Coupled Autoencoder Based Reconstruction of Images from Compressively Sampled Measurements

Kavya Gupta
Embedded Systems and Robotics,
TCS Research and Innovation, India
gupta.kavya@tcs.com

Brojeshwar Bhowmick
Embedded Systems and Robotics,
TCS Research and Innovation, India
b.bhowmick@tcs.com

***Abstract*— This work addresses the problem of reconstructing images from their lower dimensional random projections using Coupled Autoencoder (CAE). Traditionally, Compressed Sensing (CS) based techniques have been employed for this task. CS based techniques are iterative in nature; hence inversion process is time-consuming and cannot be deployed for the real-time reconstruction process. These inversion processes are transductive in nature. With the recent development in deep learning – autoencoders, CNN based architectures have been used for learning inversion in an inductive setup. The training period for inductive learning is large but is very fast during application. But these approaches work only on the signal domain and not on the measurement domain. We show the application of CAE, which can work directly from the measurement domain. We compare CAE with a Dictionary learning based coupling setup and a recently proposed CNN based CS reconstruction algorithm. We show reconstruction capability of CAE in terms of PSNR and SSIM on a standard set of images with measurement rates of 0.04 and 0.25.**

Keywords — inverse problem, reconstruction, inductive learning, transfer learning, autoencoders

I. Introduction

Image reconstruction is an important vision task in scene understanding, where there is an urge to acquire as low measurements as possible at the sensing level and still be able to reconstruct enough information about the scene at the decoding level. Earlier Compressive Sensing based techniques [1]-[4] have been used to recover images from their low dimensional random projections but they are hindered by various drawbacks. CS based techniques and its variants are slow; they require solving several complex optimization problems in an iterative setup. Hence, the CS-based reconstruction paradigm is not suitable for real-time applications. Additionally, current techniques are capable of giving high-quality reconstructions at high measurement rates but fail drastically as measurement rate decreases, yielding reconstructions which are not useful for any image understanding tasks.

CS involves no learning. The Dictionary Learning variants of CS learn from the signal that they have to reconstruct. Both are transductive in nature. With the advent of deep learning new inductive techniques for solving inverse problems have been proposed for various vision tasks [5]-[13].

These inductive techniques apply the transpose of the projection operator on the lower dimensional measurements to get a noisy signal. This is applied at the input and the clean version of it is applied at the output during the learning process. The deep neural network 'learns' the inversion process. The problem with all such approaches is that they can only work in the signal domain and not from the lower dimensional measurements. In this work, we show the applicability of Coupled autoencoder, where the inversion will be learned directly from the measurement domain.

Our work is based on the coupled representation learning approach. Coupled Autoencoders (CAE) [14]-[16] and Coupled Dictionary Learning (SDL) [17], [18] have been proposed before. In this work, we show Coupled Autoencoder as a framework to reconstruct images from their compressed random measurements. We learn a representation in the source domain, which in our case constitutes the lower dimensional measurements; the target domain is the signal – a representation of the signal is learned as well. There is a learned coupling map from the source (measurement domain) to the target (signal domain) representation.

Rest of the paper consists of four parts. The next section discusses the basics of CS-based reconstruction, basics of Autoencoders and coupled representation learning. Section 3 outlines the Coupled Autoencoder approach. Section 4 shows the experimental results and the last section concludes our work.

II. Literature Review

A. Compressed Sensing Based Reconstruction

The signal 'x' is collected at the sensor node. For energy efficient transmission, it needs to be compressed. Compressed Sensing (CS) compresses the signal by projecting it onto a random matrix. This is expressed as,

$$y = Ax \tag{1}$$

Here x is the signal; A is the compression matrix and y the compressed measurement.

ISBN 978-90-827970-1-5 © EURASIP 2018

This compressed measurement is transmitted where the signal is reconstructed by exploiting its sparsity in some transform domain like DCT or wavelet. This is expressed as,

$$\min_{\alpha} \|\alpha\|_1 \text{ such that} \left\| y - AS^T\alpha \right\|_2^2 \le \varepsilon \quad (2)$$

Here α is the sparse transform coefficient in domain S for the signal x, ε is a parameter that controls the data fidelity. It is assumed that the sparsifying transform is either orthogonal or tight-frame.

