IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 12, 2019, accepted January 22, 2019, date of publication January 31, 2019, date of current version February 27, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2896128

Multi-Spectral RGB-NIR Image Classification Using Double-Channel CNN

JIONGHUI JIANG[1,2], XI'AN FENG[1], FEN LIU[1], YINGYING XU[3], AND HUI HUANG[3]
[1]School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China
[2]Zhijiang College, Zhejiang University of Technology, Zhejiang 312030, China
[3]Computer Science Department, Wenzhou University, Wenzhou 325035, China

Corresponding authors: Xi'an Feng (fengxa@nwpu.edu.cn) and Hui Huang (huanghui@wzu.edu.cn)

This work was supported in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LY15F020033, Grant LSZ19F020001, and Grant LQ16G010006, in part by the Natural Science Foundation of China under Grant 61702376 and Grant 61671378, and in part by the Key Science and Technology Innovation Project of Wenzhou under Grant 2018ZG011.

ABSTRACT As 4-sensor line scan camera technology has matured, red (R), green (G), blue (B), and near-infrared (RGB-NIR) datasets have begun to appear in large numbers. The RGB-NIR data contain the rich color features of the RGB image and the sharp edge features of the NIR image. At present, in many studies, the RGB-NIR data are input directly into the processing algorithms for calculation of the 4D data; in these cases, redundant information is included, and the high correlation between the bands results in an inability to fully exploit the characteristics of the RGB-NIR data. In this paper, we propose a double-channel convolutional neural network (CNN) algorithm that takes into account the strong correlation between the R, G, and B bands in aerial images and the weaker correlation between the NIR band and the R, G, and B bands. First, the features of the RGB and NIR bands are calculated in two different CNN networks, and subsequently, feature fusion is performed in the fully connected layer. This is followed by the classification. By combining the two neural networks of RGB-CNN and NIR-CNN, the respective characteristics of the RGB-NIR data are fully exploited.

INDEX TERMS Multi-spectral, CNN, RGB-NIR, double-channel CNN.

I. INTRODUCTION

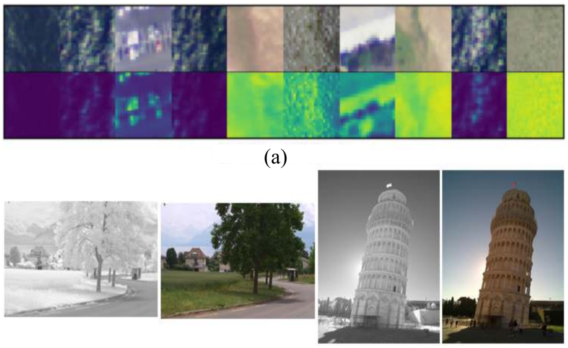
In multi-spectral red, green, blue, near-infrared (RGB-NIR) images, the visible (RGB) and near-infrared (NIR) spectral bands are captured simultaneously by a 4-sensor line scan camera [1]. The RGB spectral bands are in the visible range (400-700 nm), whereas the NIR spectral band is beyond the visible range (700-1100 nm). As a result, a scene captured with an RGB-NIR image exhibits a wide range of characteristics [2]. The combination of RGB and NIR data provides rich image features for image recognition and classification (e.g., [3]–[5]). As a result of the emergence of RGB-NIR datasets in various fields, multi-spectral RGB-NIR image classification has been widely used in video surveillance, medical imaging, satellite remote sensing, vegetation mapping, and other fields [6]–[8].

In recent years, many researchers have investigated multi-spectral image recognition and classification from different aspects. Brown and Süsstrunk proposed the MSIFT algorithm, a multispectral scale-invariant feature transform (SIFT) descriptor that, when combined with a kernel-based classifier, exceeded the performance of state-of-the-art scene recognition techniques (e.g., GIST) and their multispectral extensions [9]. Salamati et al. proposed to use visible and NIR images as input to a classifier in the form of feature vectors to classify the materials in the image. The relation between the visible and NIR information provided an improvement in the image-based machine classification. The materials were more accurately classified when the NIR information was present [10]. Miyamoto et al. concluded that the availability of high-resolution (HR) training data such as balloon-based image mosaics was useful for the classification of NIR color video images and it was found that the combination of the nadir and off-nadir video images was effective for the classification of wetland vegetation [11]. Han *et al.* proposed a convolutional neural network (CNN)-based super-resolution (SR) algorithm for up-scaling NIR images under low-light conditions using the corresponding visible images. The high-frequency (HF) components were extracted from the up-scaled low- resolution (LR) NIR

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

image and the corresponding HR visible image and were then used as multiple inputs into the CNN [12]. Researchers also have developed different types of CNN algorithms to use RGB and NIR data as input layers for remote sensing image classification, such as Alex-Net, deep CNN, and VGG [13]–[16]. Through the use of CNN algorithms, the multi-spectral RGB-NIR image recognition rate is greatly improved.

