Representation and Classification of Auroral Images Based on Convolutional Neural Networks

Qiuju Yang and Penghui Zhou

***Abstract*—Auroral forms are correlated with certain physical processes in the magnetosphere and ionosphere. It is, therefore, desirable to automatically classify the vast amount of observed auroral images and make large statistical studies. The key problem in classification tasks is image representation. In this article, using the adaptive feature learning ability of convolutional neural networks, an end-to-end auroral image classification network is proposed, which automatically classifies the auroral images observed at the Chinese Yellow River Station into four classes: arc, drapery corona, radial corona, and hotspot corona. Based on the AlexNet, our method exploits the advanced spatial transformer network (STN) and large margin Softmax (L-Softmax) loss function to extract auroral features. STN is able to learn invariance to translation, scaling, and rotation, whereas L-Softmax increases the difficulty of auroral feature learning so that it encourages the intraclass compactness and interclass separability between learned features. The proposed method was validated on the auroral image datasets by supervised classification, image retrieval, and statistical analysis of the temporal occurrence distributions of the four auroral categories. Experimental results showed that after trained on the winter auroral observations in 2003, the proposed model achieves an average classification accuracy of 93.7% on the auroral data of the following five winters (2004–2009) while maintaining high efficiency, which is superior to the previously reported articles.**

***Index Terms*—Auroral image classification, convolutional neural network (CNN), large margin Softmax (L-Softmax) loss function, spatial transformer network (STN).**

I. Introduction

AURORA is a luminous glow of the upper atmosphere caused by energetic particles that enter into the atmosphere from the magnetosphere. Different auroral types are correlated with specific Earth's magnetosphere activities, and effective auroral classification benefits the prediction of Earth's magnetosphere structure and space weather. How to accurately characterize and reasonably classify aurora has become an indispensable part of the study of space physics [1], [2].

Aurora is always analyzed in a "case study" way by space scientists [3]–[6]. These studies, however, are subjective, operator dependent, and very time consuming. Syrjäsuo and Donovan [8] first introduced the computer vision technique into auroral image classification and classified auroras into arcs, patchy auroras, omega bands, and north–south structures according to their shape information. Since then, a lot of automatic auroral studies, including auroral image representation, classification, retrieval, and segmentation, have been emerging [9]–[15]. Wang *et al.* [16] proposed to combine the local binary pattern operator and a delicately designed block partition scheme to characterize auroral morphology (shape and texture). Rao *et al.* [17] described the color variants of scale-invariant feature transform features for performing the automated classification of all-sky camera images into three mutually exclusive classes: aurora, no aurora, and cloudy. Yang and Hu [18] utilized the Weber local descriptor to represent and classify auroral images into the arc, drapery corona, radial corona, and hotspot. By combining multiple handcrafted features (grayscale, structural, and textural features) extracted from auroral images, Zhong *et al.* [19] proposed an auroral image classification method based on multifeature latent Dirichlet allocation.

Above auroral image classification techniques include two separate procedures: feature extraction and classifier design. The extracted features represent the intensity, shape, and texture of the aurora and play an important role in auroral image classification. Ideally, they should be distinctive and, at the same time, robust to a variety of possible image transformations (e.g., rotation, scaling, and translation). However, hand designing an effective feature is always a lengthy process that requires considerable expertise to delicately design according to specific data and tasks.

Recently, deep learning based algorithms have quickly dominated most vision-based tasks due to their discriminative power [21]. Image representations based on the convolutional neural network (CNN) have attracted increasing attention in the community and demonstrated impressive performance [34], [38], [39]. The main difference between CNNs and traditional machine learning methods is that CNNs learn features directly from the data without an additional manual feature extraction process. So far, several CNN models have been studied for the applications of various computer vision tasks, including OverFeat [28], AlexNet [32], GoogLeNet [35], VGGNet [36], ResNet [37], Inception [20], and their variants. Razavian *et al.* [29] used pretrained OverFeat to extract features and proved the off-the-shelf features yielded better results than the handcrafted features on various computer vision tasks.

These advances have been quickly brought into the field of auroral image classification. Wang and Yang [40] introduced AlexNet to automatically classify the dayside auroral images of

Manuscript received September 17, 2019; revised December 27, 2019; accepted January 21, 2020. Date of publication January 27, 2020; date of current version February 13, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 41504122, and in part by the Fundamental Research Funds for the Central Universities under Grant GK201903021. *(Corresponding author: Qiuju Yang.)*

The authors are with the School of Physics and Information Technology, Shaanxi Normal University, Xi'an 710119, China (e-mail: yangqiuju@snnu.edu.cn; zhoupenghui@snnu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2020.2969245

Yellow River Station (YRS) into the arc, drapery corona, radial corona, and hotspot. Clausen and Nickisch [30] utilized the pretrained Inception-v4 to automatically classify auroral images into clear/no aurora, cloudy, moon, arc, diffuse, and discrete from the Oslo Auroral THEMIS (OATH) dataset. Niu *et al.* [10] proposed a weakly supervised semantic segmentation method to achieve joint pixel-level localization of the key local structure and image-level classification of the auroral images. Han *et al.* [25] proposed a multisize kernels CNN with eye movement guided task-specific initialization to classify auroral images into the arc, drapery corona, and radiation corona.

However, there is a gap between the above classification methods and the actual application. On the one hand, the aurora is rich in spatial variations and often lacking well-defined interclass boundaries. The extracted CNN features for auroral image classification should be rotation invariant and discriminative [42]–[44]. But as shown in [40] and [30], the auroral features obtained by AlexNet and Inception-v4 are not satisfactory. On the other hand, the number of auroral images is very huge and increasing every winter. Therefore, the method for auroral image classification must be efficient and easy-to-implement before it has practical value. However, the training of the patch scale model in [10] is a very time-consuming process, and obtaining the eye movement annotation in [25] is extremely expensive.

This article aims to develop a practical method to classify the vast amount of existing auroral images automatically. A new auroral image classification model based on CNN architectures is explored according to the unique characteristics of the aurora. First, we compare the three CNN models (AlexNet, VGG16, and Inception-v4) used in the previous automatic analysis of auroral images [23]–[24] and [30], and the simplest AlexNet is chosen as the backbone for auroral image classification in consideration of accuracy and efficiency. Second, considering the large spatial difference of auroral images (see Fig. 3), the spatial transformer network (STN) [22] is inserted before AlexNet. STN can adaptively transform and align the images according to the classification task (including translation, scaling, rotation, and more generic warping), which can make the model pay more attention to the region of interest (ROI) and improve the classification accuracy (such as [41]). Third, as a natural phenomenon, the continuous evolution of the aurora gives it abundant intraclass forms while lacking discriminative interclass boundaries. The most frequently occurring arc auroras last a long time and exhibit high intraclass appearance variance, whereas there is little interclass difference when auroras transit from one type to another as they are always gradually changing. In view of this, the large margin Softmax (L-Softmax) loss function [27] is adopted to optimize the auroral image classification model. L-Softmax loss function simultaneously maximizes the intraclass compactness and interclass separability. We pretrain the classification model on ImageNet and fine-tune it with our auroral image data.

The main contributions of our work can be summarized as follows.

1) We propose an easy-to-implement yet effective CNN model for the auroral image classification, which can help analyze the huge auroral image datasets.
2) The proposed classification model is domain knowledge based and its submodules, STN and L-Softmax loss function, are chosen according to the unique characteristics of the aurora.
3) In order to meet the actual demands in auroral physics, different from previous experimental settings, the proposed model is trained on one-winter auroral observations and tested on the following five-winter observations.
4) Extensive experiments are carried out to validate the effectiveness of the proposed model, which suggests the practical application value of our method to the automatic classification of the huge amount of auroral images.

