IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 23, 2021, accepted March 30, 2021, date of publication April 16, 2021, date of current version April 30, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3073845

Training Images Generation for CNN Based Automatic Modulation Classification

WEI-TAO ZHANG[1,2], DAN CUI[1], AND SHUN-TIAN LOU[1], (Member, IEEE)
[1]School of Electronic Engineering, Xidian University, Xi'an 710071, China
[2]Research Institute of Advanced Remote Sensing Technology, Xidian University, Xi'an 710071, China

Corresponding author: Wei-Tao Zhang (zhwt-work@foxmail.com)

This work was supported by the National Natural Science Foundation of China under Grant 62071350.

ABSTRACT Convolutional neural network (CNN) models have recently demonstrated impressive classification and recognition performance on image and video processing scope. In this paper, we investigate the application of CNN to identifying modulation classes for digitally modulated signals. First, the received baseband data samples of modulated signal are gathered up and transformed to generate the constellation-like training images for convolutional networks. Among the resulting training images, the proposed convolutional gray image is preferred for network training and inference because of the lower computational burden. Second, we propose to use a multiple-scale convolutional neural network (MSCNN) as the classifier. The skip-connection technique is deployed for mitigating the negative effect of vanishing gradients and overfitting during the network training process. Numerical simulations have been carried out to validate the effectiveness of the proposed scheme, the results show that the proposed scheme outperforms the traditional algorithms in terms of classification accuracy.

INDEX TERMS Convolutional neural network, automatic modulation classification, deep learning.

I. INTRODUCTION

In modern wireless communication systems, the modulation type of transmitted signal is mandatory for a receiver to successfully demodulate the original transmitted message. Conventional way of doing this involves sending a header or pilot signal along with the original message to inform the receiver about the modulation type. However, such approach incurs penalty in terms of bandwidth utilization and data throughput. While intelligent receivers can mitigate this drawback by intelligently pre-processing the received signal to identify the modulation type of the transmitted signal with no need of prior knowledge. This has led to huge interest in developing automatic modulation classification (AMC) techniques [1], which is actually an intermediate step between signal detection and demodulation. It is usually preferred in adaptive modulation scenarios such as software defined radio (SDR) and cognitive radio (CR) [2], where the transmitted modulation can be dynamically chosen such that the spectral efficiency is constantly optimized. Now AMC has found a variety of applications in both commercial and military fields such as spectrum management, surveillance and threat analysis.

The associate editor coordinating the review of this manuscript and approving it for publication was Gustavo Callico.

Over the past few years, numerous methods for AMC were proposed in the literature [3]. They can be mainly divided into two categories, namely, Likelihood-Based (LB) methods [4]–[6] and Feature-Based (FB) methods [7]–[11]. The former treats AMC as a hypothesis testing problem, where the exact or approximated likelihood function of the incoming signal is calculated and compared with a threshold value. Note that if the probability density function (pdf) of the received signal is known and identical to the actual pdf of the parameters, the LB method yields the optimal solution in terms of correct classification rate (CCR) since it minimizes the probability of false classification. However, LB methods usually suffer from model mismatch with respect to carrier frequency and phase offsets [8]. In addition, the LB methods always have a high computational complexity. So LB solution serves as an upper bound performance benchmark of any classifier, while they are commonly discarded in practical use.

FB methods are popular in the practical implementations because of the less complexity involved. Most of FB methods usually consist of two steps, the first step involves extracting features from the received signal. In the second step, a linear or nonlinear classifier is designed to perform the classification. Numerous features with their respective merits and defects were proposed. Among them, the most used are high-order cumulants [6]–[8], wavelet transform [9], and cyclic statistics [10]. As for the classifier, machine learning algorithms, such as support vector machines (SVM) [11], K-nearest neighbor (KNN) [8], and artificial neural networks [10], have been widely studied for inference. Whereas these methods were developed and optimized for some environments, they suffer from performance degradation for mismatch between extracted feature and classifier because the feature selection procedure and classification are independent of each other. The quality of the whole AMC relies on both the performance of classification algorithm and the ability of the features to differentiate between the constellations of a given set. Obviously, the features that are insensitive to the inherent parameters of the received signal such as the phase and frequency offsets, the synchronization, and the noise are preferred. Unfortunately, such properties are rarely achieved by manually designed feature under various conditions.

More recently, studies have shown that deep neural networks (DNN) can learn from the complex data structures and achieve superior classification accuracy [12]. This makes them an obvious choice in AMC problem because of the much denser modulation schemes used in the modern communication system. Kim used a fully connected model with three hidden layers [13]. To feed the DNN model, twenty-one features are computed from the received data samples based on power spectrum density and cumulants. Ali proposed a fully connected DNN model based on autoencoder with non-negativity constraints [14], where the input features are fourth order cumulants. Note that the fully connected DNN model always involves too many free parameters to be trained, which usually results in high computational load for network learning and inference. In addition, the above DNN model used in AMC only serves as a classifier, which is still independent of feature extraction.

Recently, the convolutional neural network is more popular for modulation classification to overcome some obstacles of traditional machine learning algorithms. As for CNN based AMC, the feature extraction procedure is incorporated into CNN model, the model extracts feature from data autonomously, then the challenging task of manual feature selection can be avoided. CNN-based methods can be roughly divided into two categories according to the input of network. One was trained using IQ component signals, while the other was trained using image-based constellation diagrams. For example, a complex-valued network [15] considered the correlation between the real and imaginary parts of signal, is proposed to demonstrate the high potential for AMC and validate the superior performance compared with the real-valued network. However, a complex-model has a higher computational complexity because of the plenty of complex valued multiplications involved. Huyunh-The [16] proposed a cost-efficient convolutional neural network (MCNet) for AMC, whose input is IQ components and network architecture is built with several specific convolutional blocks to concurrently learn the spatiotemporal signal correlations via different asymmetric convolution kernels. Although the accuracy performance was good, thousands of parameters were used in the network, which is still large relative to the small IQ length. Kim *et al.* [17] proposed a novel CNN architecture for AMC with low computational complexity compared with MCNet. The proposed model showed good performance in the SNR range from −4dB to 20dB.

