COMPRESSIVE AND CODED IMAGE RECOVERY USING DEEP RECURRENT PRIORS

A Project Report

submitted by

AKSHAT DAVE

in partial fulfilment of requirements
for the award of the dual degree of

BACHELOR OF TECHNOLOGY AND MASTER OF TECHNOLOGY

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY MADRAS

JUNE 2017

THESIS CERTIFICATE

This is to certify that the thesis titled **Coded and compressive image recovery using deep recurrent priors**, submitted by **Akshat Dave**, to the Indian Institute of Technology, Madras, for the award of the dual degree of **Bachelor of Technology and Master of Technology**, is a bona fide record of the research work done by him under our supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Prof. Kaushik Mitra
Research Guide
Assistant Professor
Dept. of Electrical Engineering
IIT-Madras, 600 036

Place: Chennai

Date: 8th June 2017

ACKNOWLEDGEMENTS

I would like to begin by thanking my guide, Prof. Kaushik Mitra for all his help, support and patience throughout the course of my project work. His approach and dedication to research has been incomparable and has always inspired me to keep pushing myself. All the students in the Computational Imaging Lab have been immensely cooperative, friendly and accommodating and have greatly inspired my own work. I am grateful to Qualcomm, for awarding the Qualcomm Innovation Fellowship award to our team. I would especially like to thank my team mate Anil for helping me out throught the project.

Were it not for the infallible faith of my parents and sister, I would not be present here. They have been a constant support throughout my life and I am forever grateful to them for their blessing and encouragement.

Lastly, I would like to give a huge shout-out to every one of my friends who have made the experience at IIT Madras, a treasured one. A special thank you to Ankit, Akash, Ayush, Shubham and Sapana for countless memories and experiences, that will stay with me forever.

ABSTRACT

KEYWORDS: Computational imaging, generative models, deep learning, LSTMs, MAP inference

Reconstruction of images from compressively sensed and coded measurements is an ill-posed problem. In this paper, we leverage the recurrent generative model, RIDE, as an image prior for reconstruction. Recurrent networks can model long-range dependencies in images and hence are suitable to handle global multiplexing in reconstruction from compressive and coded imaging. We perform MAP inference with RIDE using back-propagation to the inputs and projected gradient method. We propose an entropy thresholding based approach for preserving texture in images well. We apply our method for reconstruction from measurements obtained by three novel computational imaging cameras: Single Pixel Camera LiSens and FlatCam. Our approach shows superior reconstructions compared to recent global reconstruction approaches like D-AMP and TVAL3 on both simulated and real data.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

RIDE	Recurrent Image Density Estimator
MCGSM	Mixture of Conditional Gaussian Scale Mixtures
LSTM	Long Short Term Memory
MAP	Maximum Aposteriori Principle
SPC	Single Pixel Camera
CS	Compressive Sensing
MR	Measurement Rate
SSIM	Structural Similarity Index
PSNR	Peak Signal to Noise Ratio

CHAPTER 1

Introduction

Imaging in the non-visible region of the spectrum has a plethora of applications owing to its unique properties (Hansen and Malchow, 2008). For example, improved penetration of infrared waves through fog and smog enables imaging through scattering media. However, prohibitive sensing costs in the non-visible range have limited its widespread use[1]. Many works have proposed Compressive Sensing (CS) (Baraniuk, 2007; Donoho, 2006) as a viable solution for high-resolution imaging beyond the visible range of spectrum (Duarte *et al.*, 2008; Sankaranarayanan *et al.*, 2012; Chen *et al.*, 2015). Compressive sensing theory states that signals exhibiting sparsity in some transform domain can be reconstructed from much lower measurements than sampling at Nyquist rate (Donoho, 2006). Lesser the number of measurements lesser is the cost of sensing. The single-pixel camera (SPC) is a classical example of CS framework (Duarte *et al.*, 2008). In SPC, a single photo diode is used to capture compressive measurements and then reconstruct back the whole scene.

A challenge faced by CS reconstruction algorithms is to recover a high dimensional signal from a small number of measurements. This ill-posed nature of the reconstruction makes data priors essential. Often, signals exhibit sparse structure in some transform domain. For example, natural images in the domain of wavelets, DCT coefficients or gradients. Initially, reconstruction methods exploited this prior knowledge about the signal structure thereby restricting the solution set to desired signal (Li *et al.*, 2013; Sankaranarayanan *et al.*, 2012; Chen *et al.*, 2015; Wang *et al.*, 2015). However, using these simple sparsity based priors at very low measurement rates results in low-quality reconstructions (see TVAL3 reconstruction in fig. 5.2). This is due to their inability to capture the complexity of natural image statistics. On the other hand, data-driven approaches have been proposed recently to handle the complexity (Aghagolzadeh and Radha, 2012; Kulkarni *et al.*, 2016; Mousavi *et al.*, 2015). They led to successful results in terms of reconstruction. But all of these approaches handle only local multiplexing i.e

[1]Megapixel sensors in short-wave infrared, typically constructed using InGaAs, cost more than USD 100k.

measurements are taken from image patches and recovery is also done patch wise. This is not appealing for classical SPC framework as such since measurements are acquired through global multiplexing.

To address these problems, in this work we propose to use a data-driven global image prior, RIDE, proposed by Theis et al. (Theis and Bethge, 2015) for CS image recovery. RIDE uses recurrent networks with Long Short Term Memory (LSTM) units and is shown to model the long-term dependencies in images very well. Also, being recurrent it is not limited to patch size, hence can handle the global multiplexing in SPC, LiSens and FlatCam. Our contributions are as follows:

- We propose to use RIDE 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 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 the ability of the model.

CHAPTER 2

Prior Work

2.1 Role of Signal Priors

Image data priors have played a significant role for signal reconstruction from ill posed problems which are very common in image processing and computational photography. Initially such image priors were constructed through empirical observation of data statistics, for example TV norm minimization, sparse gradient prior (Levin *et al.*, 2007) and sparsity of coefficients in wavelet domain (Portilla *et al.*, 2003). On the other hand, many methods were proposed to learn the priors directly from data such as dictionary learning (Aharon *et al.*, 2006), mixture models like GMMs (Zoran and Weiss, 2011) and their variants GSMs (Portilla *et al.*, 2003), conditional models like Mixture of Conditional GSMs (MCGSM) (Theis *et al.*, 2012), undirected models like Field of Experts (FoEs) (Roth and Black, 2005). In dictionary learning an overcomplete set of basis is learnt by representing natural image patches as sparse linear combination of these basis. It has been successfully applied for many image processing tasks (Mairal *et al.*, 2008*b*; Aharon *et al.*, 2006). On the contrary, rest of the approaches explicitly model the data distribution by maximizing likelihood. GMMs are quite popular image patch priors and have been used for restoration tasks like image denoising and deblurring (Zoran and Weiss, 2011) where it gives competitive results compared to state-of-the-art methods like BM3D (Dabov *et al.*, 2009) and KSVD (Aharon *et al.*, 2006). FoEs (Roth and Black, 2005) is another popular model which is a Product of Experts (PoEs) with the desirable property of translational invariance making it a whole image prior. It has been used for image inpainting and denoising.

