DEEP NEURAL NETWORKS FOR LOW-RESOLUTION PHOTON-LIMITED IMAGING

Omar DeGuchy,[†] Fabian Santiago,[†] Mario Banuelos,[⋆] and Roummel F. Marcia[†]

[†]Department of Applied Mathematics, University of California, Merced, Merced, CA 95343 USA
[⋆]Department of Mathematics, California State University, Fresno, Fresno, CA 93740 USA

ABSTRACT

In this paper, we implement deep learning methods to recover downsampled noisy signals often present in compressed sensing applications. As an alternative to relying on previously established optimization based algorithms, we implement stacked denoising autoencoders and convolutional neural networks to perform signal reconstructions. Moreover, we propose a Poisson autoencoder inverting network (PAIN) architecture to reconstruct compressed signals imposed with Poisson noise. We observe less computational costs associated with this method while improving on reconstructions from a traditional stacked denoising autoencoder and remaining competitive with a more complex architecture in terms of Mean Squared Error (MSE). We train all proposed architectures on the MNIST dataset and establish deep neural networks as a reconstruction method.

Index Terms— Deep Learning, Photon-limited imaging, Poisson noise, Autoencoders

1. INTRODUCTION

Applications in image reconstruction and compressed sensing require recovering a true signal from noisy and undersampled linear measurements. These signals – often sparse in some basis – allows for the use of penalty based algorithms to promote sparsity in its reconstruction [1, 2]. This modality is typically found in applications such as medical imaging and night vision where measurements at the photon detector are corrupted by Poisson noise and thus modeled using the Poisson distribution [3].

Under this process of photon-limited imaging, we seek to reconstruct sparse signals from noisy low-dimensional observations. Previous methods for solving the signal reconstruction problem include its reformulation into an optimization problem and use iterative methods in order to arrive at a solution [4, 5, 6, 7, 8, 9]. Deep neural network architectures have been used to effectively extract features from similar signals through the use of autoencoders and convolutional neural networks (CNN) [10, 11, 12]. In this paper, we explore the use of various architectures in deep learning techniques as applied to the area of compressed sensing and establish their effectiveness in the field of photon-limited imaging.

This work was supported by NSF Grant IIS-1741490 and the UC Merced Fletcher Jones Fellowship.

2. PROBLEM FORMULATION

In the context of photon limited imaging, the arrival of photons at the detector are modeled by the following inhomogeneous Poisson process

$$\mathbf{y} \sim \text{Poisson}(\mathbf{Af}^*),$$

where $\mathbf{y} \in \mathbb{Z}_+^m$ is the observation vector whose entries consists of photon counts, $\mathbf{f}^* \in \mathbb{R}_+^n$ is the true signal and $\mathbf{A}_+^{m\times n}$ is the system matrix projecting the true signal to the observation space with $m \ll n$. Our interest is in recovering the higher dimensional signal $\mathbf{f}^*$ given the lower dimensional observation vector $\mathbf{y}$. Existing recovery methods use the maximum likelihood principle to maximize the probability of observing the vector $\mathbf{y}$. Furthermore, under the assumption that $\mathbf{f}^*$ is sparse, a sparsity promoting penalty term is incorporated in the reconstruction $\widehat{\mathbf{f}}$. These iterative algorithms require tuning parameters associated with the choice and enforcement of penalties and also require a substantial number of iterations to recover the signal [13, 14, 15].

We seek to avoid the complications associated with the current iterative optimization methods by solving the sparsity-promoting Poisson reconstruction problem using a variety of deep learning architectures. We accomplish this by training neural networks to process the low dimensional input $\mathbf{y}$ and provide a reconstruction of the true signal $\mathbf{f}^*$. Recently, deep learning techniques have been implemented separately for image reconstruction from downsampled observations and for Poisson denoising problems [16, 17, 18]. The novelty of the proposed architecture is that it solves both problems simultaneously as is required in many photon-limited applications.

