Supplementary Materials for A Bayesian Perspective on the Deep Image Prior

Zezhou Cheng Matheus Gadelha Subhransu Maji Daniel Sheldon
University of Massachusetts - Amherst
{zezhoucheng, mgadelha, smaji, sheldon}@cs.umass.edu

We provide the proof of Theorem 1 and show additional visualizations of the denoising and inpainting results.

Proof. Let $r = t_2 - t_1$. First, observe that

$$\bar{x}(t_1)^T \bar{x}(t_2) = \sum_{k=1}^{n} \left(\sum_{i=0}^{d-1} x(k, t_1 - i) x(k, t_2 - i) \right).$$

where n is the number of channels in the input, d is the filter width. By stationarity, the terms $x(k, t_1 - i)x(k, t_2 - i)$ are identically distributed with expected value $K_x(r)$, even though they are not necessarily independent. Since the channels are drawn independently, the parenthesized expressions in the sum over k are iid with mean $dK_x(r)$. According to Equation (2) in the main text, we have

$$\begin{aligned} K_z^{\text{erf}}(t_1, t_2) &= \mathbb{E}_x V_{\text{erf}}\big(\bar{x}(t_1), \bar{x}(t_2)\big) \\ &= \mathbb{E}_x \frac{2}{\pi} \sin^{-1} \frac{\sigma_u^2 \bar{x}(t_1)^T \bar{x}(t_2)}{\sigma_u^2 \|\bar{x}(t_1)\| \|\bar{x}(t_2)\|}. \end{aligned}$$

Consider the sequence $Y_n = (a_n, b_n, c_n)$

$$a_n = \frac{1}{n}\left(\bar{x}(t_1)^T \bar{x}(t_2)\right), \quad b_n = \frac{1}{n}\left(\bar{x}(t_1)^T \bar{x}(t_1)\right), \quad c_n = \frac{1}{n}\left(\bar{x}(t_2)^T \bar{x}(t_2)\right). \tag{1}$$

By the strong law of large numbers we know the sequences a_n, b_n and c_n each converge almost surely to their expected values $dK_x(t_1, t_2)$, $dK_x(t_1, t_1)$ and $dK_x(t_2, t_2)$, respectively. Since K_x is stationary this is equal to $dK_x(r)$, $dK_x(0)$, and $dK_x(0)$, respectively, where d is the filter width. Thus we have that Y_n converges almost surely to $\bar{y} = \big(dK_x(r), dK_x(0), dK_x(0)\big)$.

Consider the continuous function $g(y) = \frac{2}{\pi} \sin^{-1}\left(\frac{a}{\sqrt{bc}}\right)$ for $y = (a, b, c)$. Applying the continuous mapping theorem, we have that

$$y_n \xrightarrow{a.s.} \mu \implies g(y_n) \xrightarrow{a.s.} g(\mu) \tag{2}$$

Thus,

$$K_z^{\text{erf}}(t_1, t_2) \xrightarrow{a.s.} \frac{2}{\pi} \sin^{-1} \frac{K_x(r)}{K_x(0)}.$$

□

Additional visualizations. Figure 1 and 2 present results for various inputs using various inference schemes for image denoising and image inpainting respectively.

References

[1] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep Image Prior. In *Computer Vision and Pattern Recognition (CVPR)*, 2018. 3