2.2 Deep Nets for Image Processing

Many recent approaches have been proposed to use feed forward deep networks for image reconstruction problems. Burget et al. (Burger *et al.*, 2012) used Multilayer per-

ceptrons (MLP) for image patch denoising performing on par with BM3D. Mao et al. (Mao *et al.*, 2016) used very deep convolutional encoder-decoder with skip connections for image denoising even handling different levels of Gaussian noise. It has surpassed BM3D's performance. Xu et al. (Xu *et al.*, 2014) have used convnets for image deconvolution. Kulkarni et al. (Kulkarni *et al.*, 2016) trained a convnet, termed as ReconNet, to recover image from compressed measurements of image patches with measurements being as low as 1%. Although these feed forward discriminative models are very fast at run time, their application is limited to the task they are trained for. Burget et al. (Burger *et al.*, 2012) reported difficulty in generalizing a MLP network trained at a particular noise level for different levels of Gaussian noise. Mao et al. (Mao *et al.*, 2016) handle this but at the cost of a huge network. ReconNet proposed for CS signal recovery requires the network to be trained again and again for each different sensing matrix and at each different measurement rate.

2.3 Deep Generative Models

Owing to the inherent problems posed by discriminative models, recently much effort has gone into building generative models such as, Generative Adversarial Nets (GAN) (Goodfellow *et al.*, 2014), Variational Auto Encoders (VAE) (Kingma and Welling, 2013), Pixel Recurrent Neural Networks (PixelRNN) (van den Oord *et al.*, 2016) and Recurrent Image Density Estimator (RIDE) (Theis and Bethge, 2015). GANs learn the ability to generate a plausible sample from the distribution of natural images. VAE provides a probabilistic framework for both encoding data to latent representation and decoding from it. Auto regressive models like RIDE model the current pixel distribution conditioned on the causal context where Spatial Long Short Term Memory (SLSTM) (Graves, 2012) units are used to obtain the contextual summary. PixelRNN is also an auto regressive model like RIDE but with much more complex architecture achieving the state-of-the-art performance in terms of loglikelihood scores. Apart from being expressive, RIDE and PixelRNN come with added advantages. Their directed nature facilitates the computation of exact likelihood. Also, these priors being auto regressive aren't limited to patch size, as is the case with discriminative and even non deep generative models. This is very useful particularly in cases like single pixel camera where the reconstruction has to take account of global multiplexing and patch based methods

can't be used directly.

Among these deep generative 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. PixelRNN although models the distribution, it discretizes the distribution of a pixel to 256 intensity values resulting in optimization difficulties. In this work we extend RIDE as an image prior for reconstruction problems in compressive sensing and image inpainting.

2.4 Inpainting

Image inpainting has been previously attempted with image priors. FoEs were applied to remove scratches or unwanted effects like text from an image. Theis et al. (Theis *et al.*, 2012) used conditional model MCGSM for image inpainting. Dictionary learning (Mairal *et al.*, 2008*a*) has also been proposed for image inpainting although not ideal since it is patch based. A multiscale adaptive version of dictionary learning (Mairal *et al.*, 2008*b*) is shown to perform well.

2.5 Coded and Compressive Imaging

2.5.1 Single pixel camera

SPC (Duarte *et al.*, 2008) is a compressive sensing framework (Candès *et al.*, 2006), where the goal is to reconstruct the image back from a very less number of random linear measurements. Typically this is an ill-posed problem and hence we need to use signal priors. Initially algorithms were proposed to minimize the l_1 norm assuming sparsity in the domain of wavelet coefficients, DCT coefficients or gradients (Li *et al.*, 2013). Later class of algorithms known as approximate message passing (AMP) algorithms (Donoho *et al.*, 2009) (Metzler *et al.*, 2014) use off-the-shelf denoiser to iteratively refine their solution. ReconNet is another recent method using CNNs. But 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.

2.5.2 LiSens

LiSens (Wang *et al.*, 2015) is a novel compressive imaging camera which replaces the single photo-diode in SPC with a 1-D line sensor. Here, all the rows in the scene array are multiplexed in parallel with the same line code. As a result, instead of multiplexing the entire scene into a single measurement, it captures a vector of measurements. For this an entire row of the DMD array is optically mapped to a single pixel in the linear sensor using a cylindrical lens. LiSens provides very high measurement rates than SPC, which are comparable to that of a full frame sensor, while capturing small fraction of measurements. Priors are essential for reconstruction from the compressed measurement vector. (Wang *et al.*, 2015) show reconstructions using Total Variational norm as prior.

2.5.3 FlatCam

FlatCam (**?**) is a novel lensless camera consisting of a coded mask on top of the sensor array. This architecture provides a thin, light and flexible form factor to the camera. Unlike conventional lens based camera which captures the entire scene on the sensor, FlatCam captures a coded representation of the scene. Due to the mask, each pixel on the sensor sees a linear combination of light from multiple scene elements. As a result, the image captured on the sensor looks nothing like the actual scene, necessitating the need for a reconstruction algorithm.

CHAPTER 3

Background on RIDE

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 follows:

$$p(\mathbf{x}) = \prod_{ij} p(x_{ij}|\mathbf{x}_{<ij}, \boldsymbol{\theta}_{ij}) \tag{3.1}$$

where $\boldsymbol{\theta}_{ij}$ are distribution parameters at that location. By making the Markov assumption we can limit the extent of $\mathbf{x}_{<ij}$ to a smaller neighbourhood. Another valid assumption is stationarity of the data which results in sharing the same parameters $\boldsymbol{\theta}$ across all locations ij, thus achieving translational invariance.