This paper presents a double-channel CNN algorithm to improve the classification accuracy of RGB images and NIR images. Because the R, G, and B bands are strongly correlated, whereas the NIR band is not strongly correlated to the information in the R, G, and B bands, the direct use of the RGB and NIR images as input layers into a CNN network does not provide the full advantages of the different features in the RGB-NIR images. Redundant information is included and mutual interference between the features will also occur. Considering that the color information is richer in the RGB image, whereas the NIR image provides more edge information, we developed a double-channel CNN algorithm to capture the different features of the RGB images and NIR images. First, different source image features are convolved using two independent CNN networks and are then pooled; subsequently, the two networks are fused based on features in the fully connected layer. Finally, the loss value is calculated, followed by target identification and classification.


(a)

(b)

FIGURE 1. Examples of the RGB-NIR dataset and the SAT-4 and SAT-6 dataset. (a) SAT-4 and SAT-dataset. (b) RGB-NIR dataset.

II. MULTI-SPECTRAL RGB-NIR IMAGE FEATURE ANALYSIS

The RGB image consists of the three color channels of R, G, and B. The NIR band is located in the electromagnetic spectrum between the visible and mid-infrared bands; it has a wide wavelength range and provides clear image information, even in low light conditions. With the advent of the 4-sensor RGB-NIR camera, we can capture RGB-NIR data simultaneously. Figure 1(a) shows an RGB-NIR remote sensing dataset (SAT-4 and SAT-6 dataset); Fig.1(b) shows simultaneously acquired RGB and NIR data captured by an RGB-NIR camera (RGB-NIR dataset). It is observed that the RGB and NIR images reflect different characteristics of the same target.

The literature [9], [17] suggests that the correlation between the NIR and the R, G, and B bands is significantly lower than the correlation between the individual bands. In this study, we use the mutual information in the bands to determine the correlation between the RGB and NIR data. The mutual information is a measure of the statistical correlation between two random variables. It can also be interpreted as the correlation between the two types of images. The expression is as follows:

$$I(X,Y)=\sum_{x\in X}\sum_{y\in Y}p(x,y)\log(\frac{p(x,y)}{p(x)p(y)}) \tag{1}$$

We used 100 random RGB-NIR images to determine the mutual information (correlation) between the G, R, B, and NIR bands. Figure 2(a), (b), and (c) show the correlations between the G, R, and B bands. Fig. 2(d), (e), and (f) show the correlations between the G, R, B, and NIR bands. It is observed that most of the values in the R-G, R-B, and G-B relationships are concentrated in the range of 0-20. The distributions of R_NIR, B_, and G_NIR are quite scattered and the correlation is very weak.

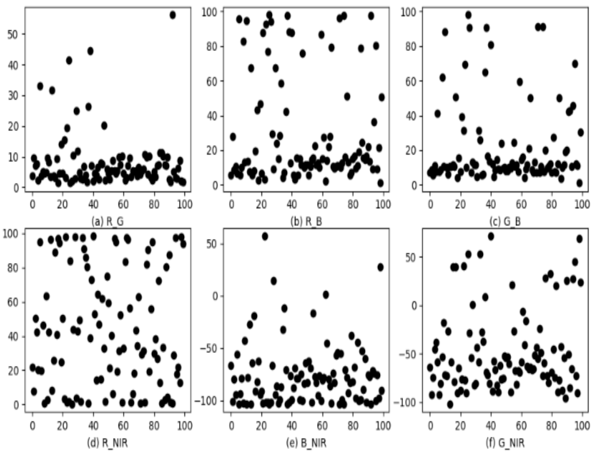


FIGURE 2. Correlation between the R, G, B, and NIR bands; the pixels are sampled from 100 images.

The results in Fig. 2 indicate that if the RGB-NIR data are input directly to the CNN, the amount of information from the features is increased, thereby improving the recognition rate. At the same time, it also causes mutual interference between features and the image information of the R, G, B, and NIR bands cannot be fully utilized.

