COMPRESSIVE IMAGE RECOVERY USING RECURRENT GENERATIVE MODEL

*Akshat Dave, Anil Kumar Vadathya, Kaushik Mitra**

Department of Electrical Engineering
IIT Madras

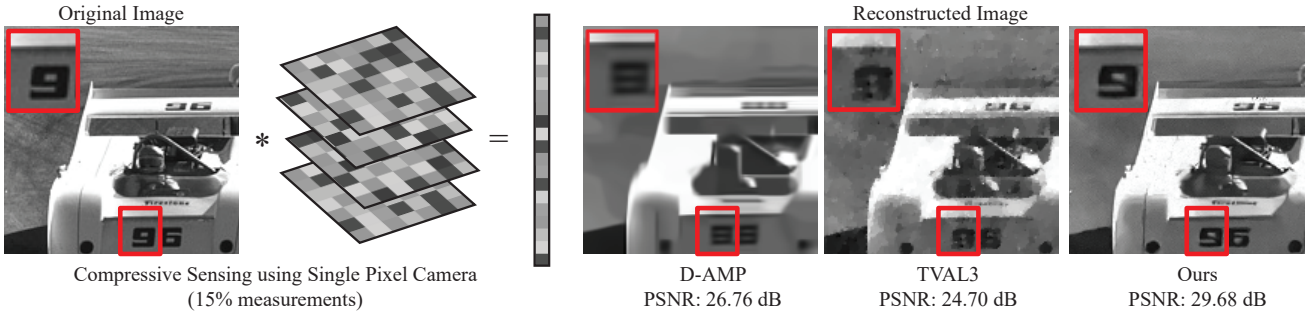


Fig. 1: We propose to use a deep generative model, RIDE [1] , as an image prior for compressive signal recovery. Since RIDE models long-range dependency in images using spatial LSTM, image recovery is better than other competing methods.

ABSTRACT

In this paper we leverage the recurrent generative model, RIDE, for applications like image inpainting and compressive image reconstruction. Recurrent networks can model long range dependencies in images and hence are suitable to handle global multiplexing in reconstruction from compressive imaging. We perform MAP inference with RIDE as prior using back-propagation to the inputs and projected gradient method. We propose a entropy thresholding based approach for preserving texture well. Our approach shows comparable results for image inpainting task. It shows superior results in compressive image reconstruction compared to traditional methods D-AMP and TVAL3 which uses global prior of minimizing TV norm.

Index Terms— Compressive imaging, recurent generative models, image inpainting,

1. INTRODUCTION

It is a well known fact that natural images are sparse in some transformed basis like DCT and Wavelet etc. Compressive sensing (CS) theory states that such sparse signals can be reconstructed from much lower compressed measurements than sampling at Nyquist rate []. This cuts down the cost of sensing, especially for imaging in the non-visible region of the spectrum where sensors are much costlier. Imaging in non-visible spectrum has applications in industrial inspection, surveillance, anti-counterfeiting and many more [2]. Given the potential advantages, CS is a promising solution for its practical viability. The single-pixel camera (SPC) is a classical example of CS framework [3]. In SPC, a single photo diode is used to capture compressive measurements and then reconstruct back the whole scene. It is extended for SWIR imaging []. SPC is commercially made available from InView Corporation.

In SPC, we take M linear measurements of the scene $\mathbf{x} \in \mathbb{R}^N$, with $M < N$ reconstruction of the signal $\mathbf{x}$ is ill-posed making data priors essential. Initially, priors based on empirical observations of natural image statistics were proposed for CS reconstruction. For example, priors assuming sparsity in the domain of wavelets [4], DCT coefficients and gradients [5]. Most widely used minimum total-variation (TV norm) prior is based on smooth local variations in natural images [6, 7]. However, using these priors at low measurement rates results in low quality reconstructions (see TVAL3 reconstruction in fig. 1). This is due to their inability to capture the complexity of natural image statistics. On the other hand, data driven approaches based on dictionary learning [8] and deep neural nets [9, 10] have also been proposed. Although they capture the complexity both the approaches handle only local multiplexing on image patches and are not suitable for global multiplexing scenario as in case of SPC. To address these problems, in this work we prospose to use a generative model, RIDE, proposed by Theis et al. [1] as the prior for CS image recovery. RIDE is both data driven and because of re-