Compared with IQ component-based model, image-based model is more elegant for AMC, because it can provide the visualization of modulation categories. Huang *et al.* proposed a compressive CNN (CCNN) for AMC [18], where multiple images (called RCs and CGCs) are utilized as the input of the network. The CGCs further considered the two-dimensional probability distribution of signal samples on the basis of RCs. However, there are still two limitations. First, the location of each sample within a grid region is not considered. Second, the impact of each sample in a region on its neighboring pixel is ignored. To handle these problems, Doan [19] leverages a bivariate histogram and an exponential decay mechanism to obtain gray-scale constellation image. Meanwhile, a novel CNN model, namely FiF-Net, is introduced for modulation classification. However, the computation burden of generating an image is higher since it must compute the distances between sample points and the center of each pixel. In [20], modulation classifiers are developed based on transfer learning of classical ResNet-50 and Inception V2 deep learning model, where the classifiers are trained with color images generated through the constellation density of the masked signal. The constellation density matrix (CDM) based modulation classification algorithm is proposed to identify the orders of different modulation categories. Despite exploiting more explicitly discriminative features of constellation diagrams, modeling a modulation classifier from color image suffers from poor performance along increment of QAM at lower SNR level. In [21] the transfer learning of classical AlexNet and GoogleNet are adopted for AMC using multiple constellation-like image data sets. As for the CNN based AMC, the training images generation is crucial for model learning. Unfortunately, the constellation diagrams generated from data samples are binary images with limited resolution, or enhanced gray (color) images with high computational load. Moreover, classical large CNN models are actually inappropriate for AMC problem since the constellation diagrams of the incoming signals are relatively simple images with uniform background, training of these models based on constellation-like images probably result in overfitting.

In this paper, we focus our attention on the CNN based AMC problem. In order to overcome the aforementioned drawbacks, the constellation-like convolutional gray images are generated for model training, which exhibit better representation than binary images and other existing gray images. Moreover, the multiple-scale convolutional neural network (MSCNN) with dropout is proposed as the classifier.

II. PROBLEM FORMULATION

We assume that the radio frequency (RF) signal in the receiver is preprocessed such that the received waveform consists of samples of prefiltered and pulse shaping digital signal in multipath fading channel. The oversampled data point reads

$$r(n) = \alpha e^{(j2\pi f_o n T_s + \varphi_o)} \sum_{l=0}^{L-1} s(l)h(n - lT - n') + w(n) \quad (1)$$

where α represents the channel attenuation factor, $\{s(l)\}_{l=0}^{L-1}$ are L complex transmitted symbols drawn independently from a finite alphabet constellation, $h(n)$ represents the overall effect of pulse shaping filter and physical channel, $w(n)$ is additive white Gaussian noise with power ${\sigma_w}^2$, f_o is the carrier frequency offset due to the impairment between transmitter and receiver, φ_o is the phase offset, n' represents the propagation delay, T is the symbol period, T_s is the sampling period, then the oversampling rate is given by $\rho = T/T_s$.

In this paper, we assume that the channel is flat fading or be properly equalized such that $h(n)$ is negligible and the parameters T and n' are assumed to be known. The goal of AMC is to identify which modulation scheme has been utilized with the knowledge of N received samples $\mathbf{r} = [r(1), \ldots, r(N)]$. A CNN based AMC technique for adaptive modulation system is shown in Fig. 1, where the imaging method for data conversion and the training of CNN model are critical points. In the sequel, the focus will be on these two points.

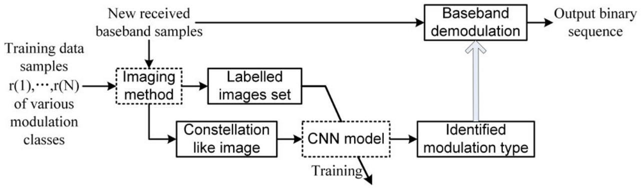

New received baseband samples
Baseband demodulation
Output binary sequence
Training data samples r(1),⋯,r(N) of various modulation classes
Imaging method
Labelled images set
Constellation like image
CNN model
Identified modulation type
Training

FIGURE 1. The architecture of CNN based AMC technique for adaptive modulation system.

III. TRAINING IMAGES GENERATION

The received signal in equation (1) can be represented by its constellation diagram through mapping signal samples into scattering points on a complex plane. Note that the complex plane is infinite, while the signal samples represented by the scattering points are distributed within a certain area in the complex plane. Moreover, the amplitude of the received signal varies from different channel responses and modulation types, which makes the selection of appropriate area for constellation diagram more difficult. If the selected area is too small, some signal samples may be excluded from the image. On the contrary, if the area is too large, signal samples may crowd a small region, which makes it difficult to discriminate the higher order modulation types.

In order to solve this problem, we compensate for the arbitrary channel attenuation by normalizing the received complex baseband samples as $(r(n) - \mu_r)/\sigma_r$, where μ_r and σ_r are the mean and standard deviation of received samples. After that, the signal samples are distributed in a relatively fixed area, which is convenient for us to choose an appropriate complex plane. We select a 6 × 6 complex plane, assuming a typical signal-to-noise ratio (SNR) range from 0 to 15 dB.

A. BINARY IMAGE

According to the distribution of signal samples in constellation diagram, a pixel limited binary image is straightforward. In this case, the selected complex plane is uniformly divided into grids, which correspond to pixels in the resulting binary image. Naturally, if the grids contain signal sample points, the corresponding pixels are set 1, otherwise 0. Then the constellation diagram of the signal is converted to binary image. However, there might be multiple samples that crowded one pixel, in this case the pixels with one or more sample points are treated identically. So binary image is unable to provide an accurate representation of the distribution of signal samples.

B. GRAY IMAGE

For a pixel of binary image, the number of sample points in corresponding grid has been ignored, which degraded the resulting image quality. In order to improve the representation accuracy of the pixels with multiple sample points, the binary image can be upgraded to gray image by regarding the number of sample points as the weight coefficient for the pixels with multiple sample points. The multiple pixels with different number of sample points are shown in Fig. 2. For a gray image, the pixels 3, 6, 13, 14 will have the weight coefficients 1, 2, 4, 3 respectively, which can be normalized to form the intensity value for these pixels.

