Compressed Learning: A Deep Neural Network Approach

Amir Adler
McGovern Institute for Brain Research
MIT, Cambridge, MA
adleram@mit.edu

Eve Zisselman
Electrical Engineering Department
Technion, Haifa 32000, Israel
eve.zisselman@gmail.com

Michael Elad
Computer Science Department
Technion, Haifa 32000, Israel
elad@cs.technion.ac.il

***Abstract*—Compressed Learning (CL) is a joint signal processing and machine learning framework for inference from a signal, using a small number of measurements. This paper presents an end-to-end deep learning approach for CL, in which a network composed of fully-connected layers followed by convolutional layers perform the linear sensing and non-linear inference stages. During training, the sensing matrix and the inference operator are *jointly* optimized, leading to a significant advantage compared to existing methods. For example, at a sensing rate of 1% (only 8 measurements of 28×28 pixels images), the classification error for the MNIST handwritten digits dataset is 6.46% compared to 41.06% with CL state-of-the-art.**

***Index Terms*—compressed learning, deep learning.**

I. Introduction

CL [1] is a mathematical framework that combines Compressed Sensing (CS) [2], [3] with machine learning. CL has diverse applications including image classification [4], reconstruction-free action recognition [5], [6], [7], acquisition of dynamic scenes [8], least-squares regression [9], watermark detection [10], prediction of protein–protein interactions [11], targets classification [12], [13], and hyperspectral image analysis [14]. The theoretical study in [1] revealed that direct inference from CS measurements is feasible with high classification accuracies. In particular, this work provided analytical results for training a linear Support Vector Machine (SVM) classifier in the CS domain[1] $\mathbf{y} = \Phi\mathbf{x}$, and it was proved that under certain conditions the performance of a linear SVM classifier operating in the CS domain is almost equivalent to the performance of the best linear threshold classifier operating in the signal domain. A different approach, termed *smashed filters*, was presented in [12], in which a generalized maximum-likelihood criterion was employed to design matched filters for target detection in the measurements domain of compressive cameras such as the single-pixel camera [15]. This approach was extended and termed *smashed correlation filters* for activity recognition by [5], and for face recognition by [6]. A deep neural network (DNN) approach was introduced by [4], in which random and Hadamard sensing matrices were employed for image classification in the CS domain. This work utilized convolutional networks that operate on the image domain, and used the following projected measurement vector as the input to the network[2]:

$$\mathbf{z} = \Phi^T \mathbf{y} \in \mathbf{R}^N. \tag{1}$$

By training a network similar to LeNet [16] for classifying MNIST digits images, and using the projected measurement $\mathbf{z}$ rather than the true image $\mathbf{x}$, outstanding classification results were obtained by [4], which significantly outperform the smashed filters approach at sensing rates as low as $R = 0.01$. This approach was also successfully verified for the challenging task of classifying a subset of the ImageNet dataset, consisting of 1.2 million images and 1,000 categories.

[1] $\Phi \in \mathbf{R}^{M \times N}$ is the sensing matrix, $\mathbf{x} \in \mathbf{R}^N$ is the signal, and $R = M/N$ is the sensing rate ($M \ll N$).

[2] Instead of the true image, and after reshaping it to $\sqrt{N} \times \sqrt{N}$ pixels.

II. The Proposed Approach

Our approach provides an end-to-end DNN solution to CL, in which the sensing matrix Φ is *jointly* optimized with the inference operator. Our choice is motivated by the outstanding success of convolutional networks for the task of compressive image classification [4], which employed a random sensing matrix (with Gaussian entries) for classifying the MNIST [16] dataset, and a Hadamard matrix for classifying a subset of the ImageNet dataset. In our approach, the first layer learns and performs the sensing matrix $\widetilde{\Phi}$ stage, and the subsequent layers perform the non-linear inference stage. Note that the second fully-connected layer performs a similar operator to (1), however, a different matrix $\Psi \in \mathbf{R}^{N \times M}$ is learned. Once the network is trained, the sensing stage (i.e. the first hidden layer) can be detached from the subsequent inference layers, into two separate elements of a CL system. The proposed DNN architecture includes the following layers[3]: (1) an input layer with N nodes; (2) a CS fully-connected layer with NR nodes, $R \ll 1$ (its weights form the sensing matrix); (4) a fully-connected re-projection layer that expands the output of the sensing layer to the original image dimensions N; (6) a convolution layer with kernel sizes of 5×5, and 6 feature maps; (8) maxpooling layer which selects the maximum of 2×2 feature maps elements, with a stride of 2 in each dimension; (9) a convolution layer with kernel sizes of 5×5, and 16 feature maps; (11) maxpooling layer which selects the maximum of 2×2 feature maps elements, with a stride of 2 in each dimension; (12) reshape operator that reshapes the 16 4×4 max-pooled features maps into a 256-dimensional vector; (13) a fully connected layer of 256 to 120 nodes; (15) a fully connected layer of 120 to 84 nodes; and (17) a SoftMax layer with 10 outputs. We have trained[4] the proposed network from the training images of the MNIST dataset, using the Stochastic Gradient Descent algorithm with a learning rate of 0.0025 and 100 epochs. Classification error performance was evaluated for sensing rates in the range of $R = 0.01$ to $R = 0.25$, and averaged over the 10,000 MNIST test images. Classification error results are summarized in Table I, and reveal a consistent advantage of the proposed approach, which increases significantly for lower sensing rates.