Now each factor in the above equation can be modeled by a mixture of GSMs with shared parameters $\boldsymbol{\theta}$ which makes it Mixture of Conditional Gaussian Scale Mixtures (MCGSM) as proposed by (Theis *et al.*, 2012),

$$p(x_{ij}|\mathbf{x}_{<ij}, \boldsymbol{\theta}) = \sum_{c,s} p(c, s|\mathbf{x}_{<ij}, \boldsymbol{\theta}) p(x_{ij}|\mathbf{x}_{<ij}, c, s, \boldsymbol{\theta}), \tag{3.2}$$

Where the sum is over components and scales,

$$\begin{aligned} p(c, s|\mathbf{x}_{<ij}) &\propto \exp(\eta_{cs} - 0.5 * e^{\alpha_{cs}} \mathbf{x}_{<ij}^T \mathbf{K}_c \mathbf{x}_{<ij}), \\ p(x_{ij}|\mathbf{x}_{<ij}, c, s) &= \mathcal{N}(x_{ij}; \mathbf{a}_c^T \mathbf{x}_{<ij}, e^{-\alpha_{cs}}) \end{aligned} \tag{3.3}$$

In MCGSM, Markov assumption was made and the past context $\mathbf{x}_{<ij}$ was actually limited to a small causal neighborhood. However natural images exhibit long range correlations and any smaller neighbourhood fails to capture them. On the other hand increasing the neighbourhood leads to dramatic increase in number of parameters. In order to take into account such dependencies (Theis and Bethge, 2015) have proposed to use two dimensional Spatial Long Short Term Memory (LSTMs) (Graves, 2012) units

for summarizing the causal context through their hidden representation $\mathbf{h}_{ij}$ at location ij as,

$$\mathbf{h}_{ij} = f(\mathbf{x}_{<ij}, \mathbf{h}_{i-1,j}, \mathbf{h}_{i,j-1}) \tag{3.4}$$

where f is a complex non linear function with memory elements analogous to physical read, write and erase elements thus giving it the ability to model the long term dependencies in sequences. This formulation results in replacement of the finite context $\mathbf{x}_{<ij}$ in conditional modeling equation (3.2) with $\mathbf{h}_{ij}$, thus bringing in the summary of entire causal context. Thus, the complete model is specified as follows:

$$p(\mathbf{x}) = \prod_{ij} p(x_{ij}|\mathbf{h}_{ij}, \boldsymbol{\theta}) \tag{3.5}$$

$$p(x_{ij}|\mathbf{h}_{ij}, \boldsymbol{\theta}) = \sum_{c,s} p(c, s|\mathbf{h}_{ij}, \boldsymbol{\theta}) p(x_{ij}|\mathbf{h}_{ij}, c, s, \boldsymbol{\theta}), \tag{3.6}$$

Using Recurrent Image Density Estimator (RIDE) (Theis and Bethge, 2015) have achieved one of the state-of-the-art results in terms of log-likelihood scores. For more details we recommend the reader to go through (Theis and Bethge, 2015).

CHAPTER 4

Compressive Image Recovery Using RIDE

Here we consider the problem of image restoration from linearly compressed measurements $\mathbf{y} = A\mathbf{x} + \mathbf{n}$, where the linear transformation A is a $M \times N$ with $M < N$, $\mathbf{n}$ is noise in the observation with known statistics.

4.1 MAP Inference via Backpropagation

Sequential sampling of the conditional factors has been used by RIDE to generate image samples from the joint distribution (Theis and Bethge, 2015). On similar lines, one method to do inference is to sample from the posterior distribution. But here sequential sampling is not possible and we have to resort to Markov Chain Monte Carlo methods such as Gibbs sampling which are computationally expensive even for smaller image sizes. Hence, we use Maximum-A-Posteriori principle to find the desired image $\hat{\mathbf{x}}$,

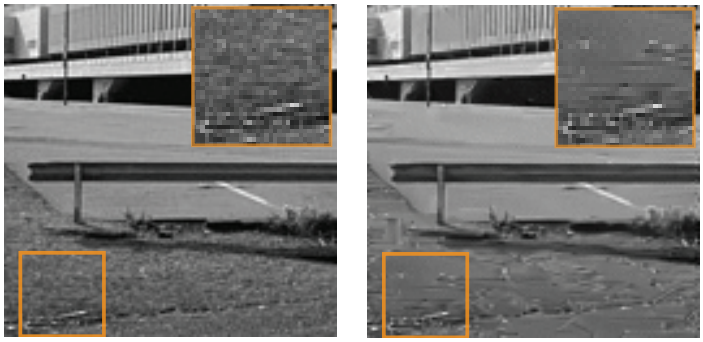
$$\hat{\mathbf{x}} = arg\max_{\mathbf{x}} p\left(\mathbf{x}|\mathbf{y}\right) = arg\max_{\mathbf{x}} p\left(\mathbf{x}\right) p\left(\mathbf{y}|\mathbf{x}\right) \tag{4.1}$$

The prior term $p\left(\mathbf{x}\right)$ is specified by the generative model (3.5),(3.6) and the likelihood is given by $p(\mathbf{y}|\mathbf{x}) \propto exp(-||\mathbf{y} - A\mathbf{x}||^2/\sigma^2)$ for the isotropic Gaussian noise case.

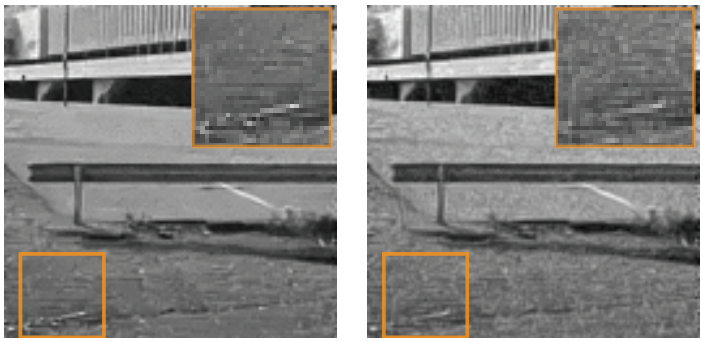
We apply gradient ascent to the net posterior distribution in order to obtain the reconstructed image. After log transforming the product in (4.1), the gradient with respect to the prior is given by:

$$\frac{\partial \log p(\mathbf{x})}{\partial x_{ij}} = \sum_{k \geq i, l \geq j} \frac{\partial \log p(x_{kl}|\mathbf{h}_{kl}, \boldsymbol{\theta})}{\partial x_{ij}} \tag{4.2}$$

Due to the recurrent nature of the model, each pixel through its hidden representation can contribute to the likelihood of all the pixels that come after it in forward pass.



Original Image w/o threshold, PSNR: 28.4 dB



w/ 3.5, PSNR: 29.90 dB w/ 3.0, PSNR: 26.97 dB

Figure 4.1: Compressive sensing image reconstructions from 30% measurements obtained by varying entropy thresholds. The texture of the magnified patch is recovered better with the threshold.

In a similar fashion during backward pass the gradient from each pixel propagates to all the pixels prior to it in the sequence. Gradients with respect to log-likelihood are much easier to evaluate is given by:

$$\nabla_{\mathbf{x}} \log p(\mathbf{y}|\mathbf{x}) \propto 2A^T(\mathbf{y} - A\mathbf{x}) \tag{4.3}$$

Using these gradient formulations, we can do gradient ascent for maximizing the log posterior with a momentum parameter for quick convergence.

$$\hat{\mathbf{x}}_{t+1} = \hat{\mathbf{x}}_t + \eta \nabla_{\mathbf{x}} \log(p(\mathbf{x})p(\mathbf{y}|\mathbf{x})) \tag{4.4}$$

Where η is the learning rate parameter.

