Lossless Compression of Curated Erythrocyte Images Using Deep Autoencoders for Malaria Infection Diagnosis

Hongda Shen, W. David Pan, Yuhang Dong, and Mohammad Alim
Dept. of Electrical and Computer Engineering
University of Alabama in Huntsville
Huntsville, AL 35899, USA

***Abstract*—While autoencoders have been used as an unsupervised machine learning technique for classification and dimensionality reduction of the input data, they are lossy in nature when used alone in data compression. In this work, we proposed an image coding scheme by using stacked autoencoders, where the reconstruction residuals were entropy-coded to achieve lossless compression. As a case study, we compressed labeled red blood cell images from a database curated by pathologists for malaria infection diagnosis. Specifically, we trained two separate stacked autoencoders to automatically learn the discriminative features from input images of infected and non-infected cells. Subsequently, the residuals of these two classes of images were coded by two independent Golomb-Rice encoders. Testing results showed that this deep learning approach provided remarkably higher compression on average than several other lossless coding methods including JPEG-LS, JPEG 2000 lossless mode, and CALIC.**

I. Introduction

Malaria is a life-threatening disease caused by parasites that are transmitted to people through the bites of infected female Anopheles mosquitoes. According to the report released by the World Health Organization (WHO), there were 214 million cases of malaria in 2015 and 438,000 deaths [1]. There has been a growing interest in building an efficient automated malaria diagnostic system [2] to deliver resources remotely to those underdeveloped areas where malaria has a marked predominance. In most cases, malaria can be diagnosed by manual examination of the microscopic slides. Whole slide imaging (WSI), which scans the conventional glass slides in order to produce digital slides, is the most recent imaging modality being employed by pathology departments worldwide. WSI data provide a direct access for pathologists to diagnose malaria. Fig. 1 shows four human erythrocyte (red blood cell) samples segmented from a whole slide image. Whole slide images have very high resolutions and multi-layer display feature, which allow pathologists to see more details; however, the images tend to be very large, easily taking up multiple gigabytes. This often is a bottleneck for remote diagnostic applications with limited network bandwidths and inadequate storage space. To this end, compression techniques provide a good solution. While lossy compression methods have been studied for whole slide images [3], [4], information loss may have a negative impact on doctors' diagnosis, and thus lossless compression is often preferred on whole slide images [5].

Among many lossless image compression methods, JPEG-LS/LOCO-I [6] is often used as the benchmark. JPEG-LS employs a simple but effective predictor called Median Edge Detector (MED), followed by an adaptive Golomb-Rice Coder (GRC) [7]. Context based Adaptive Lossless Image Codec (CALIC) [8] is another benchmark algorithm. In contrast to MED of JEPG-LS, the Gradient Adjusted Predictor (GAP) was employed by CALIC with a dedicated context modeling scheme. In [9], an Edge Directed Predictor (EDP) was proposed in order to better model the structure of the image data using Least-Square (LS) based adaptation. Besides the above well-known methods, there are a wide variety of lossless compression methods that are based on different mathematical models and optimized for different applications. However, most of these approaches focus on spatial redundancy removal within an individual image, without considering cross-image correlations. Learning-based methods have been exploited on this regard [10]. However, they mostly focus on the lossy compression. In a recent work [11], a memory-assisted approach utilizes Principal Component Analysis (PCA) to take into account both intra- and inter-image redundancies, achieving an improved lossless compression efficiency over traditional methods. Specifically, the following two-stage scheme was proposed: 1) Train the "memory" model based on a large number of similar images (learning); 2) Use the trained model to compress the new incoming data samples (testing). Since the training stage can extract the common features shared by the data, inter-image redundancy can be removed in addition to individual image de-correlation.

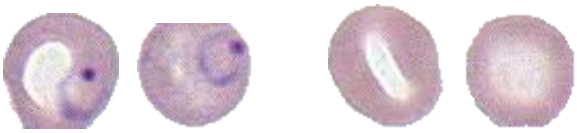


Fig. 1: Red blood cell samples: The two cells on the left are malaria infected, and two cells on the right are normal (non-infected).

