Compressed Classification from Learned Measurements

Robiulhossain Mdrafi and Ali Cafer Gurbuz
Department of Electrical and Computer Engineering
Mississippi State, MS 39762
{rm2232,gurbuz}@ece.msstate.edu

Abstract

This work proposes a deep compressed learning framework inferring classification directly from the compressive measurements. While classical approaches separately sense, reconstruct signals, and apply classification on these reconstructions, we jointly learn the sensing and classification schemes utilizing a deep neural network with a novel loss function. Our approach employs a data-driven reconstruction network within the compressed learning framework utilizing a weighted loss that combines both in-network reconstruction and classification losses. The proposed network structure also learns the optimal measurement matrices for the goal of enhancing classification performance. Quantitative results demonstrated on CIFAR-10 image dataset show that the proposed framework provides better classification performance and robustness to noise compared to the tested state of the art deep compressed learning approaches.

1. Introduction

Typical approaches transform the signals to a sparse domain after acquiring them at a high rate, at least as directed by the Nyquist rate for further processing. Compressive sensing (CS) [3, 7, 11] proposes to acquire linear projections of the signal onto a lower dimensional space to show that if the inner distances of signals are not disturbed much by this projection, the original signals can still be reconstructed. CS theory provides the theoretical guarantees, as well as the algorithms, for successful reconstruction of these sparse signals from small number of linear projections.

The final goal in many area is not to reconstruct the signal but rather apply a signal processing task such as detection, estimation or classification. For compressively sensed signals, a typical implementation of such an effort requires a two stage approach. First, the signal is reconstructed and second, the inference is done on the reconstructed signals. This two-stage approach although allowing already developed inference techniques to be easily used on reconstructed signal domains, it requires the burden of heavy computational complexity due to the nonlinear reconstruction stage of CS. Hence, approaches that will allow inference and learning directly from compressed data domain with computational advantages are utmost important.

Direct inference in CS framework is not entirely a novel concept where the effect of inference from compressive measurements is also studied in parallel with the introduction of CS. In [13], CS projects is studied for M-ary hypothesis testing. The work in [8] introduced smashed filters to show that we can perform classification task in CS settings if we statisfy Johnson-Lindenstrauss Lemma [10, 26]. After the evolution of smashed filters in CS-based classification, the term 'compressed learning' (CL) was introduced in [6] where it was theoretically shown a linear support vector machine (SVM) classifier operating on the compressed domain performs almost as well as the best linear classifier operating on the original signal space. In [5], a family of deterministic CS measurement matrices (MM) are analyzed and presented in terms of CL results. The work in [9] provides some guarantees on detection, classification, estimation, and filtering taks in compressed domain for a wide variety of signal classes.

The deep neural networks (DNN) based CL approaches [2, 19, 31] utilizes compressive measurements obtained with a given MM to generate a proxy image as the input to a convolutional neural network (CNN) followed by fully connected (FC) and softmax layers for the classification task. In [2, 31], the DNN uses two FC layers to denote the sensing MM and adjoint operator prior making the proxy image to learn the MMs where the work in [19] uses a fixed Gaussian MM. In [28], several updates are proposed on the network of [2] including new ReLU activations, dropout, and a regularized loss function including mutual coherence of the learned measurement matrix. However, the classification performance of the existing DNN based CL approaches is low due to the quality of the proxy image used as the input to their classification networks.

Motivated by the successful enhanced results on utilizing data driven deep learning based structures to reconstruct signals from their compressed measurements [17,18,20–22, 30] to learn the mapping from low dimensional data space to the original signal space for the given signal while also providing much lower reconstruction time given a trained model, this paper proposes a new deep joint compressed learning (DJCL) framework that incorporates a deep reconstruction network within the CL structure to optimize a novel weighted loss function that combines classification and reconstruction losses driving the learning for the full network. While inferring directly from the compressed measurements, the DJCL can also derive optimal set of measurements for the classification task. The proposed network is trained and tested using CIFAR-10 [16] dataset. The obtained results are compared with CL based approaches as well as separately reconstructing images and applying classification on the reconstructed images. Our results show that proposed deep joint compressed learning framework performs better than the state of the art deep CL approaches surpassing separate reconstruction and classification levels.

The main contributions of this paper can be listed as follows:

- A new deep compressed learning framework is proposed for direct inference from lower number of linear measurements jointly utilizing DNN structures for the learning of measurements, reconstruction, and classification schemes.

- A novel loss function that is a weighted combination of cross-entropy and mean square error is proposed to train the combined deep network structure. This novel loss results in a higher classification accuracy than using only cross-entropy for the network for compressed learning framework.

- The proposed DJCL structure is a general framework that is flexible enough to employ most of the existing high performing reconstruction and classification networks within its structure.

- Task dependent measurement matrices are learned for the specific goal of classification.

- Detailed performance analysis in comparison to state of the art deep CL approaches are provided for a variety of measurement rates.

The rest of the paper is organized as follows: The proposed DJCL framework is detailed in Section 2. The dataset, experimental settings, and training and testing results of the proposed method with the compared approaches have been presented in Section 3. Finally, conclusions are drawn in Section 4.

