Article

Heterogeneous Deep Model Fusion for Automatic Modulation Classification

Duona Zhang [1], Wenrui Ding [1], Baochang Zhang [1], Chunyu Xie [1], Chunhui Liu [1], Jungong Han [2] and Hongguang Li [1,*]

[1] School of Beihang University, Beijing 100083, P. R. China; zhangduona@buaa.edu.cn; ding@buaa.edu.cn; bczhang@buaa.edu.cn; yuxie@buaa.edu.cn; liuchunhui2134@buaa.edu.cn

[2] School of Computing & Communications, Lancaster University, Lancaster LA1 4WA, U.K.; jungonghan77@gmail.com

* Correspondence: lihongguang@buaa.edu.cn; Tel.: +86-10-82317391

Abstract: Deep learning has recently attracted much attention due to its excellent performance in processing audio, image, and video data. However, few studies are devoted to the field of automatic modulation classification (AMC). It is one of the most well-known research topics in communication signal recognition, which remains challenging for traditional methods due to the complex disturbance from other sources. This paper proposes a heterogeneous deep model fusion (HDMF) method to solve the problem in a unified framework. The contributions include: 1) The convolutional neural network (CNN) and long short-term memory (LSTM) are combined by two different ways without prior knowledge involved; 2) A large database, including eleven types of single-carrier modulation signals with various noises as well as a fading channel, is collected with various signal-to-noise ratios (SNRs) based on a real geographical environment; and 3) Experimental results demonstrate that HDMF is super capable of coping with the AMC problem, and achieves much better performance when compared with the independent network. The source code and the database will be publically available.

Keywords: Deep learning; automatic modulation classification; classifier fusion; convolutional neural network; long short-term memory

1. Introduction

Communication signal recognition is of great significance for several daily applications, such as operator regulation, signal feature map generation, and user identification. One of the main objectives of signal recognition is to detect the communication resources, which ensures the reliability of communications. To achieve this objective, automatic modulation classification (AMC) is indispensable because it can help users identify the modulation mode within a frequency band, which benefits the communication reconfiguration and electromagnetic environment analysis. AMC plays an essential role in obtaining digital baseband information from the signal when only limited knowledge about the parameters is available. Such a technique is widely used in both military and civilian applications, e.g., intelligent cognitive radio and anomaly detection [1]-[2], which have attracted much attention from researchers in the past decades.

Basically, existing AMC algorithms can be divided into two main categories [3], namely, likelihood-based (LB) methods and feature-based (FB) methods. LB methods require calculating the likelihood function of received signals for all modulation modes and then make decisions in accordance with maximum likelihood ratio test [3]. LB methods usually generate accurate classification results but suffer from heavy computational cost. Alternatively, a traditional FB method consists of two parts, namely, feature extraction and classifier, where classifier identifies

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 January 2018 doi:10.20944/preprints201801.0097.v1

digital modulation modes in accordance with the effective feature vectors extracted from the signals. As opposite to the LB methods, the FB methods are computationally light but may not be theoretically optimal. To date, several FB methods have been validated effective on the AMC problem. For instance, they successfully extract features from various time-domain waveforms, such as cyclic spectrum [4], high-order cumulant [6], and wavelet coefficients. Afterwards, a classifier is used for final classification based on features mentioned above. With the development of learning algorithms, the performances have been improved, such as from the shallow neural network [7] and decision tree to the support vector machine (SVM). Recently, deep learning is widely applied to audio, image, and video processing, facilitating the applications such as face recognition and voice discrimination [8]. However, a few works are done based on deep learning in the field of communication.

Although researchers have developed various algorithms to implement AMC of digital signals, these methods are suitable for simple communication equipment and struggle in the real-world applications where more complicated equipment is in use, because: 1) they cannot handle complex disturbance from other sources; 2)they usually separate feature extraction and classification process so that the information loss is inevitable; and 3) those methods must use distributed receivers to collect in-phase and quadrature signals, which costs additional storage space and bandwidth. In this paper, we propose to realize AMC using the convolution neural networks (CNNs) [9], long short-term memory (LSTM) [10], and their fusion model to directly process the time-domain waveform data.

CNNs exploit spatially-local correlation by enforcing a local connectivity pattern between neurons of adjacent layers. The convolution kernels are also shared in each sample for the rapid expansion of parameters caused by the fully connected structure. Sample data are still retained in the original position after convolution such that the local features are well preserved. Despite its great advance in spatial feature extraction, CNNs could not model the changes in time series well. As is known to us, the temporal property of data is important for AMC applications. As a variant of recurrent neural network (RNN), LSTM uses the gate structure to realize the information transfer of the network in time sequence, which reflects the depth in time series. Therefore, LSTM has a super capacity to process the time series data.

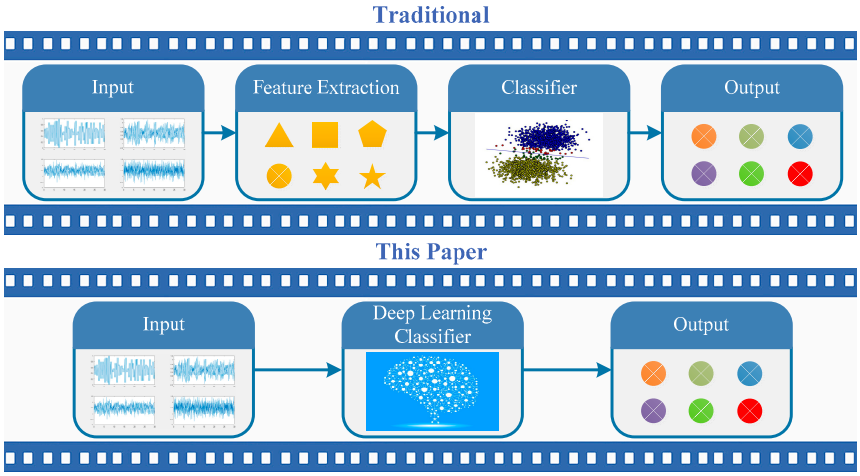


Figure 1. Illustration of the traditional and classifier methods in this study for AMC. The traditional method needs to extract features as preprocessing and suffers from the perturbation caused by high computational complexity and effective information loss. By contrast, the classifier based on deep learning is used to process signal data directly in this study. AMC is implemented more efficiently with a heterogeneous deep model fusion (HDMF) method.

