RandNet: Deep Learning with Compressed Measurements of Images

Thomas Chang*, Bahareh Tolooshams*, and Demba Ba
School of Engineering and Applied Sciences, Harvard University, Cambridge, MA

HARVARD
John A. Paulson
School of Engineering
and Applied Sciences

Introduction

We introduce a class of neural networks, termed **RandNet**, for learning representations using **compressed random measurements** of data.

RandNet, a ReLU auto-encoder, is specifically designed for training deep networks from **random projections** of the original data.

Training with compressed data offers **memory** efficiency, and in the case of **sparse measurements**, it provides **computational** efficiency.

In **unsupervised** settings, RandNet performs **dictionary learning** using compressed data. In **supervised** settings, RandNet offers minimal loss in classification accuracy.

Dictionary Learning (DL)

- Generative model of the signal $\mathbf{y}^j \in \mathbb{R}^N$

$$\mathbf{y}^j = \mathbf{A}\mathbf{x}^j + \mathbf{v}^j$$

where $\mathbf{A} \in \mathbb{R}^{N \times p}$ (dictionary), $\mathbf{x}^j \in \mathbb{R}^p$ (sparse code).

- Optimization problem

$$\min_{\substack{(\mathbf{x}^j)_{j=1}^J \\ \mathbf{A}}} \sum_{j=1}^{J} \frac{1}{2}\|\mathbf{y}^j - \mathbf{A}\mathbf{x}^j\|^2 + \lambda \|\mathbf{x}^j\|_1 \;\; \text{s.t.} \;\; \|\mathbf{a}_i\|_2 = 1$$

- Sparse coding step

$$\mathbf{x}^{j(l)} = \arg\min_{\mathbf{x}^j} \frac{1}{2}\|\mathbf{y}^j - \mathbf{A}^{(l-1)}\mathbf{x}^j\|^2 + \lambda \|\mathbf{x}^j\|_1$$

- Dictionary update step

$$\mathbf{A}^{(l)} = \arg\min_{\mathbf{A}} \sum_{j=1}^{J} \frac{1}{2}\|\mathbf{y}^j - \mathbf{A}\mathbf{x}^{j(l)}\|^2 \quad \text{s.t.} \;\; \|\mathbf{a}_i\|_2 = 1$$

Compressed DL

- Given compressed measurements

$$\mathbf{r}^j = \mathbf{\Phi}\mathbf{y}^j$$

where $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ is a measurement matrix with measurement ratio $\beta = \frac{M}{N}$ and compression factor $\gamma = \beta s$ where s is sparsity of rows of $\mathbf{\Phi}$.

- Optimization problem

$$\min_{\substack{(\mathbf{x}^j)_{j=1}^J \\ \mathbf{A}}} \sum_{j=1}^{J} \frac{1}{2}\|\mathbf{r}^j - \mathbf{\Phi}\mathbf{A}\mathbf{x}^j\|^2 + \lambda \|\mathbf{x}^j\|_1 \quad \text{s.t.} \;\; \|\mathbf{a}_i\|_2 = 1$$

References

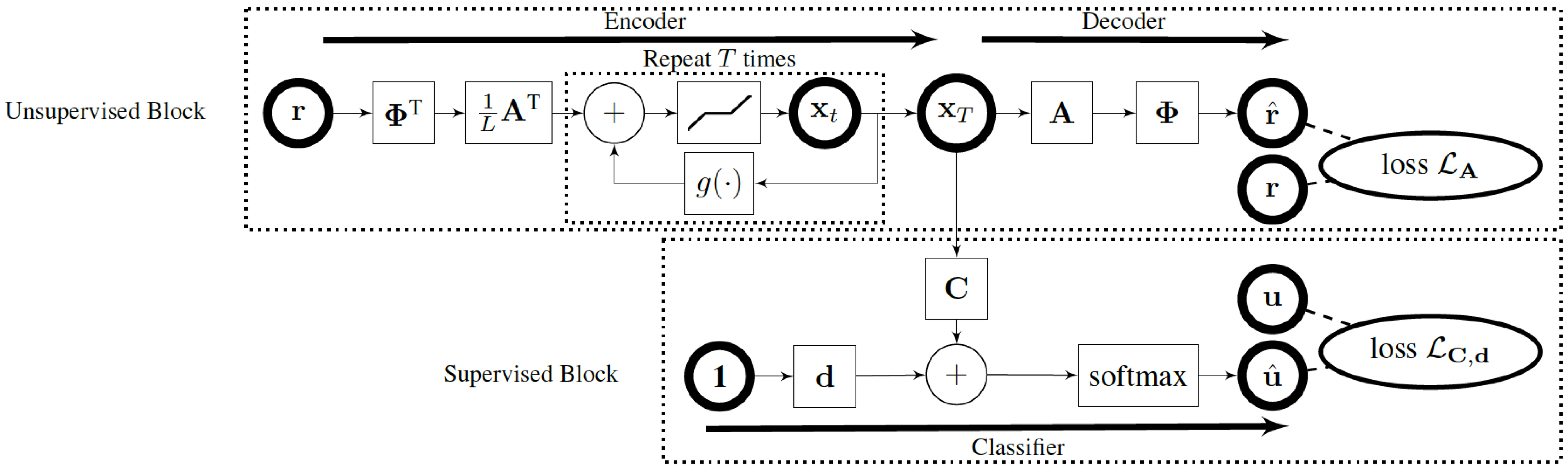
[1] B Tolooshams, S Day, and D Ba, "Scalable convolutional dictionary learning with constrained recurrent sparse auto-encoders," in Proc. IEEE 28th International Workshop on Machine Learning for Signal Processing, pp. 1-6, 2018.