3. DEEP LEARNING ARCHITECTURES

We propose three different neural network configurations with the purpose of recovering data from noisy low dimensional observations. One architecture employs the use of a convolution neural network (CNN) while the other two take advantage of the structure of autoencoders. As is common in these types of networks, all three implementations were trained using backpropagation and the mean squared error (MSE) was used as a loss function.

Poisson Inverting ConvolutionS (PICS). The first implementation is a CNN based on the neural network known as DeepInverse [17]. The DeepInverse architecture was not intended to denoise the signal, but rather recover signals from compression. In [17] the disparity between the dimension of the observation and the true signal is circumvented by the use of the proxy signal $\hat{x} = \mathbf{A}^T y$. The proxy is fed into the neural network and padding is used to keep the signal dimension consistence throughout. Instead of assuming explicit access to a measurement matrix, we added an extra fully connected layer in place of the proxy signal. This transformation serves two purposes. First, it increases the size of the observation vector $\mathbf{y}$ to the dimensions of the original signal. Secondly, it allows us to learn the transformation to the compressed space. There are three primary layers to this CNN architecture. The first consists of 64 filters of size $11 \times 11 \times 1$ with the last dimension pertaining to the depth of the filter . The next layer consists of 32 filters each with a dimension of $11 \times 11 \times 64$. The final layer consists of one $11 \times 11 \times 32$ filter to output the original image. After each layer a ReLU nonlinearity is applied to the output. The network was originally trained in the literature using 64×64 natural images and the preliminary results showed that this method was not accommodating to our database. Alternatively, the structure was modified using 2×2 filters while the number of filters was modified to reflect the scale of the input data and the required final output. This resulted in a slightly different architecture that was allowed to scale with the size of the input vector. Because these changes were not reflected in the literature, we will refer to the modified network as Poisson Inverting ConvolutionS (PICS).

Stacked Denoising Autoencoders (SDA). The authors in [16] use stacked denoising autoencoders to learn and recover the structure of sparse signals, again the intention of the method did not include a denoising component. This type of architecture was introduced in [19] as a way to make the learning capabilities more robust by introducing noise to the input before each autoencoding layer. The architecture involves stacking a decoder before encoding and decoding once again to arrive at the dimension of the true signal. Layers are differentiated by the dimensions of the weight matrix and bias vectors. After the dimension has been increased or decreased, the sigmoid function is applied to the output. The Stacked Denoising Autoencoder (SDA) structure implicitly resolves the increase in dimension eliminating the need for modifications.

Poisson Autoencoder Inverting Network (PAIN). The final architecture we propose in this work is Poisson Autoencoder Inverting Network (PAIN). Similar networks have been effective in the application of image compression [20]. The architecture is similar to SDA, the difference being that each decoding and encoding layer consists of multiple layers. Under this structure, compression and decompression is done gradually. The intuition is that the image must past through more layers and therefore the compression is refined at each layer, making the encoding more impervious to noise. In order to adapt the architecture to the sparse Poisson reconstruction problem, we required two modifications. The first change involves initializing the weight matrix using a truncated normal distribution instead of random samples from a normal distribution. The truncated normal initialization eliminates values more than 2 standard deviations from the mean. This immediately improved the architecture's ability to denoise the given signal. The network also utilizes a single layer decoder to initialize the process. The network starts with a layer that boosts the dimension of the observation to the dimension of the true signal. The next layer is a double encoder that reduces the dimension of the true signal to a length of 256 and then to the dimension of the observation. The dual layer decoder then brings the dimension back to 256 and finally to the dimension of the true signal (see Fig. 1).