III. DOUBLE-CHANNEL CNN MODEL

The double-channel CNN Model is an improvement of the traditional CNN model. Currently, this method is being applied to image comparisons, such as fingerprint analysis, medical image analysis, facial recognition, etc. [18], [19]. Bromley *et al.* [20] proposed a two-branch network based on the Siamese network for signature authentication. Different from SIFT, the two-branch network allows patch1 and patch2 to extract feature vectors through two networks; subsequently, a similarity loss function is applied to the two feature vectors in the last layer and the network training is combined, thereby improving the precision. Zagoruyko and Komodakis [21] and Hamester *et al.* [18] improved the two-channel network based on the Siamese network and the spatial pyramid pooling (SPP) proposed by He *et al.* [22] for a similarity comparison of large images. The method achieved good results.

In this paper, the double-channel CNN represents an improvement based on the two-branch network. The features are fused in the fully connected layer of the CNN, which is a different approach from the feature comparison of the 2-branch network. The double-channel CNN is used for target recognition and classification of the multi-spectral image. In the double-channel CNN, the two neural networks are completely independent. The weights of the convolutional layer and the pooling layer are also independent. In the fully connected layer, the features are merged and the classification loss function derived based on the joint features of the RGB images and NIR images. The flowchart of the process is shown in Fig. 3.

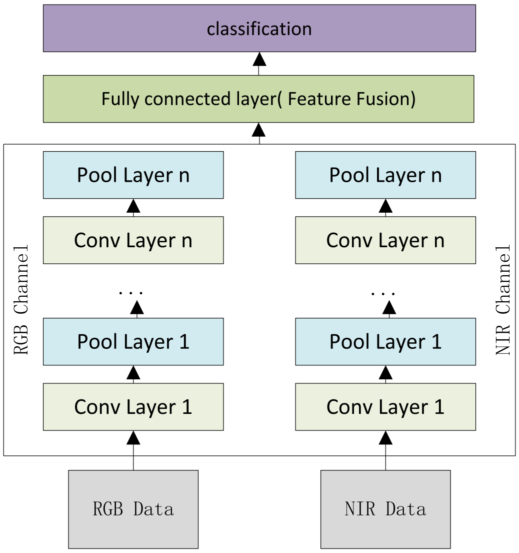


FIGURE 3. **Double-channel CNN model.**

In the double-channel CNN, the RGB image is represented as $x \in R^{m \times m}$ and the NIR Image is defined as $n \in R^{\mathrm{N} \times N}$. The convolutional layer is expressed as:

The feature maps of the upper layer are convolvedwith a learnable convolution kernel and then an output function map is obtained by an activation function f. Each output map is the value of a combined convolution of multiple input maps, as shown in Equation (2):

$$\begin{cases} \mathrm{conv}(x_j^l) = f(\sum_{i \in M_j} w_{ij}^l \times x_i^{l-1} + b_j^l) \\ conv(n_j^l) = f(\sum_{i \in N_j} \omega_{ij}^l \times n_i^{l-1} + \delta_j^l) \end{cases} \tag{2}$$

where l represents the current layer, M_j and N_j represent the set of selected input maps; each output map has an extra offset b and weight w. For a particular output map, the convolution kernel that convolves each input map is different.

Each convolution layer l is connected to a pooling layer $l+1$. For the sub-sampling layer, there are N input maps and there are N output maps but each output map is smaller, as shown in Equation (3):

$$\begin{cases} x_j^{l+1} = lrn(\beta_j^{l+1} down(x_j^l) + b_j^{l+1}) \\ n_j^{l+1} = lrn(\varphi_j^{l+1} down(n_j^l) + \delta_j^{l+1}) \end{cases} \tag{3}$$

where *down*() represents a downsampling function. A typical operation generally consists of summing all the pixels of a different $n \times n$ block of the input image. In this manner, the output image is reduced by n times in both dimensions. Each output map corresponds to its own multiplicative bias β and an additive bias b. The lrn() function (local response normalization) is a method to improve the accuracy during deep learning. The principle of the local response normalization is to mimic the inhibition of the adjacent neurons by biologically active neurons.

The fully connected layer converts all two-dimensional (2D) feature maps into inputs for a fully connected one-dimensional (1D) network. When entering the final 2D feature maps into a 1D network, a very convenient method is to join all the output feature maps into a long input vector, as shown in Equation (4).

$$\begin{cases} x^l \rightarrow [X_1, X_2, X_3, \ldots, X_i] \\ n^l \rightarrow [N_1, N_2, N_3, \ldots, N_i] \end{cases} \tag{4}$$