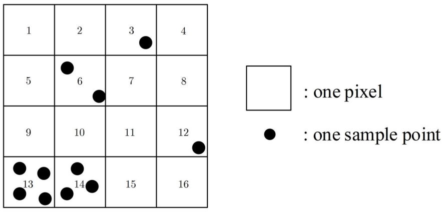


FIGURE 2. Sample ponits and pixels.

C. ENHANCED GRAY IMAGE

Although the number of samples in each pixel is considered in the gray image, the impact of each sample in a pixel on its neighboring pixels is neglected. Hence, an enhanced gray image is developed, which takes into consideration the distances between sample points and centroids of pixels. Concretely, it adopts an exponential decay model, $o_{ij} = \sum_{n=1}^{N} \theta^{-\lambda d_{(n,ij)}}$, where o_{ij} represents the cumulative impact of all received sample points on (i, j)th pixel, $d_{(n,ij)}$ is the distance between the centroid of (i, j)th pixel and sample point $r(n)$, θ is the base of exponential function, and λ is the decay factor. o_{ij} can be normalized to form the intensity values to generate an enhanced gray image.

D. CONVOLUTIONAL GRAY IMAGE

The enhanced gray image greatly improves the image quality for the subsequent classification procedure. However, it has two limitations. Firstly, for the pixels located in the boundary between two adjacent constellation points, there are usually rare signal sample points available in the corresponding grids, then the boundary pixels commonly have very small intensity values relative to the constellation point pixel, which is useful to identify the different constellation diagrams. Whereas, the enhanced gray imaging model still compute the cumulative impact of all data samples on the boundary pixels. This results in a dim boundary between two adjacent constellation point, and hence it is difficult to identify the higher order modulations in noisy channel. Secondly, according to the enhanced gray imaging model, the computation of intensity of each pixel involves the distance from the corresponding grid to all data samples and the subsequent N exponential operation. When the number of signal samples N is large or a higher resolution is preferred, the computational burden of enhanced gray image will be prohibitive for the generation of a large amount of training images for deep model.

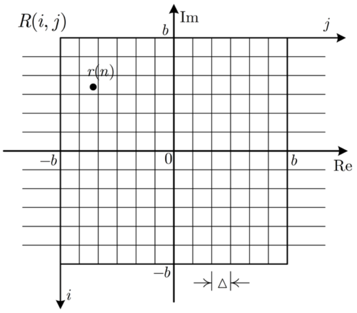


FIGURE 3. Resulting image in natural coordinate system and graphic coordinate system.

In order to overcome the aforementioned drawbacks, we put forward a convolutional gray image generation method, which is based on the simple convolution operation of local gray image. First, let us build a gray image $R(i,j)$ using the received complex-valued signal $r(n)$ with N data samples. Let b denote the selected boundary of the complex plane, and Δ denote the grid size, which actually represents the resolution of the resulting image. Fig. 3 shows the image in natural coordinate system and graphic coordinate system. For a certain sample point $r(n)$, it contribute to the corresponding pixel by

$$R(i,j) \longleftarrow R(i,j) + 1 \tag{2}$$

where the graphic index i and j are related to data sample $r(n)$ by

$$i = \left\lceil \frac{b - Im[r(n)]}{\Delta} \right\rceil \tag{3a}$$

$$j = \left\lceil \frac{b + Re[r(n)]}{\Delta} \right\rceil \tag{3b}$$

where notation $\lceil x \rceil$ denotes the smallest integer that is greater than or equal to x. After that, the gray image can be obtained by the following normalization procedure

$$R(i,j) \longleftarrow R(i,j)/p \tag{4}$$

where p is the maximum of R.

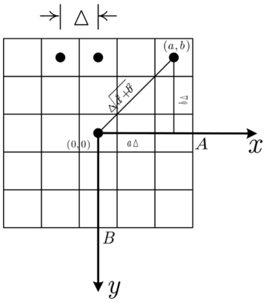


FIGURE 4. Schematic diagram of convolution filter.

Second, note that the constellation diagram of a modulated signal actually represents the cluster of data samples. So it is appropriate to take into account only the impact of surrounding data samples to the selected pixel, which represents the local features for the modulated signal. In order to efficiently calculate the locally clustered gray image, we propose to use a convolution kernel W, which is shown in Fig. 4. It can be seen that the stride of the convolution filter is simply set to the image resolution Δ, and the size of the filter is determined by positive integer A and B. The filter weight coefficients can be evaluated by

$$W(a,b) = \theta^{-\lambda d_{ab}} \tag{5}$$

where $d_{ab} = \Delta\sqrt{x^2+y^2}$ is the Euclid distance between the (a,b)th element of filter and its centroid. The other parameters are similar to that used in an enhanced gray image. The reason why we choose the aforementioned filter is based on the following. The enhanced gray image is actually obtained by passing the signal samples through a 2D filter with infinite size, because the impact of all data samples on the certain pixel are evaluated. However, it is inappropriate to use such a large filter since only the data samples that belong to the certain constellation point contribute to the corresponding pixel. So a 2D filter with finite size, say [A, B], is preferred, which is expected to solve the dim boundary problem in enhanced gray images. The computation of filter coefficients in (5) reflects the fact that the data sample points closed to the certain pixel have a greater impact than those far away from the pixel.

Third, for the complex-valued received signal $r(n)$, the pixels of convolutional gray image can be computed by

$$I(i,j) = \sum_{a=-A}^{A}\sum_{b=-B}^{B} W(a,b)R(i+a,j+b) \tag{6}$$

where $I(i,j)$ represents the intensity of the pixel (i,j). By performing a convolution operation instead, the computation

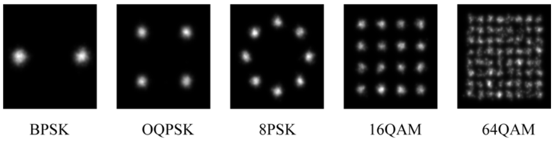


FIGURE 5. Convolutional gray image for five modulation categories.

complexity for generating a higher resolution image will be greatly decreased compared to enhanced gray image. Moreover, faster computation of (6) can be implemented via fast Fourier transform in frequency domain. The resulting convolutional gray images for different modulations are depicted in Fig. 5.