2. Proposed Learning Structure

The proposed deep compressed learning network structure along with recent DNN based CL techniques [19, 31] has been illustrated in Figure 1. It can be seen that the proposed model differs from existing approaches in several fronts. First, while existing approaches only generate a proxy image through a single FC layer, proposed approach utilizes a reconstruction network and produces an enhanced input to the classification network. Second, existing approaches train network with only a cross entropy loss while we propose a weighted loss of reconstruction and classification costs to train the proposed model. The proposed DJCL framework is flexible to use most of the existing reconstruction or classification networks within and it jointly utilizes its sensing, reconstruction, and classification stages as detailed in the following subsections:

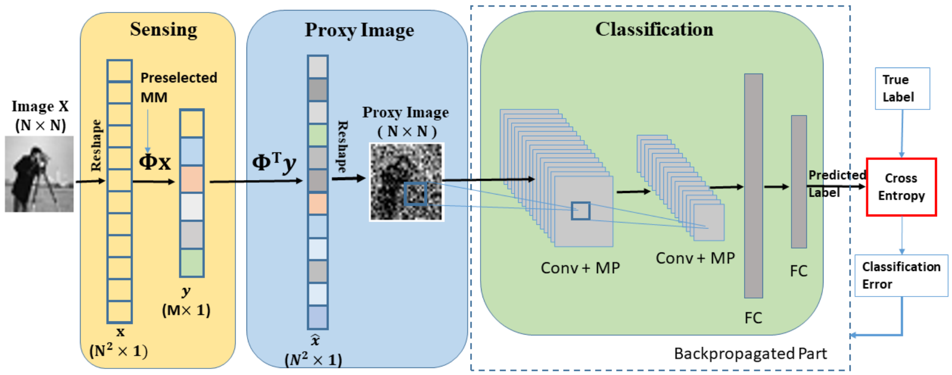
2.1. Sensing: Learning Measurements for Classification

The sensing system in the proposed DJCL framework as shown in Figure 1(c) utilizes a reshaping and a fully connected layer (FC_1) with linear activations modelling the sensing system to acquire the data from the original signal to the compressed domain. Since only linear activations are used in FC_1 layer, the output of FC_1 fully models the $\mathbf{y} = \mathbf{\Phi x}$, CS data acquisition process where entries of MM, $\mathbf{\Phi}$ are the weights used in FC_1 layer where $\mathbf{x}$ is obtained from via reshaping the original signal e.g. image $\mathbf{X}$. Learning the parameters of this layer corresponds to learning a linear MM suitable for CS for the task of classification.

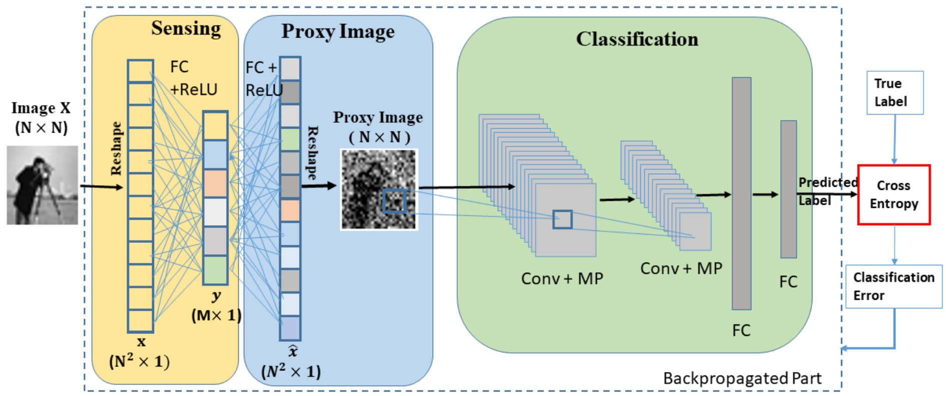
2.2. Reconstruction Network

Reconstruction network is the second part of the proposed DJCL framework where we use the compressed measurements $\mathbf{y}$ from the sensing stage to reconstruct a signal estimate $\widehat{\mathbf{X}}$, which will be fed as an input to the classification network. The DNN based CL in [19] only uses adjoint operators $\mathbf{\Phi^T}$ to generate proxy images by reshaping $\hat{\boldsymbol{x}} = \mathbf{\Phi^T}\boldsymbol{y}$ as inputs to the following classification stages. Learning in [19] is only in the classification network where this pseudo image is used as the input. In [2, 31], a single FC layer with ReLu activation is used to imitate the adjoint operator by learning to create a pseudo images.

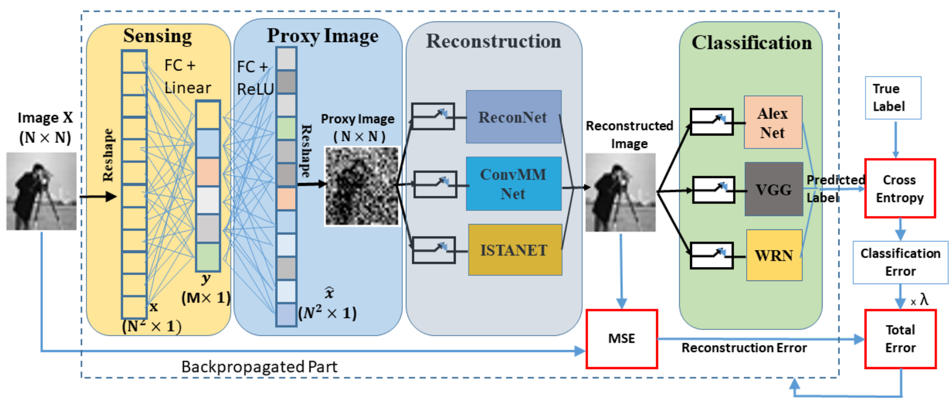
However, in recent years, studies on DL based signal recovery from compressed measurements led to several successful DNN structures [17, 18, 22–24, 27, 30]. These DNN models show enhanced signal recovery performance for the class of signals as they are trained on with much less computational complexity compared to classical CS recovery approaches. We propose to utilize a DL based reconstruction network that maps the compressed measurements to the original signal domain in our CL framework as opposed to