The remainder of this article is organized as follows. Section II outlines the overall framework of the auroral image classification model and briefly introduces the STN and L-Softmax loss function. Section III presents the description of auroral image classification mechanism and dataset setup. The experiments and results are reported in Section IV. We conclude this article in Section V.

II. STN and L-Softmax-Based AlexNet for Auroral Image Representation

Fig. 1 shows the whole network architecture of the auroral image classification model. The spatial transformer network (STN) is embedded in the input layer of the AlexNet for the spatial transformation of auroral images, which enables the network to automatically learn more effective auroral regions in the training process. In addition, the L-Softmax loss function is utilized to optimize the network so that the network can learn more discriminatory auroral features. In the following, we will abbreviate the model to STN-Lsoftmax-AlexNet.

A. Spatial Transformer Network

Although CNN is a powerful classification model, it is still affected by the spatial diversity of data. Jaderberg *et al.* [22] proposed a new learning module, spatial transformer network, to solve this problem. STN can be trained to pay more attention to the ROIs. STN does not require the calibration of key points and can adaptively transform the input images according to a certain task (including translation, scaling, rotation, and other geometric transformations to select the most relevant regions). When the spatial variance of the input data is large, STN can be added to the existing CNN architecture to improve the classification accuracy of the model. In this article, the STN is inserted in the input layer of the AlexNet network (see Fig. 1), and the auroral images are affine transformed into the new images with strong characterization ability. As shown in Fig. 2, from the input U to the output V, STN consists of the localization network, grid generator, and sampler [22].

The first part localization network is a custom CNN architecture that computes the generated two-dimensional (2-D) affine transformation parameter θ. We define it as two convolutional layers, a max-pooling layer, a fully connected layer, and a regression layer. The network takes the input auroral images $U \in R^{H \times W}$ with width H and height W and outputs the 2×3

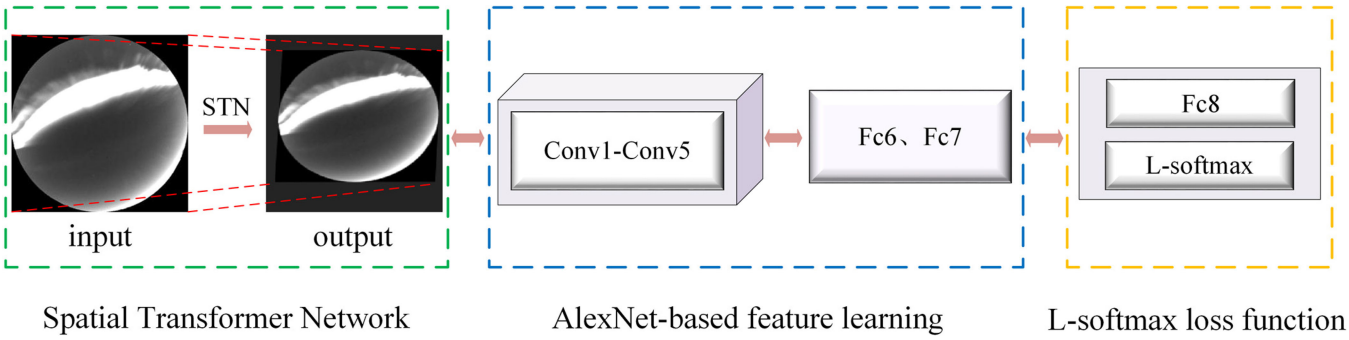

Fig. 1. Architecture of STN-Lsoftmax-AlexNet.

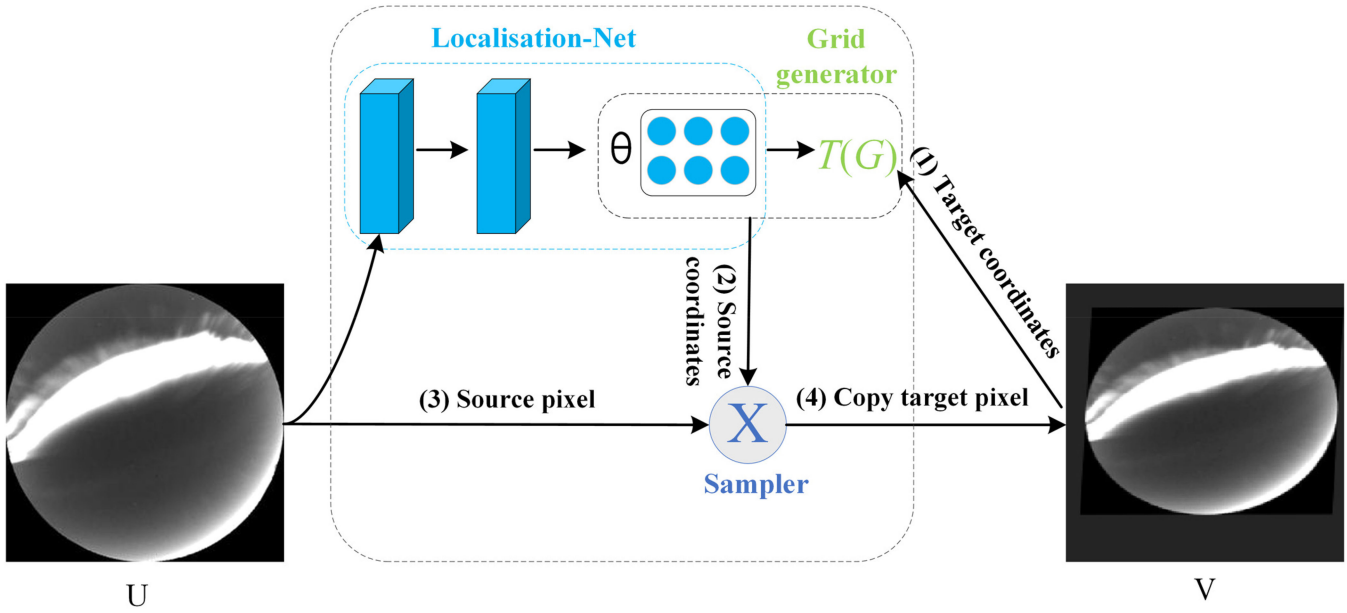

Fig. 2. STN, (1)–(4) are executed sequentially.

dimensional matrix transformation parameters θ. The second part grid generator is used to map every coordinate (x_i^t, y_i^t) in the target image V to the coordinate (x_i^s, y_i^s) in the source image U. The correspondence between (x_i^t, y_i^t) and (x_i^s, y_i^s) can be written as

$$(x_i^s, y_i^s) = T_\theta(G_i) = A_\theta \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} = \begin{pmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{pmatrix} \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} \tag{1}$$

where T_θ is a 2-D affine transformation function that is composed of the transformation parameter θ obtained in the first part [22]. A_θ is the transformation matrix, and G_i is the mapping space grid coordinates, where i represents the ith pixel point in the target image V. The third part sampler fills the pixel value of the coordinate (x_i^t, y_i^t) in the target image V using both the pixel values of the source image U and the coordinate (x_i^s, y_i^s) obtained in the second part. Since some of the coordinates mapped to the source image U may be decimals, we estimate the pixel values according to each surrounding pixel by bilinear interpolation.