Our previous work [5] studied regions of interest (ROIs) extraction from whole slide images and achieved very large

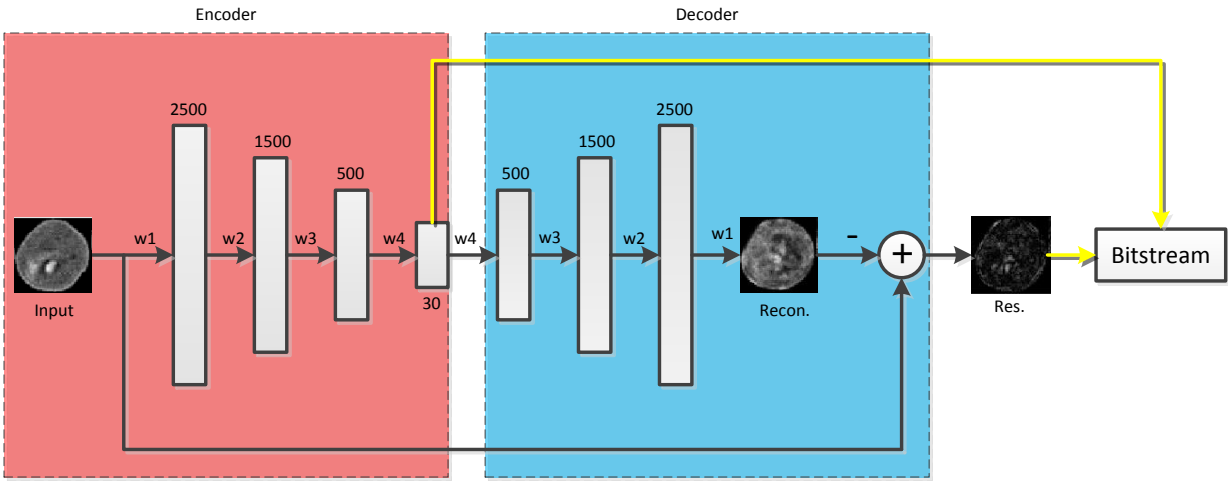


Fig. 2: The lossless image compressor.

lossless compression on the ROIs, which are typically red blood cells in the case of malaria diagnosis. From Fig. 1, we can see that infected cells differ from normal cells in that the infected cells have the purple "ring form" characteristic of a parasite being present. Therefore, if this common feature shared by all infected cells can be "learned" automatically, then we can remove the inter-image redundancy to further improve the compression performance. For this sake, we considered deep learning techniques [12]. While most deep learning approaches mainly focus on pattern classification problems [13], [14], the Stacked Autoencoder (SAE) based on Restricted Boltzmann Machines [15] has been proposed to reduce the dimensionality of the data, with the capability to extract discriminative features of the input data automatically. It was shown in [15] that, even in very low dimensions, the SAE networks (as a non-linear method in general) provide much better separation of input data belong to different classes than the PCA approach. Therefore, we proposed a new lossless compression method for images. As a case study, we designed a four-layer SAE network to "learn" the inherent features of the input images, which were taken from a set of labeled red blood cell images curated by a group of pathologists we are collaborating with. The reconstructed images using these learned low-dimensional representations were then used as the approximation of the input images. To achieve lossless compression, we employed a Golomb-Rice Code (GRC), which is a computationally efficient coding scheme, to code the residual images. Note that this approach is different from that used in [11]. To the best of our knowledge, this might be the first attempt to achieve lossless data compression by using a deep learning technique.

The rest of this paper is organized as follows. Section II gives a brief introduction to autoencoders and presents the proposed lossless image compression scheme based on a stacked autoencoder network. Section III gives the testing results by comparing against three state-of-the-art image lossless compression methods. Section IV concludes this paper with a discussion of further work.

