LISTA-NET: LEARNABLE INTERPRETABLE OPTIMIZATION-INSPIRED DEEP NETWORK FOR IMAGE COMPRESSIVE SENSING

Komal Krishna Mogilipalepu

Indian Institute of Science, India.

ABSTRACT

This work is based on the paper ISTA-Net [7], which focuses on reconstructing natural images by combining traditional optimization-based methods and recent network-based ones. Specifically, [7] developed a novel structured deep network, dubbed ISTA-Net, inspired by Iterative Shrinkage-Thresholding Algorithm(ISTA) for optimizing a general ℓ_1 norm CS reconstruction model. Moreover, considering that the residuals of natural images are more compressible, an enhanced version of ISTA-Net in the residual domain, dubbed ISTA-Net$^+$, is derived and compared with the state-of-the-art methods. This report discusses the current work and produces interesting results by considering additional learnable parameters $\mathbf{\Phi}$ in the ISTA-Net model.

***Index Terms*—** Compressive sensing,Deep Learning.

1. INTRODUCTION

Mathematically, the purpose of CS reconstruction is to infer the original signal $\mathbf{x} \in R^N$ from its randomized CS measurements $\mathbf{y} = \mathbf{\Phi x} \in R^M$. Here, $\mathbf{\Phi} \in R^{M\times N}$ is a linear random projection (matrix). Given the linear measurements $\mathbf{y}$, traditional image CS methods usually reconstruct the original image $\mathbf{x}$ by solving the following (generally convex) optimization problem:

$$\min_{\mathbf{x}} \ \frac{1}{2}\|\mathbf{\Phi x} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{\Psi x}\|_1, \tag{1}$$

where $\mathbf{\Psi x}$ denotes the transform coefficients of $\mathbf{x}$ with respect to some transform $\mathbf{\Psi}$ and the sparsity of the vector $\mathbf{\Psi x}$ is encouraged by the ℓ_1 norm with , which stands for adding all the absolute values of the entries in a vector. λ being the (generally pre-defined) regularization parameter. Since natural images are typically non-stationary, the classic fixed domains (DCT, wavelet , and gradient domain) usually result in poor reconstruction performance.So, In the proposed ISTA-Net method we learn the transform matrix by considering as an operator using CNN's. ISTA-Net borrows insights from traditional optimization methods to allow for interpretability in its network design and it utilizes the structural diversity originating from the CS domain. Extensive experiments demonstrate that ISTA-Net significantly outperforms the existing optimization-based and network-based CS methods, even when compared against methods that are designed for a specific domain.

2. TECHNICAL DETAILS

ISTA solves the CS reconstruction problem in Eq. (1) by iterating between the following update steps:

$$\mathbf{r}^{(k)} = \mathbf{x}^{(k-1)} - \rho\mathbf{\Phi}^\top(\mathbf{\Phi x}^{(k-1)} - \mathbf{y}), \tag{2}$$

$$\mathbf{x}^{(k)} = \arg\min_{\mathbf{x}} \ \frac{1}{2}\|\mathbf{x} - \mathbf{r}^{(k)}\|_2^2 + \lambda\|\mathbf{\Psi x}\|_1. \tag{3}$$

Instead of the hand-crafted transform $\mathbf{\Psi}$ in Eq. (1), ISTA-Net adopts a general nonlinear transform function to sparsify natural images, denoted by $\mathcal{F}(\cdot)$, whose parameters are learnable.

$$\min_{\mathbf{x}} \ \frac{1}{2}\|\mathbf{\Phi x} - \mathbf{y}\|_2^2 + \lambda\|\mathcal{F}(\mathbf{x})\|_1. \tag{4}$$

By solving Eq. (4) using ISTA, Eq. (2) is unchanged while Eq. (3) becomes

$$\mathbf{x}^{(k)} = \arg\min_{\mathbf{x}} \ \frac{1}{2}\|\mathbf{x} - \mathbf{r}^{(k)}\|_2^2 + \lambda\|\mathcal{F}(\mathbf{x})\|_1. \tag{5}$$