This paper proposes a heterogeneous deep model fusion (HDMF) method to solve the AMC problem in a unified framework. The framework is shown in Figure 1. Different from conventional

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 January 2018 doi:10.20944/preprints201801.0097.v1

methods, AMC does not need to rely on other methods to extract features. In addition, the modulation modes can be obtained directly on the basis of the previous training model. Such improvement helps the communication system to overcome the shortcoming that cognition based on a separate feature and classification process and enhance classification accuracy. We use CNNs and LSTM to process the time domain waveforms of modulation signal. Eleven types of single-carrier modulation signal samples (e.g., MASK, MFSK, MPSK, and MQAM) added with additive white Gaussian noise (AWGN) and a fading channel are generated under various signal-to-noise ratios (SNRs) based on an actual geographical environment. Two kinds of HDMFs based on the serial and parallel modes are proposed to increase the classification accuracy. The results show that HDMFs achieved much better results than the single CNN or LSTM method, when SNR is in the range of 0–20 dB. In a summary, the contributions are as follows:

1) CNNs and LSTM are fused based on the serial and parallel modes to solve the AMC problem, thereby leading to two HDMFs. Both are trained in the end-to-end framework, which can learn features and make classifications in a unified framework.

2) The experimental results show that the performance of the fusion model has been significantly improved compared with the independent network and also traditional wavelet+SVM. The serial version of HDFM achieves much better performance than the parallel version.

3) We collect communication signal data sets, which approximate the transmitted wireless channel in the actual geographical environment. Such datasets are very useful for training networks like CNNs and LSTM.

The rest of this paper is organized as follows. Section II briefly introduces the related works. Section III introduces the principle of digital modulation signal and deep learning classification methods. Section IV presents the experiments and analysis. Section V summarizes the paper.

2. Related Works

AMC is a typical multi-classification problem in the field of communication. This section briefly introduces several feature extraction and classification methods in the traditional AMC system. The CNN and LSTM models are also presented.

2.1. Conventional works based on separated feature and classifiers

Traditionally the feature and classifier are separately built for an AMC system. For example, the envelope amplitude of signal, the power spectral variance of signal, and the mean of absolute value signal frequency, was extracted in [11] to describe the signal from several different aspects. Yang and Soliman used the phase probability density function for AMC [12]. Meanwhile, traditional methods usually combine instantaneous and statistical characteristics. Shermeh used the fusion of high-order moments and cumulants with instantaneous characteristics for AMC [13]-[14]. The features can describe the signals using both absolute and relative levels. In addition, the high-order characteristics can eliminate the effects of noise. The sixth and eighth statistics are widely used in several methods.

Classical algorithms have been widely used in the AMC system. Panagiotou et al. considered AMC as a multiple-hypothesis test problem and used decision theory to obtain the results [15]. They assumed that the phase of AWGN was random and dealt with the signals as random variables with the known probability distribution. Finally, the generalized likelihood ratio test or the average likelihood ratio test was used to obtain the classification results by the threshold. The classifiers were then used in the AMC system. In [16], shallow neural networks and SVM were used as classifiers. In [17]-[18], modulation modes were classified using CNNs with high-level abstract learning capabilities.

However, the traditional classifiers either need preprocessing to extract features or rely on the detailed prior information. This approach has led to negative influences of the classification performance.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 January 2018 doi:10.20944/preprints201801.0097.v1

2.2. CNN – based methods

Advantage of CNNs is achieved with local connections and tied weights followed by some form of pooling which results in translation invariant features. Furthermore, another benefit is that they have many fewer parameters than fully connected networks with the same number of hidden units. In [9], the authors treat the communication signal as a 2 dimensional data which similar to an image and take it as a matrix to a narrow 2D CNN for AMC. They study the adaptation of CNN to the time domain IQ data. A 3D CNN was used in [19]-[20] to process video information. The result showed that CNN multi-frames were considerably more suitable than a single-frame network for video cognition. In [21], Luan et al propose a Gabor Convolutional Networks, which combines Gabor filters and CNN model, to enhance the resistance of deep learned features to the orientation and scale changes. Recently, Zhang et al apply one-two-one network to compression artifacts reduction in remote sensing [22]. This motivates us to solve the AMC problem.

2.3. LSTM – based methods

Various models have been used to process sequential signal, such as hidden semi-Markov models [23], conditional random fields [24], and finite-state machines [25]. Recently, RNN became well-known with the development of deep learning. As a special RNN, LSTM has been widely used in the field of voice and video because of its ability to handle gradient disappearance in traditional RNNs. It has the less conditional independence hypothesis compared with the previous models and facilitates integration with other deep learning networks. Researchers have recently combined spatial/optical flow CNN features with vanilla LSTM models for global temporal modeling of videos [26]-[30]. These studies have demonstrated that deep learning models have a significant effect on action recognition [27], [29], [31] and video description [30], [32]. But to our best of knowledge, the fusion of CNN and LSTM is never investigated to solve the AMC problem.

3. Heterogeneous Deep Model Fusion

3.1. Digital modulation signal description

The received signal in the communication system can be expressed as follows:

$$y(t) = x(t) \cdot c(t) + n(t), \tag{1}$$

where $x(t)$ is the efficient signal from the transmitter, $c(t)$ represents the transmitted wireless channel on the basis of the actual geographical environment, and $n(t)$ denotes the AWGN. The digital modulation signals $x(t)$ can be expressed as follows:

$$x(t) = (A_c + jA_s)e^{j(2\pi ft+\theta)}g(t-nT) = (A_c \cos(2\pi ft+\theta) - A_s \sin(2\pi ft+\theta))g(t-nT), 0 \le t \le NT, \tag{2}$$

where A_c and A_s are the amplitudes of the in-phase and quadrature channel, respectively; f stands for the center frequency; θ is the initial phase of the carrier; and $g(t-nT)$ represents the digital sampling pulse signal. In the case of ASK, FSK, and PSK, A_s is zero. In accordance with the digital baseband information, ASK, FSK, and PSK change A_c, f, and θ in the range of $0-M$, $1-M$, and $0-2\pi/M$, respectively, with time. By contrast, QAM fully utilizes the orthogonality of the signal. After dividing the digital baseband into I and Q channels, the information is integrated into two identical frequency carriers with phase difference of 90° using ASK modulation mode, which significantly improves the bandwidth efficiency.