The fusion $[X_1, X_2, X_3, \ldots, X_i]$ and $[N_1, N_2, N_3, \ldots, N_i]$ are expressed as Equation (5):

$$(X, Y) \rightarrow \begin{bmatrix} X_1, N_1 \\ X_2, N_2 \\ X_3, N_3 \\ \ldots \;,\; \ldots \\ X_i, \;\; N_i \end{bmatrix}$$

or

$$(X, Y) \rightarrow [X_1, X_2, X_3, \ldots X_i, N_1, N_2, N_3, \ldots, N_I]^T \tag{5}$$

Based on the fully connected layer, we calculate the final output of a num_classes_sized vector as $[Y_1, Y_2, Y_3, \ldots, Y_t]$. Subsequently, a softmax classification is performed based on the output and the prediction result is as shown in Equation (6).

$$soft\ \max(Y_t) = \frac{\exp(Y_t)}{\sum \exp(Y_t)} \tag{6}$$

Finally, the error function between the predicted and actual values of the model is determined. Through neural network back-propagation, each neuron is continuously trained to update the network weights and offset values so that the error gradient is reduced and the error is reduced; the model is continuously optimized, as defined in Equation (7):

$$loss(H_{y'}(y)) = mean(-\sum_t y'_t \log(y_t)) \tag{7}$$

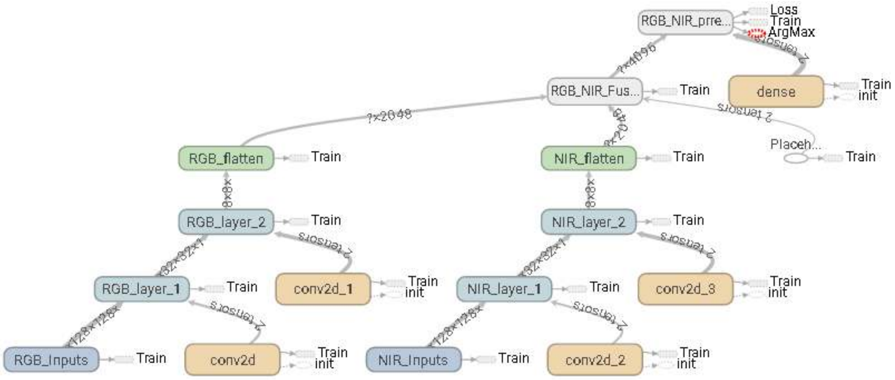


FIGURE 4. Tensorboard visualization of the double-channel CNN.

where y'_t refers to the value of the actual t label; y_t is the actual t element in the output vector $[Y_1, Y_2, Y_3, \ldots, Y_t]$ of softmax; *mean*() is the one to be averaged over the vector.

IV. DATA SOURCE

The experimental data consisted of the RGB-NIR dataset [9] and the SAT-4 and SAT-6 airborne dataset [23].

The RGB-NIR dataset consisted of 477 images in 9 categories captured in RGB and NIR. The images were captured using separate exposures from modified SLR cameras using visible and NIR filters. The original size of the images in the RGB-NIR dataset is 1024 x 680 or 512 x 768 window size. For more info on the NIR photography, please see the references below. The scene categories are country, field, forest, indoor, mountain, old building, street, urban, and water.

The SAT-4 and SAT-6 images were extracted from the National Agriculture Imagery Program (NAIP) dataset, which consists of 330,000 scenes spanning the Continental United States (CONUS); it covers different landscapes such as rural areas, urban areas, densely forested regions, mountainous terrain, small to large water bodies, agricultural areas, etc. The images consist of 4 bands, i.e., R, G, B, and NIR; a 28 x 28 window size was used to obtain images with varied information.

The RGB-NIR dataset is available at http://ivrg- www.epfl.ch/supplementary_material/cvpr11/nirscene1.zip

The SAT-4 and SAT-6 dataset is available at https://drive.google.com/uc?id=0B0Fef71_vt3PUkZ4YVZ5WWNvZWs&export=download.

V. EXPERIMENTAL RESULTS AND DISCUSSION

The experiment was conducted by using the Python 3.6 and TensorFlow platforms and the double-channel CNN model. The validity of the double-channel CNN model is verified by calculating the classification accuracy, the loss function, and the degree of overfitting. The performance of the double-channel CNN model algorithm is determined by comparing the recognition rate with the single-channel CNN model for the same length and the same parameter settings. We also compared the results with that of the classification algorithms used by other researchers for the same datasets.

