

6 Appendix

6.1 Validity of Eq. (14)

In the iterative process (13) that solves the problem (9a), it is formed by the alternating iteration of two points: r^t and x^t . In Eq. (13a) (which is always called the gradient descent step), the step size β is always chosen from $(0, \frac{1}{L_f}]$ [14][4][6], where $f(x) = \frac{1}{2} \|y - Ax\|_2^2$ and L_f is the smallest Lipschitz constant of $\nabla f(x)$. Indeed, $L_f = \|A\|_2^2$, where $\|A\|_2$ is the spectral norm of A . From Eq. (13a), we can observe that $r^t - x^{t-1} = \beta A^T (y - Ax^{t-1})$, which implies

$$\|r^t - x^{t-1}\|_2^2 = \beta \|A^T (y - Ax^{t-1})\|_2^2 \quad (29a)$$

$$\leq \beta \|A^T\|_2^2 \|y - Ax^{t-1}\|_2^2 \quad (29b)$$

$$= \beta \|A\|_2^2 \|y - Ax^{t-1}\|_2^2 \quad (29c)$$

$$\leq \frac{1}{L_f} \|A\|_2^2 \|y - Ax^{t-1}\|_2^2 \quad (29d)$$

$$= \|y - Ax^{t-1}\|_2^2. \quad (29e)$$

As the problem (9a) is convex provided Ψ is a linear operator, the PGD (proximal gradient descent) algorithm (which is the iterative process (13)) is global-convergence [4][3]. Let $\bar{x} = \lim_{t \rightarrow \infty} x^{t-1}$, by inequality (29), we have

$$\lim_{t \rightarrow \infty} \|r^t - x^{t-1}\|_2 \leq \|y - A\bar{x}\|_2. \quad (30)$$

On the other hand, if the penalty coefficient λ of problem (9a) is small enough, the magnitude of the regularization term $\lambda \sum_{i=1}^n \frac{|x_i|}{(|\bar{c}_i| + \varepsilon_i)^{1-q}}$ is smaller than $f(x) = \frac{1}{2} \|y - Ax\|_2^2$. This implies the optimal solution \bar{x} to the problem (9a) leads to a relatively small value $\frac{1}{2} \|y - A\bar{x}\|_2^2$. In other words, if λ is small enough, by inequality (30), we can validate Eq. (14) as

$$\bar{r} = \lim_{t \rightarrow \infty} r^t \approx \lim_{t \rightarrow \infty} x^{t-1} = \bar{x}. \quad (31)$$

Moreover, the two points r^t and x^{t-1} are close if t is sufficiently large, $\beta \in (0, \frac{1}{L_f}]$, and λ is sufficiently small. Under the circumstance, instead of choosing $\bar{c} = x^t$ as the solution to the problem (9b), the approximation

$$\bar{c} \approx r^t, \quad (32)$$

where r^t is the iterative point of the proximal gradient descent algorithm that solves the problem (9a), is adopted. The benefits of using the approximate optimal solution instead of the exact optimal solution to problem (9b) will be described in Sec. 3.3.