As one of the most common noise, AWGN is always true whether or not the signal is in the communication system. The power spectrum density is a constant at all frequencies, and the noise amplitude obeys the Gauss distribution.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 January 2018 doi:10.20944/preprints201801.0097.v1

3.2. CNNs

CNNs are a hierarchical neural network that contains convolution, activation, and pooling layers. In this study, the input of the CNN model is the data of signal time-domain waveform. The difference among the classes of modulation methods is deeply characterized by the stacking of multiple convolutional layers and nonlinear activation. Different from the CNN models in the image domain, we use a series of one-dimensional convolution kernels to process the signals.

Each convolution layer is composed of a number of kernels with the same size. The convolution kernel is common in each sample; thus, each kernel can be called a feature extraction unit. This method of sharing parameters can effectively reduce the number of learning parameters. Moreover, the feature extracted from convolution remains in the original signal position, which preserves the temporal relationship well within the signal. In this paper, ReLU is used as the activation function. We do not use the pooling layer for dimensionality reduction because the amount of signal information is relatively small.

3.3. LSTM

Traditional RNNs are unable to connect the information as the gap grows. The vanishing gradient can be interpreted as the forgetting of the human brain. LSTM overcomes this drawback using gate structures that optimize the information transfer among memory cells. The particular structures in memory cells include the input, output, and forget gates. An LSTM memory cell is shown in Figure 2.

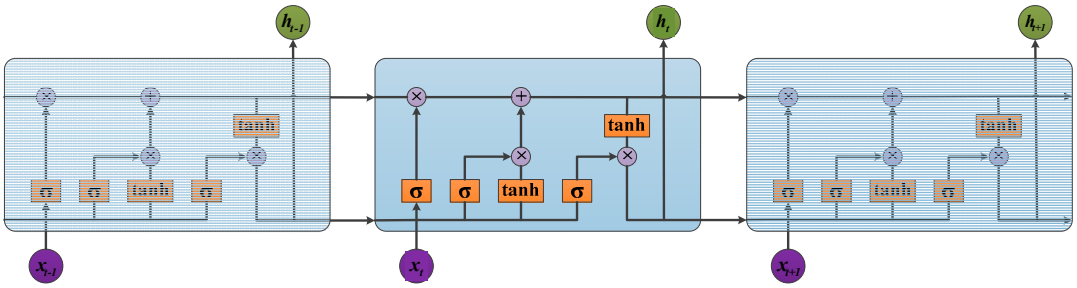


Figure 2. LSTM memory cell structure.

The iterating equations are as follows:

$$f_t = sig\,\mathrm{mod}(W_f \cdot [h_{t-1}, x_t] + b_f), \tag{3}$$

$$i_t = sig\,\mathrm{mod}(W_i \cdot [h_{t-1}, x_t] + b_i), \tag{4}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \tag{5}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t, \tag{6}$$

$$o_t = sig\,\mathrm{mod}(W_o \cdot [h_{t-1}, x_t] + b_o), \tag{7}$$

$$h_t = o_t \cdot \tanh(C_t), \tag{8}$$

where W is the weight matrix; b is the bias vector; i, f, and o are the outputs of the input, forget, and output gates, respectively; C and h are the cell activations and cell output vectors, respectively; and $sig\,\mathrm{mod}$ and $\tanh$ are nonlinear activation functions.

Standard LSTM usually models the temporal data in the backward direction but ignores the forward temporal data, which has a positive impact on the results. In this paper, a method based on bidirectional LSTM (Bi-LSTM) is exploited to realize AMC. The core concept is to use a forward and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 January 2018 doi:10.20944/preprints201801.0097.v1

a backward LSTM to train a sample simultaneously. Similarly, the architecture of Bi-LSTM network is designed to model the time domain waveforms from past and future.

3.4. Fusion model based on CNN and LSTM

The HDMFs are established based on the fusion model in serial and parallel ways to enhance the classification performance. The specific structure of the fusion model is shown in Figure 3.