Because the RGB-NIR dataset consists of raw image data with different images sizes of 1024 × 680 or 512 × 768 window size, the data required preprocessing. Reference [9] tested different compressed dimensions of RGB-NIR raw images and compared the recognition rates. The experimental results demonstrated that an image size of 128 x 128 resulted in good performance. Therefore, in this study, the images were compressed to a uniform 128 x 128 size using the Tensorflow bilinear interpolation algorithm without loss of image quality. We used the TFrecord method integrated into the TensorFlow software to classify the RGB-NIR dataset images data into nine categories. The data was randomly extracted using the shuffle_batch method and was used as input into the double-channel CNN model for calculation. The SAT-4 and SAT-6 airborne datasets are standardized and there was no need to preprocess the data. The training and test data were directly input into the mat data and used as input for the model for calculation.

Prior to the fully connected layer, the double- channel CNN model consisted of two LeNet-5 [24]. The parameters used for the convolution and pooling in the double channel CNN architecture are the same as the parameters of the LeNet-5. Figure 4 shows the TensorBoard visualization in TensorFlow of the processing flow of the RGB-NIR dataset using the double-channel CNN model. In the RGB bands, the convolution layers use a 5x5 convolution kernel. The activation function is the 'ReLu'. The pooling layers utilize 4x4 regions for pooling and the step lengths are 4x4; the data are normalized using local response normalization. In the NIR band, the convolution layers use a 3x3 convolution kernel. After the convolution is completed, the 'ReLu' activation function is selected. The pooling layer uses 4x4 regions for pooling and the step lengths are 4x4; the data are then normalized using local response normalization. A smaller convolution kernel is used for the NIR band than the RGB band to improve the edge information of the NIR image. Finally, the fully connected layer vector feature is fused into a 1D vector using the TensorFlow concat function and then the image is classified. The fully connected layer uses the net dropout method to randomly discard 60% of the neurons to avoid overfitting.

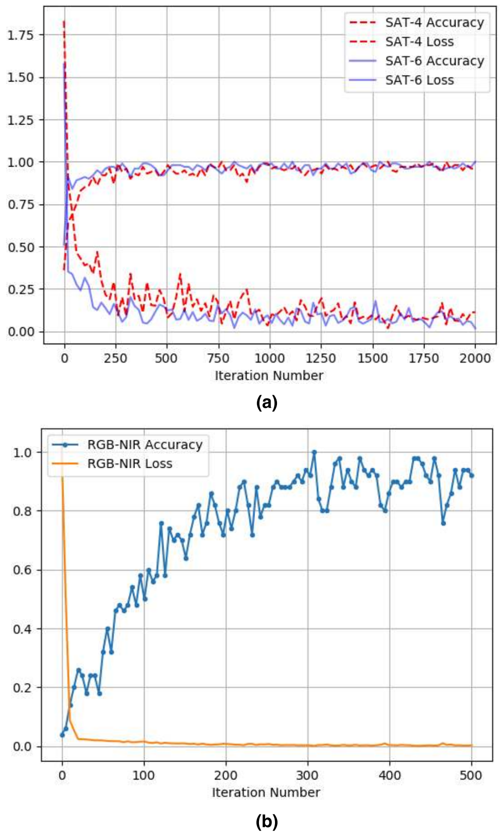


FIGURE 5. Accuracy and Loss of the RGB-NIR dataset and the SAT-4 and SAT-6 airborne dataset. (a) SAT-4 and SAT-6 airborne dataset. (b) RGB-NIR dataset.

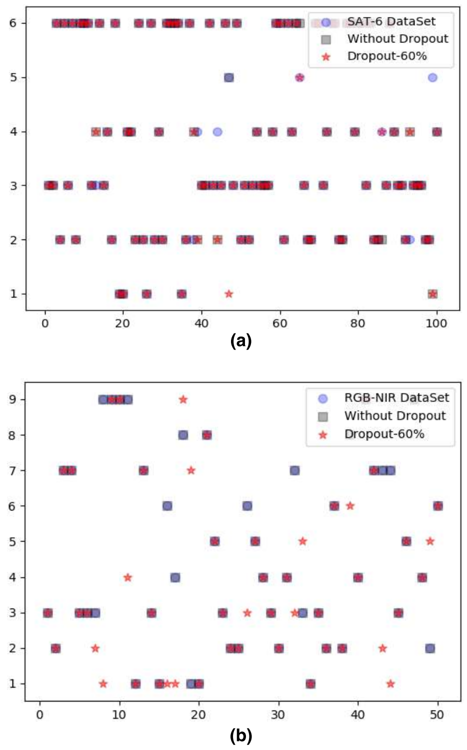


FIGURE 6. Net dropout test results of the RGB-NIR dataset and the SAT-4 and SAT-6 airborne dataset. (a) SAT-6 airborne dataset. (b) RGB-NIR dataset.

