A NEW MAX-MIN CONVOLUTIONAL NETWORK FOR HYPERSPECTRAL IMAGE CLASSIFICATION

S. K. Roy[1], *M. E. Paoletti*[2], *J. M. Haut*[3], *E. M. T. Hendrix*[4], *A. Plaza*[2]

[1]Dept. of Computer Science and Engineering, Jalpaiguri Government Engineering College. India.
[2]Dept. of Technology of Computers and Communications, University of Extremadura, Spain.
[3]Dept. of Communication and Control Systems, National Distance Education University, Spain.
[4]Dept. of Computer Architecture, Universidad de Málaga, Spain.

ABSTRACT

Convolutional neural networks (CNNs) are a noteworthy tool for the classification of hyperspectral images (HSIs). CNNs apply non-linear activation functions to learn data patterns. One of them is the rectified linear unit (ReLU), which is a piece-wise linear function with a value which is the input if positive and zero otherwise. As a result, it is computationally efficient and tends to show good convergence behaviour. Nonetheless, its performance suffers from the so-called dying ReLU effect. This is usually managed by introducing more convolution layers increasing the depth of the model followed by a ReLU non-linearity layer that may hamper the convergence of network and produce a low classification accuracy due to data degradation. In order to alleviate these issues and transmit more information after the activation layer from the convolutional block, this paper develops a new end-to-end supervised feature learning framework called MaxMin-CNN, which works with sub-cubes of the original HSI data and successively applies 3-D MaxMin convolutional filters to improve the discrimination ability of the obtained spectral-spatial features by doubling the feature maps over all the convolutional layers. The new model gradually increases the heterogeneity of high-level spectral-spatial features across the MaxMin convolutional layers, enhancing the performance of HSI classification and reducing the model depth while preserving the classification performance. In order to validate the model, we report experiments over three widely used HSI datasets: Indian Pines, University of Pavia and University of Houston. The results reveal that the proposed MaxMin-CNN achieves a classification comparable to state-of-the-art classification models.

Index Terms— Convolutional neural networks (CNNs), deep learning, hyperspectral images (HSIs), remote sensing, image classification.

This work has been funded by Spanish state grant RTI2018-095993-B-I00, in part financed by the European Regional Development Fund (ERDF).

1. INTRODUCTION

With the ultimate goal of providing , Hyperspectral images (HSIs) provide rich spectral-spatial information collecting (for the same observation area) a huge data cube composed by several images, captured as different measurements along the electromagnetic spectrum from the visible to the short-wave infrared. In this context, HSIs provide a characterization of materials, capturing their chemical and physical properties within the captured region. Therefore, they ar aplied in a wide range of applications, including earth observation, agriculture, vegetation monitoring, urban analysis, and crop analysis. Among all, land cover classification using HSIs is one of the most popular research topics, where spectral pixels are classified into one of several pre-defined land cover classes. However, the high-dimensionality and the intrinsic structure of HSI pixels increase the representation complexity of HSIs and, due to the lack of enough labeled training samples, the classification of HSIs is quite challenging.

In the past two decades, traditional machine learning methods, such as support vector machines (SVMs) [1, 2], Bayesian models [3], and k-nearest neighbor (KNN) [4], have been widely used for HSI classification due to the discrimination power of handcrafted spectral signatures. However, due to the complex design of modern remote sensing sensors, the captured HSI data may contain both noise and spectral redundancy or correlations, along with a lack of sufficient training samples, thus limiting the HSI classification performance of traditional pattern recognition and machine learning methods.

Recently, deep learning methods have been successfully adopted in the HSI domain and proved effective for image classification tasks [5, 6]. Among various developments, stacked autoencoders (SAEs) and deep belief networks (DBNs) are the most popular unsupervised techniques to transform spectral-spatial features for HSI classification. Convolutional neural networks (CNNs) [7] have shown to be effictive automatic feature extractors, able to handle the spatial relation among pixels and images. Therefore, they have become state-of-the-art approaches for HSI classification [8].

