IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP ’20)

ConFirmNet: Convolutional FirmNet and Application to Image Denoising and Inpainting

Praveen Kumar Pokala, Prakash Kumar Uttam, and Chandra Sekhar Seelamantula

Department of Electrical Engineering
Indian Institute of Science
Bangalore-12, India

May 6, 2020

Email: css@iisc.ac.in

SPECTRUM LAB

Outline

1. Motivation and Applications
2. Sparse Coding and *Learned* Sparse Coding
3. Convolutional Sparse Coding and *Learned* Convolutional Sparse Coding
4. *Minimax-Concave* Penalty (MCP)
5. MCP-Regularized Convolutional Sparse Coding (MC^2SC)
6. Learned MC^2SC
7. Application to Image Denoising and Inpainting.

Motivation and Applications

1. Standard convolutional sparse coding algorithms are not fast [Sreter et al., ICASSP, 2018].
2. Convex-relaxation of standard CSC [Garcia and Wohlberg, TCI, 2018].
3. Convex-relaxation for sparse coding results in biased estimates [Selesnick, TSP, 2017].
4. Objective: To develop an efficient non-convex CSC formulation.
5. Train an efficient sparse-coding network that results in high-quality reconstruction.
6. Image denoising and inpainting applications.

Sparse Coding

1. The actual sparse coding problem (NP-hard):

$$\min \|\mathbf{X}\|_0 \quad \text{s.t.} \quad \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_{\mathrm{F}}^2 \leq \epsilon.$$

2. Convex relaxation of the above problem:

$$\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_{\mathrm{F}}^2 + \lambda\|\mathbf{X}\|_1.$$

Pros:

1. Convex cost with global convergence guarantees.
2. Fast algorithms.

Cons:

1. Biased amplitude estimates [Selesnick, TSP, 2017].
2. Inefficient for large number of patches.
3. Loss of spatial continuity.
4. Lack of translation-invariance.

Sparse Coding: l1 minimization

- LASSO minimization problem [Candès and Plan, Annals Stat., 2009]:

$$\min_{\mathbf{X}} \tfrac{1}{2}\|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_{\mathrm{F}}^2 + \lambda\|\mathbf{X}\|_1,$$

  where λ is the regularization parameter.

- Iterative Soft-Thresholding Algorithm (ISTA) update [Daubechies et al., CPAM, 2004]:

$$\mathbf{X}^{k+1} = \Gamma_{\lambda\eta}\left((I - \eta\mathbf{D}^{\mathrm{T}}\mathbf{D})\mathbf{X}^k + \eta\mathbf{D}^{\mathrm{T}}\mathbf{Y}\right),$$

  where $\Gamma_{\lambda\eta}$ is the element-wise soft-thresholding operator; η is the step-size.

- Faster ISTA update [Beck and Teboulle, SIAM, 2009]:

$$\mathbf{Y}^{k+1} = t_{k+1}\mathbf{X}^{k+1} + \tfrac{t_k - 1}{t_{k+1}}(\mathbf{X}^{k+1} - \mathbf{X}^k),$$

  where $\mathbf{X}^{k+1}$ is the ISTA update and $t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}$.

LISTA: Learned ISTA

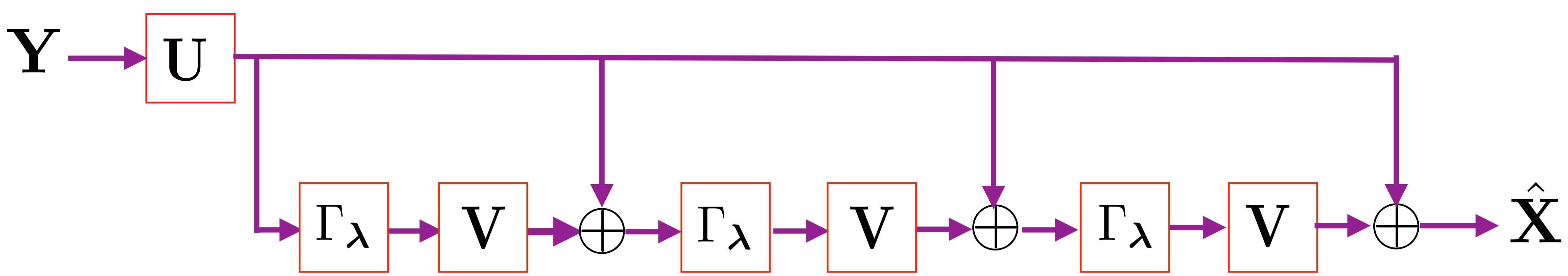
- Learned ISTA [Gregor and LeCun, ICML, 2010]: Proximal operator and affine transformations made learnable: $\{\mathbf{U}, \mathbf{V}, \boldsymbol{\lambda}\}$

$$\mathbf{X}^{k+1} = \Gamma_{\boldsymbol{\lambda}} \left(\mathbf{V}\mathbf{X}^{k} + \mathbf{U}\mathbf{Y} \right).$$

- Loss function:

$$\mathcal{L}(\mathbf{X}^{\star}, \hat{\mathbf{X}}; \mathbf{U}, \mathbf{V}, \boldsymbol{\lambda}) = \frac{1}{2}\|\mathbf{X}^{\star} - \hat{\mathbf{X}}\|_{\mathrm{F}}^{2}.$$

- Three-layered LISTA architecture is shown below:



LISTA: Deep-unfolding architecture

Limitations of LISTA

- Poor support recovery, biased amplitude estimates, and lack of interpretability between sparsity and shrinkage [Pokala et al., ICASSP., 2020].
- Patch-based method: Loss of spatial continuity, inefficient for large number of patches, and lack of translation-invariance [Sreter and Giryes, ICASSP, 2018].

Minimax-Concave Penalty Sparse Regularization

- Minimax-concave regularization for sparse recovery [Zhang, Annals Stat., 2010]:

$$\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_{\mathrm{F}}^2 + \lambda\|\mathbf{X}\|_{\mathrm{MC}},$$

  where λ is the regularization parameter.