Figure 5 shows the statistical results of the accuracy and loss of the RGB-NIR dataset and the SAT-4 and SAT-6 airborne dataset using the double-channel CNN model. Figure 5(a) shows the accuracy and loss of the SAT-4 and SAT-6 airborne dataset. The SAT-4 and SAT-6 images have a window size of only 28x28; therefore, a smaller convolution kernel and pooling regions are used. In the RGB and NIR bands, a 3x3 convolution kernel is used, the pooling layer uses 2x2 regions for pooling, and the step lengths are 2x2. The other parameters are the same as shown in Fig. 4. In the model, the batch size is 100, 2,000 iterations are performed, and the loss and accuracy values of the model are recorded every 20 times. Figure 5(b) shows the accuracy and loss of the RGB-NIR dataset. The model batch size is 50, 500 iterations are performed, and the loss and accuracy values of the model are recorded every 10 iterations. The results demonstrate that the loss gradually decreases and the accuracy rate increases as the number of iterations increases, indicating that the double-channel CNN model exhibits good performance and is valid.

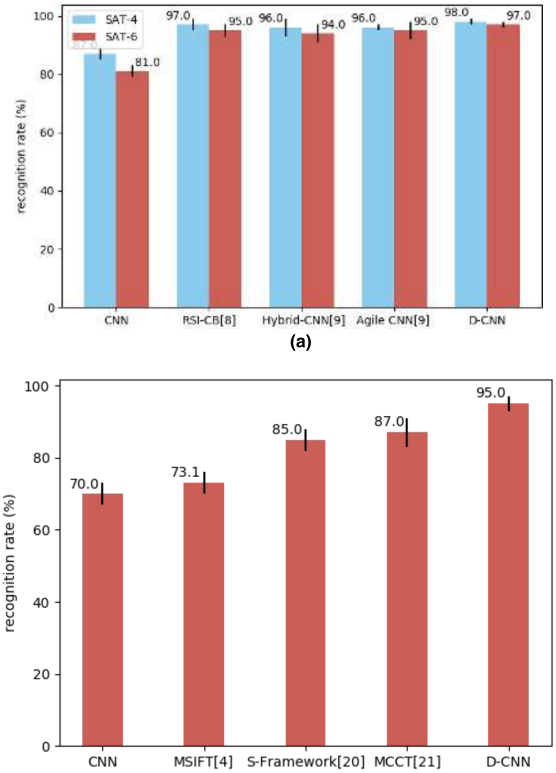
Figure 6 shows the net dropout test results of the two datasets; the objective is to determine test whether the model is over-fitted. Figure 6(a) shows that, after 1000 iterations and at a batch size of 100, the SAT-6 accuracy is about 95% without net dropout and the net dropout accuracy is about 93%. We extracted 100 SAT-6 test data and imported them into the model for testing; without net dropout, the accuracy was 92% and with dropout, the accuracy was 93%. The experiments indicated that there was a small degree of over-fitting in the SAT-6 data set when there was no net dropout. Figure 6(b) shows the overfitting test results for the RGB-NIR dataset. There were only 477 images in the RGB-NIR dataset. Therefore, after 230 random extractions of the data without net dropout, the recognition rate was 100%. The model completely identified all internal data and over-fitting has been happened. The recognition rate with the net dropout was about 75%, which represents the ambiguity of the data. The test results show that the double-channel CNN effectively avoids the overfitting of the model by using the net dropout, as shown in Fig. 6.

The double-channel CNN model consists of two independent single-channel CNNs. To evaluate the difference between the double-channel CNN model and the single-channel CNN model, we determined the recognition rates of both methods. The lengths and parameter settings of the single-channel CNN model are basically the same as those of the double-channel CNN model. We used the RGB-NIR data as the input layer for the single-channel CNN calculation (Table 1). It was found that for the same number of iterations, the recognition rate was significantly higher for the double-channel CNN model than the single-channel CNN model.

We compared the classification results of the double-channel CNN model with those of some recent classification algorithms, as shown in Fig. 7. Figure 7(a) shows the comparison of the classification accuracy for the SAT-4

TABLE 1. Recognition rates of the single-channel CNN and double-channel CNN.