B. L-Softmax Loss

It is important to choose an appropriate loss function for a specific task. Softmax loss has been widely adopted by many CNNs due to its simplicity, probabilistic interpretation, and superior performance. Specifically, Softmax function is used to transform the prediction of the *j*th class for the *i*th input by

$$\sigma_i(f_{y_i}) = \frac{\exp(f_{y_i})}{\sum_{j=1}^{k} \exp(f_{y_j})}, \quad i = 1, 2, \ldots, k \tag{2}$$

where y_i is the sample class label, and f_{y_i} is usually the activations of a fully connected layer and can be written as

$$f_{y_i} = W_{y_i}^T x_i + b_{y_i}. \tag{3}$$

Softmax turns the predictions into the nonnegative values and normalizes them to get a probability distribution over classes. The probability that data x belongs to class i is called likelihood.

$$o_i = \sigma_i(f_{y_i}). \tag{4}$$

Therefore, the original Softmax loss can be written as

$$L = -\log(oi) = \frac{1}{N}\sum_i L_i = \frac{1}{N}\sum_i -\log\left(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}}\right) \quad (5)$$

where f_j denotes the class score of the jth element ($j \in [1, k]$, k is the number of classes), and N is the number of training data. If we omit the constant b_{y_i} in f_j, f_j, it can also be formulated as

$$f_j = ||W_j||\,||x_i||\cos(\theta_j), \quad 0 \le \theta_j \le \pi \quad (6)$$

where θ_j is the angular margin. Then, the Softmax loss can further be defined as

$$L_i = -\log\left(\frac{e^{||W_{y_i}||\,||x_i||\psi(\theta_{y_i})}}{\sum_j e^{||W_j||\,||x_i||\cos(\theta_j)}}\right). \quad (7)$$

A standard CNN can be viewed as a convolutional feature learning machine that is supervised by the Softmax loss [27].

Although the Softmax loss function has been widely used, it does not explicitly encourage discriminative learning of features. To solve this problem, Liu *et al.* [27] generalized the Softmax loss to a more general large-margin Softmax (L-Softmax) loss in terms of angular similarity, which leads to potentially larger angular separability between learned features and, thus, generates more discriminative features. The intuition behind L-Softmax is simple. Consider a binary classification and a sample x from class 1, the original Softmax is to force $||W_1||\,||x_1||\cos(\theta_1) > ||W_2||\,||x_2||\cos(\theta_2)$ to classify x correctly. Instead, L-Softmax requires $||W_1||\,||x_1||\cos(m\theta_1) > ||W_2||\,||x_2||\cos(\theta_2)$, $(0 \le \theta_1 \le \pi/m)$, where m is a positive integer. Since m is a positive integer, and the cosine function is monotonically decreasing on the intervals [0, π]; thus, cos(mx) is less than cos(x). Therefore, the new classification criteria is a stronger requirement to correctly classify x, producing a more rigorous decision boundary for class 1.

Formally, the L-Softmax loss is defined as

$$\mathrm{L-Softmax} = -\log\left(\frac{e^{||W_{y_i}||\,||x_i||\psi(\theta_{y_i})}}{e^{||W_{y_i}||\,||x_i||\psi(\theta_{y_i})} + \sum_{j\neq y_i} e^{||W_j||\,||x_i||\cos(\theta_j)}}\right) \quad (8)$$

in which

$$\psi(\theta) = \begin{cases} \cos(m\theta), & 0 < \theta \le \frac{\pi}{m} \\ D(\theta), & \frac{\pi}{m} < \theta \le \pi \end{cases} \quad (9)$$

where m is an integer that controls the classification margin [27]. With larger m, the classification margin becomes larger and the learning objective becomes harder. Specifically, when $m = 1$, the L-Softmax loss becomes identical to the original Softmax loss. $D(\theta)$ is a monotonically decreasing function and $D(\pi/m) = \cos(\pi/m)$. By adjusting the margin m between classes, a relatively difficult learning objective with an adjustable margin will be defined, which can effectively avoid overfitting.

III. Auroral Classification Mechanism and Dataset Setup

So far, there is no uniform classification scheme for auroral images. Generally, optical auroras observed on the ground can be classified into two broad categories: discrete and diffuse auroras with structured forms or relatively homogenous luminosity, respectively [31]. Discrete auroras can be further classified into arc and corona according to their spatial morphology. Based on the observations acquired by the three-wavelength (427.8, 557.7, and 630.0 nm) all-sky imageries (ASIs) at YRS in Ny-Ålesund, Svalbard, Hu *et al.* [26] further classified the dayside coronal auroras into the radial corona, drapery corona, and hotspot.

Considering that the auroral data used in this article were obtained by the ASI during daytime at YRS, we classify the auroral images according to the classification scheme proposed in [26], which was also applied in [13], [16], and [18]. Specifically, the dayside discrete auroras were classified into the arc, drapery corona, radial corona, and hotspot categories according to the spectral and morphological characteristics. Examples and characteristics of each type are given in Fig. 3.

The optical instruments at YRS capture photoemissions at 427.8, 557.7, and 630.0 nm during the winter season with a time resolution of 10 s. The ASI auroral images observed at YRS are available at http://www.chinare.org.cn/uap/database. In consideration of the image characteristics, we concentrated on the dayside aurora at 557.7 nm from December 2003 to February 2009. To better focus on the study of auroral image classification, the images that do not contain auroral structures or were captured under bad weather conditions (e.g., auroral structures were severely covered by clouds) were eliminated by human visual inspection. Specifically, the dataset consists of three parts.

1) *ASI8K* contains 8000 images (2913 arc images, 1771 drapery corona images, 1640 radial corona images, and 1676 hotspot images, respectively) from December 2003 to February 2004, which is used to train the classification network. Specifically, the ratio of training and validation is 4:1; that is, there are 6400 training images and 1600 validation images, respectively, in each training epoch.
2) *ASI2K* contains 2184 images with class labels in which each category has an approximately equal number of images from December 2004 to February 2009 (different winters from the training images). It is used in the testing phase to evaluate the performance of the proposed method.
3) *ASI399K* contains 399 515 auroral images from 2004 to 2009 without the manual labels. Almost all the observed images are selected except those who have no auroral activity or only have diffuse auroras or under cloudy weather. It is used to make a statistical study of the temporal occurrence distributions of the four auroral categories.

In order to facilitate the following experiments, all images in the dataset were preprocessed in the same manner as in [16], including subtracting system noise and cutting off the outer regions of images, where significant wide-angle distortion happened and might contain YRS lights, and auroral images were cropped from 512×512 to 440×440 pixels finally. To benefit the future research work in this field, all datasets used in this article are released at https://github.com/wszph/Aurora_data.