*All the authors acknowledge the Qualcomm Innovation Fellowship 2016, India by Qualcomm India.

current nature handles the global multiplexing in SPC well. A line about LSTM and modeling the long term dependencies. Our contributions are as follows:

- We utilize RIDE's ability as an image prior to model long term dependencies for reconstructing compressively sensed images.
- We use backpropagation to inputs while doing gradient ascent for MAP inference.
- We formulate random image inpainting as a special case of compressive sensing recovery problem and using our MAP framework, show better results than multiscale KSVD approach.
- We hypothesize that the model's uncertainty in prediction can be related to the entropy of component posterior probabilities. By thresholding the entropy, we enhance texture preserving ability of the model.

2. RELATED WORK

Single Pixel Camera: Signal reconstruction from CS measurements is an ill-posed problem and hence we need to use signal priors. Initially algorithms were proposed to look for a sparse solution in l_1 norm. They assume sparsity in the domain of wavelet coefficients, DCT coefficients or gradients. Later class of algorithms known as approximate message passing (AMP) algorithms [11, 12] use off-the-shelf denoiser to iteratively refine their solution. ReconNet is another recent method using CNNs [9]. But as mentioned eariler, it can only handle local multiplexing since it is a patch based approach. Here we propose to do compressive image reconstruction with recurrent generative model, RIDE as the image prior. Since it is not patch limited, we can handle global multiplexing.

Deep Generative Models: Recently advances with deep neural nets have led to powerful deep generative models. These include Generative Adversarial Nets (GAN) [13], Variational Auto Encoders (VAE) [14], Pixel Recurrent Neural Networks (PixelRNN) [15] and Recurrent Image Density Estimator (RIDE) [1]. Among these contemporary models we find RIDE particularly suitable as low level image prior for our tasks involving Bayesian inference. GANs don't model the data distribution and VAE doesn't provide the exact likelihood measure. PixelRNN although models the distribution, it discretizes the distribution of a pixel resulting in optimization difficulties. RIDE models continuous distribution and gives exact likelihood thus facilitating gradient ascent for optimization. Also, RIDE being auto regressive isn't limited to patch size, as is the case with discriminative and even non deep generative models like dictionary learning. This is very useful particularly in cases like SPC where the reconstruction has to take account of global multiplexing and patch based methods can't be used directly.

3. BACKGROUND

Let $\mathbf{x}$ be a gray scale-image and x_{ij} be the pixel intensity at location ij then $\mathbf{x}_{<ij}$ describes the causal context around that pixel containing all x_{mn} such that $m \leq i$ and $j < n$. Now the joint distribution over the image can be factorized as,

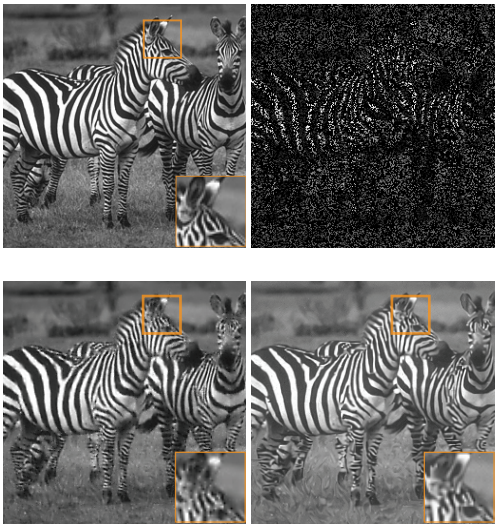
$$p(\mathbf{x}) = \prod_{ij} p(x_{ij}|\mathbf{x}_{<ij}), \tag{1}$$