B. Stacked Denoising Autoencoder

An autoencoder is a self-supervised neural network, i.e. the input and the output are same. There is an encoder W that projects the input to a hidden representation and a decoder (W') that reverse maps the representation to the output (= input). Mathematically this is expressed as,

$$X = W'\varphi(WX) \quad (3)$$

Here φ is a non-linear activation function.

Given the training samples (X), the encoder and the decoder are estimated by minimizing the Euclidean cost function.

$$\min_{W,W'} \left\| X - W'\varphi(Wx) \right\|_F^2 \quad (4)$$

Deeper architectures can be learned by nesting autoencoders inside each other. There are multiple encoders followed by an equal number of decoders. The learning is expressed as,

$$\arg\min_{W_1,\ldots,W_L,W_1',\ldots,W_L'} \left\| X - g \circ f(X) \right\|_F^2 \quad (5)$$

where $g = W_1'\phi(W_2'\ldots W_{L'}'(f(X)))$ and $f = \phi(W_L\phi(W_{L-1}\ldots\ldots\phi(W_1X)))$.

Traditionally autoencoders have been used to pre-train deep neural networks [19]. Studies in the recent past [8]-[10] have shown that autoencoders can be used for 'learning to solve' inverse problems; especially simple inverse problems like denoising and deblurring. During the training phase, the corrupt signal (noisy/blurry) is input to the autoencoder and the corresponding clean signal is at the output. From a large volume of data, the stacked autoencoder learns the inversion operation. During operation the corrupt signal is input, the stacked autoencoder is expected to clean it.

For images, this idea has been extended to reconstruction. A corrupted version from compressive measurements can be obtained via application of the transpose of the measurement operator, i.e. $\hat{x} = A^T y$. This corrupt version is used at the input and the clean one at the output during training.

C. Coupled Representation Learning

The main idea in coupled representation learning is to learn a basis (and corresponding coefficients) for the two domains - source and target, such that the coefficients from one domain can be linearly mapped to the other. It has been used for solving a variety of problems in image synthesis, e.g., single image super-resolution, photo-sketch synthesis, cross-spectral (RGB-NIR) face recognition, RGB depth classification etc. It has also been used for trans-lingual information retrieval [20].

III. Coupled Autoencoders

In this work, we show the applicability of Coupled Autoencoders to reconstruct images directly from their lower dimensional measurements/representation. The autoencoders for source and target will be learned along with the linear mapping (a coupling) between the two. Learning for all the variables influence each other in the training process. The source autoencoder uses a lower dimensional version of the images and the target autoencoder uses the corresponding ground truth images. The coupling learns to map the representation from the source (measurement domain) to the target (signal domain). Mathematically this is expressed as,

$$\min_{W_{DS},W_{ES},W_{DT},W_{ET},M} \left\| X_S - W_{DS}W_{ES}X_S \right\|_F^2 + \left\| X_T - W_{DT}W_{ET}X_T \right\|_F^2 + \lambda \left\| W_{ET}X_T - MW_{ES}X_S \right\|_F^2 \quad (6)$$

The first term $\left\| X_S - W_{DS}W_{ES}X_S \right\|_F^2$ is the standard autoencoder formulation for the source. Here W_{DS} corresponds to the decoder and W_{ES} is the encoder. The second term $\left\| X_T - W_{DT}W_{ET}X_T \right\|_F^2$ is the autoencoder for the target; W_{DT} and W_{ET} are the decoder and the encoder weights respectively. $\left\| W_{ET}X_T - MW_{ES}X_S \right\|_F^2$ is the coupling term. The linear map M couples the representation from the source to that of the target.