- Iterative Firm-Thresholding Algorithm (IFTA) update [Pokala et al., ICASSP, 2019]:

$$\mathbf{X}^{k+1} = \mathcal{F}_{\lambda\eta}\left((I - \eta\mathbf{D}^{\mathrm{T}}\mathbf{D})\mathbf{X}^k + \eta\mathbf{D}^{\mathrm{T}}\mathbf{Y}; \gamma\right),$$

  where $\mathcal{F}_{\lambda\eta}$ is the element-wise firm-thresholding operator, η is the step-size and γ is the parameter of minimax-concave penalty.

- IFTA: nearly unbiased estimates, slow convergence, and lack of global convergence guarantees.

Learned Sparse Coding: FirmNet

- Deep-unfolding of IFTA ······▶ FirmNet
- FirmNet [Pokala et al., ICASSP, 2019] learnable parameters $\{\mathbf{U}, \mathbf{V}, \boldsymbol{\lambda}, \boldsymbol{\gamma}\}$:

$$\mathbf{X}^{k+1} = \mathcal{F}_{\boldsymbol{\lambda}}\left(\mathbf{V}\mathbf{X}^{k} + \mathbf{U}\mathbf{Y}; \boldsymbol{\gamma}\right).$$

- Loss function (δ is a small positive number) :

$$L(\mathbf{X}^{\star}, \hat{\mathbf{X}}) = \begin{cases} \frac{1}{2}\|\mathbf{X}^{\star} - \hat{\mathbf{X}}\|_{\mathrm{F}}^{2}, & \text{for } \|\mathbf{X}^{\star} - \hat{\mathbf{X}}\|_{\mathrm{F}} \leq \delta, \\ \delta\left(\|\mathbf{X}^{\star} - \hat{\mathbf{X}}\|_{\mathrm{F}} - \delta/2\right), & \text{Otherwise.} \end{cases}$$

- Overcomes limitations: Inaccurate support recovery and biased estimates for amplitudes.
- Limitation: Patch-based method

Solution: Use a convolutional structure.

Convolutional Sparse Coding

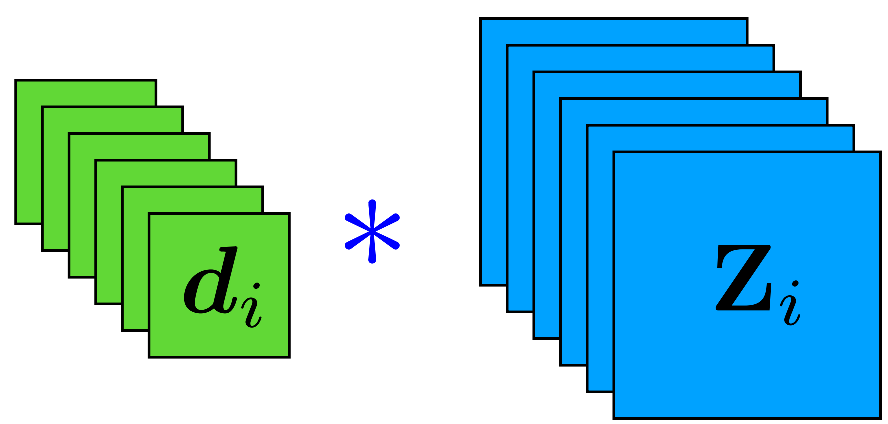
- Convolutional sparse coding (CSC) operates on the whole image.
- CSC overcomes the limitations of patch-based approaches.
- CSC employs a convolutional dictionary.

$$\min_{\mathbf{d},\mathbf{Z}=\{\mathbf{Z_i}\}_{i=1}^{m}} \|\mathbf{X} - \sum_{i=1}^{m} \boldsymbol{d}_i * \mathbf{Z}_i\|_{\mathrm{F}}^2 + \lambda \sum_{i=1}^{m} \|\mathbf{Z}_i\|_1$$

$$\text{s.t. } \|\boldsymbol{d}_i\|_2^2 \leq 1 \quad i = 1 \text{ to } m.$$

- Given $\mathbf{X}$, find kernels $(\boldsymbol{d}_i)$ and sparse feature maps $(\mathbf{Z}_i)$, $\forall i \in \{1, m\}$.

s.t. $\mathbf{X} = \sum_{i=1}^{m}$



Convolutional Extension of ISTA

$$\mathbf{Z}^{k+1} = \Gamma_{\lambda\eta}\left((I - \eta\mathbf{D}^{\mathrm{T}}\mathbf{D})\mathbf{Z}^{k} + \eta\mathbf{D}^{\mathrm{T}}\mathbf{Y}\right)$$

ISTA update

$\mathbf{D}$ is a concatenation of Toeplitz matrices.

$$\mathbf{Z}^{k+1} = \Gamma_{\lambda\eta}\left(\mathbf{Z}^{k} - \eta\hat{\boldsymbol{d}} * \boldsymbol{d} * \mathbf{Z}^{k} + \eta\hat{\boldsymbol{d}} * \mathbf{Y}\right)$$

$$\hat{\boldsymbol{d}} = \boldsymbol{d}^{\mathrm{T}}$$

Convolutional ISTA update

Convolutional LISTA

- Convolutional LISTA [Sreter and Giryes, ICASSP, 2018]: Replaces convolutional kernels and parameters of the nonlinearity with the learnable parameters: $\{\boldsymbol{w}_e, \boldsymbol{w}_d, \boldsymbol{\lambda}\}$.

$$\mathbf{Z}^{k+1} = \Gamma_{\lambda\eta}\left(\mathbf{Z}^k - \eta\hat{\boldsymbol{d}} * \boldsymbol{d} * \mathbf{Z}^k + \eta\hat{\boldsymbol{d}} * \mathbf{Y}\right)$$

$$\hat{\boldsymbol{d}} = \boldsymbol{d}^{\mathrm{T}}$$

Convolutional ISTA update

$$\mathbf{Z}^{k+1} = \Gamma_{\boldsymbol{\lambda}}\left(\mathbf{Z}^k - \boldsymbol{w}_e * \boldsymbol{w}_d * \mathbf{Z}^k + \boldsymbol{w}_e * \mathbf{Y}\right)$$