4.2 Tricks used for inference

4.2.1 Four directions

Joint distribution (3.5) 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 the inference on crops from randomly sampled BSDS test images, we observe that the convergence rate is roughly 2 times faster when considering four directions.

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 (3.6)) 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 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 evaluated through posterior entropy given as,

$$H(i,j) = -\sum_{c,s} p(c,s|\mathbf{x}_{<ij}, x_{ij}) \log(p(c,s|\mathbf{x}_{<ij}, x_{ij})) \tag{4.5}$$

If the point lies on an edge, the posterior entropy is low as there are only certain selected components which can explain that edge. Whereas, if the point lies in a flat or textured patch, the posterior entropy is high and the point is equi-probable to come from different components and scales. Therefore, to reduce blurring we maintain a threshold on posterior entropy above which we clip the gradients to zero. 4.1 shows the effect of entropy constraint on the texture reconstruction.

4.3 Compressive Image Recovery

To demonstrate the effectiveness of our method, we consider the problems of image inpainting and compressive and coded imaging. In image inpainting our goal is to recover the missing pixels from a randomly masked image. We estimate the missing pixels by maximizing the prior over missing pixels, keeping the observed pixels constant. This is done by updating the gradients for only missing pixels. We have used the above mentioned entropy based gradient thresholding to avoid blurring the texture region.
For SPC, we formulated the MAP inference as,

$$\hat{\mathbf{x}} = arg\max_{\mathbf{x}} p(\mathbf{x}) \;\; s.t. \;\; \mathbf{y} = \Phi\mathbf{x} \tag{4.6}$$

Formulation for LiSens will also be similar , except for the fact that y will be a two dimensional matrix as each measurement is a one dimensional vector. For FlatCam, the measurements are obtained as $\mathbf{y} = \Phi_L \mathbf{x} \Phi_R^T$
For SPC, 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.

$$\hat{\mathbf{x}}_k = \mathbf{x}_{k-1} + \eta \nabla_{\mathbf{x}_{k-1}} p(\mathbf{x}) \tag{4.7}$$

$$\mathbf{x}_k = \hat{\mathbf{x}}_k - \Phi^T \left(\Phi\Phi^T\right)^{-1} (\Phi\hat{\mathbf{x}}_k - \mathbf{y}) \tag{4.8}$$

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

For the noisy measurements $\mathbf{y}$ will not exactly satisfy the constraint $\mathbf{y} = \Phi\mathbf{x}$. So, we cannot enforce hard constraints using the projected gradient method. Hence, we instead apply soft constraints by adding the term $\lambda\|\mathbf{y} - \Phi\mathbf{x}\|$ to the cost function for gradient ascent.
For LiSens and FlatCam, we use the soft constraints method. Since y is a matrix here we consider Frobenius norm $\|\mathbf{y} - \Phi\mathbf{x}\|_F$ instead of the $L2$ norm.
Inpainting can be considered a compressive sensing problem where each column of A matrix is 1-sparse. Projecting the gradients to Nullspace of A implies fixing the data gradients of the given points to be zero. We have proved this in Appendix A

CHAPTER 5

Experiments

For training the RIDE model we have used publicly available Berkeley Segmentation dataset (BSDS300). Following the instincts from (Theis and Bethge, 2015), 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[1]. 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 (Kingma and Ba, 2014) 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 in this paper. Also, we have used entropy based gradient thresholding 4.2.2 with threshold 3.5, to avoid blurring the texture regions in all the experiments. In order to accommodate for boundary issues we remove a two pixel neighbourhood around the image for PSNR and SSIM calculation in all the experiments. For a fair comparison, we also do the same for the reconstructions of TVAL3 (Li *et al.*, 2013) and D-AMP (Metzler *et al.*, 2014).

5.1 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 (Mairal *et al.*, 2008*b*), which is an improvement over the KSVD algorithm, see Figure 4.2. 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.

[1] https://github.com/lucastheis/ride/

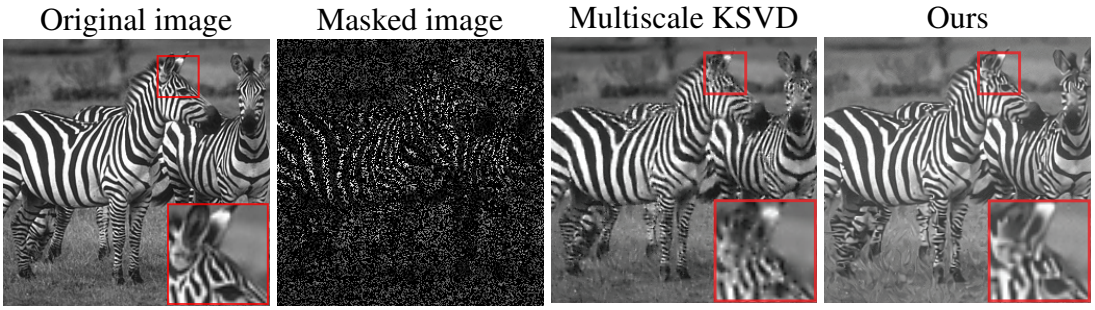


Figure 5.1: Inpainting comparisons: We compare our approach with the multiscale dictionary learning approach (Mairal *et al.*, 2008*b*). Our method is able to recover the sharp edges better than the multiscale KSVD approach, as is evident in the zoomed region around zebra's eye. This is because our method is a global prior as compared to the patch-based multiscale KSVD approach. The numbers mentioned below the figures are PSNR(left) and SSIM(right)

5.2 Single Pixel Camera

5.2.1 Results on simulated data

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 Figure 5.1. 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 300 iterations for 0.4, 0.3 and 0.25 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 5.1 and Figure 5.2. 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 ten-

car monument building statue bird

Figure 5.2: Randomly selected image crops of size 160x160 from BSDS300 test dataset used for CS reconstruction.

dency to over-smooth the image, whereas TVAL3 adds blotches to even the smooth parts of the image.

5.2.2 SPC with noise

To analyze the robustness of our framework with noise, we add different levels of Gaussian noise to the measurements obtained in the simulated case and obtain the reconstructions. The optimal value of λ is empirically found out at different noise levels. Here we report our results in terms of average PSNR values over the same set of five images shown in Figure 5.1 at different measurement rates. We can see that we are better than other methods at lower noise levels whereas at higher noise levels our performance drops slightly.

5.2.3 Real Image Reconstruction

Here we consider the real measurements acquired from a single pixel camera using Fast Walsh Hadamard transform (FWHT) as ϕ matrix. Figure 5.2 depicts the reconstructions obtained in this case for the measurement rates of 15% and 30%. It can be observed that our method provides superior reconstructions similar to the simulated case. Since we don't have original image here, we take reconstruction from D-AMP at 100% measurements as the ground truth. Using this we evaluate the PSNR and SSIM metrics.