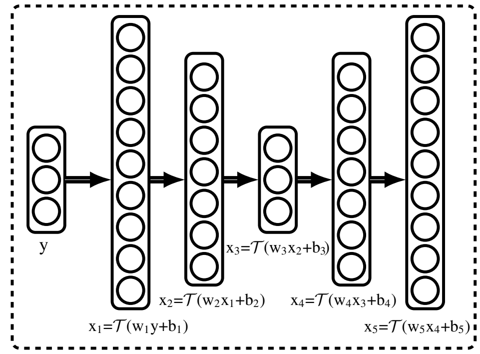
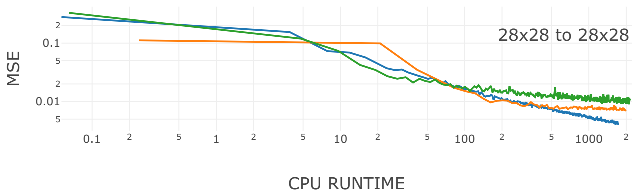


Fig. 1: Our proposed Poisson Autoencoder Inverting Network (PAIN). Each encoder and decoder consists of two layers. This framework also incorporates the sigmoidal activation function $\mathcal{T} = 1/(1 + e^{-x})$.

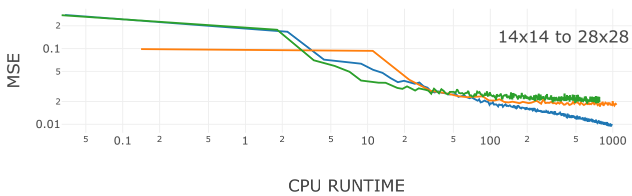
4. NUMERICAL EXPERIMENTS

The proposed architectures PICS, PAIN and SDA were all implemented using the open source machine learning language Tensorflow. Training and testing of the neural networks was performed using a quad core Intel i7-6700 CPU on a local PC with 64 GB of ram. The networks were trained using the stochastic gradient descent method known as RMSprop [21, 22].

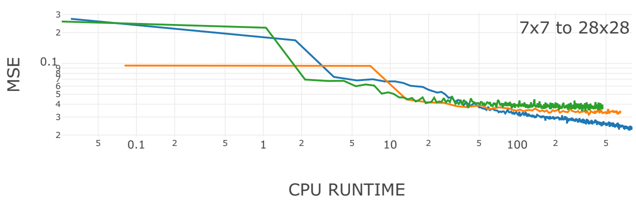
MNIST Dataset. The MNIST data set first used in [23] was altered in order to create pairs of signals to fit the Poisson model. The dataset consists of 70,000 28×28 images of handwritten numbers from 0-9 and their associated classification labels. From this dataset, 60,000 examples are used for training and 10,000 examples are used for testing. The reconstruction problem does not make use of the labels because



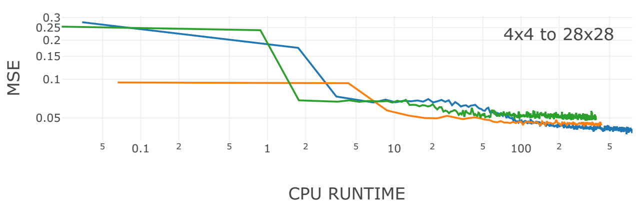
(a)



(b)



(c)



(d)

Fig. 2: The evolution of the MSE versus CPU runtime for PAIN (our proposed method, in blue), PICS (in orange), and SDA (in green) on a log-log scale for the purpose of reconstructing 28×28 images from (a) 28×28, (b) 14×14, (c) 7×7, and (d) 4×4 Poisson realizations. All CPU time is recorded in seconds.

we are not concerned with classification. For the purposes of this paper, we consider the original 28×28 image as the true signal $\mathbf{f}^*$. The associated observation vectors are created by taking the mean average of blocks of pixels, taking care that the size of the blocks reduce the size of the image without the need for padding. We then impose Poisson noise on the downsampled signal. Data sets were created by pairing true signals with observational signals of varying size. Under this structure the neural network is expected to train on a set of noisy observations with a fixed dimension ($n \times n$ with $n \in \{4, 7, 14, 28\}$) and reconstruct the full image (28×28).