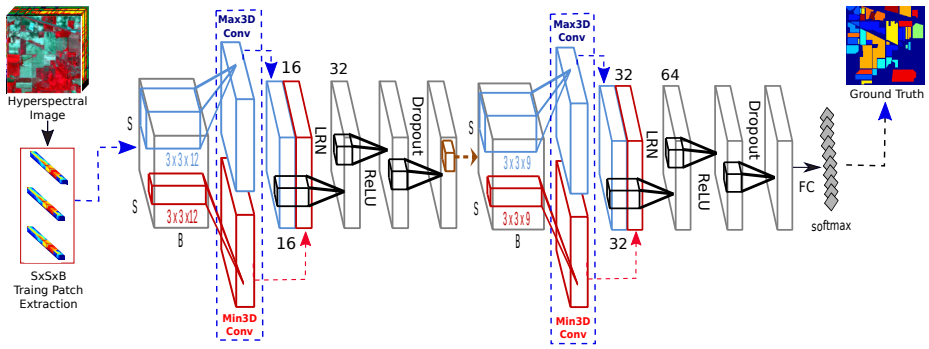


Fig. 1. New 3-D MaxMin CNN for HSI classification.

In order to jointly extract spectral-spatial feature representations from the input data, 3-D convolutions have received special attention in HSI classification [9] due to their ability to process multidimensional input arrays. In this sense, Zhong *et al.* proposed a spectral-spatial residual network (SSRN), improving the obtained performance by introducing 3-D convolutions (coupled with residual blocks) to reduce the vanishing gradient problem that arises while extracting joint spectral-spatial features for classification [10]. The SSRN model performance can be further improved using a spectral-spatial attention mechanism [11]. Moreover, Paoletti *et al.* introduced the deep pyramid residual network in order to gradually increase the feature maps of each residual block to process the spectral–spatial features uncovered by standard CNNs, involving more locations as the model depth increases while balancing the workload among all units [12]. Deep capsule networks have been successfully adapted to extract spectral-spatial patterns from 3-D raw HSI datasets for classification [13]. In addition to those deep structures, that process spectral-spatial features simultaneously, these features can also be processed separately and then joined together. For instance, 3-D and 2-D CNNs are sequentially combined to model more robust spectral-spatial features, and the resulting HybridSN has attracted significant attention in the community [14]. Recently, Dong *et al.* introduced cooperative spectral-spatial attention dense networks to re-calibrate the extracted feature maps for HSI classification [15]. The above mentioned networks provide satisfactory performance. However, the increasing number of convolutional layers may hamper the network convergence and produce below standard results. As a result, the deep network may suffer from a large number of trainable parameters implying overfitting problems when few training samples are available.

Existing CNN models successively apply several linear combination filter blocks followed by non-linear activation functions, to capture non-linear relations within the data. At the end, the CNN has a fully connected layer for the final classification. The rectified linear unit (ReLU) [16] is one of the applied activation function. In the literature [17] it can be observed that the standard CNN architecture performs well when the ReLU activation functions mitigate negative information pieces obtained from convolutional feature maps. To allow transmission of some negative information through the network, the original ReLU function can be slightly modifying into the parametric ReLU (PReLU). The network requires propagating both positive and negative direction information. However, due to the strong negative detection ability of the activation functions, the developed networks may fail to propagate all necessary information, which can limit classification performance of the network. The intrinsic complexity of HSI data requires transmission of both positive and negative responses from the convolutional layers to achieve correct classification. To overcome these limitations, this paper introduces a new Max-Min convolutional neural network that prevents the network from learning the opposite filter for HSI classification. The new network gradually increases the heterogeneity of high-level spectral-spatial features across the MinMax convolutional layers, which helps to enhance HSI classification performance of the network, where positive and negative learned features reduce the depth of the network significantly while preserving classification performance.

The rest of this paper is organized as follows. The new `MinMax-CNN` is introduced in Sect. 2. Section 3 reports experimental results. Finally, conclusions are given in Sect. 4

2. A NEW CLASSIFICATION FRAMEWORK

`MaxMin-CNN` aims to propagate more information through its successive convolutional blocks by successfully enabling the convolutional filters and their negative counterparts to detect both positive and negative patterns in the convolutional feature maps. A graphical overview of the new `MaxMin-CNN` model is shown in Fig. 1. As can be observed, the developed network consists of two 3-D MaxMin convolution blocks that are followed by a local response normalization module and the ReLU activation function, respectively. A detailed description of this architecture is given in the following subsections.