IV. DEEP NETWORK FOR AMC

In modern communication systems, reliability is one of the most import indexes to evaluate the system performance, which demands a well-performed AMC model in terms of classification performance. Hence, the multiple-scale architecture of a modulation classification convolutional neural network, named MSCNN, is proposed to learn modulation patterns from constellation like uniform background images. The network architecture is presented in Fig. 6. The deep network is specifically designed with several convolutional blocks associated with skip connections, in which each block comprises various asymmetric convolutional layers and comprised of one convolutional layer, followed by one batch normalization layer and ReLU activation function. The proposed network is capable of analyzing the multi-scale feature map correlations exhaustively to promisingly improve the accuracy of modulation classification under poor conditions with the cheaper computational cost. At the beginning of network, an input layer configured by the size of 240 × 240 × 1 to be compatible with the volume size of resulting image is followed by a process block with 64 kernels of size 3 × 3 to acquire coarse features. With 2 × 2 kernels, the first pooling layer is able to reduce the size of the feature map to optimize the extraction of the image characteristics with the stride of (2, 2). Subsequently, two layers of process layers organized in parallel, called pre-block as illustrated in Fig. 7(a), use an asymmetric kernel matrix of kernel sizes 3 × 1 and 1 × 3 corresponding to vertical and horizontal kernels, respectively, instead of 3 × 3 to decrease the number of trainable parameters. After that, the network consists of three modules for deeply mining more explicitly discriminative features at multi-scale feature maps. Each module has two sophisticated process blocks, called M-block and M-block-drop respectively, which are cascaded along the network backbone. For details, M-block is configured by three process layers with different kernels, 3 × 1, 1 × 3, 1 × 1 kernels arranged in parallel, at which all feature maps are then merged in the depth dimension at the output of each block via depth-wise concatenation layer. It is worth noting that the reason why the spatial dimension of feature maps remains unchanged at the output of M-block is that all kernels are applied with stride (1, 1). Meanwhile, another dimension-reduced version of M-block, named M-block-drop, is given with the same structure of M-block, except a dropout layer is carried out following a 3 × 1 process layer as shown in Fig. 7(b). Notably, different from the traditional CNN model, where the maximum pooling operation is applied in convolution blocks, a dropout layer (rather than a pooling operation) follows every convolutional block instead. This modified architecture not only implements the down-sampling of the feature maps, but also improves the robustness of the model against the various additive noise. Moreover, it enables to prevent the network training process from overfitting. M-block-drop is also applied immediately after the pre-block to quickly diminish the dimension of feature maps and subsequently reduce the computational burden of following layers. As a result, the feature maps go through the M-block-drop, whose dimension before reaching the concatenation layer will be halved. In order to be compatible with output of dropout layer when performing depth concatenation, two remaining layers are deployed with stride of (2, 2). Each block has two 1 × 1 convolutional layer: one for feature extraction and another on the top for reduction of the channel dimension. The module is finalized with M-block-drop. By following this architecture, the spatial size of output feature volume halves for every module.

To improve the accuracy performance of AMC model for mitigating the negative effect of vanishing gradient problem caused by popular ReLU activation function in a relatively deep network and maintain the informative identity of previous layers, skip-connection technique is deployed for associating M-blocks via an element-wise addition layer as described in Fig. 6. Unlike the traditional structure of network, skip-connection mechanism allows the network to learn the integrated information. At the end of network, the feature maps of the last M-block are gathered with its input by a depth concatenation layer. It is obvious that multiple scale features extracted in each block and the informative identity maintained throughout the network via skip-connection are jointly synthesized to enrich the AMC model. MSCNN can overcome the problems of vanishing gradients and overfitting during the network training process. The network is finalized with an average pooling layer with the pool size of (2, 2), a fully connected layer (where the number of hidden nodes is identical to the number of modulation categories considered for classification), and a softmax layer arranged sequentially after the fully connected layer. The detailed configurations of MSCNN are given in Table 1.

V. RESULTS

The simulation settings are as follows: 1) complex baseband modulated signal are obtained from the output of additive white Gaussian noise (AWGN) channel with four noise level (the corresponding SNR is 0dB, 5dB, 10dB, 15dB); 2) 1000 data samples are collected to generate the gray image, enhanced gray image and convolutional gray image respectively, the binary image is ignored because of its low

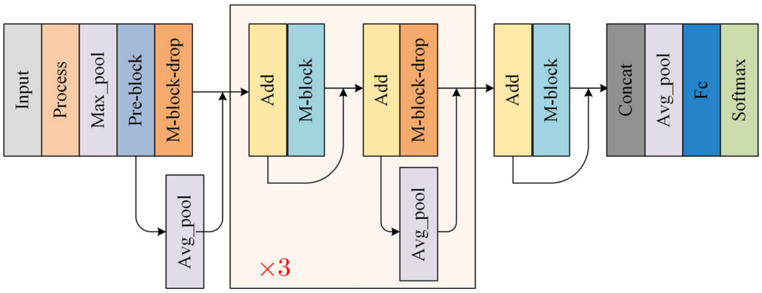
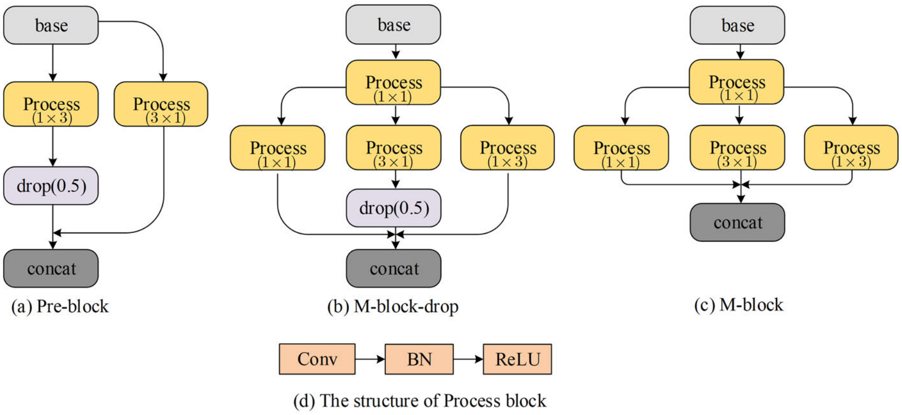


FIGURE 6. The overall network architecture of MSCNN.