II. Lossless Compression using Stacked Autoencoders

A. Stacked Autoencoders

Autoencoders [15] and their variants [16] are in essence artificial neural networks that perform unsupervised learning on the input data [17]. In the encoding phase, low-dimensional representations of the input data are learned through training the neural network. Next, these learned representations (so called low-dimensional "codes") are used to reconstruct the original data (decoding). The training algorithm will seek to optimize the neural network by minimizing the reconstruction loss as a cost function on sufficiently large amount of data. Moreover, a deep neural network can be constructed by concatenating multiple autoencoders (refer to Fig. 2 for the four serially connected autoencoders in the "Encoder" component of the Compressor). This would allow for a hierarchical representation of the data through a multi-layer architecture. In [15], the Restricted Boltzmann Machine (RBM) was used as an autoencoder, which serves as a building block of a deep autoencoder network. Each RBM was pretrained and unrolled. Then back propagation was carried out to fine-tune the entire stacked autoencoder based on Cross-Entropy or Mean Square Error (MSE) as the cost function. Results reported in [15] indicate superior performance of SAE compared to the standard PCA and its variants. SAE takes into account the nonlinearity of the image data and is able to extract features in a hierarchical manner. All these properties allow for very large dimensionality reduction of the input image, while preserving well the discriminative features. Thus we propose a lossless compression scheme, where the SAE is used as an aggressive yet lossy image encoder. The very low-dimensional code (30-point vector in this work), as well as the residual (the difference between the original image and the reconstructed version) will be coded to achieve lossless compression.

Here we distinguish the term "compressor" from the term "encoder" to avoid confusion. The input to the compressor is the image to be losslessly compressed, and its output is a compressed bitstream. In contrast, the encoder is the first component of the compressor (see Fig. 2). The job of the encoder is to reduce the input image to a very low dimensional vector through the "encoding" function of the stacked autoencoders. The second component of the compressor is the "decoder", whose job is to reconstruct an approximate version of the original image by going through the reversed operations of the stacked autoencoders. The residual between the original image and its approximation needs to be entropy coded (yellow line in Fig. 2) to produce the compressed bitstream. On the other hand, the decompressor takes as its input the compressed bitstream and the coded version of the low-dimensional vector. The "decoder" component of the compressor is reused in the decompressor to produce an approximation of the original image by feeding forward through the SAE decoder layers. The approximation, combined with the residual recovered from the compressed bitstream, will be used to reconstruct the original image losslessly. In Fig. 2, each pair of gray bars represents a RBM with its corresponding weight shown [15]. In the training stage, a sufficiently large number of images are fed to the multi-layer SAE as training data. Our autoencoder consists of four layers of RBMs with 2500, 1500, 500 and 30 neurons at each layer. Eventually, the encoder of the SAE generates a 30-point *code* of the input image. Note that the weights of the decoder are assumed known at both the compressor and decompressor sides in advance so that the residual image and the approximation image can be recovered for end-to-end lossless compression.

B. The Separate Coding Scheme

We experimented with training the SAE using a set of images mixed with infected and non-infected cells images. The SAE trained this way was found to offer lower compression than an alternative scheme, where we train two SAEs, one on infected cell and the other on non-infected cell images. This is expected since common features shared by the cells of the same class can be more easily learned by the SAE training with images of that particular class. The flowchart of the entire compression process is shown in Fig. 3.

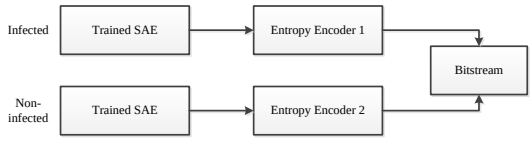


Fig. 3: Separate compression of images based on their classes.

In order to determine the discriminative capability of the stacked autoencoders, we modified the SAE network architecture into 2500-1500-500-2 and trained it on 1,000 infected and 1,000 non-infected cell images mixed together. As a result, an input image (a 2,500-point vector) was reduced to a mere two-dimensional vector. Fig. 4 shows the resulting

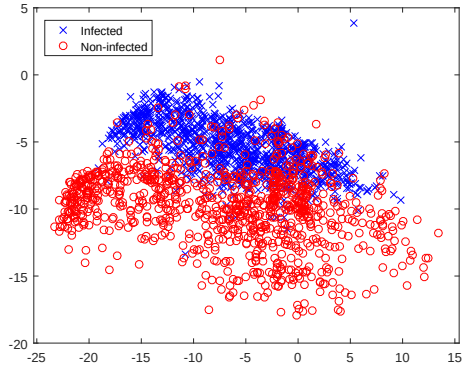


Fig. 4: The distribution of the two-dimensional codewords for the infected (blue) and non-infected cells (red).

two-dimensional codes, demonstrating the ability of the deep autoencoders in extracting the discriminative features from the data. Furthermore, we can see that the infected cells are more concentrated in their cluster than the non-infected cells. This implies that the images of the infected cells might benefit more from the proposed compression scheme than the non-infected cells.