2.1. 3-D Max-Min Convolution with Local Response Norm

In CNNs, the convolutional filter bank plays an important role as it allows to learn more discriminative features while preserving the sparsity constraint. Moreover, CNNs are also known to be automatic extractors of high level feature representations from raw images [17]. To extract joint spectral-spatial discriminative features, 3-D convolution operations are adopted as the basic building block of the `MaxMin-CNN` model.

The 3-D convolution ($Conv3D$) receives in layer $(\ell+1)$ raw HSI cubes $X_j^l \in \mathcal{R}^{S\times S\times B}$ as input. The layer consist of $x^{(\ell+1)}$ trainable filters of size $k^\ell \times k^\ell \times d^\ell$ with a stride of size (s_1, s_1, s_2) in the spatial (height, width) and spectral depth dimensions. The size of generated layer $\ell+1$ in the 3-D convolutional feature map is $S^{\ell+1}\times S^{\ell+1}\times B^{\ell+1}$, where the spatial height and width are given by $S^{\ell+1} = [1 + (S^\ell - k^{\ell+1})/s_1]$ and the channel depth is given by $B^{\ell+1} = [1 + (B^\ell - d^{\ell+1})/s_2]$. Feature map λ based on layer $(\ell+1)$ in the 3-D convolution with local response normalization can be mathematically defined as:

$$\begin{aligned} X_\lambda^{\ell+1} &= ReLU\Big(\sum_{j=1}^{x^\ell} \mathcal{F}_{LRN}(X_j^\ell) * W_\lambda^{\ell+1} + b_\lambda^{\ell+1}\Big) \\ \mathcal{F}_{LRN}(X_{x,y}^\ell) &= X_{x,y}^\ell / (\alpha \sum_{v=\max(0,i-n/2)}^{\min(B-1,i+n/2)} (X_{x,y}^\ell)^2 + \epsilon)^\beta \end{aligned} \tag{1}$$

where $\mathcal{F}_{LRN}(.)$ represents the local response normalization (LRN) operation on layer $(\ell+1)$ of feature cube X_j^ℓ, having B feature channels. The hyperparameter α is a normalization parameter, β is a contrast constant, ϵ is a small number to avoid dividing by zero, and n is the size of the neighborhood and can be determined from the training set, $W_\lambda^{\ell+1}$ and $b_\lambda^{\ell+1}$ are layer $(\ell+1)$ kernel parameter weights and bias of the filter bank λ, ReLU$(z) = \max(0, z)$ and * denotes the 3-D convolution operation.

The nonlinear ReLU plays a key role in Eq. (1). It provids an identity mapping for the non-negative values and is zero otherwise. Therefore, it has a strong negative detection property. In order to transmit more information from the convolutional layer, it is important to forward both the positive and negative patterns through the ReLU nonlinearity. This is possible if the network allows both the filter $W^{\ell+1}$ and its negative counterparts $-W^{\ell+1}$ in convolution layer $(\ell+1)$ to learn the discriminative sparsity constraint. Let the negative filter be represented by $-W^{\ell+1}$ and qualify all the input on X_j^ℓ as follows:

$$X_j^\ell * -W^{\ell+1} = -(X_j^\ell * W^{\ell+1}) \tag{2}$$

It can be observed in Eq. (2) that, if a pattern is filtered by $-W^{\ell+1}$ with a strong detection on X_j^ℓ, then -$(X_j^\ell * W^{\ell+1})$ will be sufficiently high and positive, whereas $X_j^\ell * W^{\ell+1}$ will be sufficiently low and negative. In this way, it preserves both positive and negative generated information from the convolutional feature maps using either filter $W^{\ell+1}$ or $-W^{\ell+1}$. It is obvious that the number of convolutional feature maps becomes double in each convolution block, and enables the convolutional layer to use less filters in the network. Eq. (1) can be rewritten for the negative filter bank $-W^{\ell+1}$ as follows:

$$X_\lambda^{\ell+1} = ReLU\Big(\sum_{j=1}^{x^\ell} \mathcal{F}_{LRN}(X_j^{\ell)} * -W_\lambda^{\ell+1} + b_\lambda^{\ell+1}\Big) \tag{3}$$