We solve equation (6) using the variable splitting technique. The first step is to introduce proxies for the representations, i.e. $Z_S = W_{ES}X_S$ and $Z_T = W_{ET}X_T$ forming the Lagrangian. A better approach to handle this is to use the Split Bregman technique [21]. Bregman relaxation variables B_S and B_T are introduced such that the value of the hyper-parameter need not be changed; the relaxation variable is updated in every iteration automatically so that it enforces equality at convergence.

The Split Bregman formulation is,

$$\min_{W_{DS},W_{ES},W_{DT},W_{ET},M} \left\| X_S - W_{DS}W_{ES}X_S \right\|_F^2 + \left\| X_T - W_{DT}W_{ET}X_T \right\|_F^2 + \lambda \left\| Z_T - MZ_S \right\|_F^2 + \eta\left(\left\| Z_S - W_{ES}X_S - B_S \right\|_F^2 + \left\| Z_T - W_{ET}X_T - B_T \right\|_F^2 \right) \quad (7)$$

ISBN 978-90-827970-1-5 © EURASIP 2018

Equation (7) is solved by alternating minimization and the detailed solution can be found in [16]. We have used two stopping criteria. Iterations stop when a specified maximum number of iterations is reached or when the objective function converges to a local minimum. By convergence, we mean that the value of the objective function does not change much in successive iterations.

During operation/testing, we have the compressed version Y in the measurement/source domain. From this, we will find the corresponding target domain image by,

$$X = W_{DT} M W_{ES} Y \tag{8}$$

IV. Experimental Results

In this work, experiments are done following the protocol as in [11]. The training image set consists of 91 images [22] and testing image set consists of 11 standard images. Only the luminance component of the RGB images was used to extract the patches of size 33x33 for training the network. These patches serve as the target domain/ground truth for the architecture. The CS measurements of the patches were obtained by multiplying with the measurement matrices Φ provided by the authors of [11]. For a given measurement rate, measurement matrix, Φ was first generated by a random Gaussian matrix of appropriate size, followed by orthonormalizing of the rows. This version of patches serves as the corresponding source domain in the architecture. During test time we use the learned network to recover the luminance channel of the test images.

We show the results of two measurements rates (MR) = 0.25 and 0.04. The patch size is 33x33 = 1089 pixels per block, the compressed measurements version consists of 272 and 43 pixels per block respectively.

There are two parameters to be tuned in the proposed method – λ and η. They were trained on separate validation set. The final values used here are λ=0.1 and η=1. The weights for the signal and the measurement domain are perfectly determined. The iterations were run till the objective function converged to local minima. By convergence, we mean that the value of the objective function does not change much in successive iterations. Coupled Autoencoder formulation just takes a couple of minutes to train on 91 images.

We compare the proposed method with two other techniques. The first one is another Coupled representation learning architecture; Coupled Dictionary Learning (CDL) [18] which has been used extensively for various tasks in vision and other is a recent CNN based technique [11] which was primarily designed for reconstruction. ReConNet outperforms SDA [23], a widely used technique for solving inverse problems in vision and three iterative CS image reconstruction techniques TVAL3 [24], NLR-CS [25] and D-AMP [26]. Hence, we refrain from repeating those comparisons here and show the comparison with only ReConNet. We tried our best to optimize CDL on the given dataset and for ReConNet we use the results provided by the authors for a fair comparison.

Our technique outperforms both Coupled Dictionary learning and ReConNet for image reconstruction quality measured in terms of PSNR and SSIM. Unlike the earlier studies which perform good for higher measurements rates and fail at low measurement rates, Coupled Autoencoders performs equally good at a low measurement rate of 0.04. We show the results for two measurement ratios in Table II and Table III. We get a significant improvement on both PSNR and SSIM when compared with both CDL and ReConNet. For visual corroboration, we show the results in figure 1 and figure 2.

Also, we compare the time complexity of the networks for real-time reconstruction capacity. All the experiments were run on an Intel Xeon 3.50GHz × 8 Processor, having 64 GB RAM with Nvidia Quadro 4000 GPU. Table I shows the timings in seconds for a testing image of size 256x256 for MR = 0.04. CDL takes over a minute to reconstruct a single image which can be detrimental in case of video stream which requires seconds of processing on every scene. Even as compared to ReConNet we are five times faster for the reconstruction.