(a)



(b)



(c)

Figure 1: CL block diagram for (a) [19], (b) [31] (c) Proposed Method

a single FC layer. For this goal, we specifically focus on three of the recent and comparably successful reconstruction networks; ReconNet [17, 18], IstaNet [30], and ConvMMNet [22]. Each of these reconstruction networks differ by their DNN architectures on how they perform reconstruction from the given compressed measurements. While ReconNet utilizes convolutional layers working on initial proxy images, ConvMMNet uses a cascade of an FC layer with multistage convolutional and ReLU layers trained using a weighted Euclidean loss jointly learning the measurements as well as the reconstruction scheme. ISTA-Net unrolls the iterative shrinkage-thresholding (ISTA) [4] algorithm into a multistage DNN. In general, these networks are trained with minimizing the Euclidean loss defined as the average squared reconstruction error as in (1)

$$L_R(\mathbf{\Theta}) = \frac{1}{T}\sum_{i=1}^{T} \|f(\boldsymbol{y_i}, \mathbf{\Theta}) - \mathbf{X_i}\|_F. \tag{1}$$

where T is the total number of training samples, and $f(\boldsymbol{y_i}, \mathbf{\Theta})$ is the reconstruction network model with parameters $\mathbf{\Theta}$ and the input compressed measurement samples $\boldsymbol{y_i}$. In addition to the Euclidean loss, an adversarial loss is used within a generative adversarial network (GAN) in [18] in ReconNet and ConvMMNet models. More details for network structures of each approach can be found in more detail in their respective publications.

In this study, we utilize parts of these reconstruction network models after the proxy image generation within our framework. Not all but one of these reconstruction networks can be selected to be used in final DJCL model. The goal is to create an enhanced representation of the signal as the input to the next classification stage in the network. Simulation results presented in section 3 show that the output of the reconstruction networks have significantly higher peak signal to noise ratio levels compared to proxy images generated in compared CL approaches.

Although the specified reconstruction networks are tested and compared in this study, the general framework we propose is flexible so that other reconstruction network models can also be utilized instead of the tested networks. The generated image from the reconstruction network will be the input to the next stage of the DJCL framework, which is the classification network.

2.3. Classification Network

The final stage of our CL framework is the classification part. For this stage, we utilize one of the existing state-of-the-art classification networks such as AlexNet [16], VGG [25], or Wide residual Network (WRN) [29]. AlexNet offers a baseline for DNN based object classification. It uses five convolutional layers, followed by FC layers with ReLU activations. While VGG network utilizes convolutional, pooling, and FC layers like AlexNet, it uses smaller filters with increased depth. In this work, we opt to use and compare VGG-3 blocks. For compatibility with the input size of the images from utilized dataset, we use convolutional filters of size 32, 64, and 128 respectively in consecutive layers prior to max-pooling operation. WRN is an extension of residual network (RESNet) [14] utilizing skip connections and residual blocks. In this study, a WRN of depth 28 and width 10 is utilized with the same architecture as presented in [29].

For experimental purposes, we have employed a publicly available dataset, namely CIFAR-10 [16] to train, validate, and test the proposed DJCL and the compared approaches. This dataset has been extensively used to produce the state-of-the-art results for different kinds of computer vision task [15]. The details of the dataset with the class information is given in the Section 3. Next section describes the novel loss function for training the combined sensing, reconstruction, and classification networks.

2.4. CL with Weighted Loss function

In the proposed CL approach, we jointly learn a MM that maps an original signal $\boldsymbol{x_i}$ to compressed measurements $\boldsymbol{y_i} = \mathbf{\Phi}_{\boldsymbol{S}}\boldsymbol{x_i}$ and an inference network mapping $\boldsymbol{y_i}$ to a class label $\boldsymbol{\ell_i}$ over a training set of T samples. Learning the parameters of both sensing and inference networks can be done jointly through solving an optimization problem minimizing a defined loss. One possible such loss that is also utilized in literature for CL [2, 19, 31] is the one that can measure only the distance between the predicted and true class labels through employing a negative log-likelihood function. For this case, the network parameters can be learned by solving the following optimization problem:

$$\{\widehat{\mathbf{\Phi}}_{\boldsymbol{s}}, \widehat{\mathbf{\Theta}}_{\boldsymbol{R}}, \widehat{\mathbf{\Theta}}_{\boldsymbol{C}}\} = \underset{\mathbf{\Phi}_{\boldsymbol{s}}, \mathbf{\Theta}_{\boldsymbol{R}}, \mathbf{\Theta}_{\boldsymbol{C}}}{\arg\min}\ \mathcal{L}_C(f_{\mathbf{\Theta}_{\boldsymbol{C}}}(f_{\mathbf{\Theta}_{\boldsymbol{R}}}(\mathbf{\Phi}_{\boldsymbol{s}}\boldsymbol{x_i})), \ell_i) \tag{2}$$