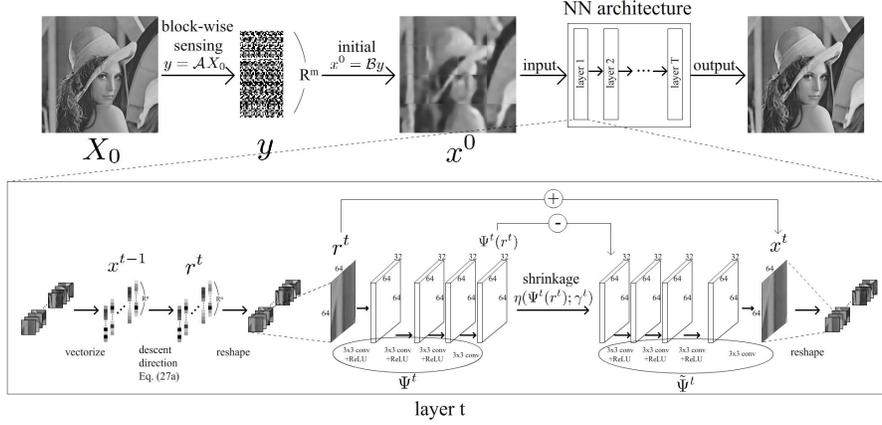


Fig. 5: The flowchart of QISTA-ImageNet. The *sensing* is the operation $y = \mathcal{A} \cdot \text{vec}(X_0)$, where $\text{vec}(X_0)$ is a vector representation of an image X_0 . The *initial solution* is obtained from Eq. (28). The *descent direction* corresponds to Eq. (27a). The *shrinkage* corresponds to the operator $\eta(\cdot; \gamma^t)$ in Eq. (27b).

6.2 Flowchart of QISTA-ImageNet

Fig. 5 illustrates the structure of QISTA-ImageNet.

6.3 Training Details

We adopted a similar training setting with SCSNet [31] (and CSNet⁺ [32]) in that the measurement matrix A is operated on the image block with size 32×32 . Thus, certain pre-processing and post-processing are required in Step 2 of QISTA-ImageNet. In Step 2, the input has a size of $64 \times 64 \times b$, where 64×64 is the size of a patch of training data and b is the training batch size. The pre-processing and post-processing in this step are as follows. To simplify the discussion, here we suppose the batch size $b = 1$.

1. Pre-processing: As the operator \mathcal{A} is unfolded by $A \in \mathbb{R}^{m \times 1024}$ (the measurement rate is $\frac{m}{1024}$) and the input x^{t-1} has a size of 64×64 , we first divided the patch x^{t-1} into 4 blocks, each of which has a size of 32×32 . Next, we vectorized each block into a vector-form \mathbb{R}^{1024} (here we have four vectors: x_1, x_2, x_3, x_4) and obtained the measurement vectors $y_i = \mathcal{A}x_i \in \mathbb{R}^m$, $i = 1, 2, 3, 4$. The input size change during pre-processing can be summarized as

$$64 \times 64 \xrightarrow{\text{divide}} 32 \times 32 \times 4 \rightarrow 1024 \times 4 \xrightarrow{\mathcal{A}} m \times 4. \quad (33)$$

2. Post-processing: For the operator \mathcal{B} , as $y - \mathcal{A}x^{t-1} \in \mathbb{R}^{m \times 4}$ and \mathcal{B} is unfolded by $A^T \in \mathbb{R}^{1024 \times m}$, we have $\mathcal{B}(y - \mathcal{A}x^{t-1}) \in \mathbb{R}^{1024 \times 4}$. After being operated

by \mathcal{B} , we reverse the pre-processing that we performed before \mathcal{A} , which leads the results back into the patch form of size 64×64 . The input size change during post-processing can be summarized as

$$m \times 4 \xrightarrow{\mathcal{B}} 1024 \times 4 \rightarrow 32 \times 32 \times 4 \xrightarrow{\text{merge}} 64 \times 64. \quad (34)$$

The pre-processing in Eq. (33) and post-processing in Eq. (34) reduce the memory usage of the measurement matrix \mathcal{A} . When an image is sensed by \mathcal{A} with the measurement rate r_0 , r_0 is fixed regardless of whether the image block size is 64×64 or 32×32 , but the required storage usage for \mathcal{A} is different. For example, if the size of a patch is 64×64 , \mathcal{A} has a size of $m \times 4096$, whereas if the size of a patch is 32×32 , \mathcal{A} has a size of $m \times 1024$.

During training, we adopted the Adam optimizer [21] with a learning rate of 0.0001. The network was trained for 120 epochs with a batch size of 64.

6.4 More Recovery and Visual Results in Sec. 4.3

As having been described in Sec. 4.3, we show more recovery (Table 4, Table 5, and Table 6) and visual results (Figure 6, Figure 7, and Figure 8) here. The SOTA methods used for comparisons with our method (QISTA-ImageNet) include ReconNet [22], [1], MS-CSNet [30], DR²-Net [19], $\{0, 1\}$ -BCSNet [32]¹, $\{-1, +1\}$ -BCSNet [32]², CSNet⁺ [32]³, SCSNet [31], AMP-Net-9-BM [43], and OPINE-Net⁺ [41].