Performance. The proposed architectures were able to perform suitable reconstructions of the test data sets. Both the MSE and the CPU runtime were used to quantify the effectiveness of the architectures during training (Fig. 2). Both PICS and PAIN achieved a lower MSE over the training sets than the SDA architecture for observation vectors in all dimensions. Furthermore, PICS completed 10,000 iterations (10,000 iterations are reported since the MSE does not show significant improvement after this number) in the same computational time required for 45,000 iterations of PAIN and SDA. This is expected as PICS is a CNN and is therefore a more computationally intensive process than the other two methods. Although the PICS architecture takes longer to complete a given number of training iterations, it reaches a lower MSE faster than SDA or PAIN. This can been seen in the restoration of 28×28 and 14×14 images. This shows that PICS is learning at a faster rate, making each training step more valuable when compared to the other architectures. While the CNN structure is initially more accurate than the autoencoder structures, a decrease in input dimension results in competitive MSE from the less complex PAIN structure. The output for a set of input digits is displayed in Fig. 4. As the amount of compression increases, the quality of the reconstruction decreases for all three types of architecture. The PAIN and PICS architectures clearly outperform the SDA when it comes to the 4×4 compression. The surprising observation is that PAIN seems to have a higher intensity

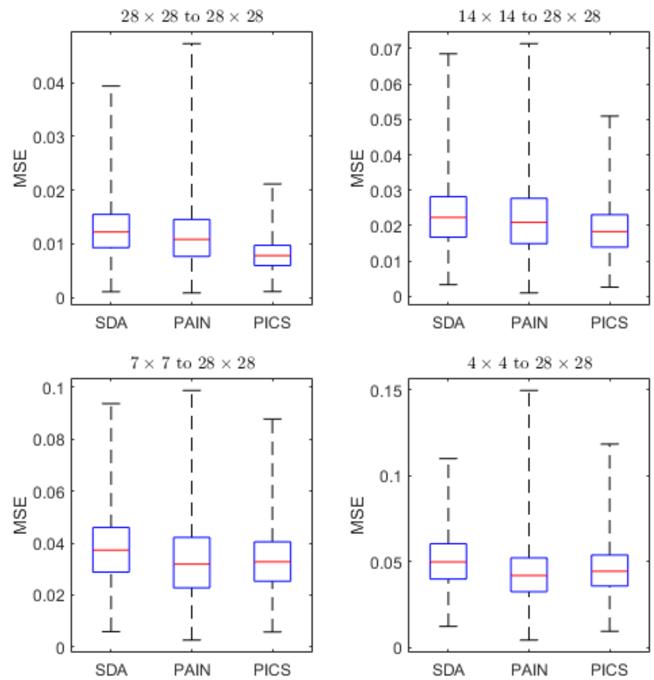


Fig. 3: Boxplots comparing mean squared error (MSE) computed using the MNIST validation images (5,000) and their reconstructions for the three architectures proposed (SDA, PAIN, and PICS). We observe that all three architectures behave similarly in terms of their MSE as the amount of compression increases.