2.2. New MaxMin Convolutional Neural Network

HSIs are acquired by encoding the same region in different spectral bands and, due to the presence of both noise and band correlation, it is desirable to consider the HSI correlation during the processing. As compared to conventional images, the intrinsic HSI complexity still limits classification performance for many existing CNN architectures. In order to extract joint spectral-spatial feature representation and propagate more information through the ReLU nonlinear function, we have designed a 3-D Max-Min convolution neural network for HSI classification. `MaxMin-CNN` takes 3-D raw HSI cubes, $X \in \mathcal{R}^{S\times S\times B}$ as input and it can be represented using $\mathcal{F}_{MaxMin}$ as:

$$\widehat{y}_{L_C} = \mathcal{F}_{MaxMin}(X, \theta) \tag{4}$$

where θ is a trainable parameter throughout the `MaxMin-CNN` network, $\widehat{y}_{L_C}$ is the number of predicted land cover classes at the $softmax$ layer, and the layer-wise design of $\mathcal{F}_{MaxMin}$ is shown in Fig. 1. The network consists of two 3-D MaxMin convolutional blocks and the initial block consists of $\{MaxMin3D \Rightarrow LRN \Rightarrow ReLU \Rightarrow Dropout\}$, while the latter one consists of $\{MaxMin3D \Rightarrow LRN \Rightarrow ReLU \Rightarrow AdaptiveAvgPool2D\}$. After the extraction of MaxMin convolutional features, we flatten them into a vector and, then, a $softmax$ function is adopted to calculate the class probability. The objective of the network is optimized using the well-known cross-entropy function, defined by:

$$Loss^{CE} = -\frac{1}{M}\sum_{m=1}^{M}\sum_{c=1}^{\ell_c} y_{L_c}^m \log(\widehat{y}_{L_c}^m) \tag{5}$$

where $y_{L_c}^m$ and $\widehat{y}_{L_c}^m$ are the actual and predicted class labels, M and L_c are the minibatch samples and the total number of land-cover classes, respectively. The parameters of the `MaxMin-CNN` network are learned through the Adam optimizer [18].

3. EXPERIMENTAL RESULTS

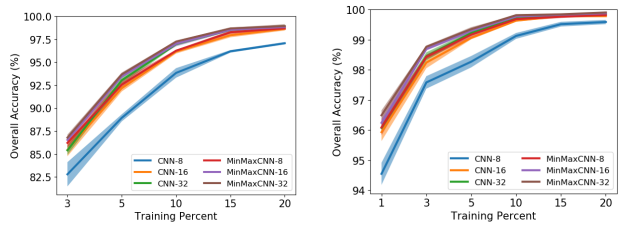
In order to evaluate the classification ability of the new network, experiments are performed with three well-known HSI datasets, i.e., **Indian Pine**s (IP), **University of Pavia** (UP), and **University of Houston** (UH).

Table 1. Architectural Details of a Basic CNN Model

Layer ID	Kernel/Neurons	LocalResponseNorm	Act. function	Dropout
MinMax3D_A	$bands \times 3 \times 3 \times 1 \times \{12, 8, 24\}$	Yes	RELU	Yes
MinMax3D_B	$bands \times 1 \times 1 \times 1 \times \{9, 3, 17\}$	Yes	RELU	No
AVG pool		Average Pooling		
FC	$n_{classes}$	No	Softmax	

Table 2. Comparison of a standard CNN with `MaxMin-CNN` using the fixed training sets available for IP, UP and UH scenes in `http://dase.grss-ieee.org`.