which is a valid factorization involving conditional ditributions. Natural images in general exhibit long range correlations. In order to model such dependencies Theis et al. [1] have proposed to use two dimensional Spatial Long Short Term Memory (LSTMs) units [16]. Spatial LSTMs summarize the entire causal context $\mathbf{x}_{<ij}$ through their hidden representation $\mathbf{h}_{ij}$, as $\mathbf{h}_{ij} = f(\mathbf{x}_{<ij}, \mathbf{h}_{i-1,j}, \mathbf{h}_{i,j-1})$, where f is a complex non linear function. f has memory elements analogous to physical read, write and erase operations thus enabling LSTMs to model long term dependencies in sequences. Now each factor in the above joint distribution is modeled using conditional Gaussian Scale Mixtures, thus the complete distribution is given by,

$$p(\mathbf{x}) = \prod_{ij} p(x_{ij}|\mathbf{h}_{ij}, \boldsymbol{\theta}), \tag{2}$$

$$p(x_{ij}|\mathbf{h}_{ij}, \boldsymbol{\theta}) = \sum_{c,s} p(c, s|\mathbf{h}_{ij}, \boldsymbol{\theta}) p(x_{ij}|\mathbf{h}_{ij}, c, s, \boldsymbol{\theta}). \tag{3}$$

For more details we recommend the reader to go through [1].



21.21 dB, 0.811 22.07 dB, 0.813

Fig. 2: Inpainting comparisons: From left to right, original image, masked image, multiscale KSVD [17], ours. The numbers mentioned below the figures are PSNR(left) and SSIM(right).

4. COMPRESSIVE IMAGE RECOVERY USING RIDE

Here we consider the problem of image restoration, $\mathbf{x} \in \mathbb{R}^N$, from linearly compressed measurements, $\mathbf{y} \in \mathbb{R}^M$ as $\mathbf{y} = \Phi\mathbf{x}+\mathbf{n}$, where $\Phi \in \mathbb{R}^{M\times N}$ is the sensing matrix with $M < N$ and $\mathbf{n} \in \mathbb{R}^M$ is noise in the observation with known statistics.

4.1. MAP Inference via Backpropagation

Compressive image recovery: Here, we use Maximum-A-Posteriori principle to find the desired image as $\hat{\mathbf{x}} = arg\max_{\mathbf{x}} p(\mathbf{x})\, p(\mathbf{y}|\mathbf{x})$. For SPC, we formulated the MAP inference as,

$$\hat{\mathbf{x}} = arg\max_{\mathbf{x}} \log p(\mathbf{x}) \;\; s.t. \;\; \mathbf{y} = \Phi\mathbf{x}. \tag{4}$$

Here we do reconstruction for the noise less case. The log-likelihood and gradients are given by the model as in Eqns. 2, 3 and 7. To optimize the above we use projected gradients method, where after each gradient update solution is projected back on to the affine solution space for $\mathbf{y} = \Phi\mathbf{x}$. Every k-th iteration consists of the following two steps,

$$\hat{\mathbf{x}}_k = \mathbf{x}_{k-1} + \eta\nabla_{\mathbf{x}_{k-1}} \log p(\mathbf{x}), \tag{5}$$

$$\mathbf{x}_k = \hat{\mathbf{x}}_k - \Phi^T\left(\Phi\Phi^T\right)^{-1}(\Phi\hat{\mathbf{x}}_k - \mathbf{y}). \tag{6}$$

The gradient with respect to the log prior is given by,

$$\frac{\partial \log p(\mathbf{x})}{\partial x_{ij}} = \sum_{k\geq i, l\geq j} \frac{\partial \log p(x_{kl}|\mathbf{h}_{kl}, \boldsymbol{\theta})}{\partial x_{ij}}, \tag{7}$$

due to the recurrent nature of the model, each pixel through its hidden representation contributes to the likelihood of all the pixels that come after it in forward pass. Hence, during backward pass the gradient from each pixel propagates to all the pixels prior to it in the sequence.