Theorem 1 *Let $X_1, ..., X_n$ be independent normal random variables with common zero mean and variance σ^2. If $\vec{X} = [X_1, ..., X_n]^\top$ and given any matrices $\mathbf{A} \in R^{m\times n}$ and $\mathbf{B} \in R^{s\times m}$, define a new random variable $\vec{Y} = \mathbf{B}ReLU(\mathbf{A}\vec{X}) = \mathbf{B}\max(\mathbf{0}, \mathbf{A}\vec{X})$. Then, $E[\|\vec{Y} - E[\vec{Y}]\|_2^2]$ and $E[\|\vec{X} - E[\vec{X}]\|_2^2]$ are linearly related, $E[\|\vec{Y} - E[\vec{Y}]\|_2^2] = \alpha E[\|\vec{X} - E[\vec{X}]\|_2^2]$, where α is only a function of $\mathbf{A}$ and $\mathbf{B}$.*

Theorem 1 can be easily extended to a normal distribution. Suppose that $\mathbf{r}^{(k)}$ and $\mathcal{F}(\mathbf{r}^{(k)})$ are the mean values of $\mathbf{x}$ and $\mathcal{F}(\mathbf{x})$ respectively, then we can make the following approximation based on **Theorem** 1:

$$\|\mathcal{F}(\mathbf{x}) - \mathcal{F}(\mathbf{r}^{(k)})\|_2^2 \approx \alpha\|\mathbf{x} - \mathbf{r}^{(k)}\|_2^2, \tag{6}$$

where α is a scalar that is only related to the parameters of $\mathcal{F}(\cdot)$. By incorporating this linear relationship into Eq. (5), we obtain the following optimization:

$$\mathbf{x}^{(k)} = \arg\min_{\mathbf{x}} \ \frac{1}{2}\|\mathcal{F}(\mathbf{x}) - \mathcal{F}(\mathbf{r}^{(k)})\|_2^2 + \theta\|\mathcal{F}(\mathbf{x})\|_1, \tag{7}$$

where λ and α are merged into one parameter θ, $\theta = \lambda\alpha$. Therefore, we get a closed-form version of $\mathcal{F}(\mathbf{x}^{(k)})$:

$$\mathcal{F}(\mathbf{x}^{(k)}) = soft(\mathcal{F}(\mathbf{r}^{(k)}), \theta). \tag{8}$$

By considering the left inverse of $\mathcal{F}(\cdot)$, denoted by $\widetilde{\mathcal{F}}(\cdot)$ such that $\widetilde{\mathcal{F}} \circ \mathcal{F} = \mathcal{I}$, Because $\mathcal{F}(\cdot)$ and $\widetilde{\mathcal{F}}(\cdot)$ are both learnable, we will enforce the ***symmetry constraint*** $\widetilde{\mathcal{F}} \circ \mathcal{F} = \mathcal{I}$ by incorporating it into the loss function during network training. Therefore, $\mathbf{x}^{(k)}$ can be efficiently computed in closed-form as:

$$\mathbf{x}^{(k)} = \widetilde{\mathcal{F}}^{(k)}(soft(\mathcal{F}^{(k)}(\mathbf{r}^{(k)}), \theta^{(k)})). \tag{9}$$

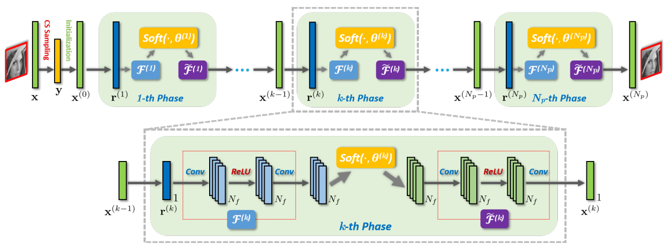

Fig. 1. Network Structure

Parameters in ISTA-Net: The learnable parameter set in ISTA-Net, denoted by $\mathbf{\Theta}$, includes the step size $\rho^{(k)}$ in the $\mathbf{r}^{(k)}$ module, the parameters of the forward and backward transforms $\mathcal{F}^{(k)}(\cdot)$ and $\widetilde{\mathcal{F}}^{(k)}(\cdot)$, and the shrinkage threshold $\theta^{(k)}$ in the $\mathbf{x}^{(k)}$ module. As such, $\mathbf{\Theta} = \{\rho^{(k)}, \theta^{(k)}, \mathcal{F}^{(k)}, \widetilde{\mathcal{F}}^{(k)}\}_{k=1}^{N_p}$, where N_p is the total number of ISTA-Net phases. All these parameters will be learned as neural network parameters.