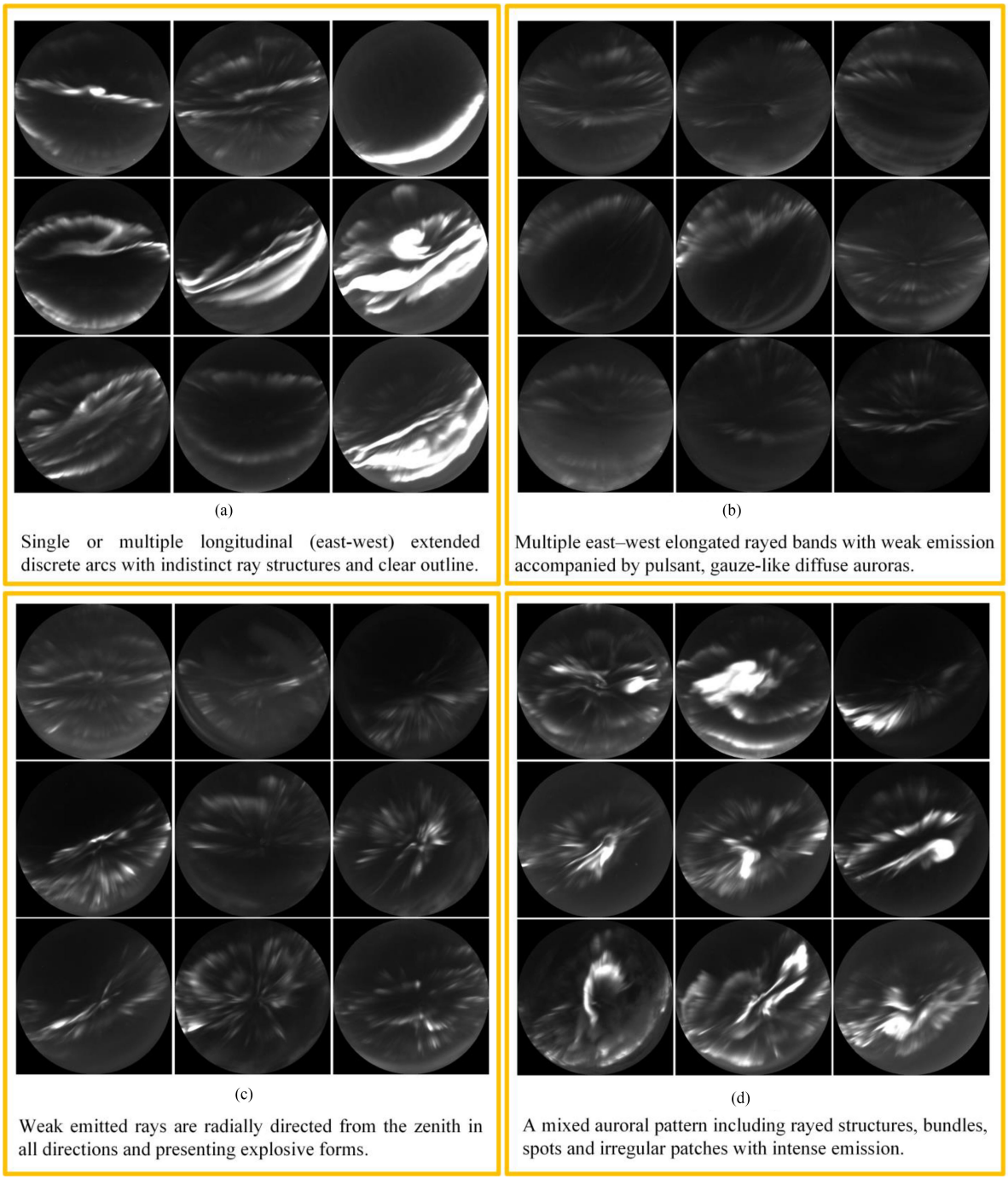

Fig. 3. Examples and characteristics of the four auroral types sampled from the dataset. (a) Arc. (b) Drapery Corona. (c) Radial Corona. (d) Hot-spot.

IV. Experiments and Results

In this section, auroral image classification and retrieval experiments were carried out on *ASI8K* and *ASI2K*. The importance of each module of the STN-Lsoftmax-AlexNet model was verified and the accuracy and efficiency of the model were compared with the existing auroral image classification methods. Besides, a statistical analysis of the temporal occurrence distribution of auroral types was further made on *ASI399K*.

A. Implementation Details

In this article, the AlexNet was pretrained on the ImageNet dataset as a feature extraction tool and all parameter settings were obtained from the Caffe toolbox [33]. Generally, the AlexNet consists of five convolutional layers (abbreviated as Conv1–Conv5) and three fully connected layers (FC6–FC8). With the increase of the layer number, the outputs represent the higher semantic features. Specifically, the input image is resized to 256×256 before fed into the AlexNet, and the output

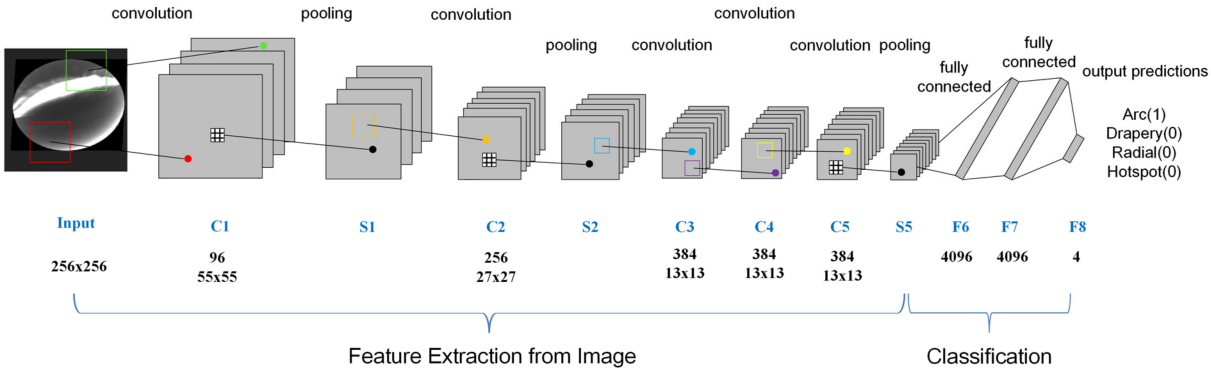

Fig. 4. Block diagram illustrates the feature extraction and classification of auroral images using AlexNet architecture. C represents the convolutional layer, S is the max-pooling layer, and F refers to the fully connected layer.

of the last FC8 layer, including four nodes, can be used as the category labels in the task of auroral image classification. Fig. 4 illustrates the diagram of the auroral image feature extraction and classification using AlexNet architecture.

The network was trained and optimized with the L-Softmax loss function. All weights were learned via backpropagation and stochastic gradient descent. Each minibatch consisted of 64 semantic regions for the training dataset or 16 for the validation dataset, which comprised the predefined four auroral categories randomly. We used an initial learning rate of 0.00003, a momentum of 0.9, and a weight decay of 0.0005. The experiments were performed on a PC with a 3.4 GHz Intel i7-6700 CPU under a Linux system.

Considering that the L-Softmax loss had difficulty in convergence, a decay factor λ was added to the learning strategy [27], which was expressed as

$$f_{y_i} = \frac{\lambda||W_{y_i}||\,||x_i||\cos(\theta_{y_i}) + ||W_{y_i}||\,||x_i||\psi(\theta_{y_i})}{1+\lambda} \tag{10}$$

where λ is a large number at the beginning of gradient descent, and it is gradually reduced during iteration. In our experiments, the initial value of λ was set to 10000 and the minimum value was set to 15 by visualizing the convergence performance of the network in the training process.

B. Experiments and Analysis

1) Supervised Classification: The following three supervised classification experiments, including base model selection, component ablation studies, and comparison with existing methods, were all trained on dataset *ASI8K* and tested on dataset *ASI2K*. The classification rate of each auroral type was calculated by dividing the number of correctly classified images by the total labeled number of that type, and the average classification accuracy was equal to the number of correctly classified images divided by the total number of testing images.

a) Base model selection: We compared the performance of three popular CNN models, AlexNet, VGG16, and Inception-v4, which have been used for the automatic analysis of auroral images before. Compared with AlexNet, VGG16 and Inception-v4 are more powerful for feature extraction because of their deeper and more complex networks. However, the deeper the network, the more likely it is to have a gradient dispersion problem and more challenging to optimize the model. The classification accuracy and running time were compared in Table I. The best results are highlighted with bold fonts. It is clear that the AlexNet achieves the best performance in terms of both average accuracy and efficiency on datasets *ASI8K* and *ASI2K*. Therefore, the following experiments are based on AlexNet.