Figure Name	Method	M.R. = 40%		M.R. = 30%		M.R. = 25%		M.R. = 15%	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Car	TVAL3	31.72	0.897	30.37	0.846	29.09	0.814	26.15	0.736
	D-AMP	34.00	0.908	32.31	0.877	30.05	0.839	24.70	0.716
	Ours	**36.05**	**0.932**	**34.24**	**0.901**	**32.91**	**0.868**	**29.58**	**0.776**
Monument	TVAL3	28.10	0.796	28.43	0.750	27.69	0.710	**26.13**	**0.611**
	D-AMP	27.33	0.740	27.90	0.707	27.19	0.665	23.05	0.460
	Ours	**32.02**	**0.881**	**29.73**	**0.809**	**28.78**	**0.766**	24.93	0.543
Building	TVAL3	28.40	0.842	26.16	0.784	25.13	0.747	22.75	0.644
	D-AMP	**36.04**	**0.961**	32.21	0.929	29.26	0.886	24.5	0.757
	Ours	34.80	0.948	**33.82**	**0.935**	**32.21**	**0.913**	**27.6**	**0.816**
Statue	TVAL3	**28.01**	0.777	**26.67**	0.712	26.08	0.675	**24.59**	0.583
	D-AMP	26.90	0.661	25.80	0.613	25.20	0.586	22.86	0.455
	Ours	27.97	**0.805**	26.59	**0.742**	**26.12**	**0.711**	24.14	**0.599**
Bird	TVAL3	32.57	0.901	31.75	0.874	30.68	0.847	28.30	0.771
	D-AMP	**38.45**	**0.970**	31.54	0.874	29.59	0.822	24.98	0.688
	Ours	37.70	0.948	**35.19**	**0.922**	**33.52**	**0.892**	**29.3**	**0.786**
Mean	TVAL3	29.70	0.833	28.68	0.793	27.73	0.759	25.58	0.670
	D-AMP	32.54	0.848	29.95	0.800	28.26	0.760	24.02	0.615
	Ours	**33.71**	**0.903**	**31.91**	**0.862**	**30.71**	**0.830**	**27.11**	**0.704**

Table 5.1: Comparisons of compressive imaging reconstructions at different measurement rates for the images shown in Figure 5.1. Our method outperforms the existing global prior based methods in most of the cases.

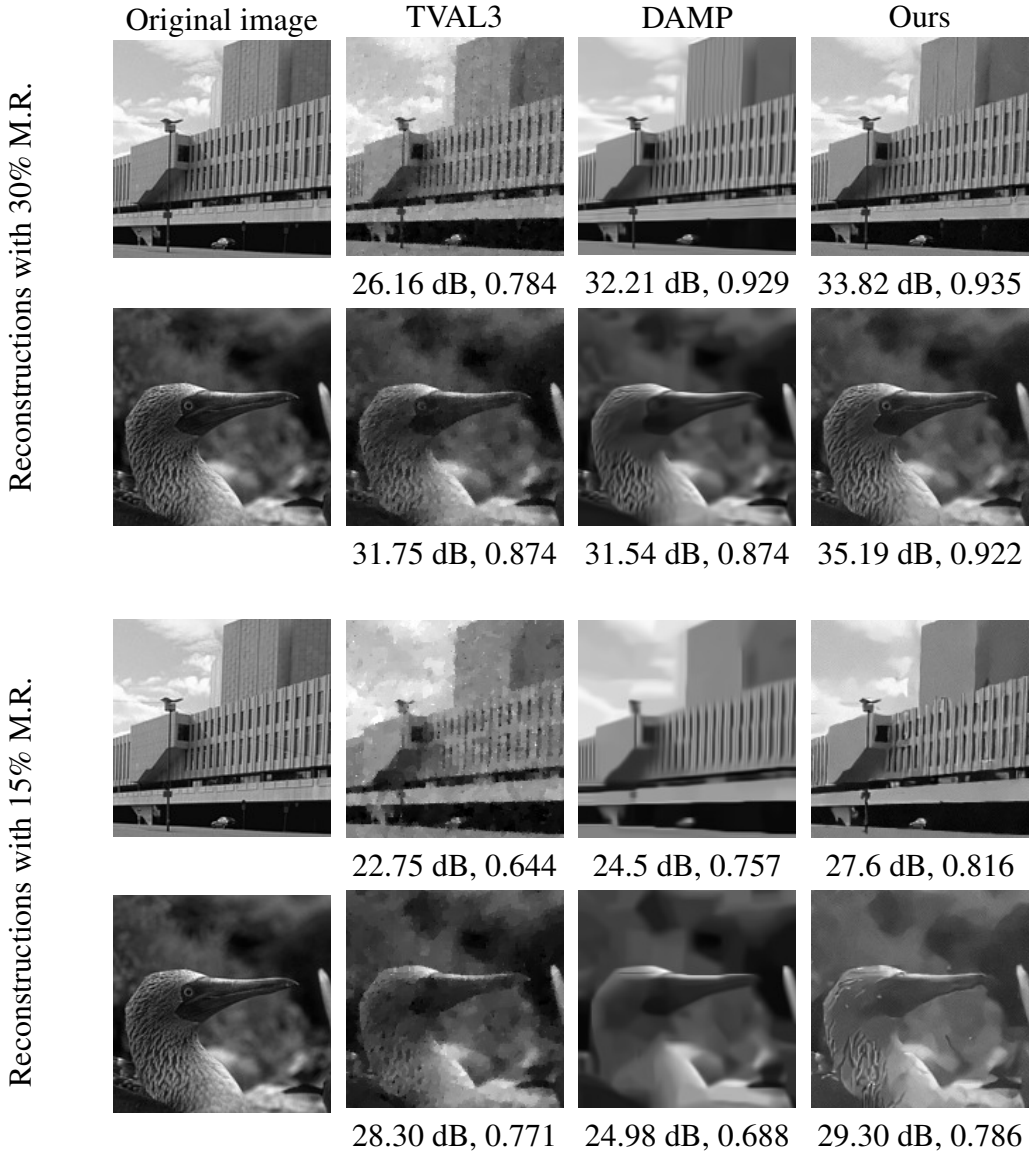


Figure 5.3: Images obtained by reconstruction from compressive measurements using D-AMP, TVAL3 and our method. Even at low measurement rates, our method preserves the sharp and prominent structures in the image. D-AMP has the tendency to over-smooth the image, whereas TVAL3 adds blotches to even the smooth parts of the image.