Method	Testing time / image (s)
CDL	65
ReConNet	0.16
CAE	0.03

Table I – Comparison of Reconstruction times

V. Conclusion

In this work, we demonstrated the applicability of Coupled Autoencoders for real-time image reconstruction from random compressed sensing measurements. We modeled the problem as an inversion learning problem. We show that by modeling the measurement as the source domain and the image as the target domain, we can learn the inversion operation via Coupled Autoencoders while simultaneously learning a mapping between the domains. This is an inductive inversion learning approach. We showed improvement over existing techniques in terms of image reconstruction quality and in terms of time complexity.

Acknowledgment

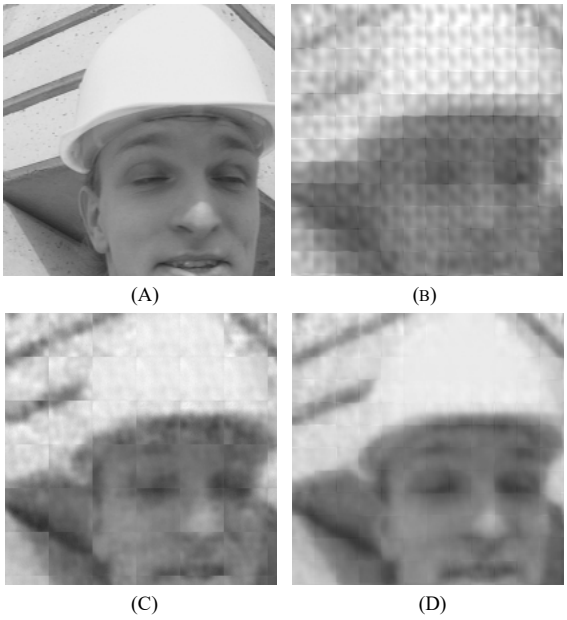
We thank Dr. Angshul Majumdar, Indraprastha Institute of Information Technology Delhi, a Research advisor with TCS Research and Innovation, for his active involvement and insights to the project.

Images	MR= 0.04	Monarch	Parrots	Barbara	Boats	Cameraman	Fingerprint	Flinstones	foreman	House	Lena	Peppers
PSNR (dB)	CDL	17.2776	20.1748	19.8658	19.6335	19.1044	17.1242	15.8905	22.0077	21.3142	20.2948	18.6010
	ReConNet	18.1921	20.3659	20.385	21.3474	19.2798	16.9281	16.375	23.6878	22.568	21.2653	19.5602
	CAE	**19.9418**	**23.0841**	**22.2775**	**23.0840**	**21.1413**	**18.4744**	**17.9238**	**27.5176**	**24.9930**	**23.4549**	**21.7059**
SSIM	CDL	0.4587	0.5572	0.4015	0.4381	0.4479	0.3262	0.3344	0.5496	0.5170	0.4879	0.4726
	ReConNet	0.5205	0.6295	0.4824	0.531	0.5314	0.4197	0.3598	0.6608	0.6087	0.5625	0.5103
	CAE	**0.6250**	**0.7597**	**0.5714**	**0.6172**	**0.6288**	**0.4914**	**0.4610**	**0.7963**	**0.7191**	**0.6714**	**0.6480**

Table II – Comparison of Proposed approach with existing techniques on PSNR and SSIM for measurement rate of 0.04