TABLE I
COMPARISON OF ALEXNET, VGG16, AND INCEPTION-V4

Type \ Method	AlexNet	VGG	Inception v4
arc accuracy	0.984	**0.998**	0.976
drapery accuracy	**0.938**	0.908	0.850
radial accuracy	**0.836**	0.812	0.706
hot-spot accuracy	**0.830**	0.824	**0.830**
Average accuracy	**0.904**	0.895	0.851
Time Cost (ms)	**10.16**	15.98	40.41

b) Component ablation studies: In this section, we present the ablation studies to isolate the effect of each module (STN and L-Softmax) of the proposed STN-Lsoftmax-AlexNet model. Fig. 5 illustrates the comparison of the classification results obtained by different methods. The dotted green line and the solid mauve line depict the performance of the original AlexNet and combined AlexNet with STN, respectively, the dotted blue lines represent that of the L-Softmax loss based AlexNet, and the solid red lines show the performance of STN-Lsoftmax-AlexNet. The following conclusions can be drawn from Fig. 5.

1) The classification accuracy of each auroral type indicates that almost all methods are easier to recognize the arc and drapery corona. In contrast, the accuracy of the hotspot aurora is relatively low. This is directly related to the morphological complexity of these auroral types, as shown in Fig. 3.
2) The classification accuracy shown by the solid lines is higher than that shown by the dotted lines, which indicates that the combination of STN and AlexNet improves the accuracy of the classification model. The average accuracy improvement is about 1%–2% over no STN structures, and the improvements are up to 4.4% and 3.5% for the

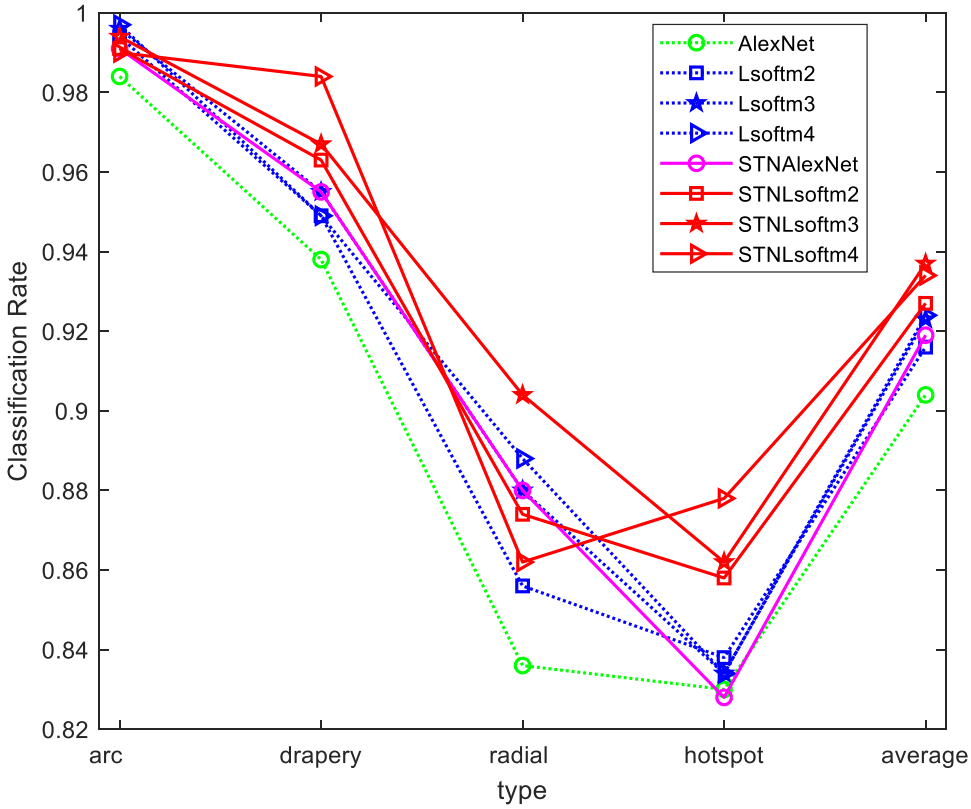

Fig. 5. Comparison of the performance of different methods.

TABLE II
COMPARISON WITH EXISTING METHODS

Class \ Method	STN-Lsoftmax-AlexNet	AlexNet [40]	Inception-v4 [30]	RSM [10]
Accuracy	**0.937**	0.904	0.851	0.8735
Time Cost (ms)	12.43	**10.16**	40.41	1233.8

hotspot and drapery coronas, respectively. It also indicates that STN is effective in the spatial transformation of the original auroral images during the training process.

3) The red lines show a higher classification rate than the mauve line and the dotted blue lines show a higher classification rate than the dotted green line. The average accuracy has been improved by about 2%, while radial coronas get an improvement of 5.2% with $m = 4$. This proves that the L-Softmax loss function can guide the network to learn more distinctive auroral features. However, as m increases, the model performance was not improved continuously. From Fig. 5, we can see that with $m = 3$, all categories achieve the best classification results. The following experiments are, therefore, based on the STN-Lsoftmax-AlexNet with $m = 3$.

c) Comparison with existing methods: We compare the performance of the proposed STN-Lsoftmax-AlexNet model ($m = 3$) in terms of both accuracy and time costs with those reported in the state-of-the-art on auroral image classification using deep learning techniques. Specifically, they are AlexNet, Inception-v4, and a region scale model (RSM) applied to the auroral image classification in [40], [30], and [10], respectively. Table II depicts the comparison results. Our method achieves much higher classification accuracy than the previous methods. In terms of running time, our method takes 12.43 ms to predict the class label of a testing image, which is 2.27 ms more than the basic AlexNet. This is related to the part of the local network of STN (see Fig. 2). And the difference of m value of the loss function has little effect on the time cost as long as the model structure is the same. Inception-v4 and RSM are much more complex networks than our method, taking 40.41 and 1233.8 ms, respectively, to predict the label of an image, which is much slower than our method. In brief, the proposed method that combines STN and L-Softmax loss function based AlexNet is both effective and efficient for auroral image classification.

Furthermore, our experimental setting is different from the previous automatic auroral image classification works (e.g., [13], [16], [19], and [30]). The training and testing data in these articles came from the same year(s), which were obtained by shuffling and randomly dividing the auroral data from one or

TABLE III
CONFUSION MATRIX OF SUPERVISED CLASSIFICATION ON TESTING IMAGES WITH STN-LSOFTMAX-ALEXNET MODEL (M = 3)

		Classification results				Classification rate	Average rate (%)
		arc	drapery	radial	hot-spot		
Manual labels	arc	**668**	1	0	3	0.994	0.937
	drapery	3	**495**	14	0	0.967	
	radial	0	28	**452**	20	0.904	
	Hot-spot	50	0	19	**431**	0.862	