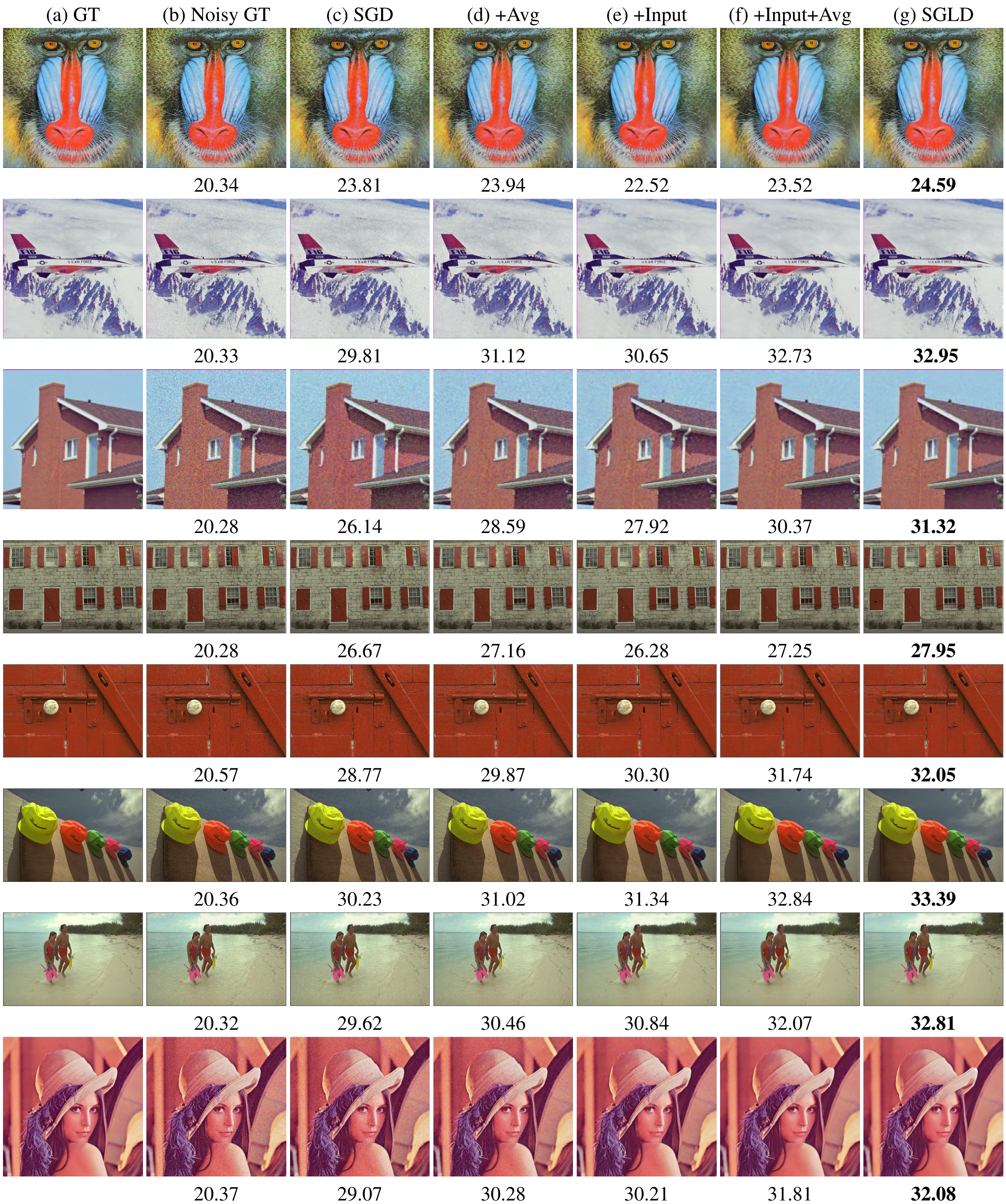

Figure 1. *(Best viewed magnified.)* **Comparison of inference schemes for image denoising.** From left to right each column shows **(a)** ground truth image, **(b)** the ground truth image corrupted with Gaussian noise of $\sigma = 25$, **(c)** inference with SGD and early stopping (iteration picked at maximum PSNR), **(d)** previous step with exponential window average, **(e)** SGD with input noise and early stopping, **(f)** the previous procedure with exponential window averaging, and **(g)** SGLD. Underneath each figure the PSNR values are listed. Note that except SGLD every procedure requires early stopping.