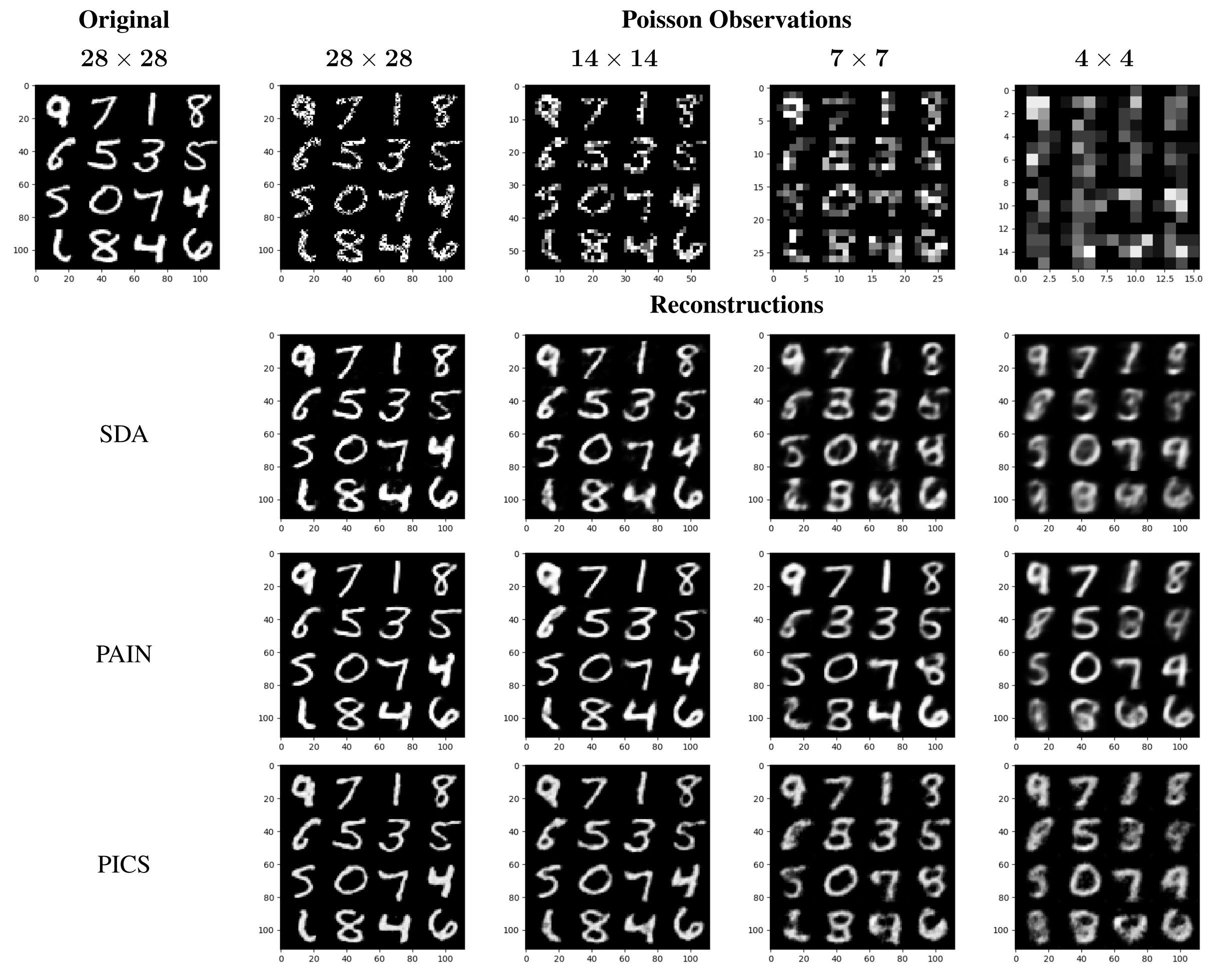


Fig. 4: The first row is comprised of 16 original MNIST images and their corresponding downsampled Poisson realizations. Given these input images, we present the reconstructions using Stacked Denoising Autoencoders (SDA), our proposed methods Poisson Autoencoder Inverting Network (PAIN) and Poisson Inverting ConvolutionS (PICS) for each dimension of images.

and smoothing effect when it comes to recreating the initial image. This is counterintuitive considering that we expect higher intensity from PICS since it uses the ReLU activation function.

The models were also tested on the 5,000 entries of the MNIST validation set. All MSE scores comparing the original signal and the reconstructed signal were computed and the results are presented in Fig. 3. We note that the MSE for all three architectures is skewed towards a lower MSE value. Also, we observe that as the amount of compression increases, the SDA and PAIN architectures become comparable to the more computationally intensive PICS architecture.

5. CONCLUSION

In this paper we implemented three deep learning architectures to solve the Poisson inverse problem. These neural networks have proven to be very effective in the reconstruction of images under the same modality. The first two networks involved modified autoencoders, while the third used a convolutional neural network. The results show that the stacked denoising autoencoder did not perform as well as the PAIN and PICS networks during training using the Mean Squared Error (MSE) as a performance metric. PAIN also has the benefit of being less computationally intensive, which could suggest that it will scale better with larger image sizes. Furthermore, PAIN has a smoothing characteristic in the reconstructions which is not reflected in the MSE. While the image smoothing could negatively effect the MSE, this property could have advantages during classification. In future work, we hope to improve on this architecture by modifying the number of layers and exploring the use of another (perhaps more appropriate) loss function. We also hope to apply these reconstruction architectures to more complex images.