Dataset	Single-channel CNN	Double-channel CNN	Iteration number
SAT -4	0.89(±0.2)	0.98(±0.1)	2000
SAT-6	0.81(±0.2)	0.97(±0.1)	2000
RGB-NIR	0.70(±0.3)	0.95(±0.2)	500



(a)

(b)

FIGURE 7. Recognition rates for different algorithms. (a) SAT-4 and SAT-6 dataset classification. (b) RGB-NIR dataset classification.

and SAT-6 dataset; the algorithms include the RSI-CB of Li *et al.* [13], the hybrid aggregation (pooling) approach by Han and Chen [14], and the agile CNN architecture by Zhong *et al.* [16]. The results show that the accuracies of the double-channel CNN model and those of the other algorithms are very close. Figure 7(b) is a comparison of the double-channel CNN model classification results of the RGB-NIR dataset and those of other researchers, including the MSIFT by Brown and Süsstrunk [9], the sensing framework by Karam *et al.* [25], and the MCCT by Rahman *et al.* [26]. The results indicate that the classification accuracy of the double-channel CNN model is significantly higher than that of the other algorithms.

VI. CONCLUSIONS

The 4-sensor RGB-NIR line scan camera is widely used in video surveillance, medical imaging, satellite remote sensing, and other fields and the simultaneous acquisition of RGB and NIR image data has become a topic of broad and current interest. Based on the correlation between the G, R, B, and NIR bands, we developed the double-channel CNN model to classify the RGB-NIR image data. The double-channel CNN model consists of two independent CNN networks, which describe the RGB and NIR image features. Feature fusion is performed in the fully connected layer and the last layer performs the classification; this configuration makes good use of the different features of the RGB-NIR images. The experimental results show that the double-channel CNN algorithm is better able to exploit the features of the RGB and NIR images than the single-channel CNN algorithm. In addition, the algorithm has certain advantages over other similar algorithms.

REFERENCES

[1] X. Soria, A. D. Sappa, and R. I. Hammoud, "Wide-band color imagery restoration for RGB-NIR single sensor images," *Sensors*, vol. 18, no. 7, p. 2059, 2018.

[2] Z. Chen, X. Wang, and R. Liang, "RGB-NIR multispectral camera," *Opt. Express*, vol. 22, no. 5, pp. 4985–4994, 2014.

[3] S. Pan, A. Chen, and P. Zhang, "Securitas: User identification through RGB-NIR camera pair on mobile devices," in *Proc. 3RD ACM Workshop Secur. Privacy Smartphones Mobile Devices*, 2013, pp. 99–104.

[4] J. Lezama, Q. Qiu, and G. Sapiro, "Not afraid of the dark: Nir-vis face recognition via cross-spectral hallucination and low-rank embedding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6807–6816.

[5] X. Cheng, Y. Tao, Y. R. Chen, and Y. Luo, "Nir/MIR dual–Sensor machine vision system for online apple stem–end/calyx recognition," *Trans. ASAE*, vol. 46, no. 2, pp. 551–558, 2003.

[6] X. Soria, A. D. Sappa, and A. Akbarinia, "Multispectral single-sensor RGB-NIR imaging: New challenges and opportunities," in *Proc. 7th Int. Conf. Image Process. Theory*, 2018, pp. 1–6.

[7] J. Zhang, M. Li, Z. Sun, H. Liu, H. Sun, and W. Yang, "Chlorophyll content detection of field maize using RGB-NIR camera," *IFAC-PapersOnLine*, vol. 51, no. 17, pp. 700–705, 2018.

[8] G. Choe, S.-H. Kim, S. Im, J.-Y. Lee, S. G. Narasimhan, and I. S. Kweon, "RANUS: RGB and nir urban scene DataSet for deep scene parsing," *IEEE Robot. Autom. Lett.*, vol. 3, no. 3, pp. 1808–1815, Jul. 2018.

[9] M. Brown and S. Süsstrunk, "Multi-spectral SIFT for scene category recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit*, Jun. 2011, pp. 177–184.

[10] N. Salamati, C. Fredembach, and S. Süsstrunk, "Material classification using color and NIR images," in *Proc. Color Imag. Conf.*, vol. 7, 2009, pp. 216–222.

[11] M. Miyamoto, K. Yoshino, and K. Kushida, "Classification of wetland vegetation using aerial photographs by captive balloon cameras and aero nir color video image, Kushiro northern wetland in Japan," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, vol. 4, Jul. 2002, pp. 1982–1984.

[12] T. Y. Han, D. H. Kim, S. H. Lee, and B. C. Song, "Infrared image super-resolution using auxiliary convolutional neural network and visible image under low-light conditions," *J. Vis. Commun. Image Represent.*, vol. 51, pp. 191–200, Feb. 2018.