[2] F. Pourkamali Anaraki and S. M. Hughes, "Compressive k-svd," in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 5469-5473, 2013.

[3] J T Rolfe and Y LeCun, "Discriminative recurrent sparse auto-encoders," in Proc. International Conference on Learning Representations, pp. 1-15, 2013.

[4] J Mairal, J Ponce, G Sapiro, A Zisserman, and F R Bach, "supervised dictionary learning," in Proc. Advances in Neural Information Processing Systems 21, pp. 1033-1040, 2009.

RandNet Architecture



$$u_k = \begin{cases} 1, & \text{if } k = c; \\ 0, & \text{otherwise.} \end{cases}$$

$$\mathcal{L}_{\mathbf{A}}(\mathbf{r}, \hat{\mathbf{r}}) = \tfrac{1}{2}\|\mathbf{r} - \hat{\mathbf{r}}\|_2^2$$

$$\mathcal{L}_{\mathbf{C},\mathbf{d}}(\mathbf{x}_T, \mathbf{u}, \mathbf{C}, \mathbf{d}) = -\mathbf{u}^{\mathrm{T}} \log \left(\frac{e^{\mathbf{C}\mathbf{x}_T + \mathbf{d}}}{\sum_i e^{(\mathbf{C}\mathbf{x}_T + \mathbf{d})_i}} \right)$$

		(a) Gaussian	(b) Sparse	(c) Identity
Memory Storage		$\boldsymbol{O(\beta N)}$	·	$O(N)$
Memory Access		·	$\boldsymbol{O(\gamma N)}$	$O(N)$
Matrix Operation	$\mathbf{\Phi}^{\mathrm{T}}$	$O(MN)$	$O(\gamma N)$	·
	$\mathbf{A}^{\mathrm{T}}$	$O(pN)$	$\boldsymbol{O(\gamma pN)}$	$O(pN)$

Table 1. Memory and computational efficiency of RandNet when using (a) Gaussian measurements, (b) sparse random measurements to compress images, and (c) no compression.

Supervised Learning - Classification on MNIST

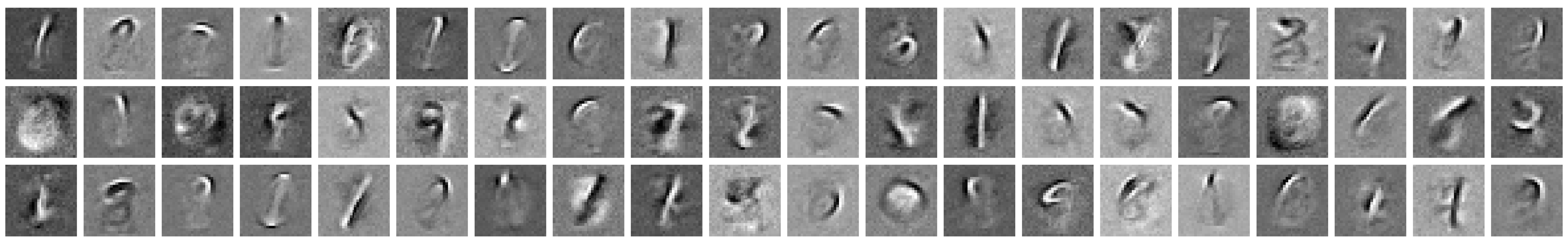


Figure 1: The 60 columns of the learned dictionary $\hat{\mathbf{A}}$ when $\mathbf{\Phi}$ is Gaussian with $\beta = 0.5$.