where $f_{\mathbf{\Theta}_{\boldsymbol{R}}}(\mathbf{\Phi}_{\boldsymbol{s}}\boldsymbol{x_i})$ is the model for the reconstruction network with parameters Θ_R, $f_{\mathbf{\Theta}_{\boldsymbol{C}}}(\cdot)$ is the model for the classification network with parameters $\mathbf{\Theta}_{\boldsymbol{C}}$. The classification loss function $\mathcal{L}_C$ is the cross entropy loss that is defined as

$$\mathcal{L}_C(\widehat{\ell_i}, \ell_i) = -\sum_{i=1}^{T}\sum_{c=1}^{C} \ell_{i,c} log S(\hat{\ell}_{i,c}). \tag{3}$$

where $S(\hat{\ell}_{i,c})$ is the soft-max layer output that gives the probability that sample i belongs to class c.

Minimizing the cross entropy loss in (2) is a natural selection, since the final goal of the CL is to obtain the best classification performance. However, we propose and show in our simulation results that reconstructing a better image estimate as the input of the classification network also increases the classification performance. Only employing the optimization in (2) does not directly force the DNN structure to create better reconstruction outputs as a middle product of the whole network. Hence, we utilize a hybrid loss that incorporates a weighted combination of the reconstruction and classification losses. The goal by injecting the reconstruction loss into the total loss is to force the reconstruction network to generate better image estimates that will lead to enhanced classification performance. Thus, the proposed CL approach solves the following minimization problem

$$\{\widehat{\mathbf{\Phi}}_{\boldsymbol{s}}, \widehat{\mathbf{\Theta}}_{\boldsymbol{R}}, \widehat{\mathbf{\Theta}}_{\boldsymbol{C}}\} = \underset{\mathbf{\Phi}_{\boldsymbol{s}}, \mathbf{\Theta}_{\boldsymbol{R}}, \mathbf{\Theta}_{\boldsymbol{C}}}{\arg\min}\ \mathcal{L}_T \tag{4}$$

where the total loss $\mathcal{L}_T$ is

$$\mathcal{L}_T = \mathcal{L}_R(f_{\mathbf{\Theta}_{\boldsymbol{R}}}(\boldsymbol{y_i}), \boldsymbol{x_i}) + \lambda\mathcal{L}_C(f_{\mathbf{\Theta}_{\boldsymbol{C}}}(f_{\mathbf{\Theta}_{\boldsymbol{R}}}(\mathbf{\Phi}_{\boldsymbol{s}}\boldsymbol{x_i})), \ell_i) \tag{5}$$

In (5), $\mathcal{L}_R$ is the mean squared reconstruction loss defined in (1), λ is the hyper-parameter that defines the ratio between $\mathcal{L}_C$ and $\mathcal{L}_R$ losses. The parameter λ can be selected over a validation set as shown in Section 3.

Note that the joint learning framework learns both a MM and an inference network including reconstruction and classification parts in the training phase. The learned MM can be detached from the combined network and it can be utilized to sense the signals. It can be seen that different MM can be learned for different purposes such as classification or reconstruction minimizing different type of losses such as $\mathcal{L}_C$, $\mathcal{L}_R$ or the total loss $\mathcal{L}_T$. After the sensing network is detached, the remaining DNN is the CL network implementing inference directly from the compressed measurements.

3. Simulation Results

In this section, we provide simulation results and quantitative analysis of the proposed DJCL framework compared with the existing deep CL approaches in [19, 31].

3.1. DataSet, Evalaution Metrics, and Learning Parameters

In this work, CIFAR-10 [16] dataset is used for simulation and quantitative analysis. CIFAR-10 dataset consists of 60000 32×32 colour images in 10 classes, with 6000 images per class. The object classes are airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck. We used the gray scale versions of all images in the dataset and out of 60000 total images, 40000 images are used for training, 10000 for validation, and the remaining 10000 are used for testing in a random manner. We present our results using two evaluation metrics. The classification results are presented using **accuracy**, while reconstruction performance is measured using **peak to signal noise ratio (PSNR)** [12]. We have selected a batch size of 32 and epoch size of 500 with the help of ADAM optimization for a varying learning rate from 0.1 to 0.0001 to determine the network parameters via gradient descent. We have used Tensorflow [1], the open source deep learning framework, for training, validation, and testing purposes. All simulations are run on a deep learning machine with 3 NVIDIA Titan RTX GPUs to carry out the training, validation, and testing tasks.

3.2. Classification on Original Image Domain

The goal of CL is to provide inference directly from the low dimensional compressed measurements. To provide a baseline for the CL performance, we first provide the classification results on the original signal domain. The original images in CIFAR-10 dataset is used without any compression. Three different classification networks, being AlexNet, VGG-3, and WRN, as described in Section 2.3 are trained and their performances are compared. All classification network parameters are initialized with random weights. The obtained accuracy over the test dataset is reported in Table 1. WRN is the best performing network over

Table 1: Classification accuracy on original images

Classification network	Accuracy
Simple Deep CNN/AlexNet	79%
VGG-3	88%
Wide Residual Network (WRN)	**97%**

original image domain among the compared techniques with 97% accuracy level. Nevertheless, we utilized both WRN and VGG-3 networks in our DJCL framework and provide results for both networks since they provide the two best accuracy results on the original signal domain.