Table 4: Average PSNR (dB) and SSIM comparisons of different methods with various measurement rates (MRs) on Set5.

Set5	PSNR	SSIM								
MR	40%		20%		10%		5%		1%	
[22]	-	-	-	-	25.98	0.734	-	-	-	-
[1]	-	-	34.55	0.939	31.31	0.894	-	-	-	-
[30]	-	-	36.26	0.950	32.82	0.909	-	-	-	-
[19]	-	-	-	-	27.79	0.798	-	-	-	-
[32] ¹	38.24	0.967	32.31	0.898	29.99	0.851	28.57	0.816	23.79	0.636
[32] ²	38.62	0.964	35.24	0.939	32.20	0.898	29.39	0.840	24.07	0.645
[32] ³	40.11	0.974	36.05	0.948	32.59	0.906	29.74	0.849	24.18	0.648
[31]	40.44	0.976	36.15	0.949	32.77	0.908	29.74	0.847	24.21	0.647
[43]	40.95	0.975	36.88	0.950	33.42	0.914	29.82	0.853	23.48	0.652
[41]	39.56	0.972	35.57	0.947	33.77	0.924	28.89	0.854	22.95	0.621
Ours	41.63	0.977	37.25	0.953	33.85	0.919	30.38	0.863	23.64	0.650

6.5 Ablation Study on Dictionary in Sec. 3.4

In this paper, the dictionary Ψ and its “left-inverse” $\tilde{\Psi}$ are adopted from Eqs. (22) and (23), respectively. Here we conducted a comparison with the dictionary

Table 5: Average PSNR (dB) and SSIM comparisons of different methods with various measurement rates (MRs) on Set14.

Set14	PSNR	SSIM								
MR	40%		20%		10%		5%		1%	
[22]	-	-	-	-	24.18	0.640	-	-	-	-
[1]	-	-	31.21	0.885	28.54	0.814	-	-	-	-
[30]	-	-	32.26	0.896	29.29	0.820	-	-	-	-
[19]	-	-	-	-	24.38	0.706	-	-	-	-
[32] ¹	34.52	0.938	29.25	0.816	27.36	0.756	26.09	0.694	22.48	0.553
[32] ²	34.81	0.934	31.55	0.880	28.78	0.805	26.67	0.724	22.74	0.562
[32] ³	36.16	0.950	32.15	0.894	29.13	0.817	26.93	0.733	22.83	0.563
[31]	36.54	0.953	32.19	0.895	29.22	0.818	26.92	0.732	22.87	0.563
[43]	37.44	0.956	33.17	0.902	29.92	0.831	27.25	0.744	22.79	0.575
[41]	36.08	0.952	31.74	0.899	29.98	0.841	26.13	0.740	22.48	0.555
Ours	38.10	0.959	33.54	0.906	30.26	0.835	27.75	0.753	22.83	0.571

Table 6: Average PSNR (dB) and SSIM comparisons of different methods with various measurement rates (MRs) on BSD100.

BSD100	PSNR	SSIM								
MR	40%		20%		10%		5%		1%	
[30]	-	-	31.15	0.874	28.61	0.786	-	-	-	-
[32] ¹	33.41	0.928	28.65	0.785	27.05	0.722	26.04	0.658	23.49	0.541
[32] ²	33.67	0.925	30.50	0.855	28.21	0.770	26.55	0.689	23.70	0.547
[32] ³	34.91	0.944	31.05	0.872	28.53	0.783	26.78	0.698	23.76	0.548
[31]	35.21	0.947	31.10	0.873	28.57	0.784	26.77	0.697	23.78	0.548
[43]	35.76	0.949	31.58	0.878	28.87	0.792	26.78	0.702	23.42	0.553
[41]	34.03	0.943	30.09	0.873	28.83	0.803	25.47	0.695	23.03	0.529
Ours	36.07	0.951	31.84	0.880	29.07	0.794	27.09	0.708	23.59	0.551