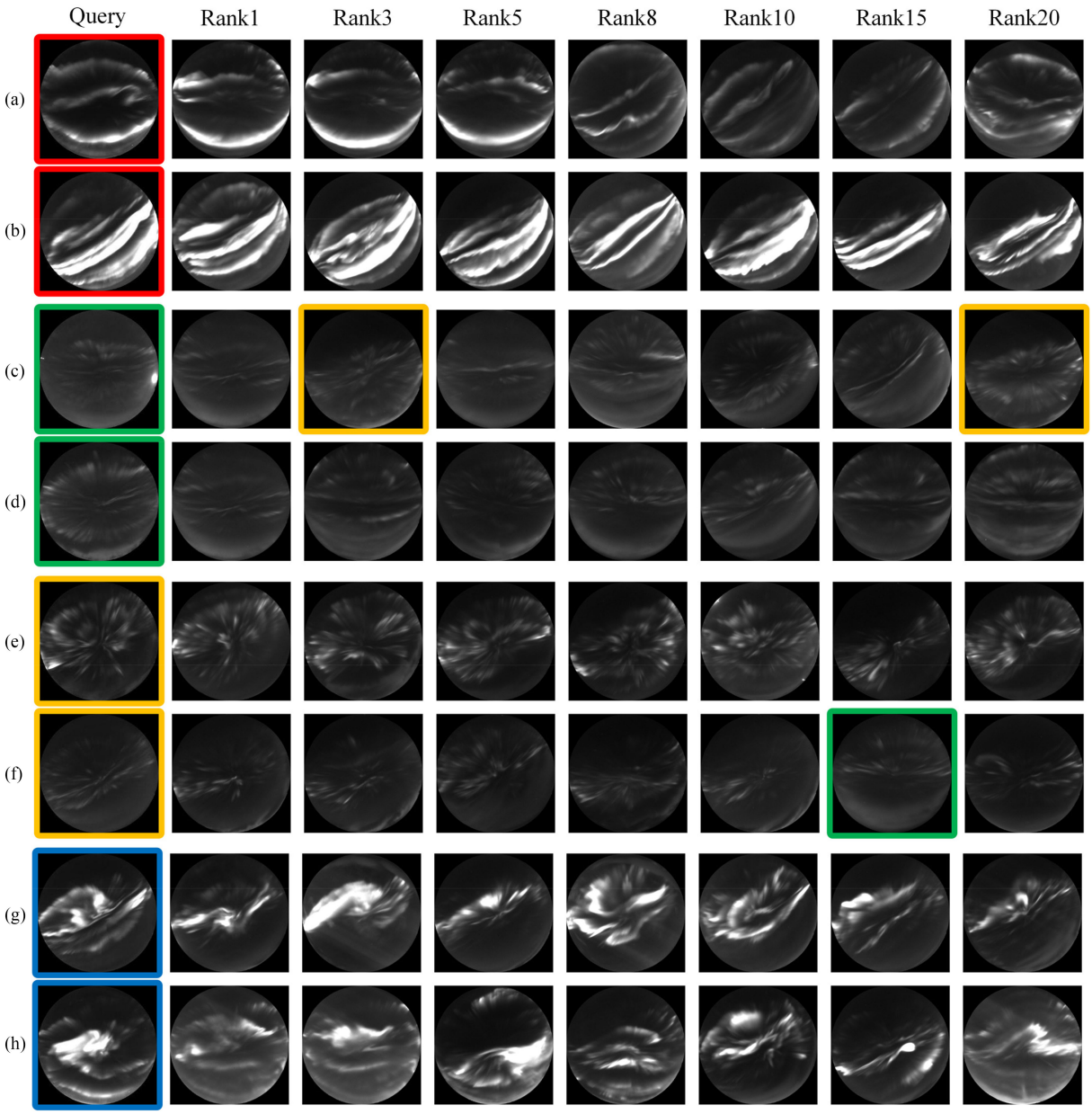

Fig. 6. Retrieval results of four types of auroral images. Query images are in the first column, where the red, green, yellow, and blue boxes denote the arc, drapery, radial, and hotspot auroras, respectively. The second–eighth columns are the returned retrieved images (ranks 1, 3, 5, 8, 10, 15, and 20, respectively.).

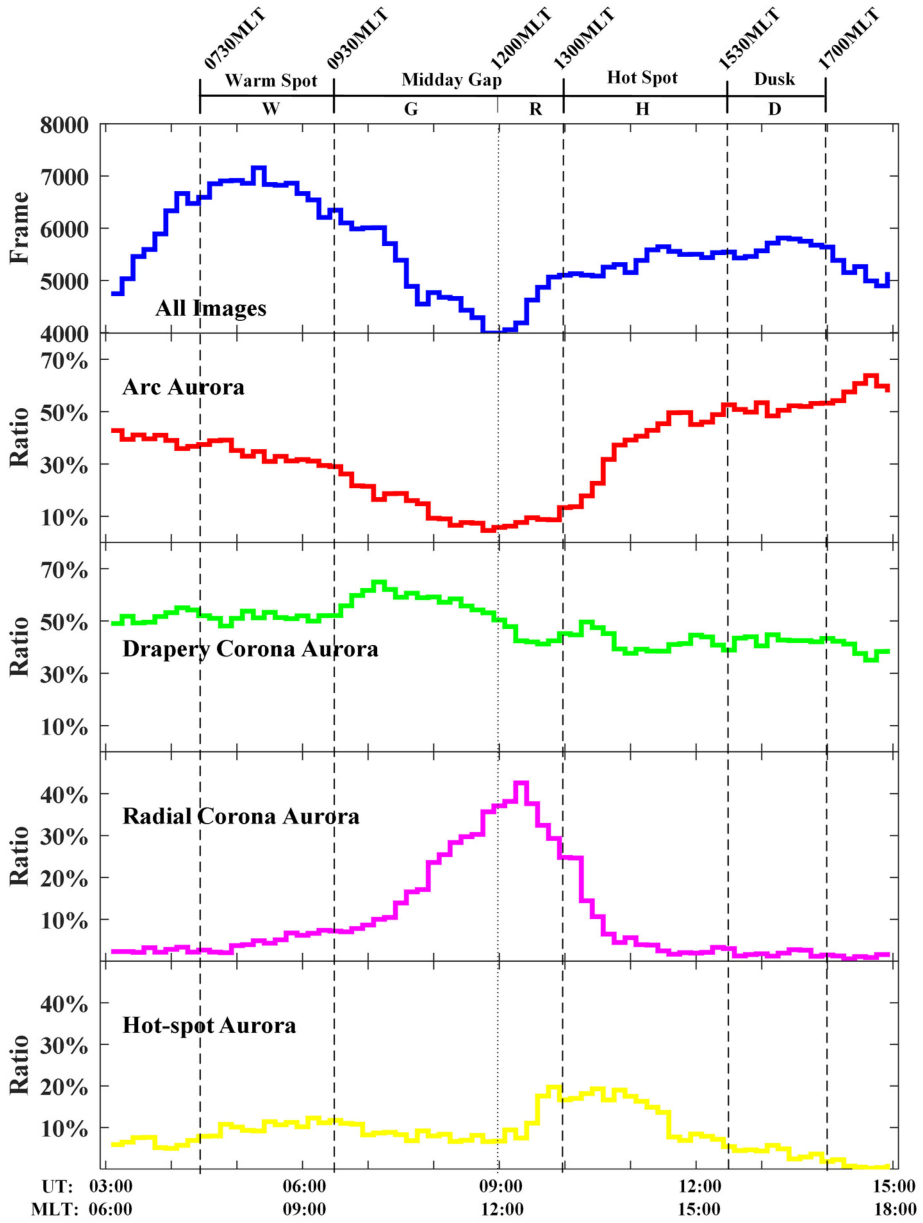

Fig. 7. Temporal occurrence distributions of dayside auroral categories.

several years. The model in this study is trained on one-year auroral data and then tested on the following five-year auroral data according to the practical needs of auroral physics. In addition, although some advanced issues, such as weakly supervised semantic segmentation, have been developed [10], the task of automatic aurora classification is more significant and practical in auroral physics. Some auroral types, such as the drapery and radial coronas, have no distinct shape boundaries to be segmented. Actually, one of the uses of aurora segmentation is to improve the accuracy of aurora classification [7], [10].

d) Classification confusion matrix: In order to quantitatively evaluate the classification effectiveness of our proposed STN-Lsoftmax-AlexNet model ($m = 3$) on each auroral type, we calculated the classification rate of each auroral type and the average classification accuracy of all testing images and depicted the confusion matrix of the classification results in Table III. The classification accuracy of the four auroral categories indicates that the arc and drapery coronas are more easily recognized, and radial coronas might be classified as drapery coronas or hotspot. The classification accuracy of hotspot auroras is somewhat low and they are likely to be classified as auroral arcs or radial coronas. The reason is that there are many complex auroral structures in the hotspot displays, such as rays, beams, spots, and irregular patches, as shown in Fig. 3(d).