Convolutional LISTA: k^{th} layer

Contributions

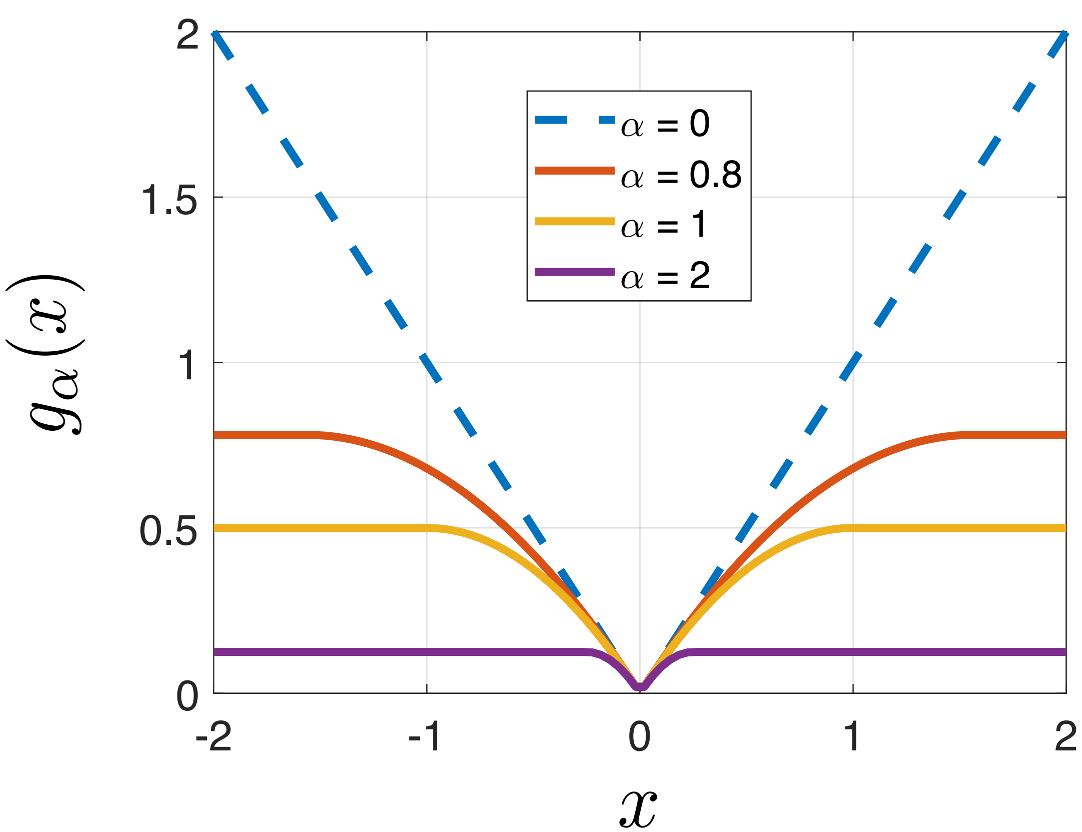
1. Minimax-concave regularized CSC (MC^2SC).
2. Convolutional IFTA (CIFTA) based on successive convex approximation (SCA).
3. Deep-unfolding of CIFTA $\dashrightarrow$ Convolutional FirmNet (ConFirmNet).
4. Auto-encoder based on Convolutional FirmNet.
5. Application to image denoising and inpainting.

Scalar Minimax-Concave Penalty

- Zhang [Annals Stat., 2010] introduced the scalar MCP, which is a non-convex relaxation to the ℓ_0 - pseudonorm:

$$g_\alpha(x) = \begin{cases} |x| - \frac{1}{2}\alpha^2 x^2, & |x| \le \frac{1}{\alpha^2}, \\ \frac{1}{2\alpha^2}, & |x| \ge \frac{1}{\alpha^2}. \end{cases}$$

- As $\alpha \to 0$, $g_\alpha(x) \to |x|$.



Scalar MCP

MCP Regularized CSC

- The minimax-concave (MC) regularized convolutional sparse coding (MC²SC):

$$\min_{\mathbf{d},\mathbf{Z}=\{\mathbf{Z_i}\}_{\mathbf{i=1}}^{\mathbf{m}}} \left\{ F(\mathbf{Z}) = \|\mathbf{X} - \mathbf{M} \odot \sum_{i=1}^{m} \boldsymbol{d}_i * \mathbf{Z}_i\|_{\mathrm{F}}^2 + \lambda \sum_{i=1}^{m} \|\mathbf{Z}_i\|_{MC} \right\}, \quad \text{s.t.} \quad \|\boldsymbol{d}_i\|_2^2 \leq 1,\ i = 1:m,$$

where $$\|\mathbf{Z}_i\|_{MC} = \sum_{w,h} \left(|\mathbf{Z}_i^{wh}| - \frac{|\mathbf{Z}_i^{wh}|^2}{2\lambda_i \gamma_i} \right) \mathbf{1}_{\{|\mathbf{Z}_i^{wh}| \leq \lambda_i \gamma_i\}} + \left(\frac{\lambda_i \gamma_i}{2} \right) \mathbf{1}_{\{|\mathbf{Z}_i^{wh}| \geq \lambda_i \gamma_i\}},$$

$\mathbf{1}$ denotes the indicator operator, $\mathbf{M} \in \{0, 1\}^{w \times h}$, $\odot$ element-wise multiplication, $\{\mathbf{X}, \mathbf{Z}_i\} \in \mathbb{R}^{w \times h}$, and $\mathbf{Z}_i^{wh} = [\mathbf{Z}_i]_{wh}$.