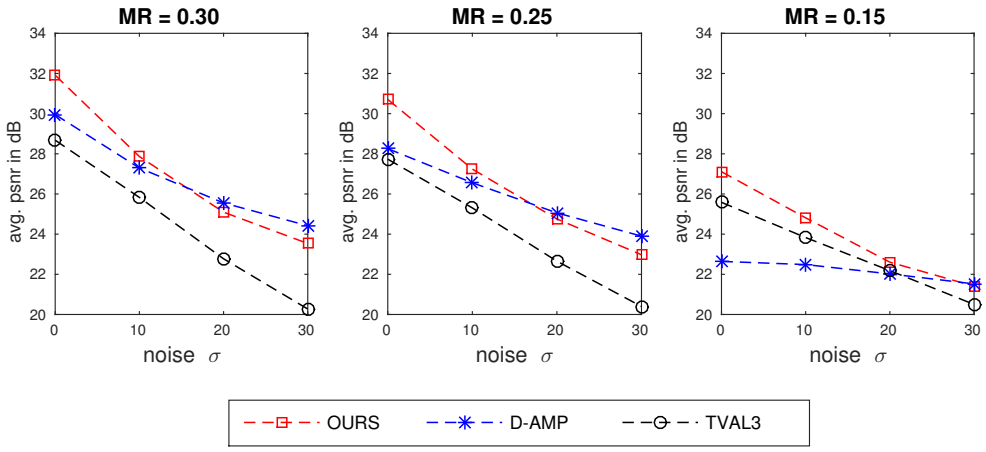


Figure 5.4: Performance of reconstructions from noisy measurements with different levels of Gaussian noise. (MR: Measurement Rate)

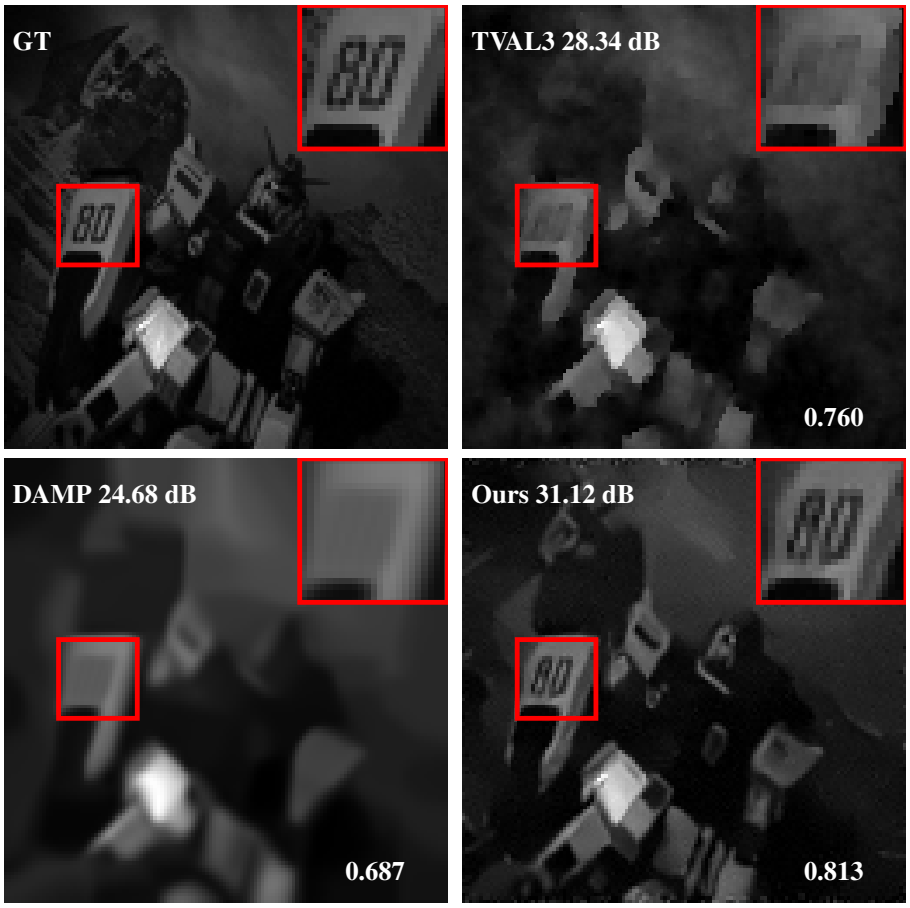


Figure 5.5: Real SPC reconstructions at 15% compression, our approach recovers the details better than others in real case also. Ground Truth (GT) is obtained from 100% reconstruction.

5.3 LiSens

For LiSens reconstructions, we simulate measurements from 160x160 test patches from BSDS dataset. ϕ matrix is constructed by selecting rows of a column permuted Hadamard

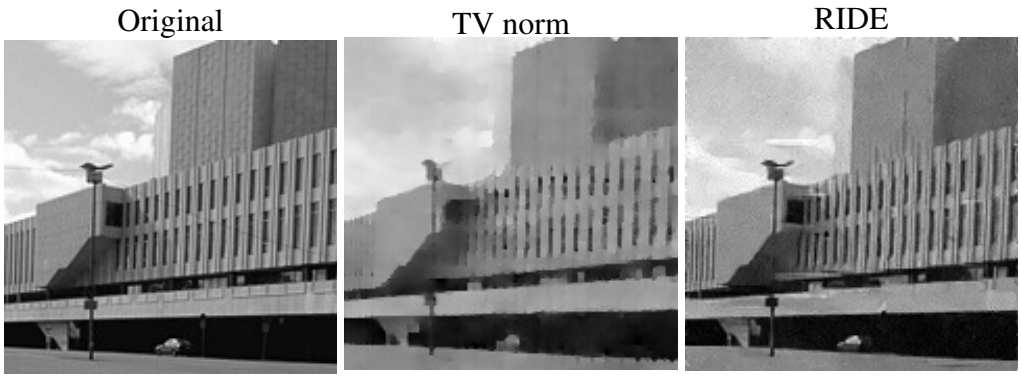


Figure 5.6: LiSens reconstructions on simulated measurements from 160x160 image with 15% measurement rate.

matrix. Soft constraints method is used for the MAP inference. The value of λ is found empirically. Figure 5.5 shows the reconstructions for 30% measurement rate. It can be observed that image obtained from TV norm minimization suffers from the same problems as with SPC. RIDE reconstruction is sharper and preserves the details well.

5.4 FlatCam

Here we show reconstructions from FlatCam on simulated and real data. For simulation we use the calibrated ϕ_L and ϕ_R matrices on 256x256 house image. Reconstructions are shown in Figure 5.6. The SVD and BM3D/SVD reconstructions have two characteristics:

- The reconstructions are bright at the centre and dark around the corners. This is because the pixels at the corner see less intensity of light.
- There are horizontal and vertical artifacts in the reconstructions. This is because the matrices ϕ_L and ϕ_R are calibrated using Hadamard patterns.