3.3. Proposed Deep Joint Compressed Learning Framework

In this part, we present the performance of the proposed deep joint CL (DJCL) framework along with the compared deep CL techniques, DCL1 ([19]) and DCL2 ([31]). All compared approaches are trained and tested for the same set of compressed measurement numbers varying from $M = 64$ to $M = 768$. Since the images are 32×32, the dimension of original signal domain is $N = 1024$ and utilized measurement rates correspond to M/N ratios of 0.0625 to 0.75. While DCL1 uses compressed measurements that are created with a random MM, DCL2 and DJCL learn the MM jointly with its CL inference. DCL1 approach uses an AlexNet like CNN in its original form. We tested its performance with both VGG-3 and WRN networks which generates comparably higher performance. DCL2 is tested with WRN as utilized in its original version. The proposed DJCL framework is implemented with a variety of choices of reconstruction and classification networks. While ReconNet, ConvMMNet, and ISTANET$^+$ are the reconstruction network choices, VGG-3 and WRN are used as classification networks. Each combination case is trained and tested over the grayscale CIFAR-10 dataset using the same set of measurement numbers. We trained each DJCL framework case with two different loss functions; either only the cross entropy loss defined in (3) or the proposed weighted total loss defined in (5). For all scenarios, all the network parameters are randomly initialized before training. Obtained accuracy results over the test datasets are shown in Table 2 for DCL1 and DCL2 and in Table 3 for DJCL.

It can be seen in Table 2 that DCL2 provides higher accuracy results compared to DCL1 for all tested number of measurements. This is because of both learning MM and applying a learnable network layer to reconstruct a proxy image as input to the classification network. However, the

Table 2: Comparison of classification accuracy for DCL1 and DCL2 approaches for tested number of measurements.

M	DCL1+VGG-3	DCL1+WRN	DCL2+WRN
64	23%	26%	29%
128	27%	31%	36%
256	32%	34%	40%
512	43%	54%	61%
768	48%	63%	66%

CL inference results from both approaches are comparably lower than proposed DJCL classification results presented in Table 3 for the same number of measurements. This result shows that utilization of a reconstruction network along with the proposed weighted loss function in DJCL framework provides much higher accuracy results compared to existing DCL1 and DCL2 approaches. One possible reason these networks to achieve lower accuracy levels is that their proxy image generations don't provide enough detailed images as input to the classification parts.

There are several important conclusions that can be observed from the DJCL results presented in Table 3. First, if the DJCL network is trained with the proposed weighted loss (WL) that combines both cross entropy loss (CEL) and reconstruction loss (RL), the achieved accuracy levels are much higher than utilizing only the CEL. DJCL framework allows using a choice of reconstruction and classification networks. It can be seen that from three possible reconstruction and two classification network combinations ISTANET$^+$ and WRN combination generally provides the best accuracy levels for both loss function cases and all tested number of measurements. It can also be observed that employing ConvMMNet also achieves similar performance for several cases. Another important observation is that achieved accuracy levels with the proposed networks and training with WL achieves similar accuracy levels obtained over original image domain for higher number of measurements. This is because proposed structure jointly reconstructs and classifies with a loss function that combines both reconstruction and classification errors in a weighted manner.

In order to understand the effect of weighting between CEL and RL, a simulation study is performed. The total loss is defined in (5) and the parameter λ controls how much CEL is added. If $\lambda = 0$, total loss is the same as only RL while for very high λ total loss is dominated by only CEL. For the case of using ISTANET$^+$ and WRN combination in DJCL framework the achieved validation accuracy levels for a set of λ values is shown in Figure 2. The reconstruction network generates an image to be the input for the classification network as a midproduct of DJCL and the PSNR of that image is also shown in the Figure 2. It can be seen

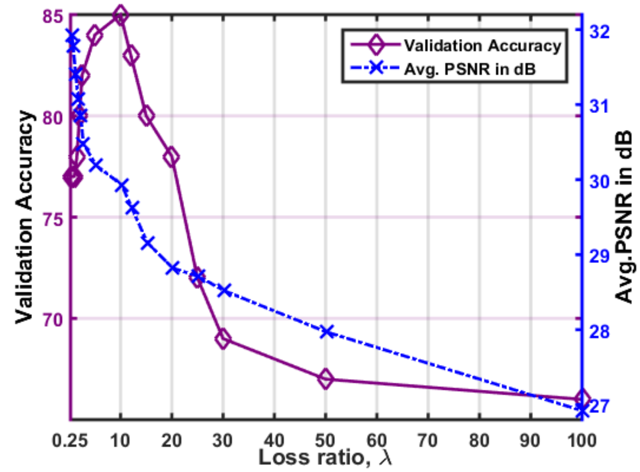


Figure 2: Effect of loss ratio parameter λ on validation accuracy and PSNR of reconstruction network output for $M = 256$.

that for smaller λ, the network focuses more on reconstruction and generates a high PSNR midproduct image but final accuracy levels are low. Increasing λ upto a level increases the achieved accuracy while sacrificing from the PSNR of the midproduct image. Although increasing λ more means for network to pay much more importance to CEL, the accuracy levels decreases since network can not generate higher PSNR images that will be input to the classification network. Using such analysis an optimal λ parameter can be selected for the weighted loss using the validation set and the performance of the selected parameter is tested with the independent test dataset. For our analysis, the parameter $\lambda = 10$ is found to be producing the highest accuracy for the validation set as seen in Fig 2.