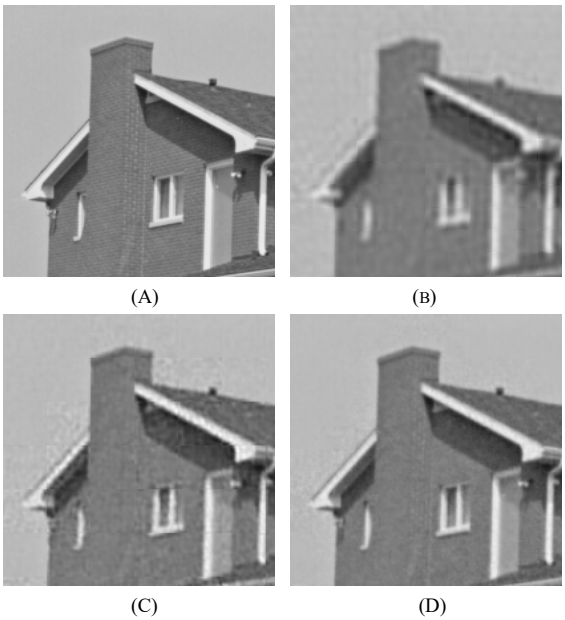
Images	MR= 0.25	Monarch	Parrots	Barbara	Boats	Cameraman	Fingerprint	Flinstones	foreman	House	Lena	Peppers
PSNR (dB)	CDL	22.2279	24.1856	23.2487	25.0479	22.4628	21.8106	19.5960	29.2850	27.3208	27.3208	23.5351
	ReConNet	24.4008	25.6906	23.2505	27.3000	23.1808	25.5837	22.6032	29.4726	28.47	26.544	24.7699
	CAE	**27.9092**	**29.6686**	**26.906**	**30.7178**	**26.6379**	**29.2478**	**25.9220**	**34.5873**	**32.7863**	**30.3342**	**28.7253**
SSIM	CDL	0.7158	0.7904	0.6198	0.6933	0.6855	0.6994	0.5497	0.8299	0.7708	0.7368	0.7254
	ReConNet	0.7721	0.7932	0.6772	0.7769	0.7052	0.8723	0.6736	0.8233	0.7701	0.7679	0.7246
	CAE	**0.8784**	**0.9117**	**0.8343**	**0.8811**	**0.8274**	**0.9412**	**0.8059**	**0.9232**	**0.888**	**0.8946**	**0.8597**

Table III – Comparison of Proposed approach with existing techniques on PSNR and SSIM for measurement rate of 0.25



(A) (B) (C) (D)

Figure 1: Image reconstruction comparison of Coupled Autoencoders with Coupled Dictionary Learning and ReConNet MR = 0.04 (A) Ground truth (B) CDL **(** PSNR: 22.0077 SSIM:0.5496) (C) ReConNet **(** PSNR: 23.6878 SSIM: 0.6608) (D) CAE **(** PSNR: 27.5176 SSIM: 0.7963)



(A) (B) (C) (D)

Figure 2: Image reconstruction comparison of Coupled Autoencoders with Coupled Dictionary Learning and ReConNet MR = 0.25 (A) Ground truth (B) CDL **(** PSNR: 27.3208 SSIM:0.7708) (C) ReConNet **(** PSNR: 28.47 SSIM:0.7701) (D) CAE **(** PSNR: 32.7863 SSIM:0.888)

ISBN 978-90-827970-1-5 © EURASIP 2018

REFERENCES

[1] Som, S., & Schniter, P. (2012). Compressive imaging using approximate message passing and a Markov-tree prior. *IEEE transactions on signal processing*, *60*(7), 3439-3448.

[2] Baraniuk, R. G., Cevher, V., Duarte, M. F., & Hegde, C. (2010). Model-based compressive sensing. *IEEE Transactions on Information Theory*, *56*(4), 1982-2001.

[3] Kim, Y., Nadar, M. S., & Bilgin, A. (2010, September). Compressed sensing using a Gaussian scale mixtures model in wavelet domain. In *Image Processing (ICIP), 2010 17th IEEE International Conference on* (pp. 3365-3368). IEEE.

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

[5] Ren, J. S., & Xu, L. (2015, January). On Vectorization of Deep Convolutional Neural Networks for Vision Tasks. In *AAAI* (pp. 1840-1846).

[6] Xu, L., Ren, J. S., Liu, C., & Jia, J. (2014). Deep convolutional neural network for image deconvolution. In Advances in Neural Information Processing Systems (pp. 1790-1798).