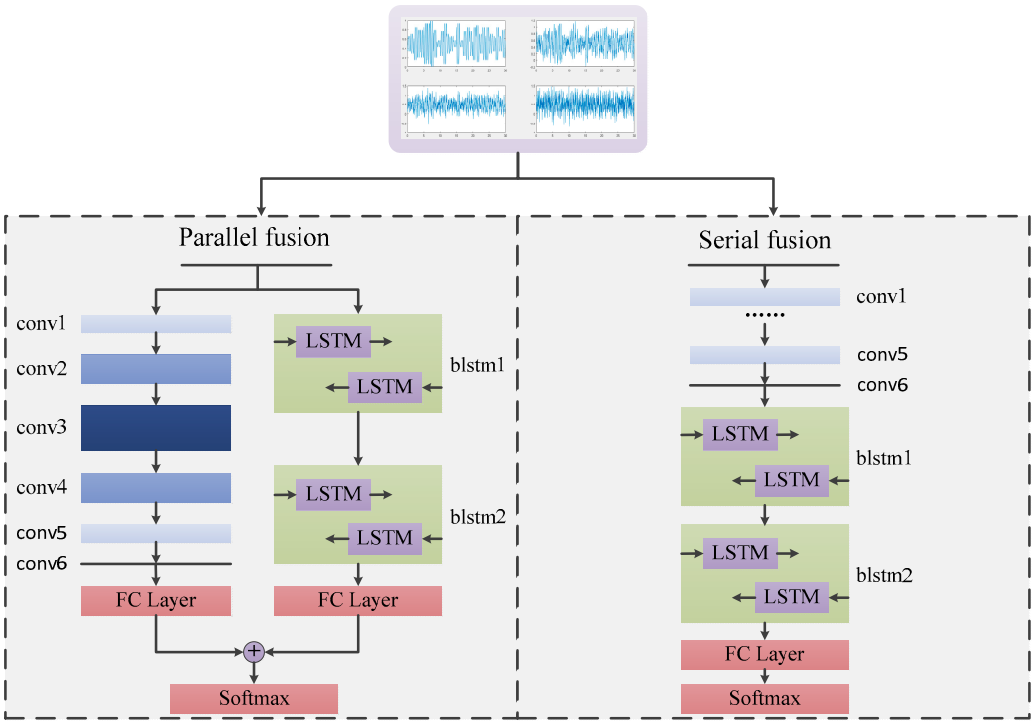


Figure 3. Fusion model structure of HDMF in parallel and series modes. We note that two HDMF models are used separately to solve the AMC problem.

The modulated communication signal has local special change characteristics. Meanwhile, the data has temporal characteristics similar to voice and video. The fusion models exploit complementary advantages on the basis of these two features.

The six layers of CNNs are used to characterize the differences between the digital modulation modes in the fusion model. The kernel numbers of the convolutional layer are different for each layer. The number of convolutional kernel in the first three layers increases gradually, which transforms the single-channel into multi-channel signal data. Such a transformation also helps to obtain effective features. Conversely, the number of convolutional kernel in the remaining layers reduces gradually. Finally, the result is restored to a single-channel data. Although the data format is same as the original signal, local features of the signal are extracted by multiple convolution kernels. This leads to the representation for the final classification based on CNNs. The remaining part of the fusion model uses the two-layer Bi-LSTM network to learn the temporal correlation of signals. The output of the upper Bi-LSTM is used as the input of the next layer.

The parallel fusion model (HDMF). The two networks are used to train samples simultaneously. The output of each network is then transformed into an 11-dimensional feature vector by the full connection layer. The resulting feature vectors represent the judgment of the modulation modes of the training samples by the two networks. We then combine the two vectors based on the sum operation as:

$$\ell_{total} = \omega_c \cdot \ell_c + \omega_l \cdot \ell_l , \tag{9}$$

and,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 January 2018 doi:10.20944/preprints201801.0097.v1

$$\omega_c + \omega_l = 1, 0 \le \omega \le 1, \tag{10}$$

The loss function of parallel fusion model consists of two parts, which are balanced by the given parameters.

Algorithm 1: Training HDMF(parallel)

1: Initialize the parameters θ_c in CNN, θ_l in LSTM, W, ω in the loss layer, the learning rate μ, and the number of iteration $t = 0$.

2: **While** the loss does not converge, **do**

3: $t = t + 1$

4: Compute the total loss by $\ell_{total} = \omega_c \cdot \ell_c + \omega_l \cdot \ell_l$.

5: Compute the backpropagation error $\frac{\partial \ell_{total}}{\partial x_i}$ for each x_i by $\frac{\partial \ell_{total}}{\partial x_i} = \omega_c \cdot \frac{\partial \ell_c}{\partial x_i} + \omega_l \cdot \frac{\partial \ell_l}{\partial x_i}$.

6: Update parameter W by $W - \mu \cdot \frac{\partial \ell_{total}}{\partial W} = W - \mu \cdot \omega_c \cdot \frac{\partial \ell_c}{\partial W} - \mu \cdot \omega_l \cdot \frac{\partial \ell_l}{\partial W}$

7: Update parameters ω_c and ω_l by $\omega_{c,l} - \mu \cdot \frac{\partial \ell_{c,l}}{\partial \omega_{c,l}}$.

8: Update parameter θ by $\theta_{c,l} - \mu \cdot \sum_i^m \frac{\partial \ell_{c,l}}{\partial x_i} \cdot \frac{\partial x_i}{\partial \theta_{c,l}}$.

9: **End while**

The serial fusion method (HDMF). It is similar to the encoder–decoder framework. In this study, the encoding process is implemented by CNNs, afterwards LSTM decodes the corresponding information. The features are extracted by the two networks, from simple representation to complex concepts. The upper convolutional layers can extract features locally. Then, the Bi-LSTM layers learn temporal characteristic from these representations.

For both kinds of fusion models, the final feature vectors are the probabilistic output of the softmax layer. The fusion models are trained in the end-to-end way even when different neural networks are used to address the AMC problem.

3.5. Implementation details and backpropagation

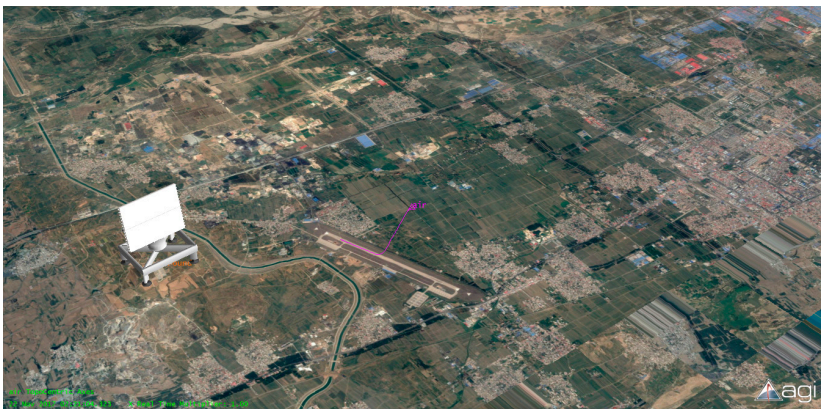


Figure 4. The geographic simulation environment.