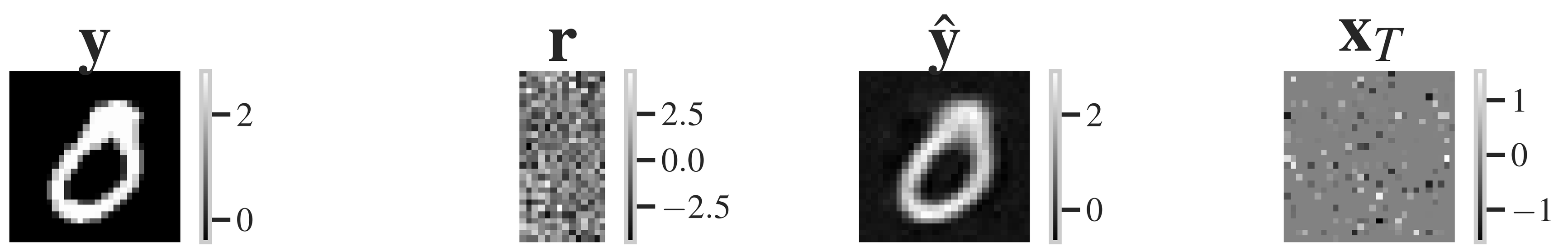


Figure 2: RandNet applied to an instance of digit 0. Left: digit in its original and compressed dimensions. Right: reconstruction and sparse representation of the digit. Note that $\mathbf{x}_T$ is highly sparse (most entries close to zero).

	RandNet	CK-SVD	DrSAE	SDL
Error Rate [%]	**(G) 1.56**	(G) 3.72	1.08	1.05
	(S) 3.16	(S) 5.20		

Table 2. MNIST classification error [%] on test dataset. (G) stands for Gaussian and (S) for row-sparse projection.

Unsupervised Learning - Dictionary Learning Performance

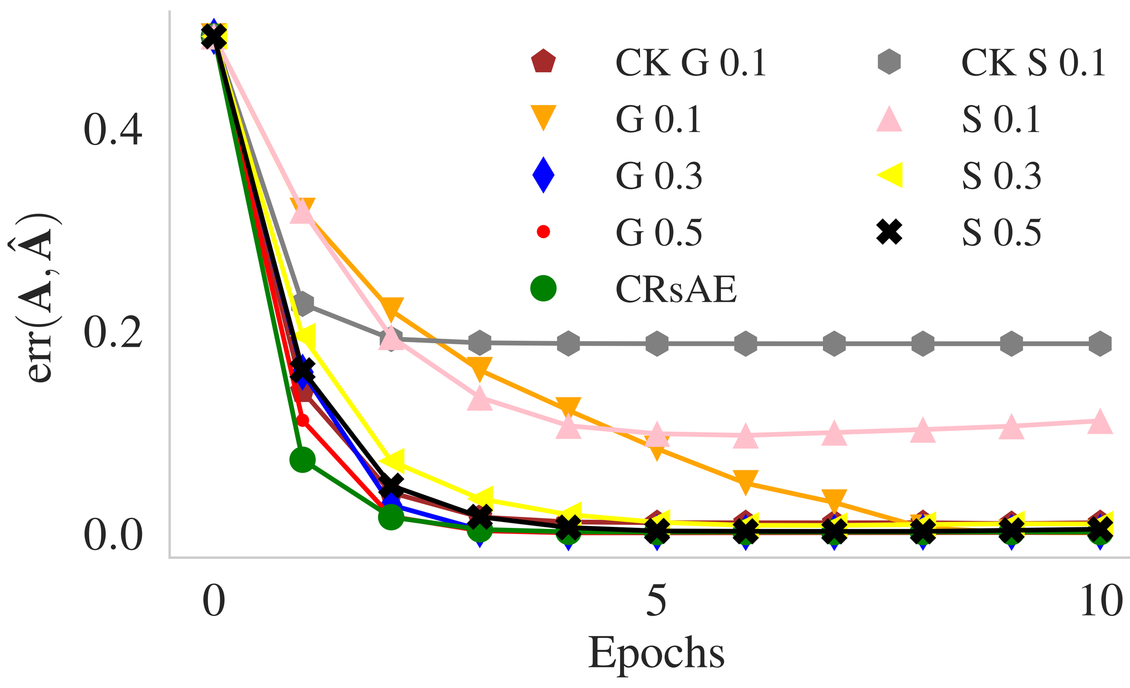


Figure 3: Error $\mathrm{err}(\mathbf{A}, \hat{\mathbf{A}})$ for RandNet, CRsAE, and Compressive K-SVD. "G 0.1" stands for Gaussian $\mathbf{\Phi}$ with $\beta = 0.1$. "S 0.5" stands for sparse $\mathbf{\Phi}$ with $\beta = 0.5$.

$$\mathrm{err}(\mathbf{A}, \hat{\mathbf{A}}) = \max_i \left(\sqrt{1 - \frac{\langle \mathbf{a}_i, \hat{\mathbf{a}}_i \rangle^2}{\|\mathbf{a}_i\|_2^2 \|\hat{\mathbf{a}}_i\|_2^2}} \right)$$