Since the residual images approximate the Laplacian distribution, Golomb-Rice codes were used to encode the residuals by using the same coding parameter estimation method as in JPEG-LS [6]. Note that while other entropy codes such as the arithmetic codes can be used to generate shorter bitstreams, they tend to have higher computational costs. Typically, the 30-point *code* consist of real numbers. We need to quantize them to integers by using direct binary coding ()with at most 12 bits/sample). The de-quantized values are then provided to the decoder to ensure strict lossless compression. So a total number of 360 bits for each image will be stored as side information with the coded bitstream for the residuals.

III. Experimental Results and Analysis

A. Data Preparation

We used a dataset of red blood cell images [18], provided and labeled by a group of pathologists from the Medical School of the University of Alabama at Birmingham (UAB), as part of the collaborative research on automated malaria infection diagnosis. To facilitate training of the deep learning network, we developed image processing algorithms to segment the original wholeslide image and cropped each individual cell into an image of 50×50 pixels. We might need to down sample some cells to make a uniform image size. All the pixel values were mapped to $[0, 1]$ prior to training. The cross-entropy was chosen as the cost function for training. To avoid overly long training time, we used only the red channel of the original color images to demonstrate the compression performance of the proposed method.

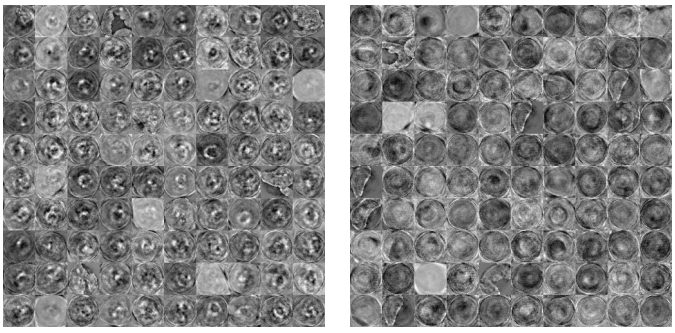
B. Experiments Setting

We chose a 2500-1500-500-30 network architecture, which was found to provide good compression on the data. Two

SAEs were trained separately, one on infected and the other on normal cell images. For each type of images, 1,000 labeled images were divided into two sets: 900 for training and 100 for testing. The number of epochs is set to 1,000 for pre-training and fine-tuning. For comparison, three lossless compression methods, including JPEG 2000 lossless mode (JP2K-LM), JPEG-LS and CALIC, were applied on the test images. The built-in JP2K-LM function of MATLAB, the implementations of JPEG-LS [19] and CALIC [20] were used. The training took about two hours to finish using Matlab running on a computer running Ubuntu 14.04.

C. Results and Analysis

The weighs obtained at the first layer of the stacked autoencoder are shown in Fig. 5a and 5b for infected and non-infected cell images, respectively. The ring forms characteristic of the infected cells are clearly visible in Fig. 5a.



(a) Images of infected cells. (b) Images of non-infected cells.

Fig. 5: Weights learned by the stacked autoencoder.

Fig. 6 demonstrates the image approximation performance of the SAE. We can see that except for the second left and the right most images, most residual images have very small values (even near the edges, where traditional intra-image predictors would likely fail). PSNR values (in dB) of the reconstructed images also indicate that SAE can produce accurate data approximation. Besides, the SAE can even learn the irregular shapes of some cells (e.g., the third image from the left), which was found to be challenging in our prior study on lossless compression of regions of interest of arbitrary shapes [21]. More accurate reconstruction leads to smaller residuals, which would very likely translate to smaller bit rates than traditional methods.

Label	SAE	JPEG-LS	JP2K-LM	CALIC
Infected	**5.1729**	5.6921	6.2320	5.4391
Normal	**5.5135**	5.9632	6.4195	5.6068

TABLE I: Comparison of average bit rates (bits/pixel) with other benchmark methods.