RIDE reconstructions do not face these issues and provide much sharper edges. However, it is unable to recover the fine texture present in the original image. Better texture recovery has been proposed as a future work by improving the model. Figure 5.7 shows the reconstructions from real world FlatCam measurements and similar inferences can be drawn from it.

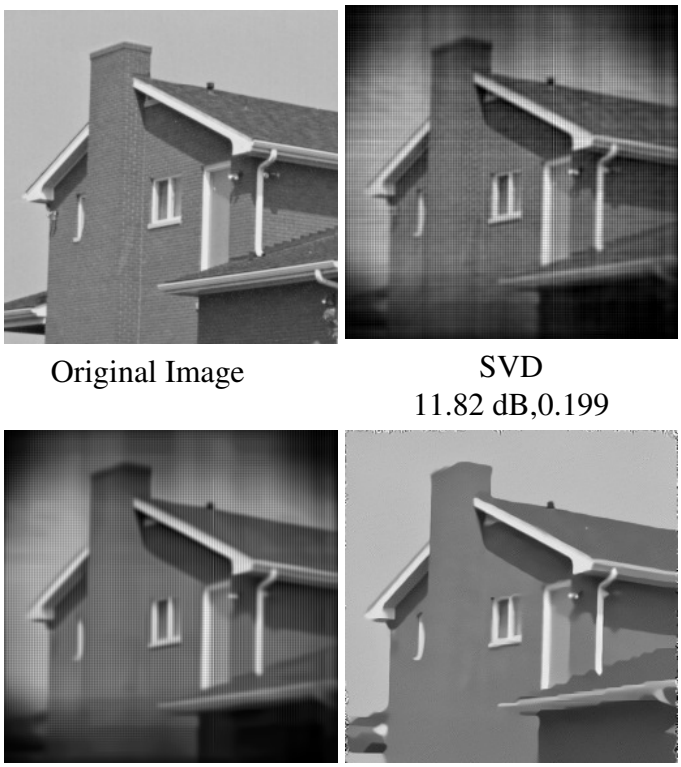


Figure 5.7: FlatCam reconstructions on simulated measurements from 256x256 image.

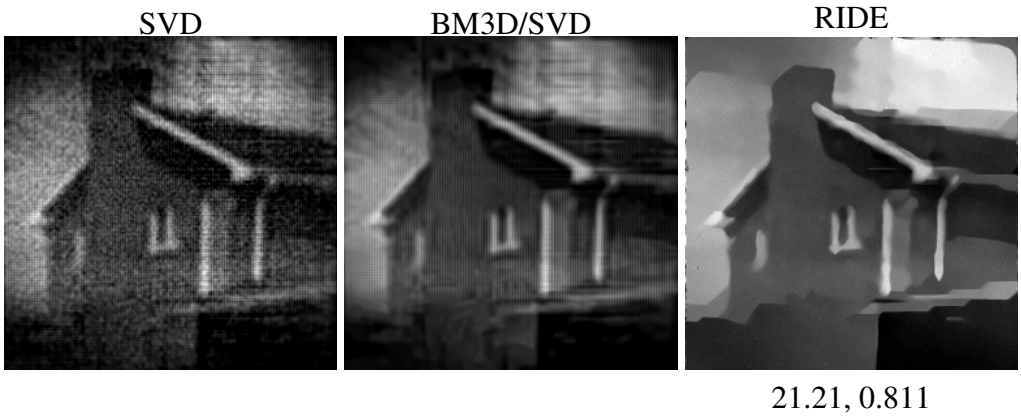


Figure 5.8: FlatCam reconstructions on real measurements

CHAPTER 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 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.

APPENDIX A

Random inpainting as compressive sensing formulation

For the case of compressive sensing recovery, we have mentioned the following equations for gradient ascent using projected gradient method :

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

$$\mathbf{x}_k = \hat{\mathbf{x}}_k - \Phi^T \left(\Phi\Phi^T\right)^{-1} \left(\Phi\hat{\mathbf{x}}_k - \mathbf{y}\right). \tag{A.2}$$

Here we prove that for inpainting an image containing random missing pixels, the iterative updates (A.1) and (A.2) simplify to gradient ascent of the prior over missing pixels while keeping the observed pixels constant. Consider $\mathbf{x}^*$ as the actual image and Φ as a binary row orthogonal CS matrix. In this case, the measurements $\mathbf{y} = \Phi\mathbf{x}$ are such that $\Phi^T\mathbf{y} = \Phi^T\Phi\mathbf{x}^*$ corresponds to the masked image $\mathbf{M} \odot \mathbf{x}^*$, where $\mathbf{M}$ is the random mask (0 for the missing pixels and 1 everywhere else) and $\odot$ denotes element-wise product.

Since $\Phi\Phi^T$ is an identity matrix, A.2 simplifies to :

$$\mathbf{x}_k = \hat{\mathbf{x}}_k - \Phi^T\Phi\hat{\mathbf{x}}_k + \Phi^T\mathbf{y}. \tag{A.3}$$

$$\mathbf{x}_k = \left(\mathbf{I} - \mathbf{M}\right) \odot \hat{\mathbf{x}}_k + \mathbf{M} \odot \mathbf{x}^* \tag{A.4}$$

Equation A.4 states that the missing pixels are updated according to gradient ascent over prior (A.1), while the known pixels are kept fixed.

APPENDIX B

Color Reconstructions

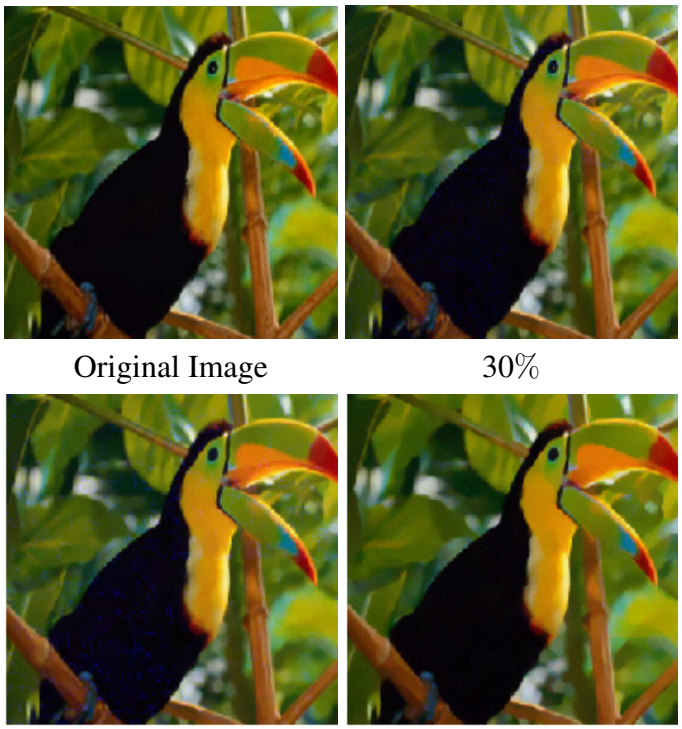


Figure B.1: Figure shows the color image reconstruction from measurements obtained through individual color channels (R, G and B) at different measurement levels.