Class	Indian Pines		University of Pavia		University of Houston	
	CNN	MinMaxCNN	CNN	MinMaxCNN	CNN	MinMaxCNN
0	39.09	**80.91**	**89.13**	88.55	**80.72**	80.63
1	69.95	**70.54**	77.63	**81.43**	**98.25**	97.22
2	78.40	**84.29**	**62.39**	61.43	**98.20**	97.80
3	33.04	**35.95**	**96.40**	94.92	**84.34**	83.58
4	**88.58**	88.50	**99.44**	99.32	**99.98**	99.89
5	96.80	**97.18**	89.83	**91.28**	92.87	**95.35**
6	0.00	0.00	91.98	**93.75**	72.45	**75.00**
7	97.51	**98.93**	**97.91**	97.14	78.86	**85.29**
8	86.67	**88.89**	**97.28**	96.19	85.16	**87.55**
9	**61.24**	60.04	-	-	**60.11**	56.56
10	78.25	**78.83**	-	-	90.83	**91.95**
11	72.44	**78.98**	-	-	95.49	**98.00**
12	91.94	**95.55**	-	-	**79.45**	77.19
13	89.14	**91.14**	-	-	99.82	**100.0**
14	**66.29**	51.68	-	-	98.5	**98.83**
15	85.00	**87.00**	-	-	-	-
OA	78.06	**79.44**	84.57	**86.17**	86.03	**86.83**
AA	70.89	**74.28**	89.11	**89.34**	87.67	**88.32**
$\kappa \times 100$	75.02	**76.59**	80.09	**81.95**	84.83	**85.70**
Parameters	**22785**	85825	**8266**	29498	**42928**	161216



(a) (b)

Fig. 2. OA results obtained by `MaxMin-CNN` and by a standard CNN for the IP (a) and UP (b) scenes (average results after 5 Monte Carlo experiments).

3.1. Classification Results

To measure the classification performance of `MaxMin-CNN`, we use three common quantitative metrics: overall accuracy (OA), average accuracy (AA), and kappa coefficient (κ). Since `MaxMin-CNN` utilizes the basic CNN model shown in Table 1 as its underlying architecture, we compare the classification performance of `MaxMin-CNN` with that of the basic model in Table 1. The used hardware is X Generation Intel® Core™i9-9940X processor with 128GB of DDR4 RAM and NVIDIA Titan RTX GPU.

Table 2 shows the OA, AA and κ coefficient achieved by the baseline CNN in Table 1 and the `MaxMin-CNN` for the IP, UP and UH datasets. It can be seen that the `MaxMin-CNN` outperforms the standard CNN in terms of the evaluated quantitative measures for the three HSI datasets. The achieved improvements of `MaxMin-CNN` in {OA, AA, κ} are $\{+1.38, +3.39, +1.57\}$, $\{+1.60, +0.23, +1.86\}$ and $\{+0.80, +0.65, 0.87\}$ for the IP, UP and UH datasets respectively compared to the baseline CNN network.

To illustrate the generalization ability of the `MaxMin-CNN` and the baseline CNN network varying the number of convolution kernels (i.e., 8, 16 and 32), we performed experiments based on randomly selected training samples of 3%, 5%, 10%, 15% and 20% of the IP dataset, whereas the chosen training set size percentages for the UP dataset are 1%, 3%, 5%, 10%, 15% and 20%. Fig. 2 depicts the obtained results. Due to the discriminative nature of the generated features, the new network can successfully propagate more information through the ReLU nonlinear function and, hence, the `MaxMin-CNN` achieves an high classification performance, even for a limited number of training samples.

Figure. 3 illustrates the generated classification maps for the UH dataset using the fixed available training set for this dataset. It can be seen that the classification map generated by `MaxMin-CNN` contains less noise and artifacts than the classification map generated by the baseline CNN.

