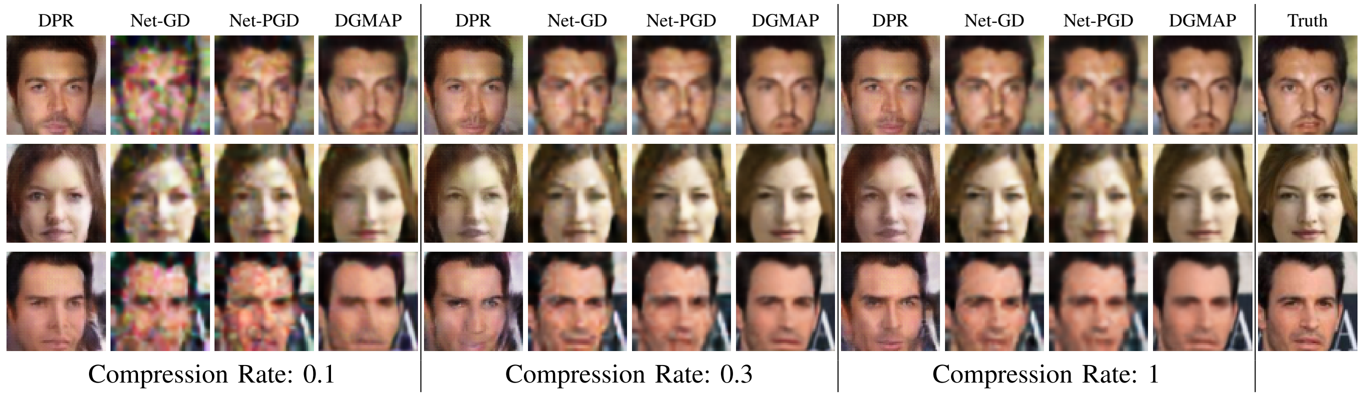


Fig. 11: Visual reconstruction results of compressive PR on CelebA dataset by different methods at different compression rates.

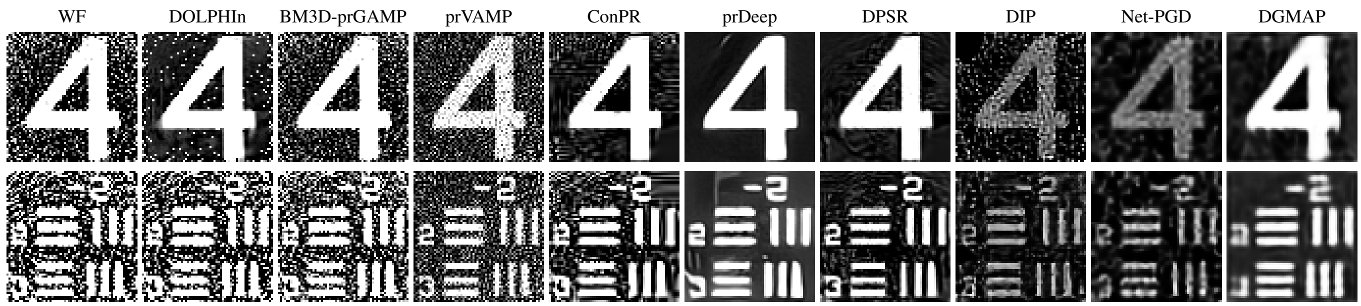


Fig. 12: Visual comparison of the results of different PR methods from the public dataset.