3.4. Separate Reconstruction and Classification

In contrary to CL, one typical implementation to achieve a classification using compressed measurements requires a two stage approach: first, the image is reconstructed from the compressed measurements using a known reconstruction technique, and second, a classification technique is utilized on the reconstructed image. While this two stage approach require to reconstruct images first, keep them in memory, and train separate reconstruction and classification, our goal is to compare the CL framework with this separate implementation of reconstruction and classification. For reconstruction, we utilize ℓ_1-minimization based basis pursuit as the CS reconstruction technique and ReconNet [18], ConvMMNet [22], and ISTANET$^+$ [30] structures for DL based reconstruction. We tested scenarios where number of compressed measurements vary from $M = 64$ to $M = 768$. The compressed measurements

Table 3: Classification accuracy for proposed DJCL framework with all tested reconstruction and classification network combinations, with cross-entropy or weighted loss functions for different number of measurements.

Acc.		Cross-Entropy Loss		Weighted Loss	
Reconst.	M	VGG-3	WRN	VGG-3	WRN
ReconNet	64	36%	42%	59%	66%
ConvMMNet		41%	49%	65%	73%
ISTANET$^+$		**43%**	**50%**	**66%**	**74%**
ReconNet	128	43%	50%	65%	72%
ConvMMNet		48%	55%	69%	79%
ISTANET$^+$		**50%**	**57%**	**70%**	**80%**
ReconNet	256	50%	57%	71%	80%
ConvMMNet		55%	61%	76%	83%
ISTANET$^+$		**57%**	**63%**	**79%**	**85%**
ReconNet	512	59%	65%	80%	86%
ConvMMNet		63%	69%	85%	86%
ISTANET$^+$		**63%**	**69%**	**86%**	**90%**
ReconNet	768	63%	72%	85%	91%
ConvMMNet		66%	74%	87%	94%
ISTANET$^+$		**67%**	**75%**	**88%**	**96%**

can be generated with a random MM ($\mathbf{\Phi_R}$) or a MM can be learned ($\mathbf{\Phi_L}$) for DL based reconstruction. Note that the DNNs here are employed for image reconstruction only and trained with minimizing only the reconstruction loss. The reconstructed images are then separately used to train WRN or VGG-3 classification networks. Both the reconstruction PSNR values and the obtained accuracy levels for compared techniques under random or learned MM are provided in Table 4. It can be seen that DL based reconstruction approaches provide better reconstruction results compared to ℓ_1-minimization. The reconstruction performance also gets better as the number of compressed measurements increases. In addition, the learned MMs provide 1-3 dB higher PSNR in average. While WRN providing better accuracy results compared to VGG-3, the accuracy increases as better reconstructions are achieved with increasing number of measurements. Higher PSNR on image reconstruction is directly correlated with higher accuracy with a correlation coefficient of 0.97 for compared techniques and measurement numbers. For the measurement number of $M = 768$, close to original image domain accuracy levels are obtained. While WRN achieved 97% accuracy on uncompressed original images, 95% accuracy can be achieved with WRN when it is trained on reconstructed images with ISTANET$^+$ for $M = 768$.

Table 4: Separate reconstruction and classification results

Method	M	PSNR(dB)		Acc. (VGG-3)		Acc. (WRN)	
		Φ_R	Φ_L	Φ_R	Φ_L	Φ_R	Φ_L
ℓ_1	64	21.08	–	51%	–	60%	–
ReconNet		21.92	22.81	53%	57%	62%	63%
ConvMMNet		23.47	24.34	59%	62%	67%	69%
ISTANET$^+$		**23.67**	**25.47**	**60%**	**64%**	**69%**	**73%**
ℓ_1	128	21.84	–	53%	–	62%	–
ReconNet		23.85	24.96	61%	64%	65%	71%
ConvMMNet		25.43	27.19	**64%**	68%	73%	77%
ISTANET$^+$		**25.61**	**28.05**	**64%**	**69%**	**74%**	**79%**
ℓ_1	256	22.57	–	55%	–	65%	–
ReconNet		25.62	28.08	64%	69%	72%	79%
ConvMMNet		28.69	31.06	70%	74%	81%	81%
ISTANET$^+$		**29.12**	**32.17**	**71%**	**77%**	**82%**	**83%**
ℓ_1	512	23.05	–	59%	–	68%	–
ReconNet		26.95	33.28	66%	78%	73%	86%
ConvMMNet		34.06	36.08	79%	82%	87%	88%
ISTANET$^+$		**34.42**	**36.95**	**81%**	**83%**	**89%**	**89%**
ℓ_1	768	28.98	–	71%	–	80%	–
ReconNet		35.31	37.81	82%	84%	89%	91%
ConvMMNet		42.97	44.68	**86%**	86%	93%	93%
ISTANET$^+$		**43.21**	**45.32**	**86%**	**87%**	**95%**	**95%**

In Figure 3, the achieved accuracy levels as a function of number of measurements for the proposed DJCL approach using the weighted loss (DJCL+WL) is compared with DCL1, DCL2, DJCL using only cross-entropy loss (DJCL+CEL) and two stage approach of separately reconstructing all images first and applying classification on them. The classification accuracy achieved on the full original image domain is also shown. It can be seen that the proposed DJCL with WL achieves significantly better than only employing CEL or compared DCL1 or DCL2 approaches. It also offers slightly better accuracy than the two stage implementation nearly achieving the original image domain classification performance for higher number of measurement cases.