2.1. Loss Function Design

Given the training data pairs $\{(\mathbf{y}_i, \mathbf{x}_i)\}_{i=1}^{N_b}$, ISTA-Net first takes the CS measurement $\mathbf{y}_i$ as input and generates the reconstruction result, denoted by $\mathbf{x}_i^{(N_p)}$, as output. We seek to reduce the discrepancy between $\mathbf{x}_i$ and $\mathbf{x}_i^{(N_p)}$ while satisfying the symmetry constraint $\widetilde{\mathcal{F}}^{(k)} \circ \mathcal{F}^{(k)} = \mathcal{I} \ \ \forall k = 1, \ldots, N_p$. Therefore, we design the end-to-end loss function for ISTA-Net as follows:

$$\mathcal{L}_{total}(\mathbf{\Theta}) = \mathcal{L}_{discrepancy} + \gamma\mathcal{L}_{constraint}, \tag{10}$$

$$\text{with: } \begin{cases} \mathcal{L}_{discrepancy} = \frac{1}{N_b N}\sum_{i=1}^{N_b} \|\mathbf{x}_i^{(N_p)} - \mathbf{x}_i\|_2^2 \\ \mathcal{L}_{constraint} = \frac{1}{N_b N}\sum_{i=1}^{N_b}\sum_{k=1}^{N_p} \|\widetilde{\mathcal{F}}^{(k)}(\mathcal{F}^{(k)}(\mathbf{x}_i)) - \mathbf{x}_i\|_2^2, \end{cases}$$

where N_p, N_b, N, and γ are the total number of ISTA-Net phases, the total number of training blocks, the size of each block $\mathbf{x}_i$, and the regularization parameter, respectively. In our experiments, γ is set to 0.01.

3. ENHANCED VERSION: ISTA-NET$^+$

Motivated by the fact that the residuals of natural images and videos are more compressible , an enhanced version, dubbed ISTA-Net$^+$, is derived from ISTA-Net to further improve CS performance. Starting from Eq. (5), we assume that $\mathbf{x}^{(k)} = \mathbf{r}^{(k)} + \mathbf{w}^{(k)} + \mathbf{e}^{(k)}$, where $\mathbf{e}^{(k)}$ stands for some noise and $\mathbf{w}^{(k)}$ represents some missing high-frequency component in $\mathbf{r}^{(k)}$, which can be extracted by a linear operator $\mathcal{R}(\cdot)$ from $\mathbf{x}^{(k)}$, $\mathbf{w}^{(k)} = \mathcal{R}(\mathbf{x}^{(k)})$. Furthermore, $\mathcal{R}(\cdot)$ is defined as $\mathcal{R} = \mathcal{G} \circ \mathcal{D}$, where $\mathcal{D}$ corresponds to N_f filters (each of size 3×3 in our experiments) and $\mathcal{G}$ corresponds to 1 filter (with size $3 \times 3 \times N_f$). By modeling $\mathcal{F} = \mathcal{H} \circ \mathcal{D}$, where $\mathcal{H}$ consists of two linear convolutional operators and one ReLU, we can replace $\mathcal{F}$ in Eq. (7) with $\mathcal{H} \circ \mathcal{D}$ to obtain:

$$\min_{\mathbf{x}} \ \frac{1}{2}||\mathcal{H}(\mathcal{D}(\mathbf{x})) - \mathcal{H}(\mathcal{D}(\mathbf{r}^{(k)}))||_2^2 + \theta||\mathcal{H}(\mathcal{D}(\mathbf{x}))||_1. \tag{11}$$

By exploiting the approximation used in Eq. (7) and following the same strategy as in ISTA-Net, we define the left inverse of $\mathcal{H}$ as $\widetilde{\mathcal{H}}$, which has a structure symmetric to that of $\mathcal{H}$ and satisfies the symmetry constraint $\widetilde{\mathcal{H}} \circ \mathcal{H} = \mathcal{I}$. Thus, the closed form of the ISTA-Net$^+$ update for $\mathbf{x}^{(k)}$ is:

$$\mathbf{x}^{(k)} = \mathbf{r}^{(k)} + \mathcal{G}(\widetilde{\mathcal{H}}(soft(\mathcal{H}(\mathcal{D}(\mathbf{r}^{(k)})), \theta))). \tag{12}$$