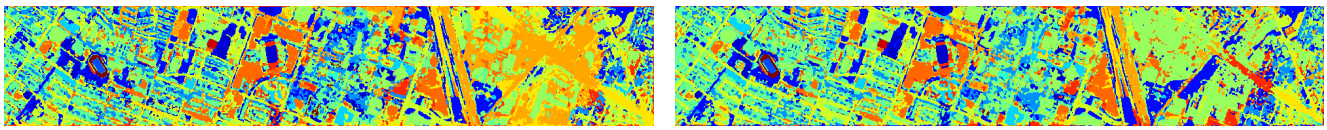
4. CONCLUSION

This paper introduces a simple yet efficient `MaxMin-CNN` model able to transmit more information through convolutional blocks. It overcomes the limitations of the ReLU function and reduces the requirement of additional convolutional layers in the network architecture by slightly increasing the number of trainable parameters. To extract discriminative spectral-spatial features from the original HSI, the nw framework uses two 3-D MaxMin convolutional blocks. This helps propagating both positive and negative information through the non-linearity layer. The intrinsic complexity of HSIs is addressed by learning both positive and negative filters. Therefore, it obtains a good classification generating twice the number of convolutional feature maps in each convolutional layer. Moreover, the new `MaxMin-CNN` network helps to reduce the number of successive convolutional blocks, since the 3-D MaxMin layer learns both filter banks. In the future, we are planing to develop residual networks with 3-D MaxMin convolution layers to achieve state-of-the-art classification performance while keeping a very small number of residual blocks.

5. REFERENCES

[1] L. Zhang, L. Zhang, D. Tao, and X. Huang, "On combining multiple features for hyperspectral remote sensing image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 3, pp. 879–893, 2011.

[2] M. E. Paoletti, J. M. Haut, X. Tao, J. P. Miguel, and



(a) CNN (86.03%) (b) MinMaxCNN (86.83%)

Fig. 3. Classification maps obtained by the baseline CNN (a) and `MaxMin-CNN` (b) for the University of Houston (UH) dataset, using the fixed available training set for this dataset.

A. Plaza, "A new gpu implementation of support vector machines for fast hyperspectral image classification," *Remote Sensing*, vol. 12, no. 8, p. 1257, 2020.

[3] Y. Bazi and F. Melgani, "Gaussian process approach to remote sensing image classification," *IEEE transactions on geoscience and remote sensing*, vol. 48, no. 1, pp. 186–197, 2009.

[4] Y. Chen, Z. Lin, and X. Zhao, "Riemannian manifold learning based k-nearest-neighbor for hyperspectral image classification," in *2013 IEEE International Geoscience and Remote Sensing Symposium-IGARSS*. IEEE, 2013, pp. 1975–1978.

[5] 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*, 2016, pp. 770–778.

[6] M. Blot, M. Cord, and N. Thome, "Max-min convolutional neural networks for image classification," in *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2016, pp. 3678–3682.

[7] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Convolutional neural networks for large-scale remote-sensing image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 2, pp. 645–657, 2016.

[8] M. Paoletti, J. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 158, pp. 279–317, 2019.

[9] A. B. Hamida, A. Benoit, P. Lambert, and C. B. Amar, "3-d deep learning approach for remote sensing image classification," *IEEE Transactions on geoscience and remote sensing*, vol. 56, no. 8, pp. 4420–4434, 2018.

[10] Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral–spatial residual network for hyperspectral image classification: A 3-d deep learning framework," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 2, pp. 847–858, 2018.

[11] M. Zhu, L. Jiao, F. Liu, S. Yang, and J. Wang, "Residual spectral-spatial attention network for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, 2020.

[12] M. E. Paoletti, J. M. Haut, R. Fernandez-Beltran, J. Plaza, A. J. Plaza, and F. Pla, "Deep pyramidal residual networks for spectral-spatial hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, no. 99, pp. 1–15, 2018.

[13] M. E. Paoletti, J. M. Haut, R. Fernandez-Beltran, J. Plaza, A. Plaza, J. Li, and F. Pla, "Capsule networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, 2018.

[14] S. K. Roy, G. Krishna, S. R. Dubey, and B. B. Chaudhuri, "Hybridsn: Exploring 3-d–2-d cnn feature hierarchy for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 2, pp. 277–281, 2019.

[15] Z. Dong, Y. Cai, Z. Cai, X. Liu, Z. Yang, and M. Zhuge, "Cooperative spectral-spatial attention dense network for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, 2020.

[16] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *ICML*, 2010.

[17] 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.

[18] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.