The geographic simulation environment is shown in Figure 4, based on which we collect our datasets. We captured the unmanned aerial vehicle communication signal data set, which is developed by us based on STK, visual studio and MATLAB. We use TensorFlow [33] to design our

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 January 2018 doi:10.20944/preprints201801.0097.v1

deep models. The Adam method [34] is used to solve our model with 0.001 learning rate. The iterations are as follows:

$$m_t = \mu \cdot m_{t-1} + (1-\mu) \cdot g_t, \tag{11}$$

$$n_t = \nu \cdot n_{t-1} + (1-\nu) \cdot g_t^2, \tag{12}$$

$$\hat{m}_t = \frac{m_t}{1-\mu^t}, \tag{13}$$

$$\hat{n}_t = \frac{n_t}{1-\nu^t}, \tag{14}$$

$$\Delta\theta = -\frac{\hat{m}_t}{\sqrt{\hat{n}_t + \varepsilon}} \cdot \eta, \tag{15}$$

where m_t and n_t are the first and second moment estimations of the gradient, which represent the estimation of $E(g_t)$ and $E(g_t^2)$, respectively; $\hat{m}_t$ and $\hat{n}_t$ are the corrections of m_t and n_t, respectively, which can be regarded as the unbiased estimation of expectation; $\Delta\theta$ is the dynamic constraint of learning rate; and μ, ν, ε, and η are constants.

The fundamental loss and the softmax functions are defined as follows:

$$\ell(x, y) = -\log(p_y), \tag{16}$$

$$p_y = \frac{e^{z_y}}{\sum_i e^{z_i}} = \frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^{n} e^{W_j^T x_i + b_j}}, \tag{17}$$

where x is the input, y is the corresponding truth label, and z_i is the input for the softmax layer. The gradient of backpropagation is calculated as follows:

$$g_t = \frac{\partial \ell}{\partial z_j} = \frac{\partial \ell}{\partial p_y} \cdot \frac{\partial p_y}{\partial z_j} = -\frac{1}{p_y} p_y (I_{jy} - p_j) = p_j - I_{jy}, \tag{18}$$

where $I_{jy} = 1$ if $j = y$, and $I_{jy} = 0$ if $j \neq y$.

4. Results

4.1. Classification accuracy of CNN and LSTM models

When CNNs and LSTM solve the AMC problem, the classification accuracies of CNNs are reported with varying convolution layer depth from 1 to 4, the number of convolution kernels from 8 to 64, and the size of convolution kernels from 10 to 40. The classification accuracies of Bi-LSTM are tested when varying layer depth from 1 to 3 and number of memory cells from 16 to 128. Bi-LSTM used in the fusion model contains two layers. The number of convolution layers is 6. The number of convolution kernels in the first three layers is 8, 16, and 32, and the size of the convolution kernel is 10. The number of convolution kernels in the remaining layers is 16, 8, and 1, and the size of the convolution kernel is 20. The Bi-LSTM model consists of two layers with 128 memory cells.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 January 2018 doi:10.20944/preprints201801.0097.v1