(a) Pre-block

(b) M-block-drop

(c) M-block

(d) The structure of Process block

FIGURE 7. Description of convolutional blocks deployed in the MSCNN. (a) the Pre-block; (b) the convolutional M-block-drop; (c) the convolutional M-block; and (d) the structure of Process block.

resolution; 3) We consider the classification of 5 modulation categories, including BPSK, OQPSK, 8PSK, 16QAM, 64QAM, each of which contain 20000 labeled images for model training and 5000 labeled images for performance test. Notably, the SNR of test dataset is different from that of the training dataset (1dB-14dB for test data), which indicates a more difficult scenario for classification model to predict the modulation categories. Therefore, the test accuracy of a trained network may not be high enough compared with the results in existing literatures.

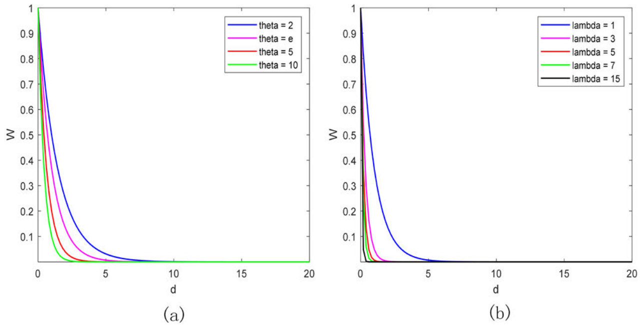


FIGURE 8. The exponential window evaluated at different θ and λ: (a) $\lambda = 1$; (b) $\theta = e$.

A. PARAMETERS SELECTION

For enhanced gray image and convolutional gray image, the parameter θ and λ play an important role in imaging process. In this section, we will discuss the parameters selection. The exponential functions (5) with different θ and λ are plotted in Fig. 8. As shown in Fig. 8, the exponential function decreases rapidly with a larger θ or λ, which have an important effect on imaging. With the increment of θ or λ, the equivalent support receptive field of convolution 2D filter shrink in imaging process. Fig. 9 shows the effect of different λ in generating convolutional gray images for BPSK modulated signal under additive white Gaussian noise. One sees that a smaller λ blurs the edge between two adjacent constellation points. Especially at lower SNR, adjacent constellation connected to each other due to the noise interference, which makes it difficult for classifier to identify the modulation type. Whereas a larger λ will produce lower resolution gray images that is similar to binary image. So θ and λ are tradeoff parameters between blurred gray image and binary image. Fortunately, these parameters can be determined using empiric values with a wide range.

Fig. 10 shows the resulting images by two different imaging methods. We see that the convolutional gray image produced a sharper change than enhanced gray image between light and dark, where sample density is different. This property enables our method to yield a clearer edge between two adjacent constellations under noisy interference scenario, which is superior in modulation classification.

TABLE 1. Detailed configuration of network architectrue.

Layer	Output Volume	Detailed Description
Input	240×240×1	
Process	240×240×64	64 conv 3 × 3
Max-pool	120 × 120 × 64	2 × 2, stride = (2,2)
Pre-block	120 × 120 × 128	$\left\{\begin{array}{l}64\mathrm{conv}1\times 3, \mathrm{stride}=(1,1)\\ 64\mathrm{conv}3\times 1, \mathrm{stride}=(1,1)\\ \mathrm{dropout}(0.5)\end{array}\right.$
Avg-pool	60 × 60 × 128	2 × 2, stride = (2,2)
M-block-drop	60 × 60 × 128	$\left\{\begin{array}{l}64\mathrm{conv}1\times 1, \mathrm{stride}=(1,1)\\ 48\mathrm{conv}3\times 1, \mathrm{stride}=(2,2)\\ \mathrm{dropout}(0.5)\\ 48\mathrm{conv}1\times 3, \mathrm{stride}=(2,2)\\ 32\mathrm{conv}1\times 1, \mathrm{stride}=(2,2)\\ \mathrm{depthcatenation}\end{array}\right.$
Add	60 × 60 × 128	element-wise addition
3 × module	8 × 8 × 128	$\left\{\begin{array}{l}32\mathrm{conv}1\times 1, \mathrm{stride}=(1,1)\\ 48\mathrm{conv}1\times 3, \mathrm{stride}=(1,1)\\ 48\mathrm{conv}3\times 1, \mathrm{stride}=(1,1)\\ 32\mathrm{conv}1\times 1, \mathrm{stride}=(1,1)\\ \mathrm{depthcatenation}\\ \mathrm{element-wise}\quad\mathrm{addition}\\ \mathrm{pool}\quad 2\times 2, \mathrm{stride}=(2,2)\\ \mathrm{M-block-drop}\\ \mathrm{element-wise}\quad\mathrm{addition}\end{array}\right.$
M-block	8 × 8 × 256	$\left\{\begin{array}{l}32\mathrm{conv}1\times 1, \mathrm{stride}=(1,1)\\ 96\mathrm{conv}1\times 3, \mathrm{stride}=(1,1)\\ 96\mathrm{conv}3\times 1, \mathrm{stride}=(1,1)\\ 64\mathrm{conv}1\times 1, \mathrm{stride}=(1,1)\\ \mathrm{depthcatenation}\end{array}\right.$
Concat	8 × 8 × 384	feature map concatenation in depth
Avg-pool	4 × 4 × 384	2 × 2, stride = (2,2)
Fc1	1 × 1 × M	fully-connpected with M neurons compatible with the modulation categories
Softmax		softmax

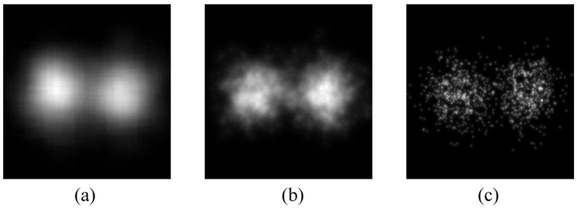


FIGURE 9. The convolutional gray image with different λ: (a) λ = 3; (b) λ = 12; (c) λ = 50.