Image inpainting: In image inpainting our goal is to recover the missing pixels from a randomly masked image. Here, we consider this as the special case of compressive imaging, where Φ is a binary matrix depending on the mask. Φ has a single 1 in every row and a column either has single 1 or all zeros. The masked image can be written as $\Phi^T\mathbf{y}$. The iterative update (6) and (7) here simplifies to gradient ascent of the prior over missing pixels, while keeping the observed pixels constant. We have included proof for this in our supplementary material.

In all of our experiments we consider row orthonormalized Φ and the term $\left(\Phi\Phi^T\right)^{-1}$ reduces to identity matrix.

4.2. Tricks used for inference

4.2.1. Four directions

Joint distribution in eqn. (1) can be factorized in multiple ways, for example along each of the four diagonal directions of an image, i.e., top-right, top-left, bottom-right and bottom-left. Gradients from different factorizations are considered at each iteration of the inference, by flipping the image in the corresponding direction. This leads to faster convergence as compared to just considering one direction. While doing inference on crops from BSDS test images, we observe that the convergence rate is twice faster with flipping.

4.2.2. Entropy-based Thresholding

While solving the MAP optimization, we observed that we can recover the edges quite well but texture regions are blurred. This happens because the RIDE model may not have the right mixture component (see Eqn. (3)) to explain the latent texture. In such cases, all the mixture components can be chosen with almost uniform probability, resulting in blurred texture. To detect such cases, in each iteration, we consider entropy of the posterior probability of scales and components in RIDE at each point as a metric to understand how confident the model is in modeling the distribution at that point. This is mathematically given as,

$$H(i,j) = -\sum_{c,s} p(c,s|\mathbf{x}_{<ij}, x_{ij}) \log(p(c,s|\mathbf{x}_{<ij}, x_{ij})),$$

if the point ij lies on an edge, the entropy is low as there are only certain selected components which can explain that edge. Whereas, if it lies in a flat or textured patch, the entropy is high and the point is equiprobable to come from different components and scales. Therefore, to reduce blurring we maintain a threshold on the entropy above which we clip the gradients to zero. Figure 3 shows the effect of entropy constraint on the texture reconstruction.

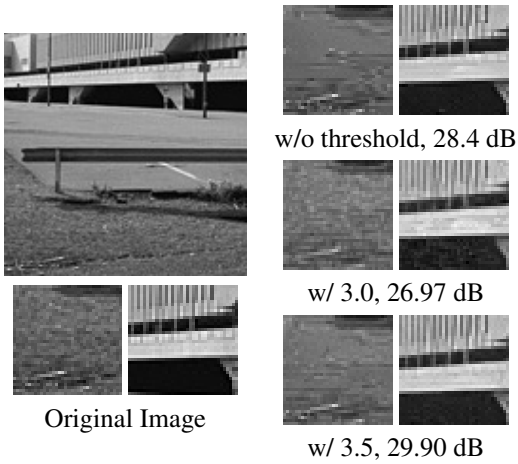


Fig. 3: Compressive sensing image reconstructions from 30% measurements obtained by varying entropy thresholds. The texture of the magnified patch is recovered better with the threshold 3.5. The numbers mentioned in dBs are PSNR of whole image.

5. EXPERIMENTS

For training the RIDE model we have used publicly available Berkeley Segmentation dataset (BSDS300). Following the instincts from [1], we trained the model with increasing patch size in each epoch. Starting with 8x8 patch we go till 22x22 in steps of 2 for 8 epochs. We used the code provided by authors of RIDE in caffe available here(https://github.com/lucastheis/ride/). We start with a very low learning rate (0.0001) and decrease it to half the previous value after every epoch. We used Adam optimization [18] for training the model. We observe that models with more than one spatial LSTM layer don't result in much of improvement for our tasks of interest. Hence we proceed with a single layer RIDE model for all the inference tasks mentioned in this paper. Also we have used entropy based gradient thresholding described above (sec. 4.2.2) to avoid blurring the texture region in all the experiments. In order to accommodate for boundary issues we remove a two pixel neighbourhood around the image for PSNR and SSIM calculations in all the experiments. For a fair comparison, we also do the same for the results of other methods.