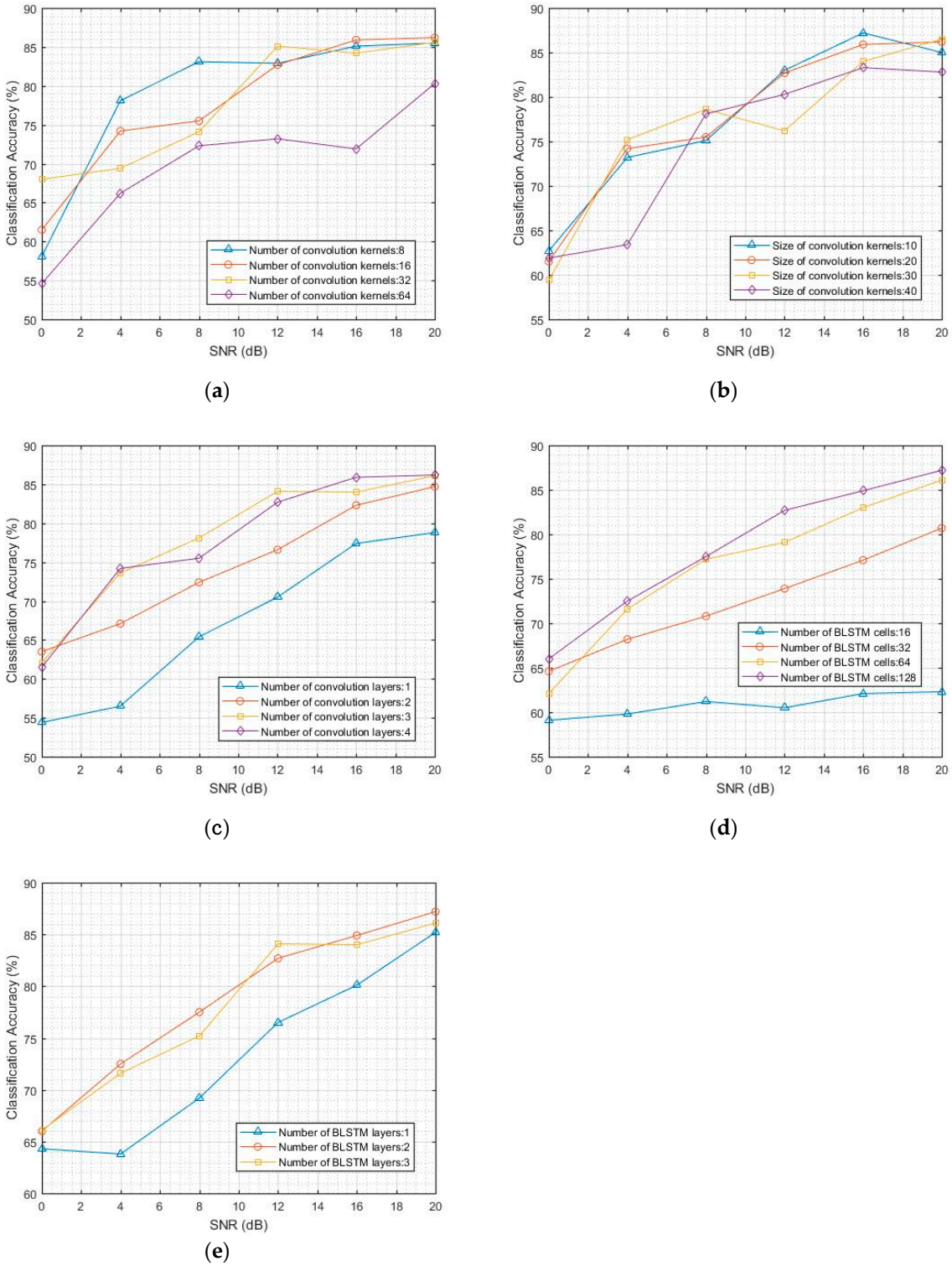


Figure 5. Classification accuracy of CNN and LSTM models. (a) Classification accuracy of CNN with the number of convolution kernels from 8 to 64; (b) Classification accuracy of CNN with the size of convolution kernels from 10 to 40; (c) Classification accuracy of CNN with the number of convolution layers from 1 to 4; (d) Classification accuracy of Bi-LSTM with the number of memory cells from 16 to 128; (e) Classification accuracy of Bi-LSTM with the number of hidden layers from 1 to 3.

When SNR is set from 0 dB to 20 dB, the classification accuracy of CNN and Bi-LSTM models is shown in Figure 5. The samples with SNR below 0 dB are not considered in this study. The classification results of the CNN models are shown in Figure 5a–c. The average classification accuracy of the CNN model for AMC can reach 75% with SNR from 0 dB to 20 dB. An excess of the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 January 2018 doi:10.20944/preprints201801.0097.v1

convolution kernels in each layer reduces the classification accuracy. The performance is better with the number of convolution kernels from 8 to 32. The CNN models with the size of convolution kernels from 10 to 40 have more or less the same classification accuracy. Increasing the number of convolution layers from 1 to 3 results in a performance boost. The classification results of the Bi-LSTM models are shown in Figure 5d–e. The results show that the Bi-LSTM model is more suitable for AMC than the CNN model. The average classification accuracy of Bi-LSTM is 77.5%, which is 1.5% higher than that of the CNN model. The performance is better with the number of memory cells from 32 to 128 than others. The Bi-LSTM models with the number of hidden layers more than 2 have essentially the same classification accuracy.

4.2. Comparison of classification accuracy between the deep learning models and the traditional method

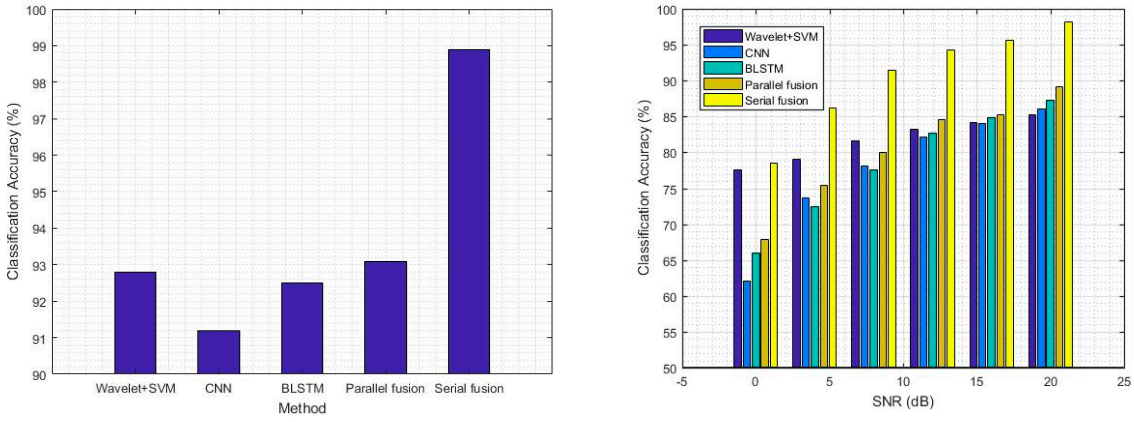
We have compared five methods, including both traditional and deep learning methods, based on the same data sets. The classification performance is as follows.

Table 1. Classification accuracy of different methods without noise.

Methods	Wavelet + SVM	CNN	Bi-LSTM	Parallel fusion	**Serial fusion**
Accuracy	92.8%	91.2%	92.5%	93.1%	**98.9%**

Table 2. Classification accuracy of different methods with SNR from 0 to 20dB

SNR Methods	20 dB	16 dB	12 dB	8 dB	4 dB	0 dB
Wavelet+SVM	85.2%	84.1%	83.2%	81.6%	79.0%	77.5%
CNN	86.1%	84.0%	82.1%	78.1%	73.6%	62.1%
Bi-LSTM	87.2%	84.9%	82.7%	77.5%	72.5%	66.0%
Parallel fusion	89.1%	85.2%	84.6%	80.0%	75.4%	67.9%
Serial fusion	**98.2%**	**95.6%**	**94.3%**	**91.5%**	**86.2%**	**78.5%**



(**a**) (**b**)

Figure 6. Comparison of classification accuracy between the deep learning models and the traditional method. (a) Classification accuracy of different methods without noise; (b) Classification accuracy of different methods with SNR from 0 dB to 20 dB.

The modified classifiers are established based on the fusion model in serial and parallel modes to increase the classification accuracy. As a result, we compare the classification accuracy of the methods on the basis of deep learning with the traditional method by using wavelet and SVM classifiers. The results are shown in Tables 1 and 2 and Figure 6. The results reveal that the fusion methods have a significant effect on improving classification accuracy. The average classification accuracy of parallel fusion model is 93% without noise, which is equal to the traditional method.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 January 2018 doi:10.20944/preprints201801.0097.v1

The classification accuracy of the parallel fusion model is 2% higher than the CNN model and 1% higher than the Bi-LSTM model. Moreover, the average classification accuracy of the serial fusion model is 99% without noise, which is 6% higher than the parallel fusion model. In fact, the fusion methods are more beneficial to the classification accuracy with the SNR from 0 dB to 20 dB compared with the noise-free situation. The average classification accuracy of the serial fusion method is 91%, which is 11% higher than the parallel fusion method.