REFERENCES

1. **Aghagolzadeh, M.** and **H. Radha**, Compressive dictionary learning for image recovery. *In Image Processing (ICIP), 2012 19th IEEE International Conference on*. IEEE, 2012. 1

2. **Aharon, M.**, **M. Elad**, and **A. Bruckstein** (2006). K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, **54**(11), 4311. 3

3. **Baraniuk, R. G.** (2007). Compressive sensing. *IEEE signal processing magazine*, **24**(4). 1

4. **Burger, H. C.**, **C. J. Schuler**, and **S. Harmeling**, Image denoising: Can plain neural networks compete with bm3d? *In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012. 3, 4

5. **Candès, E. J.**, **J. Romberg**, and **T. Tao** (2006). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, **52**(2), 489–509. 5

6. **Chen, H.**, **M. Salman Asif**, **A. C. Sankaranarayanan**, and **A. Veeraraghavan**, Fpacs: Focal plane array-based compressive imaging in short-wave infrared. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015. 1

7. **Dabov, K.**, **A. Foi**, **V. Katkovnik**, and **K. Egiazarian**, Bm3d image denoising with shape-adaptive principal component analysis. *In SPARS'09-Signal Processing with Adaptive Sparse Structured Representations*. 2009. 3

8. **Donoho, D. L.** (2006). Compressed sensing. *IEEE Transactions on information theory*, **52**(4), 1289–1306. 1

9. **Donoho, D. L.**, **A. Maleki**, and **A. Montanari** (2009). Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, **106**(45), 18914–18919. 5

10. **Duarte, M. F.**, **M. A. Davenport**, **D. Takhar**, **J. N. Laska**, **T. Sun**, **K. E. Kelly**, **R. G. Baraniuk**, *et al.* (2008). Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, **25**(2), 83. 1, 5

11. **Goodfellow, I.**, **J. Pouget-Abadie**, **M. Mirza**, **B. Xu**, **D. Warde-Farley**, **S. Ozair**, **A. Courville**, and **Y. Bengio**, Generative adversarial nets. *In Advances in Neural Information Processing Systems*. 2014. 4

12. **Graves, A.**, Neural networks. *In Supervised Sequence Labelling with Recurrent Neural Networks*. Springer, 2012, 15–35. 4, 7

13. **Hansen, M. P.** and **D. S. Malchow**, Overview of swir detectors, cameras, and applications. *In SPIE Defense and Security Symposium*. International Society for Optics and Photonics, 2008. 1

14. **Kingma, D.** and **J. Ba** (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. 14

15. **Kingma, D. P.** and **M. Welling** (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*. 4

16. **Kulkarni, K.**, **S. Lohit**, **P. Turaga**, **R. Kerviche**, and **A. Ashok**, Reconnet: Non-iterative reconstruction of images from compressively sensed measurements. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016. 1, 4

17. **Levin, A.**, **R. Fergus**, **F. Durand**, and **W. T. Freeman** (2007). Deconvolution using natural image priors. *Massachusetts Institute of Technology, Computer Science and Artificial Intelligence Laboratory*. 3

18. **Li, C.**, **W. Yin**, **H. Jiang**, and **Y. Zhang** (2013). An efficient augmented lagrangian method with applications to total variation minimization. *Computational Optimization and Applications*, **56**(3), 507–530. 1, 5, 14

19. **Mairal, J.**, **M. Elad**, and **G. Sapiro** (2008*a*). Sparse representation for color image restoration. *IEEE Transactions on image processing*, **17**(1), 53–69. 5

20. **Mairal, J.**, **G. Sapiro**, and **M. Elad** (2008*b*). Learning multiscale sparse representations for image and video restoration. *Multiscale Modeling & Simulation*, **7**(1), 214–241. vi, 3, 5, 11, 14

21. **Mao, X.-J.**, **C. Shen**, and **Y.-B. Yang** (2016). Image restoration using convolutional auto-encoders with symmetric skip connections. *arXiv preprint arXiv:1606.08921*. 4

22. **Metzler, C. A.**, **A. Maleki**, and **R. G. Baraniuk** (2014). From denoising to compressed sensing. 5, 14

23. **Mousavi, A.**, **A. B. Patel**, and **R. G. Baraniuk**, A deep learning approach to structured signal recovery. *In 2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2015. 1

24. **Portilla, J.**, **V. Strela**, **M. J. Wainwright**, and **E. P. Simoncelli** (2003). Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Transactions on Image processing*, **12**(11), 1338–1351. 3

25. **Roth, S.** and **M. J. Black**, Fields of experts: A framework for learning image priors. *In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2. IEEE, 2005. 3

26. **Sankaranarayanan, A. C.**, **C. Studer**, and **R. G. Baraniuk**, Cs-muvi: Video compressive sensing for spatial-multiplexing cameras. *In Computational Photography (ICCP), 2012 IEEE International Conference on*. IEEE, 2012. 1

27. **Theis, L.** and **M. Bethge**, Generative image modeling using spatial lstms. *In Advances in Neural Information Processing Systems*. 2015. 2, 4, 7, 8, 9, 14

28. **Theis, L.**, **R. Hosseini**, and **M. Bethge** (2012). Mixtures of conditional gaussian scale mixtures applied to multiscale image representations. *PloS one*, **7**(7), e39857. 3, 5, 7

29. **van den Oord, A.**, **N. Kalchbrenner**, and **K. Kavukcuoglu** (2016). Pixel recurrent neural networks. *arXiv preprint arXiv:1601.06759*. 4

30. **Wang, J.**, **M. Gupta**, and **A. C. Sankaranarayanan**, Lisens-a scalable architecture for video compressive sensing. *In Computational Photography (ICCP), 2015 IEEE International Conference on*. IEEE, 2015. 1, 6

31. **Xu, L.**, **J. S. Ren**, **C. Liu**, and **J. Jia**, Deep convolutional neural network for image deconvolution. *In Advances in Neural Information Processing Systems*. 2014. 4

32. **Zoran, D.** and **Y. Weiss**, From learning models of natural image patches to whole image restoration. *In 2011 International Conference on Computer Vision*. IEEE, 2011. 3

LIST OF PAPERS BASED ON THESIS

"Compressive Image Recovery using Recurrent Generative Model ", Akshat Dave, Anil Kumar Vadathya, Kaushik Mitra

1. Accepted at IEEE International Conference on Image Processing (ICIP) 2017
2. Poster presented at International Conference on Computational Photography (ICCP) 2017
3. arxiv preprint https://arxiv.org/abs/1612.04229