5.1. Single Pixel Camera

In general, the SPC framework involves global multiplexing of the scene. But the recently proposed state-of-the-art methods for signal reconstruction, like ReconNet, are designed for local spatial multiplexing and can't handle the global multiplexing case directly. Our model, using Spatial LSTMs, can reason for long term dependencies in image sequences and is preferable for such kind of tasks. We show SPC reconstruction results on some randomly chosen images from the BSDS300 test set which were cropped to 160×160 size for computational feasibility, see figures on top of table 1. We generate compressive measurements from them using random Gaussian measurement matrix with orthonormalized rows. We take measurements at four different rates 0.4, 0.3, 0.25 and 0.15. Using the projected gradient method, we perform gradient ascent for 200 iterations for 0.4 and 0.3 measurement rates. For lower measurement rates, we run gradient ascent for 400 iterations. Also, we follow the entropy thresholding procedure mentioned in section 4.2.2 with a threshold value of 3.5 which we empirically found to be good for preserving textures. In all the cases, we start with a random image uniformly sampled from $(0, 1)$. Reconstruction results for five images are shown in Table 1 and Figure 1. We were able to show improvements both in terms of PSNR and SSIM values for different measurement rates. Even at low measurement rates, our method preserves the sharp and prominent structure in the image. D-AMP has the tendency to over-smooth the image, whereas TVAL3 adds blotches to even the smooth parts of the image.

5.2. Image Inpainting

For image inpainting, we randomly removed 70% of pixels and estimated them using aforementioned inference method. We compared our approach with the multiscale adaptive dictionary learning approach [17], which is an improvement over the KSVD algorithm, see Figure 2. It is clear from the figure that our approach is able to recover the sharp edges better than the multiscale KSVD approach. This is because our method is based on global image prior as compared to the patch-based multiscale KSVD approach.

6. CONCLUSIONS AND FUTURE WORK

We demonstrate that deep recurrent generative image models such as RIDE can be used effectively for solving compressive image recovery problems. The main advantages of using such deep generative models is that they are global priors and hence can model long term image dependencies. Also using the proposed MAP formulation we can solve many other image restoration tasks such as image deblurring, superresolution, demosaicing and computational photography problems such as coded aperture and exposure. Another direction of future work would be to adapt the trained generative model to the specific image that we are interested in restoring. We

Sr. No	Method	M.R.=30%		M.R.=25%		M.R.=15%	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
1	TVAL3	29.54	0.833	28.40	0.802	26.76	0.736
	D-AMP	31.59	0.867	29.70	0.836	24.70	0.716
	Ours	**33.87**	**0.898**	**32.72**	**0.872**	**29.68**	**0.792**
2	TVAL3	26.65	0.726	25.86	0.6810	24.51	0.575
	D-AMP	26.08	0.674	25.44	0.633	23.67	0.497
	Ours	**30.03**	**0.835**	**28.85**	**0.791**	**25.59**	**0.612**
3	TVAL3	26.20	0.789	25.08	0.745	22.63	0.638
	D-AMP	31.66	**0.923**	29.09	0.883	24.5	**0.757**
	Ours	**33.02**	0.919	**31.70**	**0.897**	**25.32**	0.754
4	TVAL3	30.20	0.849	29.10	0.820	25.44	0.719
	D-AMP	30.17	0.858	28.06	0.796	25.81	0.70
	Ours	**34.70**	**0.919**	**33.37**	**0.900**	**25.88**	**0.798**
4	TVAL3	30.20	0.849	29.10	0.820	25.44	0.719
	D-AMP	30.17	0.858	28.06	0.796	25.81	0.70
	Ours	**34.70**	**0.919**	**33.37**	**0.900**	**25.88**	**0.798**
Mean	TVAL3	27.858	0.783	26.90	0.745	24.69	0.646
	D-AMP	29.07	0.787	27.51	0.7464	24.48	0.633
	Ours	**31.57**	**0.859**	**30.54**	**0.818**	**27.09**	**0.700**