The performances of the classifiers show that deep learning achieves high classification accuracy for AMC. Waveform local variation and temporal characteristics can be used to identify modulation modes. In comparison with CNN and Bi-LSTM, the performance of the HDMF methods is improved significantly because the classifiers can recognize the two features simultaneously. However, the performance of the serial fusion is considerably higher than that of the parallel fusion because the parallel method belongs to the decision-level fusion. The fusion can be viewed as a simple voting process for results. The serial method belongs to the feature-level fusion, which combines the two feature information to obtain the classification results.

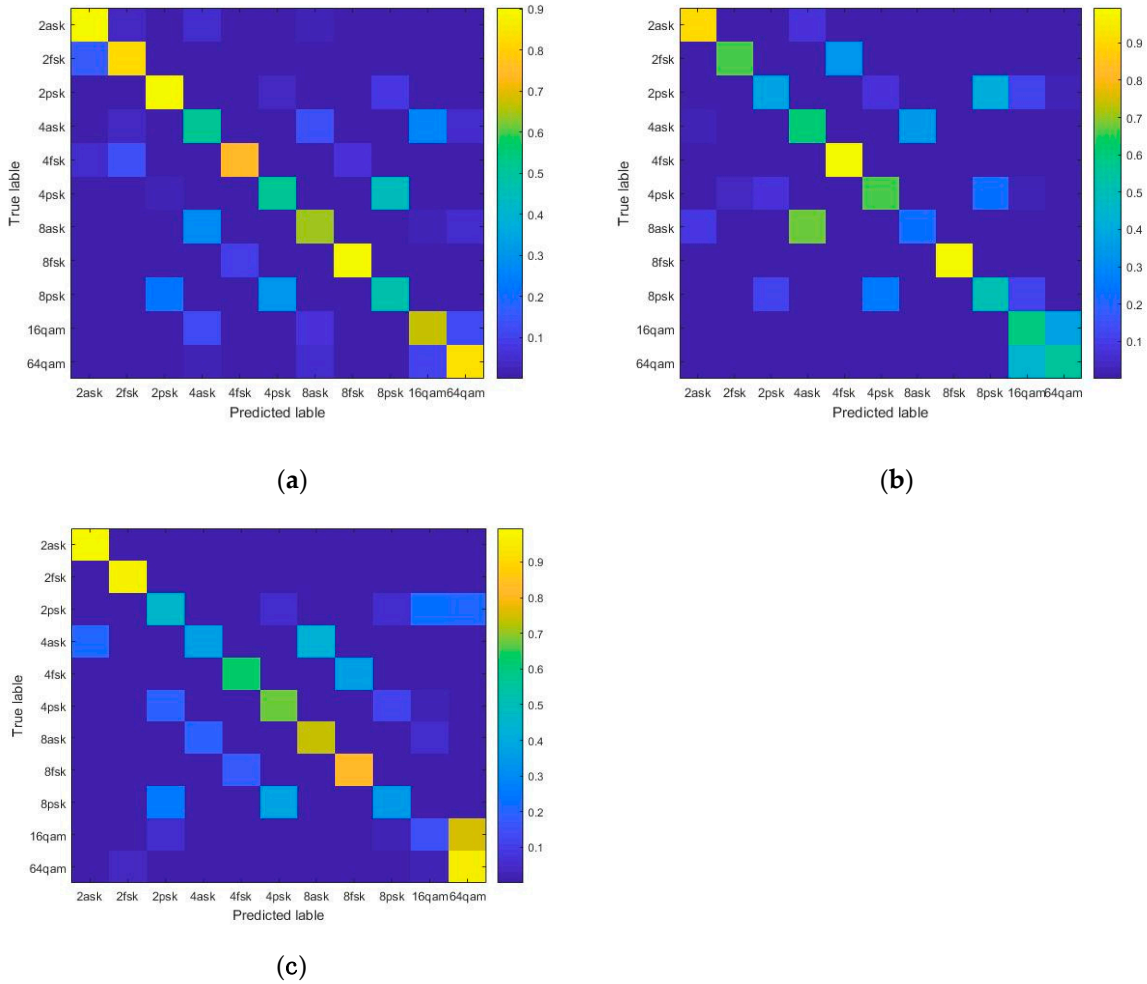


Figure 7. Confusion matrix of series fusion model. (a) Confusion matrix of series fusion model for 20 dB SNR; (b) Confusion matrix of series fusion model for 10 dB SNR; (c) Confusion matrix of series fusion model for 0 dB SNR.

In this study, the modulation mode of the samples includes two forms, namely, within-class and between-class modes. The confusion matrices show the identification results of the modulation modes by serial fusion model when the SNR is 20, 10, and 0 dB, respectively; the results are shown in Figure 7. When the SNR is 20 dB, a profound discrepancy is observed between the different

 doi:10.20944/preprints201801.0097.v1

modulation modes. The confusion result does not have the error. The decrease of SNR, PSK, and QAM is prone to misclassification within class, which is caused by the subtle differences in M-ary phase mode. Moreover, representing the phase difference by waveform amplitude is not evident. Furthermore, QAM can be considered as a combination of ASK and PSK in practice. The classifier can detect the different types of changes simultaneously even when the result is incorrect at low SNR. Therefore, only within-class misclassifications occur in the results.

5. Conclusions

In this study, we proposed the methods on the basis of deep learning to address the AMC problem in the field of communication. The classification methods are end-to-end processes, which reduce the additional steps to extract signal features compared with the traditional methods. First, the communication signal data set system is developed based on the actual geographical environment to provide the basis for related classification tasks. CNNs and LSTM are then used to solve the AMC problem compared with the traditional method. Furthermore, the modified classifiers based on the fusion model in serial and parallel modes are of great benefit to improve classification accuracy with the SNR from 0 dB to 20 dB. The serial fusion mode has the best performance compared with other modes. The confusion matrices significantly reflect the shortcomings of the classifiers in this study. We will overcome these shortcomings and further research on AMC in the future.