Although main task of the CL approaches is classification, we compare the mid-product images from proposed DJCL approach with CEL and WL cases with the proxy image from DCL2 approach in Figure 4. The direct recon-

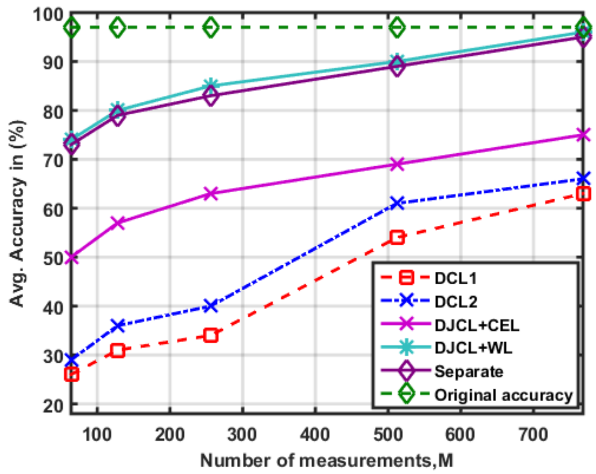


Figure 3: Average accuracy as a function of measurement number for the DJCL with CEL and WL cases compared to separate reconstruction and classification.

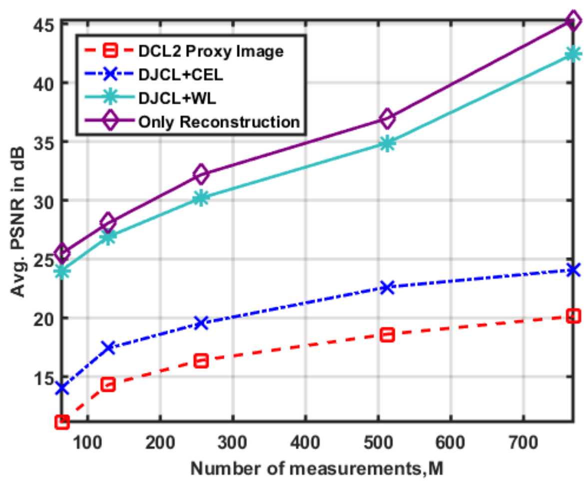


Figure 4: Average PSNR of reconstruction network outputs as a function of measurement number for compared approaches.

struction performance of ISTANET$^+$ structure with only reconstruction loss is also shown as a bound to the compared CL cases. It can be seen that proxy image of DCL2 or image output of DJCL+CEL has comparably much lower PSNR values than proposed DJCL+WL. It is expected that direct reconstruction provides best PSNR since DNNs are trained for minimizing the reconstruction loss only as also illustrated in Figure 2. While proposed DJCL+WL approach produces high PSNR images close to direct reconstruction, its final classification accuracy, which is the main task of compressed learning, is higher.

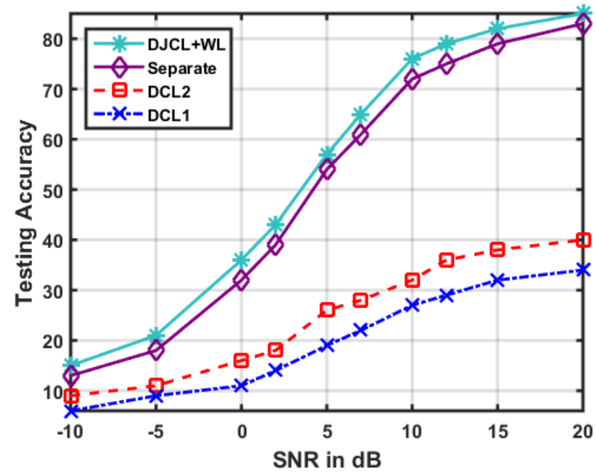


Figure 5: Accuracy as a function of SNR for compared approaches at M=256.

3.5. Noise Performance

In this part, we compare the robustness of tested approaches to additive noise in compressed measurements. White Gaussian noise (WGN) is added to compressed measurements of test dataset images with a varying level of signal-to-noise (SNR) ratios from -10dB to 20dB. The simulations are done for $M = 256$ number of measurements. The noisy compressed measurements are classified with the compared CL networks. The achieved accuracy levels for DCL1, DCL2, and the proposed DJCL framework with WL are shown in Figure 5 along with the case of separate reconstruction and classification. While every approach performs better with increasing SNR, it can be seen that the proposed DJCL outperforms compared CL approaches significantly and performs slightly better than separate implementation. All approaches nearly achieve their zero noise performances given in Tables 2 and 3 for higher SNR.