Table 1: Comparisons of compressive imaging reconstructions at different measurement rates for the images shown above. Our method outperforms the existing global prior based methods in most of the cases.

can use our entropy thresholding step to detect which part of the image is not modeled well by the generic model and then adapt the model accordingly.

7. REFERENCES

[1] Lucas Theis and Matthias Bethge, "Generative image modeling using spatial lstms," in *Advances in Neural Information Processing Systems*, 2015, pp. 1927–1935. 1, 2, 4

[2] Marc P Hansen and Douglas S Malchow, "Overview of swir detectors, cameras, and applications," in *SPIE Defense and Security Symposium*. International Society for Optics and Photonics, 2008, pp. 69390I–69390I. 1

[3] Marco F Duarte, Mark A Davenport, Dharmpal Takhar, Jason N Laska, Ting Sun, Kevin E Kelly, Richard G Baraniuk, et al., "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83, 2008. 1

[4] Aswin C Sankaranarayanan, Christoph Studer, and Richard G Baraniuk, "Cs-muvi: Video compressive sensing for spatial-multiplexing cameras," in *Computational Photography (ICCP), 2012 IEEE International Conference on*. IEEE, 2012, pp. 1–10. 1

[5] Chengbo Li, Wotao Yin, Hong Jiang, and Yin Zhang, "An efficient augmented lagrangian method with applications to total variation minimization," *Computational Optimization and Applications*, vol. 56, no. 3, pp. 507–530, 2013. 1

[6] Huaijin Chen, M Salman Asif, Aswin C Sankaranarayanan, and Ashok Veeraraghavan, "Fpa-cs: Focal plane array-based compressive imaging in short-wave infrared," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2358–2366. 1

[7] Jian Wang, Mohit Gupta, and Aswin C Sankaranarayanan, "Lisens-a scalable architecture for video compressive sensing," in *Computational Photography (ICCP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 1–9. 1

[8] Mohammad Aghagolzadeh and Hayder Radha, "Compressive dictionary learning for image recovery," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*. IEEE, 2012, pp. 661–664. 1

[9] Kuldeep Kulkarni, Suhas Lohit, Pavan Turaga, Ronan Kerviche, and Amit Ashok, "Reconnet: Non-iterative reconstruction of images from compressively sensed measurements," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 449–458. 1, 2

[10] Ali Mousavi, Ankit B Patel, and Richard G Baraniuk, "A deep learning approach to structured signal recovery," in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2015, pp. 1336–1343. 1

[11] David L Donoho, Arian Maleki, and Andrea Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18914–18919, 2009. 2

[12] Christopher A Metzler, Arian Maleki, and Richard G Baraniuk, "From denoising to compressed sensing," 2014. 2

[13] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, 2014, pp. 2672–2680. 2

[14] Diederik P Kingma and Max Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013. 2

[15] Aaron van den Oord, Nal Kalchbrenner, and Koray Kavukcuoglu, "Pixel recurrent neural networks," *arXiv preprint arXiv:1601.06759*, 2016. 2

[16] Alex Graves, "Neural networks," in *Supervised Sequence Labelling with Recurrent Neural Networks*, pp. 15–35. Springer, 2012. 2

[17] Julien Mairal, Guillermo Sapiro, and Michael Elad, "Learning multiscale sparse representations for image and video restoration," *Multiscale Modeling & Simulation*, vol. 7, no. 1, pp. 214–241, 2008. 2, 4

[18] Diederik Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014. 4