Successive Convex Approximation

- $$\mathcal{F}_{\lambda,\gamma}(\mathbf{U}) = \arg\min_{\mathbf{Z}} \underbrace{\frac{1}{2\eta}\|\mathbf{Z}-\mathbf{U}\|_{\mathrm{F}}^2 + \lambda\|\mathbf{Z}\|_{MC}}_{F(\mathbf{Z})},$$

  where $F(\mathbf{Z})$ is convex for $\gamma > 1$, unique minima exists, and it is given by,

$$\mathcal{F}_{\lambda,\gamma}(\mathbf{U})=\begin{cases}\mathbf{0}, & |\mathbf{U}| \leq \lambda\mathbf{1},\\ \dfrac{\gamma}{\gamma-1}\mathbf{1}\odot(|\mathbf{U}|-\lambda\mathbf{1})\odot\mathrm{sgn}(\mathbf{U}), & \lambda\mathbf{1} < |\mathbf{U}| \leq \gamma\lambda\mathbf{1},\\ \mathbf{U}, & |\mathbf{U}| > \gamma\lambda\mathbf{1},\end{cases}$$

where sgn is the element-wise signum operator, $|\cdot|$ is the element-wise absolute operator, and $\mathbf{1}$ is the matrix of all ones.

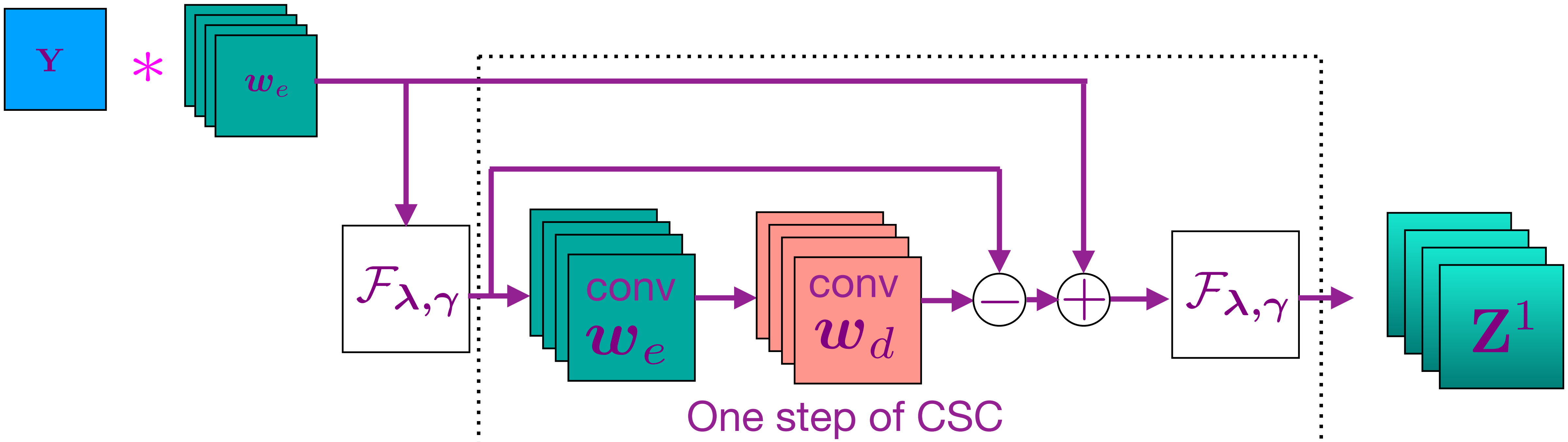
Convolutional IFTA

- MC²SC problem: solved via convex subproblems constructed by SCA.
- $\mathbf{Z}^{t+1} = \arg\min_{\mathbf{Z}} \frac{1}{2\eta} \|\mathbf{Z} - \{\mathbf{Z}^{\mathbf{t}} - \eta \nabla f(\mathbf{Z}^t)\}\|_{\mathrm{F}}^2 + \lambda \sum_{i=1}^{m} \|\mathbf{Z}_i\|_{MC}$,

  where $f(\mathbf{Z}) = \frac{1}{2}\|\mathbf{X} - \mathbf{M} \odot \sum_{i=1}^{m} \boldsymbol{d}_i * \mathbf{Z}_i\|_{\mathrm{F}}^2$ and $\mathbf{Z} = [\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_m]^{\mathrm{T}}$.
- SCA results in $\mathbf{Z}^{t+1} = \mathcal{F}_{\lambda,\gamma}(\mathbf{U})$, where $\mathbf{U} = \mathbf{Z}^{\mathbf{t}} - \eta \nabla f(\mathbf{Z}^t)$.
- SCA possesses local steepest-descent property.
- Convergence guarantees: CIFTA reaches a stationary point.

Convolutional FirmNet: Deep-Unfolding of CIFTA

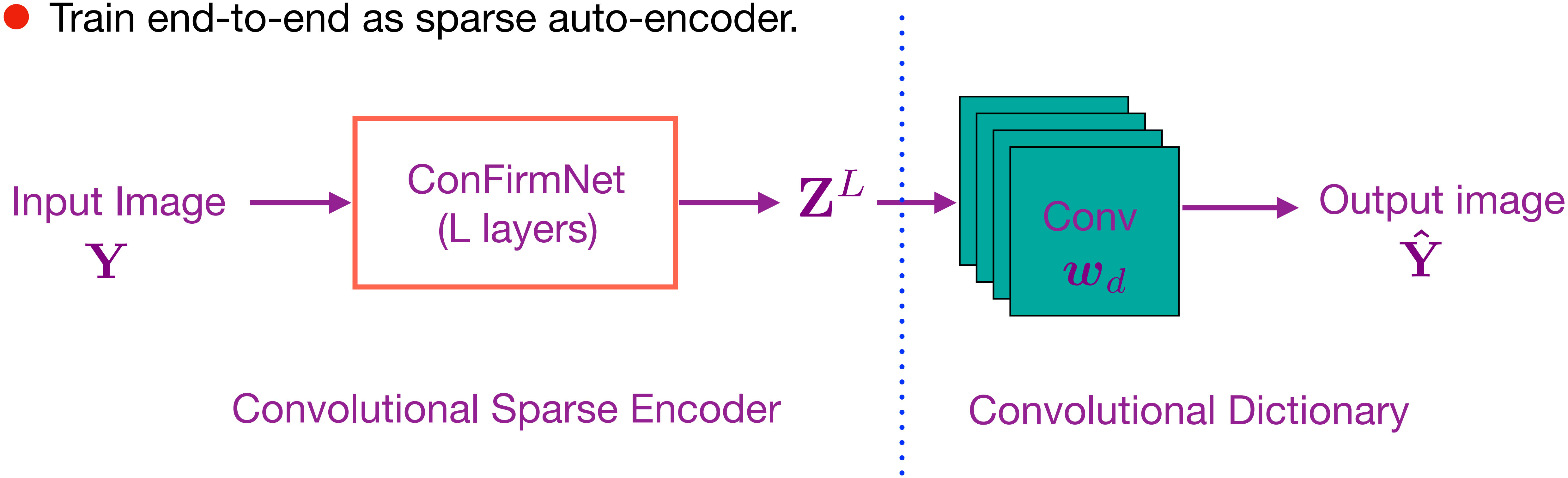
- Convolutional IFTA update: $\mathbf{Z}^{k+1} = \mathcal{F}_{\boldsymbol{\lambda},\gamma}\left(\mathbf{Z}^k - \boldsymbol{w}_e * \boldsymbol{w}_d * \mathbf{Z}^k + \boldsymbol{w}_e * \mathbf{Y}\right).$
- Convolutional FirmNet: Recurrent architecture for CSC with one layer.
- $\mathbf{Z}^1$: sparse feature maps at the output of the first layer.