Similar to ISTA-Net, each phase of ISTA-Net$^+$ also has its own learnable parameters. The learnable parameter set $\mathbf{\Theta}^+$ of ISTA-Net$^+$ is $\mathbf{\Theta}^+ = \{\rho^{(k)}, \theta^{(k)}, \mathcal{D}^{(k)}, \mathcal{G}^{(k)}, \mathcal{H}^{(k)}, \widetilde{\mathcal{H}}^{(k)}\}_{k=1}^{N_p}$. The loss function of ISTA-Net$^+$ is analogously designed by incorporating the constraints $\widetilde{\mathcal{H}}^{(k)} \circ \mathcal{H}^{(k)} = \mathcal{I}$ into Eq. (10).

- **Shared vs. Unshared:** Each phase of ISTA-Net$^+$ (N_f=32) has three types of parameters with their dimensionality listed in parentheses: step size $\rho^{(k)}$ (1), threshold $\theta^{(k)}$ (1), and transform $\mathcal{T}^{(k)} = \{\mathcal{D}^{(k)}, \mathcal{G}^{(k)}, \mathcal{H}^{(k)}, \widetilde{\mathcal{H}}^{(k)}\}$ ($32\times3\times3+32\times3\times3\times32\times2+32\times3\times3\times32\times2+1\times3\times3\times32 = 37440$). The flexibility of ISTA-Net$^+$ indicates that the same type of parameters in different phases do not need to be the same. To demonstrate the impact of this flexibility, we train several variants of ISTA-Net$^+$, where we vary the parameters that are shared among the phases. Obviously, the default *unshared* ISTA-Net$^+$ (most flexible with largest number of parameters) achieves the best performance, while the variant of ISTA-Net$^+$ that shares all parameters $(\rho^{(k)}, \theta^{(k)}, \mathcal{T}^{(k)})$ in all its phases (least flexible with smallest number of parameters) obtains the worst performance. When only $(\rho^{(k)}, \mathcal{T}^{(k)})$ or $(\theta^{(k)}, \mathcal{T}^{(k)})$ are shared, these ISTA-Net$^+$ variants register 0.75dB and 0.55dB gains over he variant with all shared parameters. Interestingly, the ISTA-Net$^+$ variant with only shared transforms $\mathcal{T}^{(k)}$ obtains very competitive PSNR results compared to the unshared variant. This indicates that further compression in ISTA-Net$^+$ parameters is possible, with limited affect on reconstruction performance.

4. RESULTS

We used 88912 images of size 33×33 for training and standard Set11, BSD68 dataset (which contains 11 and 68 gray images respectively) for testing. The measurement matrix $\mathbf{\Phi}$ is constructed by generating a random Gaussian matrix and then orthogonalizing its rows.