Acknowledgments. The authors would like to thank Prof. Harish Bhat, whose course on the mathematics of deep learning helped facilitate this research.

6. REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] 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*, 2016, pp. 449–458.

[3] D. L. Snyder and M. I. Miller, *Random Point Processes in Time and Space*, Springer, 1991.

[4] J. A Fessler and A. O. Hero, "Penalized maximum-likelihood image reconstruction using space-alternating generalized em algorithms," *IEEE Trans. Image Processing*, vol. 4, no. 10, pp. 1417–1429, 1995.

[5] F.-X. Dupé, J. M. Fadili, and J.-L. Starck, "A proximal iteration for deconvolving Poisson noisy images using sparse representations," *IEEE Transactions on Image Processing*, vol. 18, no. 2, pp. 310–321, 2009.

[6] M. A. T. Figueiredo and J. M. Bioucas-Dias, "Restoration of Poissonian images using alternating direction optimization," *IEEE Transactions on Image Processing*, vol. 19, no. 12, pp. 3133–3145, Dec 2010.

[7] Z. T. Harmany, R. F. Marcia, and R. M. Willett, "This is SPIRAL-TAP: Sparse Poisson intensity reconstruction algorithms; theory and practice," *IEEE Trans. on Image Processing*, vol. 21, no. 3, pp. 1084–1096, 2012.

[8] L. Adhikari and R. F. Marcia, "Nonconvex relaxation for poisson intensity reconstruction," in *Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing*, April 2015, pp. 1483–1487.

[9] M. Banuelos, R. Almanza, L. Adhikari, S. Sindi, and R. F. Marcia, "Sparse signal recovery methods for variant detection in next-generation sequencing data," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, 2016.

[10] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P. A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *Journal of Machine Learning Research*, vol. 11, no. Dec, pp. 3371–3408, 2010.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[12] P. Y. Simard, D. Steinkraus, J. C. Platt, et al., "Best practices for convolutional neural networks applied to visual document analysis," in *ICDAR*, 2003, vol. 3, pp. 958–962.

[13] Z. T. Harmany, R. F. Marcia, and R. M. Willett, "Sparse poisson intensity reconstruction algorithms," in *Statistical Signal Processing, 2009. SSP'09. IEEE/SP 15th Workshop on*. IEEE, 2009, pp. 634–637.

[14] D. J. Lingenfelter, J. A. Fessler, and Z. He, "Sparsity regularization for image reconstruction with poisson data," in *Computational Imaging VII*. International Society for Optics and Photonics, 2009, vol. 7246, p. 72460F.

[15] M. Banuelos, L. Adhikari, R. Almanza, A. Fujikawa, J. Sahagún, K. Sanderson, M. Spence, S. Sindi, and R. F. Marcia, "Nonconvex regularization for sparse genomic variant signal detection," in *2017 IEEE International Symposium on Medical Measurements and Applications*. IEEE, 2017, pp. 281–286.

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

[17] A. Mousavi and R. G. Baraniuk, "Learning to invert: Signal recovery via deep convolutional networks," *arXiv preprint arXiv:1701.03891*, 2017.

[18] T. Remez, O. Litany, R. Giryes, and A. M. Bronstein, "Deep convolutional denoising of low-light images," *arXiv preprint arXiv:1701.01687*, 2017.

[19] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th International Conference on Machine Learning*. ACM, 2008, pp. 1096–1103.

[20] D. Aymeric, "TF Learn," https://github.com/aymericdamien/TensorFlow-Examples/blob/master/examples/, 2017, GitHub Repository.

[21] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:1609.04747*, 2016.

[22] M. Riedmiller and H. Braun, "A direct adaptive method for faster backpropagation learning: The rprop algorithm," in *Neural Networks, 1993., IEEE International Conference on*. IEEE, 1993, pp. 586–591.

[23] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.