Acknowledgments: This work is supported by the National Natural Science Foundation of China (Grant no. 91538204).

References

1. M. Zheleva, R. Chandra, A. Chowdhery, A. Kapoor, and P. Garnett, TX miner: Identifying transmitters in real-world spectrum measurements, *IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, Sept 2015, pp. 94–105.
2. S. S. Hong and S. R. Katti, Dof: A local wireless information plane, *Proceedings of the ACM SIGCOMM 2011 Conference*, ser. SIGCOMM '11. New York, NY, USA: ACM, 2011, pp. 230–241.
3. O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, Survey of automatic modulation classification techniques: classical approaches and new trends, *IET Communications*, vol. 1, no. 2, pp. 137–156, April 2007.
4. W. A. Gardner, Signal interception: a unifying theoretical framework for feature detection, *IEEE Transactions on Communications*, vol. 36, no. 8, pp. 897–906, Aug 1988.
5. Z. Yu, Automatic modulation classification of communication signals, Ph.D. dissertation, Department of Electrical and Computer Engineering, New Jersey Institute of Technology, 2006.
6. A. V. Dandawate and G. B. Giannakis, Statistical tests for presence of cyclostationarity, *IEEE Transactions on Signal Processing*, vol. 42, no. 9, pp. 2355–2369, Sep 1994.
7. A. Fehske, J. Gaeddert, and J. H. Reed, A new approach to signal classification using spectral correlation and neural networks, *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, 2005. Nov 2005, pp. 144–150.
8. Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
9. T. J. O'Shea, J. Corgan, and T. C. Clancy, Convolutional radio modulation recognition networks, *International Conference on Engineering Applications of Neural Networks*. Springer, 2016, pp. 213–226.
10. S. Hochreiter and J. Schmidhuber, Long short-term memory, *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
11. J. Lopatka and M. Pedzisz, Automatic modulation classification using statistical moments and a fuzzy classifier, *Signal Processing Proceedings*, 2000. *WCCC-ICSP 2000*. 5th International Conference on, 2000, pp. 1500-1506 vol.3.
12. Y. Yang and S. Soliman, Optimum classifier for M-ary PSK signals, *Communications*, 1991. *ICC'91, Conference Record. IEEE International Conference* on, 1991, pp. 1693-1697.
13. A. E. Shermeh and R. Ghazalian, Recognition of communication signal types using genetic algorithm and support vector machines based on the higher order statistics, *Digital Signal Processing*, vol. 20, pp. 1748-1757, Dec 2010.
14. A. E. Sherme, A novel method for automatic modulation recognition, *Applied Soft Computing*, vol. 12, pp. 453-461, 2012.
15. P. Panagiotou, A. Anastasopoulos, and A. Polydoros, Likelihood ratio tests for modulation classification, *MILCOM* 2000. *21st Century Military Communications Conference Proceedings*, 2000, pp. 670-674.
16. M. Wong and A. Nandi, Automatic digital modulation recognition using spectral and statistical features with multi-layer perceptions, *Signal Processing and its Applications*, Sixth International, Symposium on. 2001, 2001, pp. 390-393.
17. I. A. Basheer and M. Hajmeer, Artificial neural networks: fundamentals, computing, design, and application, *Journal of Microbiological Methods*, vol. 43, pp. 3-31, Dec 2000.
18. L. S. Iliadis and F. Maris, An artificial neural network model for mountainous water-resources management: The case of Cyprus mountainous watersheds, *Environmental Modelling & Software*, vol. 22, pp. 1066-1072, Jul 2007.
19. S. Ji, W. Xu, M. Yang, and K. Yu, 3d convolutional neural networks for human action recognition. *IEEE PAMI*, 2013.
20. A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei. Large-scale video classification with convolutional neural networks. *CVPR*, 2014.
21. S. Luan, B. Zhang, C. Chen, J. Han and J. Liu. Gabor Convolutional Networks. arXiv preprint arXiv: 1705. 01450, 2017.
22. Baochang Zhang, Jiaxin Gu, Chen Chen, Jungong Han, Xiangbo Su, Xianbin Cao, Jianzhuang Liu. One-Two-One network for Compression Artifacts Reduction in Remote Sensing, *ISPRS Journal of Photogrammetry and Remote Sensing*, 2018.
23. T. Duong, H. Bui, D. Phung, and S. Venkatesh. Activity recognition and abnormality detection with the switching hidden semi-markov model. *CVPR*, 2005.

24. C. Sminchisescu, A. Kanaujia, Z. Li, and D. Metaxas. Conditional models for contextual human motion recognition. *ICCV*, 2005.
25. N. Ikizler and D. Forsyth. Searching video for complex activities with finite state models. *CVPR*, 2007.
26. N. Srivastava, E. Mansimov, and R. Salakhutdinov. Unsupervised learning of video representations using lstms. *ICML*, 2015.
27. J. Donahue, L. A. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell. Long-term recurrent convolutional networks for visual recognition and description. *CVPR*, 2015.
28. J. Y. Ng, M. J. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, and G. Toderici. Beyond short snippets: Deep networks for video classification. *CVPR*, 2015.
29. Z. Wu, X. Wang, Y. Jiang, H. Ye, and X. Xue. Modeling spatial-temporal clues in a hybrid deep learning framework for video classification. *ACM Multimedia*, 2015.
30. S. Venugopalan, M. Rohrbach, J. Donahue, R. J. Mooney, T. Darrell, and K. Saenko. Sequence to sequence - video to text. *ICCV*, 2015.
31. S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. *CVPR*, 2006.
32. L. Yao, A. Torabi, K. Cho, N. Ballas, C. Pal, H. Larochelle, and A. Courville. Describing videos by exploiting temporal structure. *ICCV*, 2015.
33. M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin et al., Tensorflow: Large-scale machine learning on heterogeneous distributed systems, arXiv preprint arXiv:1603.04467, 2016.
34. D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, *CoRR*, [Online]. Available: http://arxiv.org/abs/1412.6980.