B. EFFECT OF IMAGING SCHEMES ON 64QAM RECOGNITION

We present some simulation results to show the effect of different imaging schemes. The reason why we show the results of imaging schemes on 64QAM is that it is hard to recognize among the aforementioned modulation categories. In our simulation, the same set of complex samples is used to generate three types of images, including gray image, enhanced gray image, and convolutional gray image. For each imaging method, corresponding images are fed into MSCNN for training. Then 1000 test images are generated for 64QAM modulated signals with SNR=4dB. Table 2 records the accuracy of three imaging methods. As shown in the table, the classification accuracy improves from 73.6% to 91.9% if the convolutional gray image is utilized instead of the gray image. Notably, despite achieving the greatest accuracy of 91.9%, 64QAM suffers the misclassification with 16QAM.

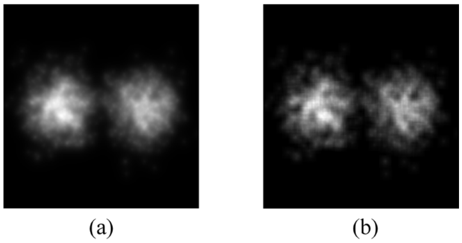


FIGURE 10. The generated images by different imaging methods: (a) the enhanced gray image; (b) the convolutional gray image.

C. COMPARISON OF COMPUTATIONAL LOAD FOR DIFFERENT IMAGING SCHEMES

In this example, we investigate the computational load of imaging schemes, because this is a very important issue for adaptive demodulation systems applicable in real time scenario. The same set of complex valued data samples is used to generate the gray image, enhanced gray image and convolutional gray image respectively. For each imaging method, we compare the impact of the number of samples and different resolutions on imaging time, where the resolutions we considered include 200 × 200, 300 × 300, 400 × 400, 600 × 600.

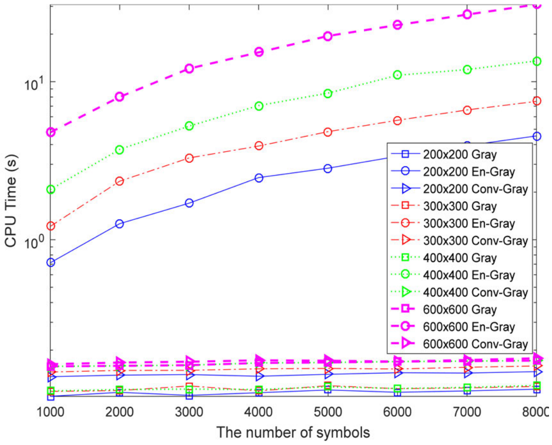


FIGURE 11. CPU time versus the number of samples under different resolution images.

Fig. 11 plots CPU time versus the number of samples for generated images with different resolutions. We see that the imaging time of enhanced gray image significantly increases

TABLE 2. Classification results of three imaging method for 64QAM with SNR = 4dB using MSCNN.

Imaging method	BPSK	OQPSK	8PSK	16QAM	64QAM	Accuracy
Gray Image	0	0	0	264	736	73.6%
Enhanced Gray Image	0	0	0	129	870	87.0%
Convolutional Gray Image	0	0	0	81	919	91.9%

when the number of samples grows. In addition, it is significantly higher than that of other imaging schemes under the same number of samples and resolution, which validates our analysis in section III. Concretely, when the number of samples is 1000 and the resolution is 200, the imaging time of enhanced gray image is more than five times of convolutional gray image. The reason for its high computational load is that it is obtained by evaluating the impact of all data samples on each pixel with repeated exponential computations. Whereas the proposed convolutional gray image avoids the repeated exponential operation, and the convolution kernel is calculated just once and shared for each pixel. Moreover, we adopt fast convolution operation with convenient implementation.

Note that the imaging time of gray image and convolutional gray image keep almost constant under different resolution when the number of samples increases. In addition, when the resolution grows, the imaging time of gray image and convolutional gray image slightly increases. We see that the computational load of convolutional gray image is slightly higher than that of gray image, the additional computational burden lies in the convolution operation, which is computationally cheaper as indicated in the gap of CPU Time between two imaging schemes.

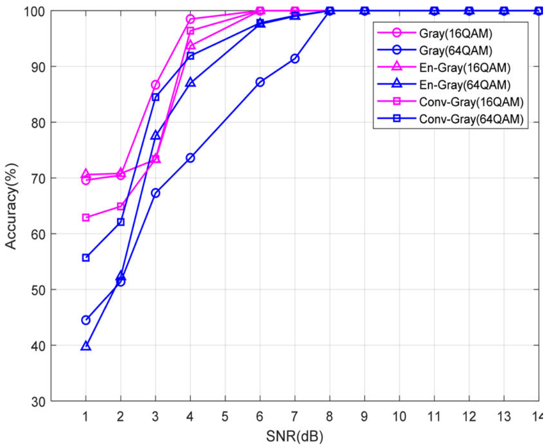


FIGURE 12. Classification accuracy versus modulation types under different images.

D. CLASSIFICATION PERFORMANCE OF MSCNN FOR DIFFERENT MODULATIONS

We report the classification accuracy of MSCNN for five modulation categories separately, where the numerical results are plotted in Fig. 12. In general, the classification accuracy increases along the increment of SNR levels. In our simulation, the classification accuracy of low order modulation categories, including BPSK, OQPSK, 8PSK, are 100% under different imaging methods. Meanwhile, MSCNN on convolutional gray image recognizes 16QAM and 64QAM signals competently with the accuracy rates of 96.4% and 91.9%, respectively, at 4dB SNR. It is observed that the classification accuracy keeps getting worse along increment of QAM due to vulnerability of high-order modulation signal. For instance, the accuracy of gray image significantly decreases over 16% when upgrading the QAM order from 16 to 64. As the worst modulation in our simulation, 64QAM suffers the confusion with 16QAM. It is well known that high-order modulations usually achieve high transmission rate in wireless communication system, but the modulation recognition of received signal will be less accurate due to the fact that the distance between scattered points distributed in a constellation map is narrower, and hence close constellation points are vulnerable with noise.