4. Conclusions

In this work, a deep joint compressed learning (CL) framework is proposed where it utilizes a deep reconstruction network within the compressed learning structure along with a novel weighted loss function to achieve classification from the low number of compressed measurements. The performance of the proposed approach is compared with the existing state-of-the-art CL approaches on CIFAR-10 image dataset. Proposed structure allows direct inference from compressed measurements with enhanced classification performance with robustness to noise compared to tested deep compressed learning approaches. In addition, optimal measurement matrices for the goal of enhancing classification performance are learned.

References

[1] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org. 5

[2] A. Adler, M. Elad, and M. Zibulevsky. Compressed learning: A deep neural network approach. *arXiv preprint arXiv:1610.09615*, 2016. 1, 2, 4

[3] R. G. Baraniuk. Compressive sensing [lecture notes]. *IEEE Signal Processing Magazine*, 24(4):118–121, 2007. 1

[4] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009. 3

[5] R. Calderbank and S. Jafarpour. Finding needles in compressed haystacks. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3441–3444. IEEE, 2012. 1

[6] R. Calderbank, S. Jafarpour, and R. Schapire. Compressed learning: Universal sparse dimensionality reduction and learning in the measurement domain. Technical report, 2009. 1

[7] E. J. Candès and M. B. Wakin. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 25(2):21–30, 2008. 1

[8] M. A. Davenport, M. F. Duarte, M. B. Wakin, J. N. Laska, D. Takhar, K. F. Kelly, and R. G. Baraniuk. The smashed filter for compressive classification and target recognition. In *Computational Imaging V*, volume 6498, page 64980H. International Society for Optics and Photonics, 2007. 1

[9] M. A. Davenport, A. K. Massimino, D. Needell, and T. Woolf. Constrained adaptive sensing. *IEEE Transactions on Signal Processing*, 64(20):5437–5449, 2016. 1

[10] A. Değerli, S. Aslan, M. Yamac, B. Sankur, and M. Gabbouj. Compressively sensed image recognition. In *2018 7th European Workshop on Visual Information Processing (EUVIP)*, pages 1–6. IEEE, 2018. 1

[11] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006. 1

[12] R. C. Gonzalez and R. E. Woods. *Digital image processing*. Prentice Hall, N.J., 2008. 5

[13] J. Haupt, R. Castro, R. Nowak, G. Fudge, and A. Yeh. Compressive sampling for signal classification. In *2006 Fortieth Asilomar Conference on Signals, Systems and Computers*, pages 1430–1434. IEEE, 2006. 1

[14] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 4

[15] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *CVPR*, volume 1, page 3, 2017. 4

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012. 2, 4, 5

[17] K. Kulkarni, 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*, pages 449–458, 2016. 2, 3

[18] S. Lohit, K. Kulkarni, R. Kerviche, P. Turaga, and A. Ashok. Convolutional neural networks for noniterative reconstruction of compressively sensed images. *IEEE Transactions on Computational Imaging*, 4(3):326–340, Sep. 2018. 2, 3, 4, 6

[19] 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)*, pages 1913–1917. IEEE, 2016. 1, 2, 3, 4, 5

[20] R. MdRafi and A. C. Gurbuz. Data driven measurement matrix learning for sparse reconstruction. In *2019 IEEE Data Science Workshop (DSW)*, pages 253–257, June 2019. 2

[21] R. MdRafi and A. C. Gurbuz. Learning to sense and reconstruct a class of signals. In *2019 IEEE Radar Conference (RadarConf)*, pages 1–5, April 2019. 2

[22] R. Mdrafi and A. C. Gurbuz. Joint learning of measurement matrix and signal reconstruction via deep learning. *IEEE Transactions on Computational Imaging*, 2020. 2, 3, 6

[23] A. Mousavi and R. G. Baraniuk. Learning to invert: Signal recovery via deep convolutional networks. In *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*, pages 2272–2276. IEEE, 2017. 2

[24] A. Mousavi, G. Dasarathy, and R. G. Baraniuk. Deepcodec: Adaptive sensing and recovery via deep convolutional neural networks. *arXiv preprint arXiv:1707.03386*, 2017. 2

[25] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 4

[26] Z. J. Xiang, H. Xu, and P. J. Ramadge. Learning sparse representations of high dimensional data on large scale dictionaries. In *Advances in neural information processing systems*, pages 900–908, 2011. 1

[27] X. Xie, Y. Wang, G. Shi, C. Wang, J. Du, and X. Han. Adaptive measurement network for cs image reconstruction. In *CCF Chinese Conference on Computer Vision*, pages 407–417. Springer, 2017. 2

[28] V. N. Xuan and O. Loffeld. A deep learning framework for compressed learning and signal reconstruction. In *5th International Workshop on Compressed Sensing applied to Radar, Multimodal Sensing, and Imaging (CoSeRa)*, pages 1–5, 2018. 1

[29] S. Zagoruyko and N. Komodakis. Wide residual networks. *arXiv preprint arXiv:1605.07146*, 2016. 4

[30] J. Zhang and B. Ghanem. Ista-net: Iterative shrinkage-thresholding algorithm inspired deep network for image compressive sensing. *CoRR*, abs/1706.07929, 2017. 2, 3, 6

[31] E. Zisselman, A. Adler, and M. Elad. Compressed learning for image classification: A deep neural network approach. In *Handbook of Numerical Analysis*, volume 19, pages 3–17. Elsevier, 2018. 1, 2, 3, 4, 5