ConFirmNet: Convolutional FirmNet with a Single Layer

Auto-Encoder Based on Convolutional FirmNet

- Train end-to-end as sparse auto-encoder.



- Auto-encoder structure based on ConFirmNet: Learn approximate sparse coding and dictionary.

Experimental Validation

- Training a sparse auto-encoder based on ConFirmNet:

  1. Filter weights are randomly initialized from a uniform distribution.
  2. Kernels of size: 3 x 3
  3. Dataset: PASCAL VOC [Evaringam et al., IJCV, 2010].
  4. Training data: Testing data = 80 : 20.
  5. Number of layers: 20
  6. $\{\boldsymbol{\lambda}, \gamma\}$ are initialized with vector of all entries equal to 0.5 and 5.0, respectively.
  7. Number of kernels: 256.
  8. Additive white Gaussian noise is considered for image denoising.

Experimental Validation

- Training cost:

$$L(\mathbf{Y}^\star, \hat{\mathbf{Y}}) = \begin{cases} \frac{1}{2}\|\mathbf{Y}^\star - \hat{\mathbf{Y}}\|_{\mathrm{F}}^2, & \text{for } \|\mathbf{Y}^\star - \hat{\mathbf{Y}}\|_{\mathrm{F}} \leq \delta, \\ \delta\left(\|\mathbf{Y}^\star - \hat{\mathbf{Y}}\|_{\mathrm{F}} - \delta/2\right), & \text{Otherwise} \end{cases}$$

  where $\delta > 0$. We set $\delta = 1$ in our experiments.

- $\mathrm{M} = \mathbf{1}\mathbf{1}^{\mathrm{T}}$ for the image denoising, where $\mathbf{1}$ denotes a 256-dimensional vector of all ones.
- $\mathrm{M} \in \{0,1\}^{256\times 256}$ is the binary mask drawn independently from a Bernoulli distribution with a probability 0.5 for image inpainting.

Image Denoising

- Model trained on noisy images (σ of 30).

Image	BM3D	ConLISTA	ConFirmNet
Peppers	29.30±0.06	29.64±0.08	**29.78±0.07**
Goldhill	29.09±0.03	29.06±0.02	**29.12±0.02**
Man	28.82±0.03	28.96±0.03	**29.01±0.02**
Lena	31.23±0.03	31.14±0.04	**31.28±0.04**
Cameraman	28.60±0.06	28.85±0.06	**29.12±0.06**
Barbara	**29.76±0.02**	27.84±0.01	28.24±0.02
House	**32.15±0.08**	31.98±0.06	32.08±0.08
Couple	28.84±0.03	28.91±0.02	**29.01±0.03**
Boats	29.05±0.03	29.02±0.03	**29.12±0.02**

- Testing scenario: Average and standard deviation of PSNR [dB] are evaluated based on 20 noise realizations.

Image Denoising: Robustness of ConFirmNet

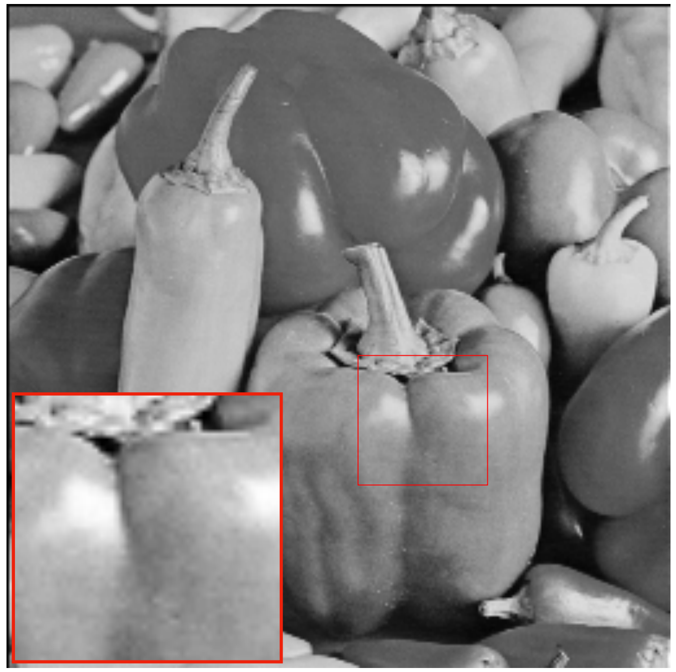
- The model is trained with a standard deviation of 20.
- Mean and standard deviation of PSNR [dB] computed based on 20 noise realizations.