As for the proposed convolutional imaging scheme, by using an appropriate kernel the convolution operation produces a gathering effect for each constellation point and creates a clear edge between two adjacent constellation points. Consequently, convolution gray image achieves a significant increment in accuracy of 64QAM compared with other images. At an SNR of 1dB, the convolutional gray image achieves 4.9% and 18.3% improvement compared to enhanced gray image and gray image respectively. It is not surprising that the enhanced gray image shows higher performance than gray image at the most of SNR levels.

E. COMPARISON OF DIFFERENT CLASSIFIERS

In this example, AMC algorithm using MSCNN model on convolutional images is compared with that using the SVM on different features extracted from the received signal, particularly SVM-7 and SVM-5 [8], and GoogleNet on three-channel images [15]. The comparison result is plotted in Fig. 13, where SVM-5 includes two sixth order and three fourth order cumulants and SVM-7 includes three fourth order cumulants and four sixth order cumulants respectively. Observing Fig. 13, the MSCNN model achieves the classification rate of 83.7% at 1dB SNR, which is better than SVM-7 and SVM-5 by approximately 5.26% and 6.84%, respectively. For two machine learning algorithms, the SVM-7 algorithm with more features employed slightly performs better than SVM-5. However, SVM-7 has a higher computational burden than SVM-5. In terms of inference

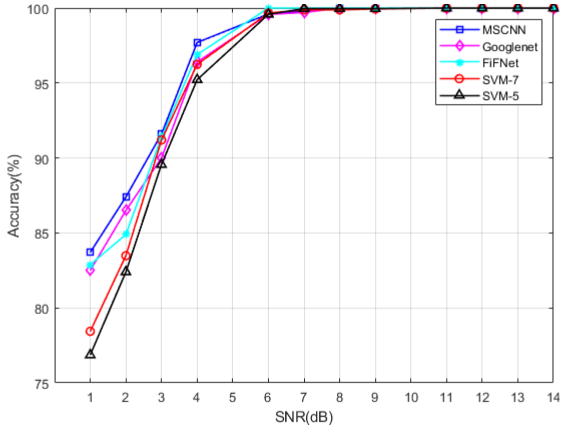


FIGURE 13. Average classification accuracy of different classifiers versus SNR.

TABLE 3. Comparison of capacity and inference time for different networks.

NetWork	Capacity (No.parameters)	Inference time (ms)
SVM-5	-	0.4
SVM-7	-	0.9
MSCNN	274K	3.0
GoogleNet	6.8M	9.6
FiFNet	416K	3.2

time, SVM-7 spends 55% more than SVM-5 because it should compute more features. It is worth noting that the manual selection of features is a critical issue and affects classification performance noticeably in classical machine learning algorithms. However, the CNN-based algorithm even performs better without manual feature selection. Subsequently, we compared our MSCNN with FiFNet [19] for constellation based modulation classification using convolutional gray images. Observing the results in Fig. 13, we see that the classification accuracy of the proposed model was better in the SNR range from 1dB to 4dB, where the proposed model improves classification accuracies of 2.5% at 2dB compared with those of FiFNet. The network capacity and average inference time are summarized in Table 3, where inference time is averaged over 5000 trials. It can be seen that MSCNN is cheaper than FiFNet by approximately 34% of capacity (aka the number of trainable parameters). However, the inference time of both networks is almost equivalent. This is because that both depth-wise concatenation and addition operations are performed many times by MSCNN. Finally, the performance of MSCNN using convolutional gray images is compared with GooleNet using three channel images. We see that MSCNN outperforms GoogleNet at the most of SNR levels. This is not surprising because GoogleNet is a standard deep network for general purpose applications, such as large scale classification with more than 1000 categories. It is probably not efficient for modulation classification using simple uniform background images. on the contrary, it will cause the problem of gradient vanish or overfitting. In terms of capacity and inference time, there is no doubt that GoogleNet is the largest one.

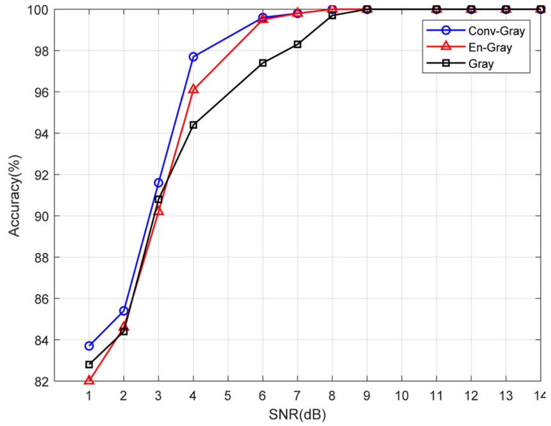


FIGURE 14. Average accuracy of three types of images versus SNR.

F. CLASSIFICATION ACCURACY OF MSCNN FOR DIFFERENT IMAGING SCHEMES

We investigate the performance of MSCNN for the generated three type images. The classification accuracy for each imaging method varied with SNRs is presented in Fig. 14. Note that the convolutional image outperforms the gray image and enhanced gray image within the SNRs less than 8dB. In lower SNR cases, the data samples belong to a certain constellation point cannot be gathered round in gray image due to noise interference, while the enhanced gray image gives rise to dim boundary problem. However, the convolution kernel with finite size is used to improve the aggregation of interfered data samples and solve the dim boundary problem, and hence improved accuracy was observed. Concretely, at an SNR of 4dB, MSCNN model on convolutional image achieves 1.6% and 3.3% improvement compared with gray image and enhanced gray image. Both convolutional gray image and enhanced gray image considered the impact of data samples on the selected pixel, but convolutional gray image performs more accurately and requires less computing resources than enhanced gray image. By deploying an appropriate kernel size, convolutional gray image achieves good trade-off between accuracy and computational cost.

VI. CONCLUSION

In this paper, a multiple-scale convolutional neural network, namely MSCNN, is proposed for constellation-based modulation classification. The network architecture consists of several processing blocks to comprehensively learn more intrinsic characteristics from constellation-like image. Meanwhile, the convolutional gray image is developed, in which convolution kernel is deployed to overcome the drawbacks in existing imaging schemes. The trained MSCNN on convolutional gray image dataset achieves the averaged classification accuracy of approximately 97.7% at 4 dB SNR. With a well-designed network and effective imaging method, MSCNN on convolutional gray image outperforms other models in terms of accuracy. For future works, the impacts of interference and frequency selective fading channel will be investigated.