As a comparison, Fig. 7a and 7b show the bit rates of each individual testing image taken from the set of infected and non-infected images. Note that image files are arranged such that their bit rates obtained by using the proposed method are

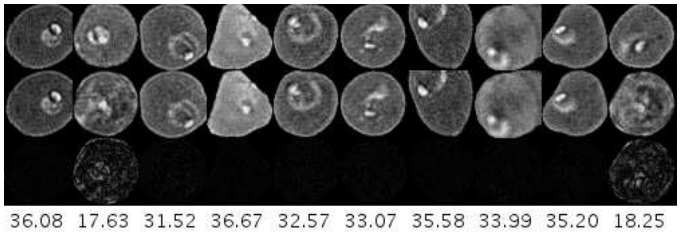


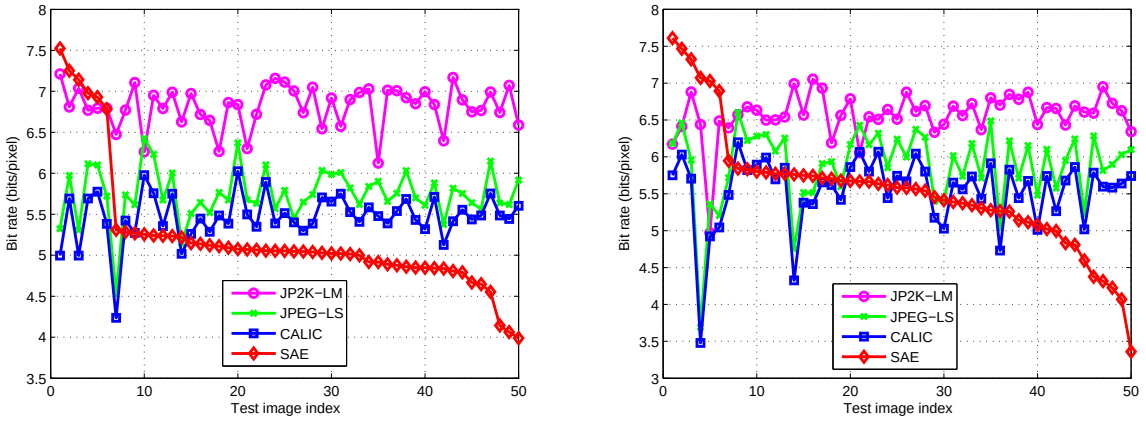
Fig. 6: Ten randomly selected infected cell images, from top to bottom: original images, reconstructed images, absolute residual images and the corresponding PSNR values. A complete list of test images and their reconstruction residuals (of infected and non-infected cells) can be seen at the website [22].

shown in a descending order for ease of comparison. Table I summarizes the average bit rates of the several methods. We can see that the proposed method based on SAE has lower average bit rates than other three benchmark methods on both infected and non-infected cell images. Particularly, for infected cell images, our method achieved $4.9\%, 9.1\%$ and 17.0% lower bit rates than CALIC, JPEG-LS and JPEG 2000-LM, respectively.

However, we can see that the SAE method gives the lowest compression among all the methods for a very small set of images, and the compression performance on normal cell images is not as good as the infected cell case. The reason might be that SAE is a "global" method, which aims to learn the common features present in all the training data, while conventional predictive lossless compression methods rely on adaptation to local statistics. For some images, intra-image correlations might be stronger than inter-image correlations, thus local method tends to compress better. On the other hand, as shown in Fig. 4, common features shared by non-infected cell images tend to be more elusive than infected cells, therefore, the advantage of using deep learning for non-infected cells becomes less pronounced. Nonetheless, infected cells typically belong to the regions of interest in a whole slide image when it comes to diagnosis of malaria infection. Hence, more efficient compression on infected cell images would be useful. In terms of computational efficiency, the proposed method only requires feedforwarding the data through the network once it is trained. It took about about 30ms on average for this testing set in our simulations. Therefore, efficient parallel implementation is feasible. In contrast, parallel implementation is difficult for most existing predictive lossless compression methods, which rely on pixel prediction based on tightly coupled local contexts.

IV. Conclusions and Future Work

We proposed an image coding scheme by using stacked autoencoders, and trained two separate stacked autoencoders to automatically learn the discriminative features from input images of malaria infected and non-infected cells. Testing



(a) Images of infected cells.

(b) Images of non-infected cells.