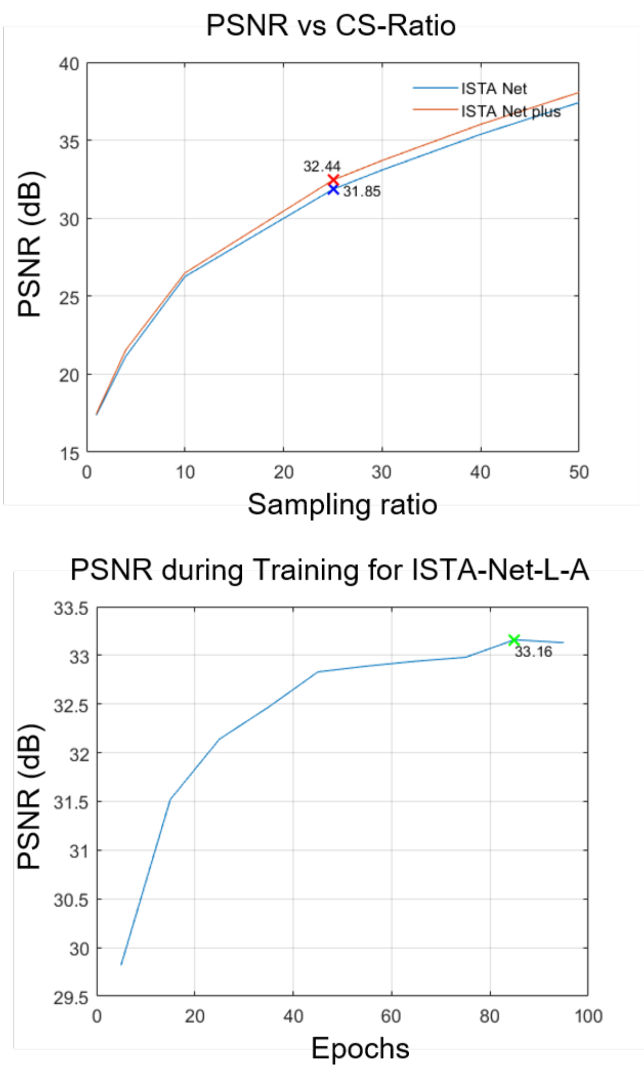

Fig. 2. Comparing PSNR curves of ISTA-NET and ISTA-NET^{+} with ISTA-Net learnable $\mathbf{\Phi}$ on Set11 dataset

5. CONCLUSION AND FUTURE WORK

We thoroughly understood the ISTA-Net paper and produced some of the results from their code. Motivated by the fact that the ISTA-Net model with unshared parameters performs well, we introduced $\mathbf{\Phi}$ as a learnable parameter in the model and compared the performance with the results shown in the

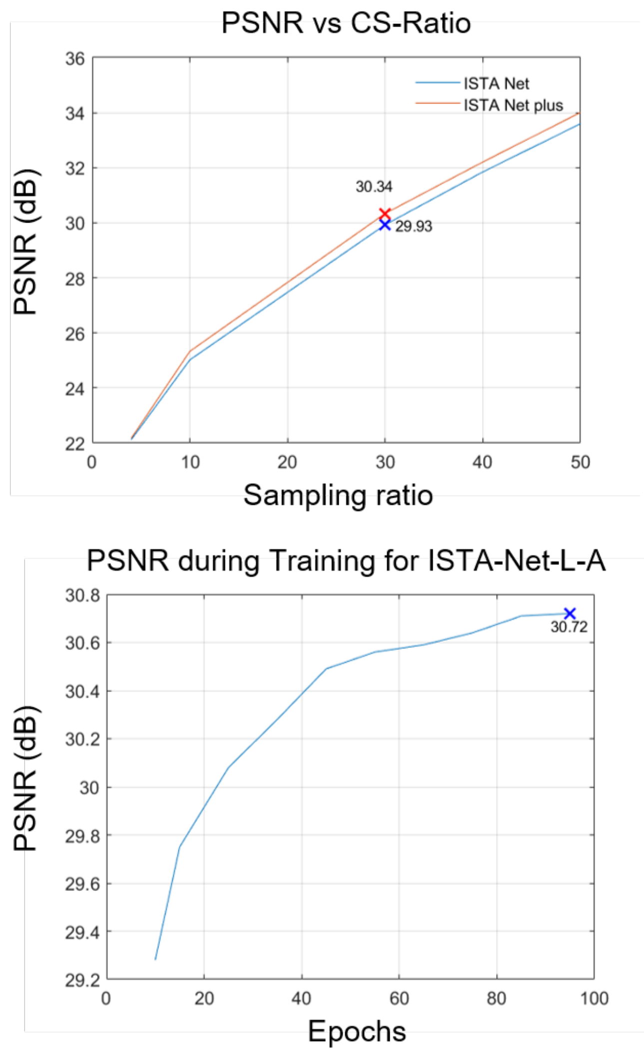

Fig. 3. Comparing PSNR curves of ISTA-NET and ISTA-NET^{+} with ISTA-Net learnable $\mathbf{\Phi}$ on BSD68 dataset

paper. We observed that ISTA-Net with learnable $\mathbf{\Phi}$ is performing better than the ISTA-Net^{+}. We have demonstrated the proposed method results on Set11 and BSD68 datasets. One can see the performance of ISTA-Net with learnable $\mathbf{\Phi}$ on other optimization methods, such as FISTA.

6. RESOURCES

1. https://github.com/jianzhangcs/ISTA-Net-PyTorch

7. REFERENCES

1. Zhang, Jian, and Bernard Ghanem. "ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.