[13] H. Li *et al.* (2017). "RSI-CB: A large scale remote sensing image classification benchmark via crowdsource data." [Online]. Available: https://arxiv.org/abs/1705.10450

[14] X.-H. Han and Y.-W. Chen, "Generalized aggregation of sparse coded multi-spectra for satellite scene classification," *ISPRS Int. J. Geo-Inf.* vol. 6, no. 6, p. 175, 2017.

[15] V. Risojević, "Analysis of learned features for remote sensing image classification," in *Proc. 13th Symp. Neural Netw. Appl.*, Nov. 2016, pp. 1–6.

[16] Y. Zhong, F. Fei, Y. Liu, B. Zhao, H. Jiao, and L. P. Zhang, "SatCNN: Satellite image dataset classification using agile convolutional neural networks," *Remote Sens. Lett.*, vol. 8, no. 2, pp. 136–145, Feb. 2016.

[17] G. French, G. Finlayson, and M. Mackiewicz, "Multi-spectral pedestrian detection via image fusion and deep neural networks," *J. Imaging Sci. Technol.*, vol. 62, no. 5, pp. 176–181, 2018.

[18] D. Hamester, P. Barros, and S. Wermter, "Face expression recognition with a 2-channel convolutional neural network," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2015, pp. 1–8.

[19] Z. Zhou, J. Shin, L. Zhang, S. Gurudu, M. Gotway, and J. Liang, "Fine-tuning convolutional neural networks for biomedical image analysis: Actively and incrementally," in *Proc. CVPR*, Jul. 2017, pp. 4761–4772.

[20] J. Bromley *et al.*, "Signature verification using a 'siamese' time delay neural network," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 7, no. 4, pp. 669–688, 1993.

[21] S. Zagoruyko and N. Komodakis, "Learning to compare image patches via convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 4353–4361.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

[23] S. Basu, S. Ganguly, S. Mukhopadhyay, R. Dibiano, M. Karki, and R. Nemani, "DeepSat—A learning framework for satellite imagery," in *Proc. ACM SIGSPATIAL*, 2015, pp. 1–10.

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

[25] L. J. Karam *et al.* (2018). "Generative sensing: Transforming unreliable sensor data for reliable recognition." [Online]. Available: https://arxiv.org/abs/1801.02684

[26] M. M. Rahman, S. Rahman, and M. Shoyaib, "MCCT: A multi-channel complementary census transform for image classification," *Signal Image Video Process.*, vol. 12, no. 2, pp. 281–289, 2018.

JIONGHUI JIANG was born in Hangzhou, Zhejiang, China, in 1981. He received the master's degree in computer software and theory from the Zhejiang University of Technology, in 2008. He is currently pursuing the Ph.D. degree with Northwestern Polytechnical University.

Since 2008, he has been an Engineer with the Zhijiang College, Zhejiang University of Technology. He has published several papers at international conferences, more than three of them were included in EI. His research interests include graphic and image processing, deep learning, remote sensing imagery, and medical image processing.

XI'AN FENG was born in Xi'an, China, in 1962. He received the master's degree in watercraft and the Ph.D. degree in signal and information processing from Northwestern Polytechnical University, in 1991 and 2004, respectively.

Since 2001, he has been teaching at Northwestern Polytechnical University. He has published more than 60 papers at international conferences, more than 30 of them were included in SCI and EI. His research interests include signal and information processing, array signal processing, underwater target recognition, underwater target tracking, underwater acoustic imaging, torpedo self-guide, and hydroacoustic confrontation and opposition.

Dr. Feng is a Senior Member of the Chinese Institute of Electronics and a member of the Chinese Acoustics Society. He received two second prizes and six third prizes for provincial and ministerial level scientific and technological progress.

FEN LIU was born in Jiaxian, Shanxi, China, in 1983. She received the master's degree in computer technology from the Xidian University of Electronic Technology, in 2011. She is currently pursuing the Ph.D. degree with Northwestern Polytechnical University.

Since 2008, she has been a Teacher with Yan'an University. She has published several papers at international conferences. Her research interests include signal and information processing, array signal processing, and artificial intelligence.

YINGYING XU received the Ph.D. degree in communication and information system from Beijing Jiaotong University, in 2014. She is currently a Lecturer with the Computer Science Department, Wenzhou University, Wenzhou, China. Her research interests include data mining, intelligent transportation, and machine learning.

HUI HUANG received the master's degree in computer software and theory from the Zhejiang University of Technology, in 2008. He is currently pursuing the Ph.D. degree with Northwestern Polytechnical University. He is the Deputy Director of the Computer Science Department, Wenzhou University, Wenzhou, China. His research interests include image processing, parallel computing, and machine learning.