design in ISTA-Net [40], which is expressed as:

$$\Psi' = \mathcal{C}'_1 \circ \text{ReLU} \circ \mathcal{C}'_0 \quad (35)$$

and

$$\tilde{\Psi}' = \mathcal{C}'_3 \circ \text{ReLU} \circ \mathcal{C}'_2 \quad (36)$$

where all \mathcal{C}'_i s, $i = 0, 1, 2, 3$ are convolutional operators. Table 7 shows that the pair $(\Psi$ and $\tilde{\Psi})$ obtains better recovery results than the pair $(\Psi'$ and $\tilde{\Psi}')$ in all testing datasets described in Sec. 4.3 under a range of measurement rates.

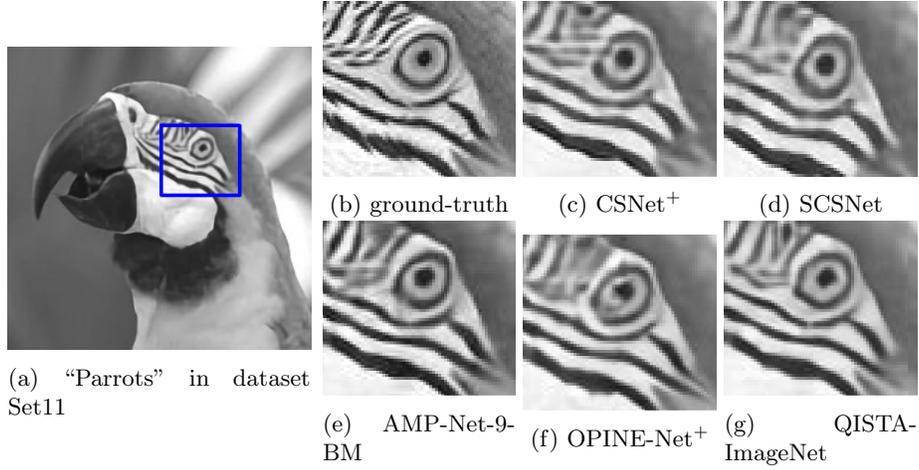


Fig. 6: Reconstruction result of (c) CSNet⁺, (d) SCSNet, (e) AMP-Net-9-BM, (f) OPINE-Net⁺, and (g) QISTA-ImageNet with 10% measurement rate.

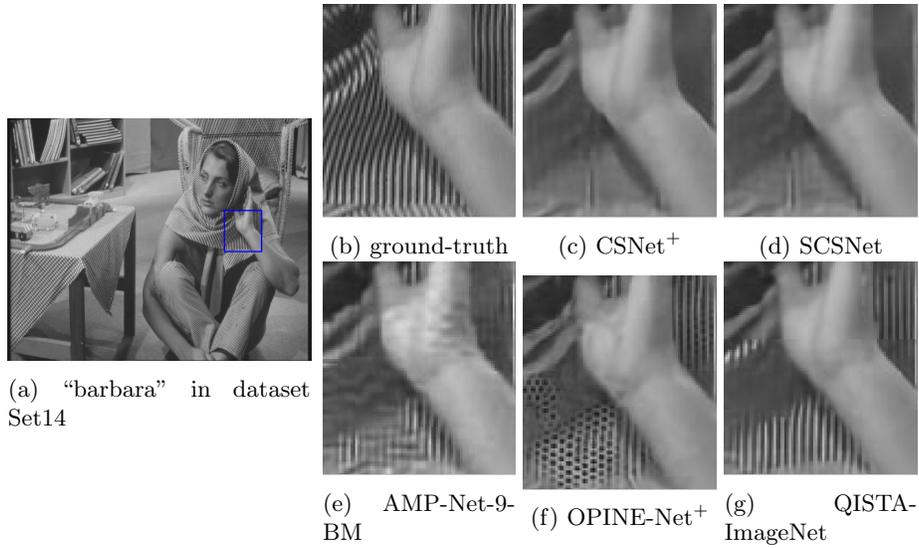


Fig. 7: Reconstruction result of (c) CSNet⁺, (d) SCSNet, (e) AMP-Net-9-BM, (f) OPINE-Net⁺, and (g) QISTA-ImageNet with 10% measurement rate.