III. Conclusions

This paper presents a novel deep learning approach to CL, in which the sensing matrix and the non-linear inference operator are jointly optimized. This approach is demonstrated to outperform state-of-the-art CL, which employs a random sensing matrix followed by a convolutional network, for the task of image classification. The proposed approach can be extended to numerous CL applications, such as detection and recognition of patterns in single- and multi-channel images and signals.

[3] The 2^{nd} linear layer and convolutional layers are followed by ReLU [17].

[4] The network was implemented in Torch7 [18], and trained on NVIDIA Titan X GPU card. A software package for reproducing the results is available at: http://www.cs.technion.ac.il/~adleram/CL_DNN_2016.zip

TABLE I: Classification Error (%) for the MNIST handwritten digits dataset vs. sensing rate $R = M/N$ (averaged over 10,000 test images):

Sensing Rate	No. of Measurements	Smashed Filters [12]	Random Sensing + CNN [4]	Proposed
0.25	196	27.42%	1.63%	**1.56%**
0.1	78	43.55%	2.99%	**1.91%**
0.05	39	53.21%	5.18%	**2.86%**
0.01	8	63.03%	41.06%	**6.46%**

REFERENCES

[1] Robert Calderbank, Sina Jafarpour, and Robert Schapire, "Compressed learning: Universal sparse dimensionality reduction and learning in the measurement domain," Tech. Rep., 2009.

[2] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[3] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, March 2008.

[4] S. Lohit, K. Kulkarni, and P. Turaga, "Direct inference on compressive measurements using convolutional neural networks," in *2016 IEEE International Conference on Image Processing (ICIP)*, Sept 2016, pp. 1913–1917.

[5] K. Kulkarni and P. Turaga, "Reconstruction-free action inference from compressive imagers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 4, pp. 772–784, April 2016.

[6] S. Lohit, K. Kulkarni, P. Turaga, J. Wang, and A. C. Sankaranarayanan, "Reconstruction-free inference on compressive measurements," in *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, June 2015, pp. 16–24.

[7] K. Kulkarni and P. Turaga, "Recurrence textures for human activity recognition from compressive cameras," in *2012 19th IEEE International Conference on Image Processing*, Sept 2012, pp. 1417–1420.

[8] Aswin C. Sankaranarayanan, Pavan K. Turaga, Richard G. Baraniuk, and Rama Chellappa, "Compressive acquisition of dynamic scenes," in *Computer Vision – ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part I*, 2010, pp. 129–142.

[9] Odalric Maillard and Rémi Munos, "Compressed least-squares regression," in *Advances in Neural Information Processing Systems*, 2009, pp. 1213–1221.

[10] Qia Wang, Wenjun Zeng, and Jun Tian, "A compressive sensing based secure watermark detection and privacy preserving storage framework," *IEEE Transactions on Image Processing*, vol. 23, no. 3, pp. 1317–1328, 2014.

[11] Ya-Nan Zhang, Xiao-Yong Pan, Yan Huang, and Hong-Bin Shen, "Adaptive compressive learning for prediction of protein–protein interactions from primary sequence," *Journal of theoretical biology*, vol. 283, no. 1, pp. 44–52, 2011.

[12] Mark A. Davenport, R Marco F. Duarte, R Michael B. Wakin, C Jason N. Laska R, Dharmpal Takhar, R Kevin F. Kelly, and R Richard G. Baraniuk R, "The smashed filter for compressive classification and target recognition," in *Proc. SPIE Symposium on Electronic Imaging: Computational Imaging*, 2007, p. 6498.

[13] Yun Li, Chinmay Hegde, Aswin C Sankaranarayanan, Richard Baraniuk, and Kevin F Kelly, "Compressive image acquisition and classification via secant projections," *Journal of Optics*, vol. 17, no. 6, pp. 065701, 2015.

[14] Jürgen Hahn, Simon Rosenkranz, and Abdelhak M Zoubir, "Adaptive compressed classification for hyperspectral imagery," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 1020–1024.

[15] J. Romberg, "Imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 14–20, March 2008.

[16] Yann Lecun, Leon Bottou, Yoshua Bengio, and Patrick Haffner, "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, 1998, pp. 2278–2324.

[17] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*. 2010, pp. 807–814, Omnipress.

[18] R. Collobert, K. Kavukcuoglu, and C. Farabet, "Torch7: A matlab-like environment for machine learning," in *BigLearn, NIPS Workshop*, 2011.