[7] Mehta, J., & Majumdar, A. (2017). RODEO: robust DE-aliasing autoencOder for real-time medical image reconstruction. Pattern Recognition, 63, 499-510.

[8] Agostinelli, F., Anderson, M. R., & Lee, H. (2013). Adaptive multi-column deep neural networks with application to robust image denoising. In Advances in Neural Information Processing Systems (pp. 1493-1501).

[9] Xie, J., Xu, L., & Chen, E. (2012). Image denoising and inpainting with deep neural networks. In Advances in neural information processing systems (pp. 341-349).

[10] Burger, H. C., Schuler, C. J., & Harmeling, S. (2012, June). Image denoising: Can plain neural networks compete with BM3D?. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on (pp. 2392-2399). IEEE.

[11] Kulkarni, K., Lohit, S., Turaga, P., Kerviche, R., & Ashok, A. (2016). Reconnet: Non-iterative reconstruction of images from compressively sensed measurements. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 449-458).

[12] Xu, K., & Ren, F. (2016). CSvideonet: A recurrent convolutional neural network for compressive sensing video reconstruction. arXiv preprint arXiv:1612.05203.

[13] Metzler, C.A., Mousavi, A. and Baraniuk, R.G., 2017. Learned D-AMP: A Principled CNN-based Compressive Image Recovery Algorithm. arXiv preprint arXiv:1704.06625.

[14] Zeng, K., Yu, J., Wang, R., Li, C., & Tao, D. (2017). Coupled deep autoencoder for single image super-resolution. IEEE transactions on cybernetics, 47(1), 27-37.

[15] Riggan, B. S., Reale, C., & Nasrabadi, N. M. (2015). Coupled auto-associative neural networks for heterogeneous face recognition. IEEE Access, 3, 1620-1632.

[16] Gupta, K., Bhowmick, B. & Majumdar, A. (2017). Motion Blur Removal via Coupled Autoencoder In Proceedings of the IEEE International Conferenece on Image processing.

[17] Huang, D. A., & Wang, Y. C. F. (2013, December). Coupled dictionary and feature space learning with applications to cross-domain image synthesis and recognition. In Computer Vision (ICCV), 2013 IEEE International Conference on (pp. 2496-2503). IEEE.

[18] Wang, S., Zhang, L., Liang, Y., & Pan, Q. (2012, June). Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on (pp. 2216-2223). IEEE.

[19] Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., & Manzagol, P. A. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. Journal of Machine Learning Research, 11(Dec), 3371-3408.

[20] Mehrotra, R., Chu, D., Haider, S. A., & Kakadiaris, I. A. Towards Learning Coupled Representations for Cross-Lingual Information Retrieval.

[21] Boyd, S., Parikh, N., Chu, E., Peleato, B., & Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, *3*(1), 1-122.

[22] Dong, C., Loy, C. C., He, K., & Tang, X. (2014, September). Learning a deep convolutional network for image super-resolution. In European Conference on Computer Vision (pp. 184-199). Springer, Cham.

[23] Mousavi, A., Patel, A. B., & Baraniuk, R. G. (2015, September). A deep learning approach to structured signal recovery. In *Communication, Control, and Computing (Allerton), 2015 53rd Annual Allerton Conference on* (pp. 1336-1343). IEEE.

[24] Li, C., Yin, W., Jiang, H., & Zhang, Y. (2013). An efficient augmented Lagrangian method with applications to total variation minimization. *Computational Optimization and Applications*, *56*(3), 507-530.

[25] Dong, W., Shi, G., Li, X., Ma, Y., & Huang, F. (2014). Compressive sensing via nonlocal low-rank regularization. *IEEE Transactions on Image Processing*, *23*(8), 3618-3632.

[26] Metzler, C. A., Maleki, A., & Baraniuk, R. G. (2016). From denoising to compressed sensing. *IEEE Transactions on Information Theory*, *62*(9), 5117-5144.

ISBN 978-90-827970-1-5 © EURASIP 2018