2) Image Retrieval: In recent decades, with the increasing number of auroral images, it is urgent to retrieve the images of interest quickly and effectively from massive auroral observations. In this section, the F7 layer of AlexNet was extracted from the trained STN-Lsoftmax-AlexNet model as the feature vector to represent each image, and the Euclidean distance between two features was used to measure the similarity between the two images. The representation ability of the proposed model is

verified by image retrieval experiments. Dataset *ASI8K* from 2003 to 2004 was used as the query images, and the retrieval experiments were carried out on dataset *ASI2K* from 2004 to 2009. We chose *ASI2K*, instead of *ASI399K*, as the retrieved dataset, because all images in it have class labels that could help to evaluate the retrieval results.

Fig. 6 shows examples of the query and retrieved images. The first column shows the query images, and the second–eighth columns are the most similar (ranks 1, 3, 5, 8, 10, 15, and 20, respectively) retrieved images by the trained model, respectively. In the query images, (a) and (b) with red boxes are auroral arcs, (c) and (d) with green boxes are drapery coronas, (e) and (f) with yellow boxes are radial coronas, and (g) and (h) with blue boxes are hotspot auroras. From the perspective of human vision, the query images and their retrieved results are very similar. Except for a few images marked with color boxes [that is, rank 3 and rank 20 in row (c) and rank 15 in row (f)], the retrieved images have the same category labels (manually labeled) as their query images. This proves that the proposed model is able to characterize the key features of auroral images. As a result, we can utilize this model to select the interested auroral images from the massive auroral observations. For the drapery and radial coronas, some retrieved images have different category labels (manual labels) from their query images, such as the images marked with yellow and green boxes in Fig. 6(c) and (f), respectively. This indicates that extracting only one layer (F7 layer) from the classification model is not enough to distinguish the two coronal auroras.

3) Occurrence Distributions: The auroral events caused by similar or repeatable physical processes share the same fundamental morphology; thus, a statistical study of the occurrence time of auroral types is important to our understanding of magnetospheric dynamics. In this section, the images in dataset *ASI399K* from 2004 to 2009 were predicted by the trained STN-Lsoftmax-AlexNet model, and the temporal distribution of the four auroral types was drawn in Fig. 7. The temporal axis between 6 and 18 MLT was divided into 240 bins of 3 min duration. At the top of these panels, four active regions proposed in [26] were denoted and partitioned by bold dashed lines, while the two states of the midday gap region were indicated by a thin dashed line. From top to bottom, the first panel shows the frame number of all predicted images occurring in each time period. The second–fifth panels show the occurrence ratio of each category, which is the result of dividing the number of images of each type within 3 min by the total number of images in that time period, respectively.

As shown in Fig. 7, dayside auroras mainly occur before noon, followed by the afternoon, and rarely occurs at noon. More arcs and drapery coronas were observed than radial coronas and hotspot auroras from 2004 to 2009 in YRS. The auroral arc has a double-peak distribution around noon and the forenoon peak is weaker than the postnoon peak. Both drapery coronas and radial coronas predominantly occur before 13 MLT but with different peak positions. The occurrence of these auroral categories dominates the different regions of the dayside oval. The drapery coronas occur more often in region G, whereas the radial coronas occupy in region R, and the hotspot auroras dominate region H. Their global distributions approximately coincide with the multiple-wavelength intensity distribution of the dayside aurora presented in [26]. This proves that our method can be used for the automatic classification of a huge amount of auroral images in practical applications.

V. Conclusion

We developed an automatic representation and classification method based on CNNs to extract auroral features and classify the auroral images into the arc, drapery corona, radial corona, and hotspot. Aurora is a natural phenomenon whose morphology continuously changes over time, and there is no obvious boundary when it gradually transits from one type to another. Therefore, a major challenge for auroral image classification is to maximize the intraclass compactness and interclass separability of auroral features. To tackle this problem, this article combines the STN and AlexNet and optimizes the network by using the L-Softmax loss function. The classification results show the following.

1) AlexNet is a good choice for the auroral image classification in terms of both accuracy and efficiency.
2) L-Softmax loss function can guide the network to learn more discriminative auroral features and, thus, improves the classification accuracy, especially that of some subcategories (e.g., radial coronas get an improvement of 5.2% with $m = 4$).
3) STN is able to actively spatially transform the input images and simplify the subsequent classification task, and thus lead to superior classification performance (the improvements are up to 4.4% and 3.5% for the hotspot and drapery coronas, respectively, as compared with no STN modules).

Auroral image retrieval experiments were also carried out to visually evaluate the representation ability of the proposed model. The retrieval results are basically in accordance with human visual judgments. In addition, to verify the generalization performance of the proposed model, a larger dataset containing nearly 0.4 million auroral images from 2004 to 2009 was tested and the category of each image was predicted. The temporal occurrence distributions of the dayside auroral types were in accordance with the empirical rules of auroral physics. In summary, the proposed model has the advantages of excellent classification accuracy, high efficiency, good generalization, and simple implementation, and thus can be truly and widely applied for aurora classification.

The proposed classification model is data-driven and does not rely on handcrafted designs, so it can also be applied to the classification of auroral images from other stations or with different classification schemes. Just providing a set of labeled data from these new stations according to a specific classification scheme, the model will be retrained and its performance can be assessed with these data. Once the vast amounts of the existing ground-based auroral data are classified, we will be able to make large statistical studies to analyze the physical mechanisms of each auroral type.

Acknowledgment

The authors would like to thank the Polar Research Institute of China for providing the auroral data.

REFERENCES

[1] T. R. Pedersen and E. A. Gerken, "Creation of visible artificial optical emissions in the aurora by high-power radio waves," *Nature*, vol. 433, no. 7025, pp. 498–500, Feb. 2005.

[2] S.-I. Akasofu, "The development of the auroral substorm," *Planet. Space Sci.*, vol. 12, no. 4, pp. 273–282, Apr. 1964.

[3] D.-S. Han, Y. Nishimura, L. R. Lyons, H.-Q. Hu, and H. Yang, "Throat aurora: The ionospheric signature of magnetosheath particles penetrating into the magnetosphere," *Geophys. Res. Lett.*, vol. 43, no. 5, pp. 1819–1827, Feb. 2016.

[4] Z.-J. Hu *et al.*, "The hemispheric conjugate observation of postnoon "bright spots"/auroral spirals," *J. Geophys. Res., Space Phys.*, vol. 118, no. 4, pp. 1428–1434, Apr. 2013.

[5] Q.-H. Zhang *et al.*, "Direct observations of the evolution of polar cap ionization patches," *Science*, vol. 339, pp. 1597–1600, Mar. 2013.

[6] Z. Xing *et al.*, "Conjugate observations of the evolution of polar cap arcs in both hemispheres," *J. Geophys. Res., Space Phys.*, vol. 123, no. 3, pp. 1794–1805 Mar. 2018.

[7] Q. Yang, D. Tao, D. Han, and J. Liang, "Extracting auroral key local structures from all-sky auroral images by artificial intelligence technique," *J. Geophys. Res., Space Phys.*, vol. 124, pp. 3512–3521, May 2019, doi: 10.1029/2018JA026119.

[8] M. T. Syrjäsuo and E. F. Donovan, "Diurnal auroral occurrence statistics obtained via machine vision," *Ann. Geophys.*, vol. 22, no. 4, pp. 1103–1113, Apr. 2004.

[9] X. Yang, X. Gao, and Q. Tian, "Polar embedding for aurora image retrieval," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3332–3344, Nov. 2015.