REFERENCES

[1] A. K. Nandi and E. E. Azzouz, "Algorithms for automatic modulation recognition of communication signals," *IEEE Trans. Commun.*, vol. 46, no. 4, pp. 431–436, Apr. 1998.

[2] F. K. Jondral, "Software-defined radio-basic and evolution to cognitive radio," *EURASIP J. Wireless Commun. Netw.*, vol. 2005, no. 3, pp. 1–9, Dec. 2005.

[3] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Survey of automatic modulation classification techniques: Classical approaches and new trends," *IET Commun.*, vol. 1, no. 2, pp. 137–156, Apr. 2007.

[4] C.-Y. Huan and A. Polydoros, "Likelihood methods for MPSK modulation classification," *IEEE Trans. Commun.*, vol. 43, nos. 2–4, pp. 1493–1504, Feb. 1995.

[5] W. Wei and J. M. Mendel, "Maximum-likelihood classification for digital amplitude-phase modulations," *IEEE Trans. Commun.*, vol. 48, no. 2, pp. 189–193, Feb. 2000.

[6] S. Huang, Y. Yao, Y. Xiao, and Z. Feng, "Cumulant based maximum likelihood classification for overlapped signals," *Electron. Lett.*, vol. 52, no. 21, pp. 1761–1763, Oct. 2016.

[7] A. Swami and B. M. Sadler, "Hierarchical digital modulation classification using cumulants," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 429–461, Mar. 2000.

[8] M. W. Aslam, Z. Zhu, and A. K. Nandi, "Automatic modulation classification using combination of genetic programming and KNN," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2742–2750, Aug. 2012.

[9] S. Kumar, V. A. Bohara, and S. J. Darak, "Automatic modulation classification by exploiting cyclostationary features in wavelet domain," in *Proc. 23rd Nat. Conf. Commun. (NCC)*, Mar. 2017, pp. 1–6.

[10] B. Ramkumar, "Automatic modulation classification for cognitive radios using cyclic feature detection," *IEEE Circuits Syst. Mag.*, vol. 9, no. 2, pp. 27–45, Jun. 2009.

[11] L. Xie and Q. Wan, "Automatic modulation recognition for phase shift keying signals with compressive measurements," *IEEE Wireless Commun. Lett.*, vol. 7, no. 2, pp. 194–197, Apr. 2018.

[12] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[13] B. Kim, J. Kim, H. Chae, D. Yoon, and J. W. Choi, "Deep neural network based automatic modulation classification technique," in *Proc. Int. Conf. Inf. Commun. Technol. Converg.*, Jeju, South Korea, Oct. 2016, pp. 579–582.

[14] A. Ali and F. Yangyu, "Automatic modulation classification using deep learning based on sparse autoencoders with nonnegativity constraints," *IEEE Signal Process. Lett.*, vol. 24, no. 11, pp. 1626–1630, Nov. 2017.

[15] Y. Tu, Y. Lin, C. Hou, and S. Mao, "Complexed-valued networks for automatic modulation classification," *IEEE Trans. Veh. Technol.*, vol. 69, no. 9, pp. 10085–10089, Sep. 2020.

[16] T. Huynh-The, C.-H. Hua, Q.-V. Pham, and D.-S. Kim, "MCNet: An efficient CNN architecture for robust automatic modulation classification," *IEEE Commun. Lett.*, vol. 24, no. 4, pp. 811–815, Apr. 2020.

[17] S.-H. Kim, J.-W. Kim, V.-S. Doan, and D.-S. Kim, "Lightweight deep learning model for automatic modulation classification in cognitive radio networks," *IEEE Access*, vol. 8, pp. 197532–197541, Nov. 2020.

[18] S. Huang, L. Chai, Z. Li, D. Zhang, Y. Yao, Y. Zhang, and Z. Feng, "Automatic modulation classification using compressive convolutional neural network," *IEEE Access*, vol. 7, pp. 79636–79643, 2019.

[19] V.-S. Doan, T. Huynh-The, C.-H. Hua, Q.-V. Pham, and D.-S. Kim, "Learning constellation map with deep CNN for accurate modulation recognition," in *Proc. GLOBECOM IEEE Global Commun. Conf.*, Dec. 2020, pp. 1–6.

[20] Y. Kumar, M. Sheoran, G. Jajoo, and S. K. Yadav, "Automatic modulation classification based on constellation density using deep learning," *IEEE Commun. Lett.*, vol. 24, no. 6, pp. 1275–1278, Jun. 2020.

[21] S. Peng, H. Jiang, H. Wang, H. Alwageed, Y. Zhou, M. M. Sebdani, and Y.-D. Yao, "Modulation classification based on signal constellation diagrams and deep learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 3, pp. 718–727, Mar. 2019.

WEI-TAO ZHANG received the Ph.D. degree in control science and engineering from Xidian University, Xi'an, China, in 2011.

He is currently an Associate Professor with the School of Electronic Engineering, Xidian University. He is also a Faculty Research Fellow with the Research Institute of Advanced Remote Sensing Technology, Xidian University. His research interests include blind signal processing, tensor analysis, and machine learning.

DAN CUI received the B.S. degree in electronic and information engineering from the Xi'an University of Science and Technology, Xi'an, China, in 2019. She is currently pursuing the Ph.D. degree with the Department of Electronic Engineering, Xidian University, Xi'an.

Her current research interest includes machine learning.

SHUN-TIAN LOU (Member, IEEE) was born in Zhejiang, China, in 1962. He received the B.Sc. degree in automatic control and the M.Sc. degree in electronic engineering from Xidian University, Xi'an, China, in 1985 and 1988, respectively, and the Ph.D. degree in navigation guidance and control from Northwest Polytechnical University, Xi'an, in 1999.

From 1999 to 2002, he was a Postdoctoral Fellow with the Institute of Electronic Engineering, Xidian University. He is currently a Professor with the School of Electronic Engineering, Xidian University. His research interests include signal processing, pattern recognition, and intelligent control using neural networks and fuzzy systems.