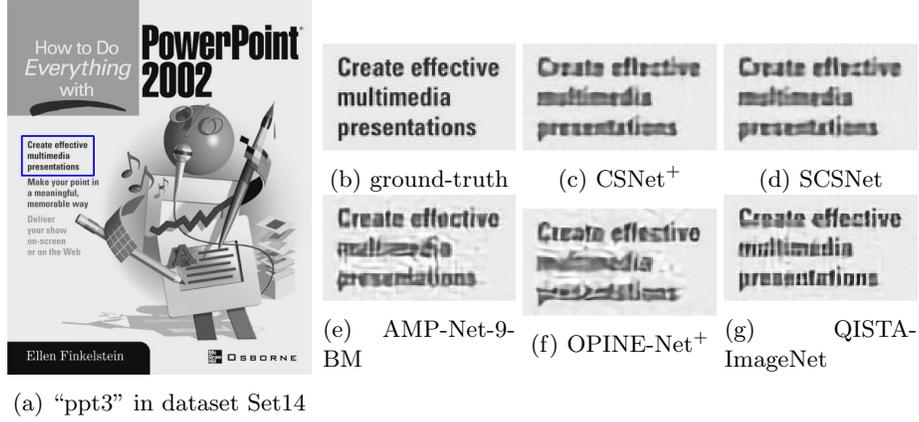


Fig. 8: Reconstruction result of (c) CSNet⁺, (d) SCSNet, (e) AMP-Net-9-BM, (f) OPINE-Net⁺, and (g) QISTA-ImageNet with 10% measurement rate.

Table 7: Ablation study on the dictionary design in terms of recovery performance. MR represents the measurement rate.

Dictionary	MR	Dataset		Set11		BSD68		Set14		BSD100	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM		
Ψ & $\tilde{\Psi}$ Ψ' & $\tilde{\Psi}'$	1%	21.30	0.5717	22.39	0.5347	22.83	0.5712	23.59	0.5506		
		21.19	0.5646	22.36	0.5312	22.75	0.5657	23.57	0.5472		
Ψ & $\tilde{\Psi}$ Ψ' & $\tilde{\Psi}'$	4%	26.07	0.7869	25.43	0.6773	27.01	0.7254	26.52	0.6816		
		25.74	0.7757	25.16	0.6713	26.64	0.7167	26.26	0.6763		
Ψ & $\tilde{\Psi}$ Ψ' & $\tilde{\Psi}'$	10%	30.01	0.8853	28.06	0.7949	30.26	0.8347	29.07	0.7942		
		29.62	0.8799	27.86	0.7896	29.98	0.8294	28.88	0.7892		
Ψ & $\tilde{\Psi}$ Ψ' & $\tilde{\Psi}'$	25%	35.41	0.9529	32.03	0.9067	34.82	0.9257	32.97	0.9050		
		34.97	0.9498	31.78	0.9033	34.49	0.9223	32.72	0.9014		
Ψ & $\tilde{\Psi}$ Ψ' & $\tilde{\Psi}'$	30%	36.64	0.9618	33.08	0.9255	35.94	0.9397	34.02	0.9240		
		36.17	0.9593	32.83	0.9224	35.56	0.9361	33.76	0.9209		
Ψ & $\tilde{\Psi}$ Ψ' & $\tilde{\Psi}'$	40%	38.84	0.9734	35.15	0.9520	38.10	0.9585	36.07	0.9508		
		38.53	0.9722	34.92	0.9503	37.82	0.9569	35.83	0.9490		
Ψ & $\tilde{\Psi}$ Ψ' & $\tilde{\Psi}'$	50%	40.87	0.9820	37.19	0.9690	40.09	0.9709	38.09	0.9684		
		40.59	0.9806	36.94	0.9678	39.73	0.9692	37.84	0.9668		