Performance comparison: Training and testing noise mismatch

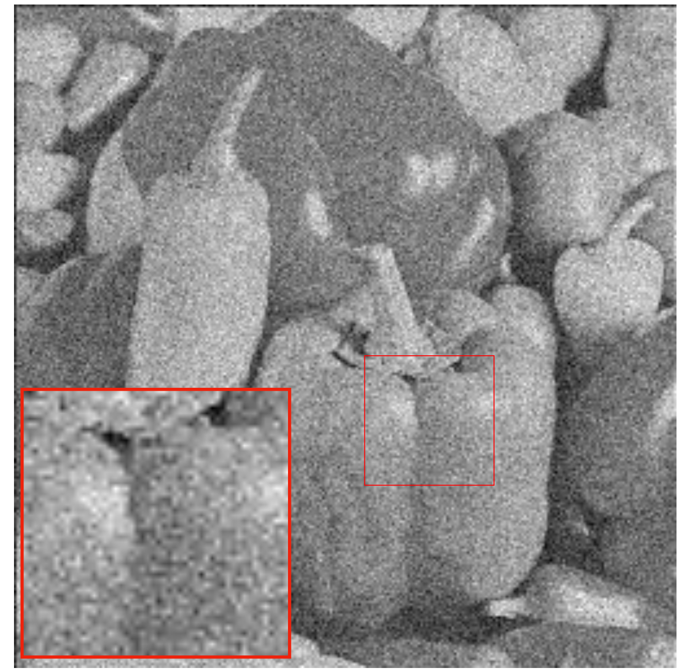
Image	Convolutional LISTA			ConFirmNet		
	$\sigma = 5$	$\sigma = 15$	$\sigma = 30$	$\sigma = 5$	$\sigma = 15$	$\sigma = 30$
Barbara	31.23±0.011	30.97±0.020	22.73±0.024	**31.85±0.011**	**31.46±0.020**	**22.96±0.028**
Goldhill	30.80±0.008	30.91±0.016	22.74±0.026	**30.83±0.007**	**30.98±0.016**	**23.03±0.027**
Man	31.01±0.008	31.14±0.026	22.76±0.028	**31.07±0.009**	**31.15±0.024**	**23.02±0.028**
Peppers	32.82±0.017	32.43±0.053	23.10±0.065	**33.03±0.023**	**32.55±0.054**	**23.22±0.061**
Lena	33.79±0.008	33.59±0.028	22.97±0.030	**34.06±0.008**	**33.74±0.029**	**23.29±0.030**
Cameraman	31.13±0.014	31.18±0.058	23.10±0.044	**31.28±0.015**	**31.28±0.056**	**23.35±0.048**
House	34.57±0.019	34.16±0.048	22.98±0.049	**34.92±0.019**	**34.39±0.048**	**23.30±0.053**
Couple	31.47±0.100	31.35±0.019	22.80±0.026	**31.63±0.009**	**31.44±0.020**	**23.00±0.024**
Boats	31.29±0.005	31.23±0.024	22.81±0.027	**31.48±0.007**	**31.37±0.021**	**23.06±0.030**

Visual Inspection

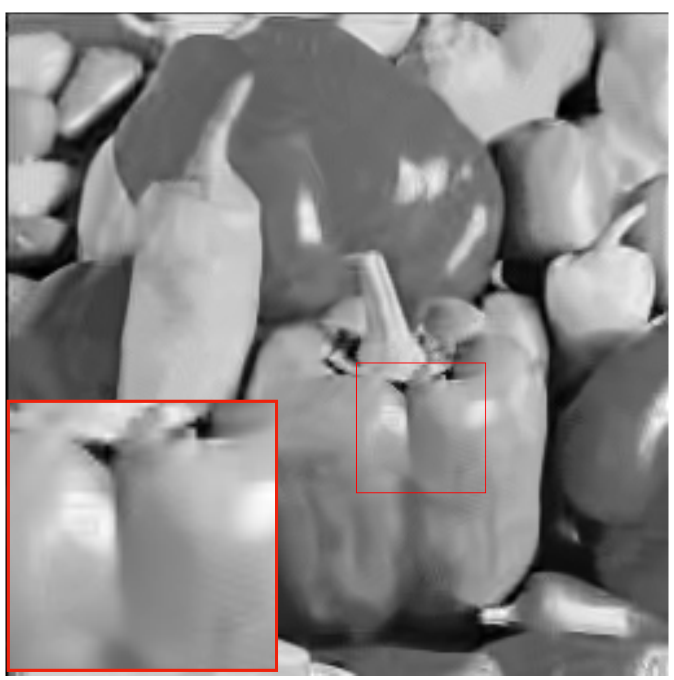
- Image denoising performance comparison for noise standard deviation of 30.