[10] C. Niu, J. Zhang, Q. Wang, and J. Liang, "Weakly supervised semantic segmentation for joint key local structure localization and classification of aurora image," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 12, pp. 7133–7146, Dec. 2018.

[11] X. Yang, X. Gao, D. Tao, and X. Li, "Improving level set method for fast auroral oval segmentation," *IEEE Trans. Image Process.*, vol. 23, no. 7, pp. 2854–2865, Jul. 2014.

[12] X. Yang, X. Gao, D. Tao, X. Li, B. Han, and J. Li, "Shape-constrained sparse and low-rank decomposition for auroral substorm detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 1, pp. 32–46, Jan. 2016.

[13] Q. Yang, J. Liang, Z. Hu, and H. Zhao, "Auroral sequence representation and classification using hidden Markov models," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 12, pp. 5049–5060, Dec. 2012.

[14] Q. Yang, J. Liang, Z. Hu, Z. Xing, and H. Zhao, "Automatic recognition of poleward moving auroras from all-sky image sequences based on HMM and SVM," *Planet. Space Sci.*, vol. 69, no. 1, pp. 40–48, Aug. 2012.

[15] X. Gao, R. Fu, X. Li, D. Tao, B. Zhang, and H. Yang, "Aurora image segmentation by combining patch and texture thresholding," *Comput. Vision Image Understanding*, vol. 115, no. 3, pp. 390–402, Mar. 2011.

[16] Q. Wang *et al.*, "Spatial texture based automatic classification of dayside aurora in all-sky images," *J. Atmos. Sol.-Terr. Phys.*, vol. 72, no. 5/6, pp. 498–508, Apr. 2010.

[17] J. Rao, N. Partamies, O. Amariutei, M. Syrjäsuo, and K. E. van de Sande, "Automatic auroral detection in color all-sky camera images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 12, pp. 4717–4725, Dec. 2014.

[18] Q. Yang and Z. Hu, "An automatic auroral classification method based on morphological characteristics (in Chinese)," *Scientia Sinica Terrae*, vol. 47, no. 2, pp. 252–260, Jan. 2017.

[19] Y. Zhong, R. Huang, J. Zhao, B. Zhao, and T. Liu, "Aurora image classification based on multi-feature latent Dirichlet allocation," *Remote Sens.*, vol. 10, no. 2, pp. 233–249, 2018.

[20] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proc. 31st AAAI Conf. Artif. Intell.*, San Francisco, CA, USA, Feb. 4–9, 2017, pp. 4278–4284.

[21] L. Zheng, Y. Yang, and Q. Tian, "SIFT meets CNN: A decade survey of instance retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 5, pp. 1224–1244, May 2018.

[22] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," in *Proc. Adv. Neural Inf. Process. Syst.*," Montreal, QC, Canada, Dec. 7–12, 2015, pp. 2017–2025.

[23] X. Yang, X. Gao, B. Song, and D. Yang, "Aurora image search with contextual CNN feature," *Neurocomputing*, vol. 281, no. 15, pp. 67–77, Mar. 2018.

[24] X. Yang, X. Gao, B. Song, N. Wang, and D. Yang, "ASI aurora search: An attempt of intelligent image processing for circular fisheye lens," *Opt. Express*, vol. 26, no. 7, pp. 7985–8000, Jun. 2018.

[25] B. Han, F. Chu, X. Gao, and Y. Yan, "A multi-size kernels CNN with eye movement guided taskspecific initialization for aurora image classification," in *Proc. CCF Chin. Conf. Comput. Vision*, Tianjin, China, Nov. 2017, pp. 533–544.

[26] Z.-J. Hu *et al.*, "Synoptic distribution of dayside aurora: Multiple-wavelength all-sky observation at yellow river station in Ny-Ålesund, Svalbard," *J. Atmos. Sol.-Terr. Phys.*, vol. 71, no. 8/9, pp. 794–804, Jun. 2009.

[27] W. Liu, Y. Wen, Z. Yu, and M. Yang, "Large-margin softmax loss for convolutional neural networks," in *Proc. 33rd Int. Conf. Mach. Learn.*, New York, NY, USA, Jun. 19–24, 2016, pp. 507–516.

[28] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," in *Proc. 2nd Int. Conf. Learn. Represent.*, Banff, AB, Canada, Apr. 14–16, 2014, pp. 1–16.

[29] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: An astounding baseline for recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit. Workshops*, Columbus, OH, USA, Jun. 23–28, 2014, pp. 512–519.

[30] L. B. Clausen and H. Nickisch, "Automatic classification of auroral images from the Oslo auroral THEMIS (OATH) data set using machine learning," *J. Geophys. Res., Space Phys.*, vol. 120, pp. 7447–7465, Sep. 2015.

[31] D.-S. Han *et al.*, "An extensive survey of dayside diffuse aurora based on optical observations at yellow river station," *J. Geophys. Res., Space Phys.*, vol. 120, pp. 7447–7465, Sep. 2015.

[32] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, Lake Tahoe, NV, USA, Dec. 3–6, 2012, pp. 1097–1105.

[33] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. 22nd ACM Int. Conf. Multimedia*, Orlando, FL, USA, Nov. 3–7, 2014, pp. 675–678.

[34] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Oct. 2016.

[35] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Boston, MA, USA, Jun. 8–10, 2015, pp. 1–9.

[36] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, San Diego, CA, USA, May 7–9, 2015, pp. 1–14.

[37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Las Vegas, NV, USA, Jun. 27–30, 2016, pp. 770–778.

[38] Y. Yu, Z. Gong, C. Wang, and P. Zhong, "An unsupervised convolutional feature fusion network for deep representation of remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 1, pp. 23–27, Jan. 2018.

[39] H. Liang, X. Sun, Y. Sun, and Y. Gao, "Text feature extraction based on deep learning: A review," *EURASIP J. Wireless Commun. Netw.*, vol. 2017, Dec. 2017, Art. no. 211.

[40] F. Wang and Q. Yang, "Classification of auroral images based on convolutional neural network," *Chin. J. Polar Res.*, vol. 30, no. 2, pp. 123–131, Jun. 2018.

[41] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification on and segmentation," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Honolulu, HI, USA, Jul. 21–26, 2017, pp. 77–85.

[42] G. Cheng, P. Zhou, and J. Han, "RIFD-CNN: Rotation-invariant and fisher discriminative convolutional neural networks for object detection," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Las Vegas, NV, USA, Jun. 2016, pp. 2884–2893.

[43] G. Cheng, C. Yang, X. Yao, L. Guo, and J. Han, "When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2811–2821, May 2018.

[44] G. Cheng, J. Han, P. Zhou, and D. Xu, "Learning rotation-invariant and fisher discriminative convolutional neural networks for object detection," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 265–278, Jan. 2019.

Qiuju Yang received the B.S. degree in electronic information science and technology from Shaanxi Normal University, Xi'an, China, in 2008, and the Ph.D. degree in pattern recognition and intelligence systems from Xidian University, Xi'an, China, in 2013.

In 2013, she joined the School of Physics and Information Technology, Shaanxi Normal University, where she is currently an Associate Professor in electronics and information systems. From 2018 to 2019, she was a Visiting Scholar with UBTECH Sydney AI Centre, University of Sydney, Sydney, NSW, Australia. Her research interests include machine learning, aurora physics, and computer vision. She has authored or coauthored more than ten papers in referred journals, including IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and *Journal of Geophysical Research: Space Physics*.

Penghui Zhou received the B.S. degree in electronic information science and technology from the Baoji University of Arts and Sciences, Baoji, China, in 2017, and the M.S. degree in electronics and communication engineering from Shaanxi Normal University, Xi'an, China, in 2019.

His main research interests include auroral image processing and analysis.