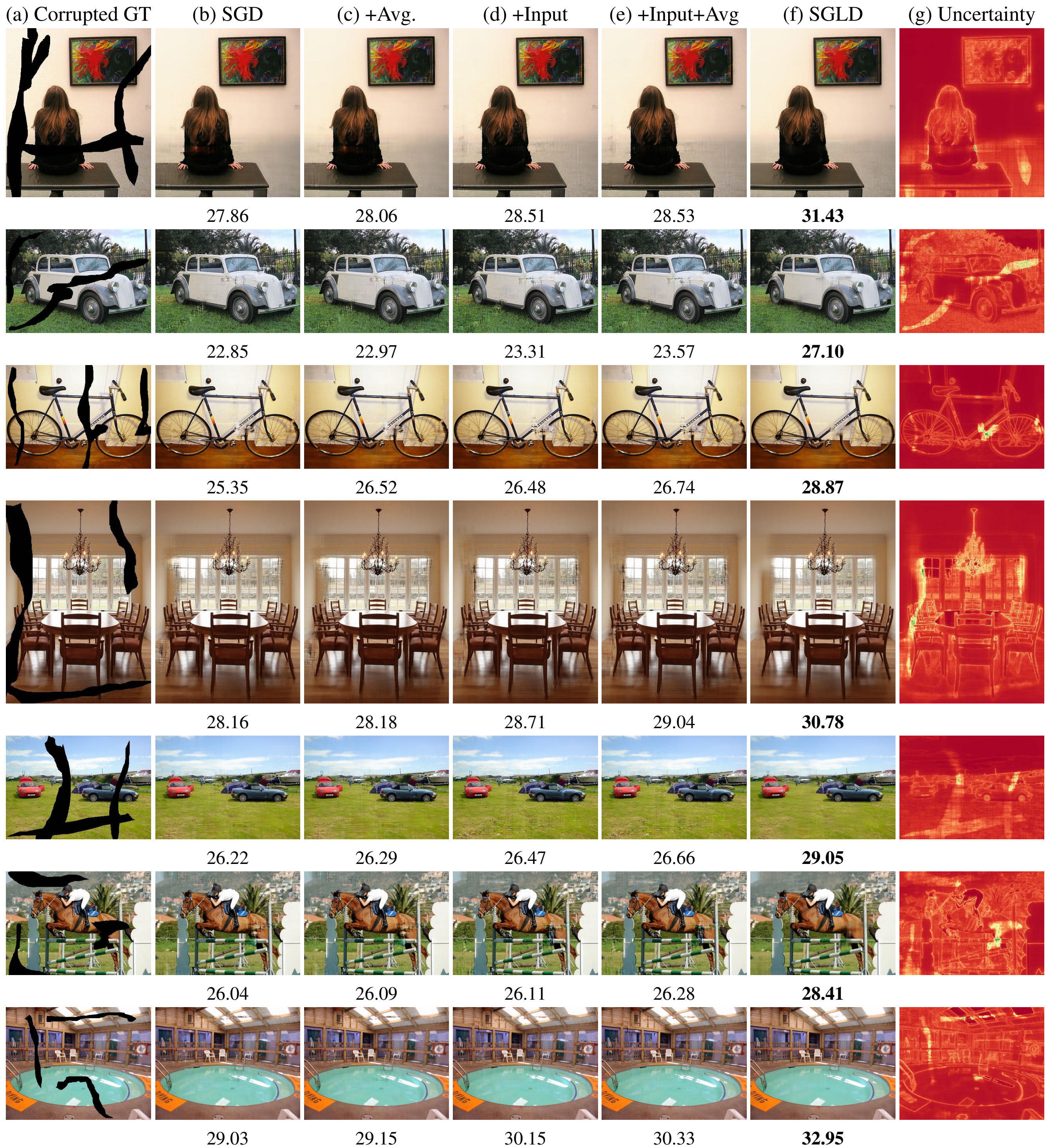

Figure 2. *(Best viewed magnified.)* **Comparison of inference schemes for image inpainting.** **(a)** The input image with parts masked out. **(b-f)** Results using various inference schemes (see the previous figure and the main text for details of the methods). **(f)** SGLD achieves higher PSNR (shown underneath each figure) than the baseline methods **(b-e)**. The early stopping iteration is set as 2000 for the first two schemes **(b, c)** and 3000 for the third and fourth methods **(d, e)**. For these experiments we use a 6-layer auto-encoder without skip connections and random noise as input as reported in [1]. **(g)** The variance estimated from the posterior samples visualized as a heat map. Notice that the uncertainty is low in the missing regions that are surrounded by areas of relatively uniform appearance such as wall, sky and swimming pool, and higher in non-uniform areas such as those near the boundaries of different object in the image.