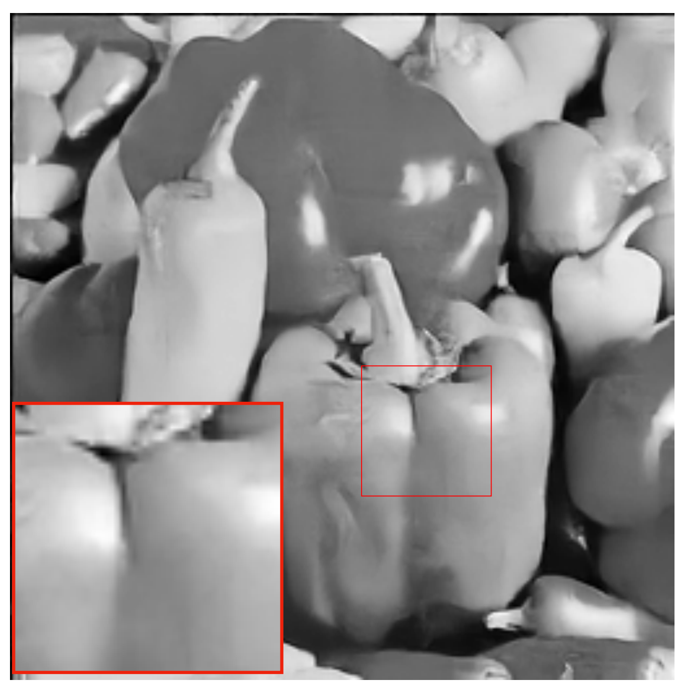
Original Image



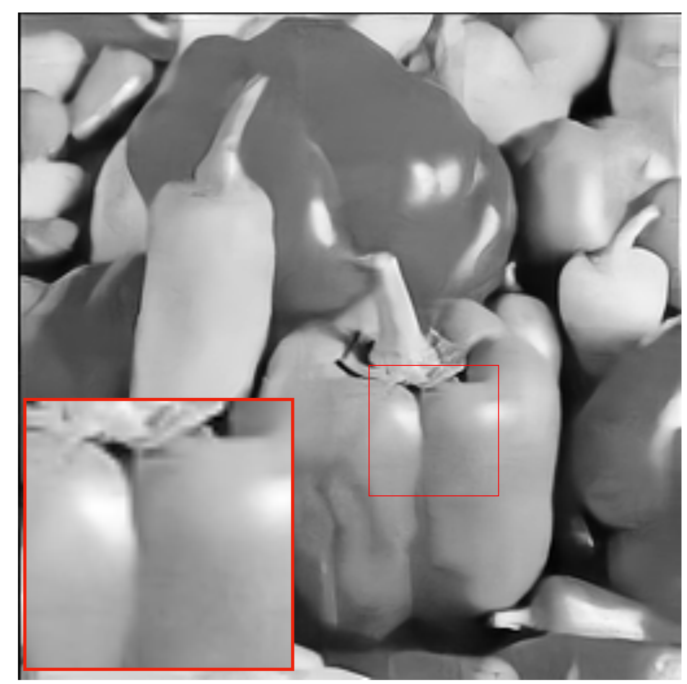
Noisy Image



BM3D, 29.24 dB



Convolutional LISTA, 29.47 dB



ConFirmNet, 29.73 dB

Image Inpainting

- Performance comparison on test images used in [Heide et al., CVPR, 2015] and [Sreter and Giryes, ICASSP, 2018].
- Binary mask probability = 0.5.

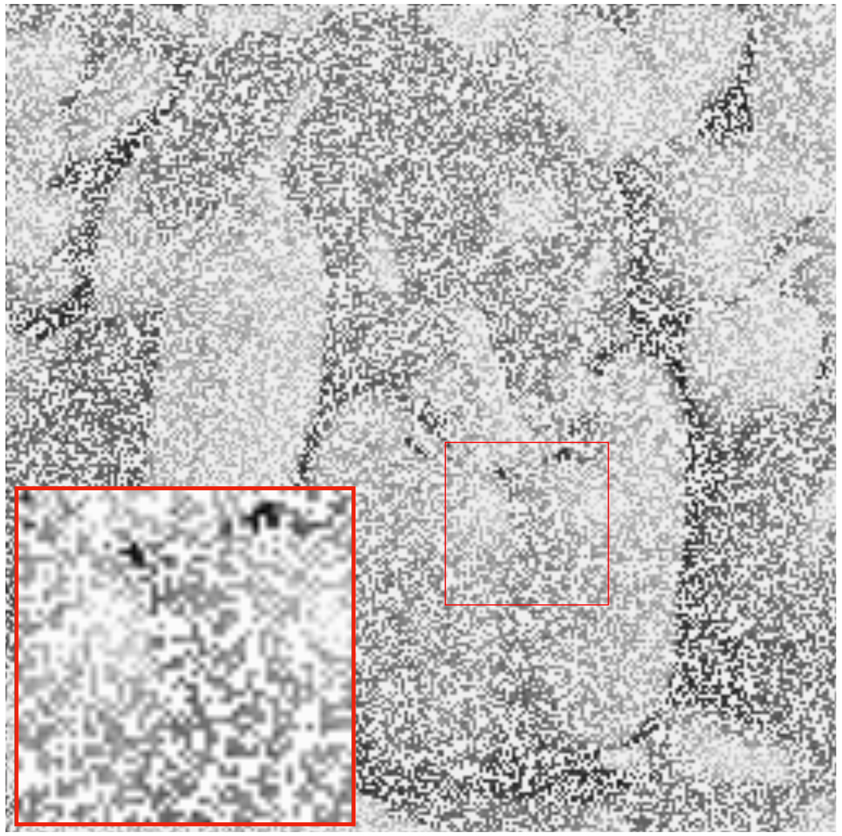
Performance comparison in terms of PSNR [dB]

Image Index*	FFCSC	Con LISTA	ConFirmNet
1	24.56±0.021	25.83±0.031	**25.86±0.029**
2	26.08±0.027	28.86±0.046	**28.89 ±0.048**
3	25.85±0.014	27.06±0.046	**27.09±0.045**
4	24.37±0.016	25.43±0.021	**25.50±0.021**
5	28.07±0.020	**29.71±0.047**	29.68±0.049
6	24.90±0.014	29.54±0.048	**29.58±0.046**
7	24.43±0.017	25.87±0.031	**26.02±0.027**
8	26.88±0.034	27.24±0.051	**27.28±0.057**
9	24.43±0.016	25.73±0.032	**25.79±0.027**
10	26.93±0.034	28.51±0.038	**28.63±0.039**