Fig. 7: Comparison of bit rates (bits/pixel) with other benchmark methods on 50 randomly selected testing images.

results showed that this deep learning approach provided remarkably higher compression on average than several other state-of-the-art lossless coding methods. Similar results have recently been obtained for other types of images (e.g., the handwritten digits from the well-known MNIST database). As the next step, we will consider a hybrid (local prediction plus global learning) approach to achieve further performance gains.

ACKNOWLEDGEMENT

We acknowledge the support from the UAH College of Engineering Mid-Career Faculty Research Award and NSF-EEC Award for REU Site. We also thank the UA Systems Collaborative Research Award, which allows us to collaborate with Dr. Allen Bryan and his colleagues in the Dept. of Pathology, Univ. of Alabama at Birmingham. We also thank the anonymous reviewers for their comments and suggestions to improve the paper.

REFERENCES

[1] Disease burden of malaria, ," http://www.who.int/mediacentre/factsheets/fs094/en/, Accessed: 2016-05-30.

[2] V. Muralidharan, Y. Dong, and W. D. Pan, "A comparison of feature selection methods for machine learning based automatic malarial cell recognition in wholeslide images," in *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, Feb 2016, pp. 216–219.

[3] T. Kalinski, R. Zwonitzer, F. Grabellus, S. Y. Sheu, S. Sel, H. Hofmann, and A. Roessner, "Lossless compression of JPEG2000 whole slide images is not required for diagnostic virtual microscopy," *American Journal of Clinical Pathology*, vol. 136, no. 6, pp. 889–895, Dec. 2011.

[4] A. Sharma, P. Bautista, and Y. Yaqi, "Balancing image quality and compression factor for special stains whole slide images.," *Analytical Cellular Pathology*, vol. 35, no. 2, pp. 101–106, 2012.

[5] Y. Dong, H. Shen, and W. D. Pan, "An interactive tool for ROI extraction and compression on whole slide images," in *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, Feb 2016, pp. 224–227.

[6] M. J. Weinberger, G. Seroussi, and G. Sapiro, "The LOCO-I lossless image compression algorithm: principles and standardization into JPEG-LS," *IEEE Trans. Image Process.*, vol. 9, no. 8, pp. 1309–1324, Aug. 2000.

[7] S. Golomb, "Run-length encodings (corresp.)," *IEEE Trans. Inf. Theory*, vol. 12, no. 3, pp. 399–401, Jul. 1966.

[8] X. Wu and N. Memon, "Context-based lossless interband compression-extending CALIC," *IEEE Trans. Image Process.*, vol. 9, no. 6, pp. 994–1001, Jun 2000.

[9] X. Li and OM. T. Orchard, "Edge-directed prediction for lossless compression of natural images," *IEEE Trans. Image Process.*, vol. 10, no. 6, pp. 813–817, Jun 2001.

[10] J. Jiang, "Image compression with neural networks - a survey," in *Signal Processing: Image Communication 14*, 1999, pp. 737–760.

[11] Z. R. Hesabi, M. Sardari, A. Beirami, F. Fekri, M. Deriche, and A. Navarro, "A memory-assisted lossless compression algorithm for medical images," in *2014 IEEE Intl. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, May 2014, pp. 2030–2034.

[12] Y. LeCun, Y. Bengio, and G. E. Hinton, "Deep learning," *Nature*, vol. 521, no. 5786, pp. 436–444, 2015.

[13] 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.

[14] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.

[15] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[16] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Advances in Neural Information Processing Systems 19 (NIPS'06)*, 2007, pp. 153–160.

[17] I. Goodfellow, Y. Bengio, and A. Courville, "Deep learning," Book in preparation for MIT Press, 2016.

[18] University of Alabama at Birmingham, "PEIR-VM," Available: http://peir-vm.path.uab.edu/about.php.

[19] "JPEG-LS software," http://www.labs.hp.com/research/info_theory/loco/, 2016 (accessed May 30, 2016).

[20] "CALIC software," http://www.ece.mcmaster.ca/~xwu/calicexe/, 2016 (accessed May 30, 2016).

[21] H. Shen, W. D. Pan, and Y. Wang, "A novel method for lossless compression of arbitrarily shaped regions of interest in hyperspectral imagery," in *Proc. 2015 IEEE SoutheastCon*, April 2015.

[22] "Link to test images," http://www.ece.uah.edu/~dwpan/pcs/test_images/.