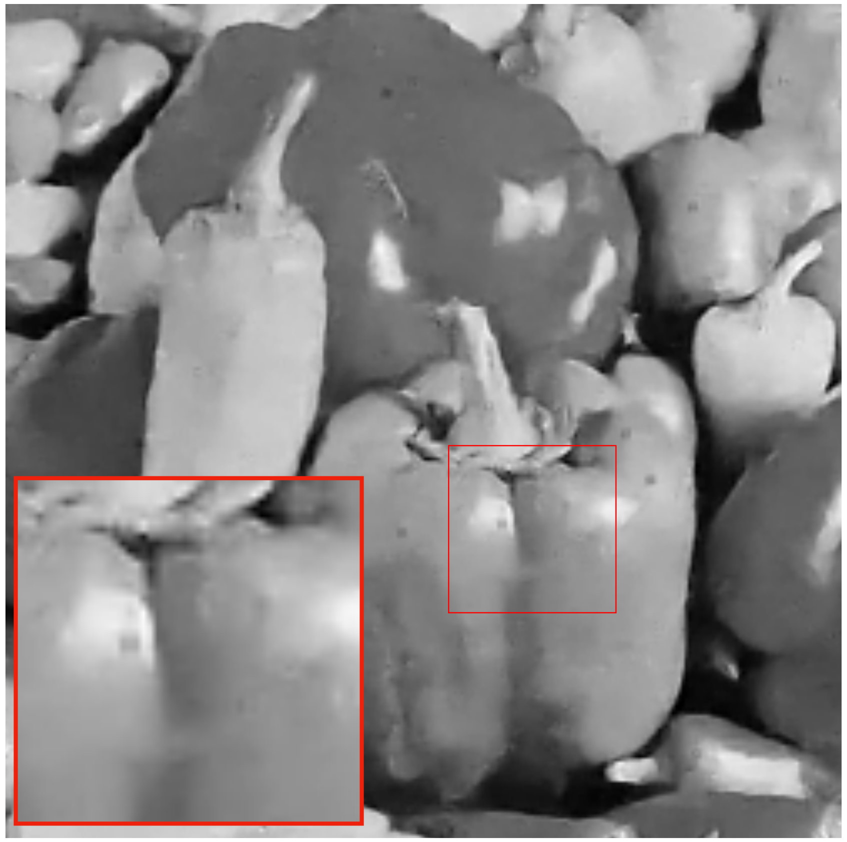
Test image index* is the same as that mentioned in [Sreter and Giryes, ICASSP, 2018].

Visual Inspection

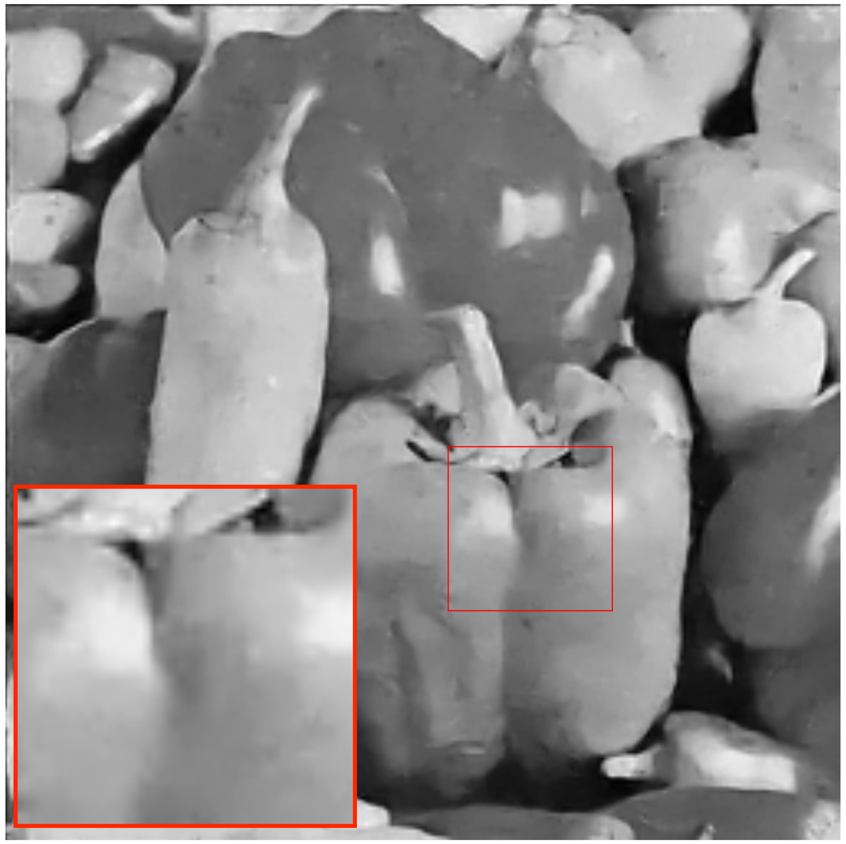
- Image inpainting; Mask probability = 0.5.



Corrupted Image



Convolutional LISTA
27.65 dB



ConFirmNet
27.87 dB

Conclusions

1. We proposed minimax-concave regularized convolutional sparse coding problem ($\mathrm{MC^2SC}$) and developed the convolutional iterative firm-thresholding algorithm (CIFTA) to solve it.

2. Deep-unfolding of CIFTA ----► Convolutional FirmNet (ConFirmNet).

3. ConFirmNet is superior to the other techniques with application to image denoising and inpainting. Further, it is robust to mismatch between the training and testing noise conditions.

4. Computational complexity comparable to that of convolutional LISTA.

Key References

[1] H. Sreter and R. Giryes, “Learned convolutional sparse coding,” *in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2191-2195, 2018.

[2] P. K. Pokala, A. G. Mahurkar, and C. S. Seelamantula, “FirmNet: A sparsity amplified deep network for solving linear inverse problems,” *in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2982-2986, 2019.

[3] K. Gregor and Y. LeCun, “Learning fast approximations of sparse coding,” *in Proceeding of the 27th International Conference on Machine Learning*, pp. 399-406, 2010.

[4] F. Heide, W. Heidrich, and G. Wetzstein, “Fast and flexible convolutional sparse coding,” *in Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 5135-5143, 2015.

[5] V. Papyan, Y. Romano, J. Sulam and M. Elad, “Convolutional dictionary learning via local processing,” *in Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 5296-5304, 2017.

[6] B. Wohlberg, “Efficient algorithms for convolutional sparse representations,” *IEEE Transaction on Image Processing*, vol. 25, no. 1, pp. 301-315, 2015.

[7] C-H. Zhang, “Nearly unbiased variable selection under minimax-concave penalty,” *The Annals of Statistics*, vol. 38, no. 2, pp. 894-942, 2010.

[8] M. Evaringam, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, “The Pascal visual object classes VOC challenge,” *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303-338, 210.

Acknowledgments

Funding Agencies:

- Ministry of Electronics and Information Technology, India
- Robert-Bosch Centre for Cyberphysical Systems

Thank You