Weakly Supervised Semantic Segmentation for Joint Key Local Structure Localization and Classification of Aurora Image

Chuang Niu, Jun Zhang, Qian Wang, and Jimin Liang, *Member, IEEE*

***Abstract*— In this paper, we propose a novel weakly supervised semantic segmentation (WSSS) method that uses image tags as supervision to achieve joint pixel-level localization of the key local structure (KLS) and image-level classification of the aurora images captured by the ground-based optical all-sky imager. First, a patch-scale model (PSM) based on the small-scale structure of aurora is designed to identify the type-specific regions for each training image. Second, a region-scale model is trained with the identified type-specific regions to coarsely localize the KLS from multiple sizes of field of view, based on which the aurora image is classified. Finally, given the predicted image type, the PSM further refines the KLS in a pixel level. By localizing KLS from coarse to fine, the proposed method captures both overall shape with a bottom–up processing and local structure details of aurora in a top–down manner. Extensive experiments on the expert labeled data sets have demonstrated the efficacy of the proposed method in benchmarking with the state-of-the-art WSSS methods.**

***Index Terms*— Aurora image analysis, bag of visual words (BoVW), convolutional neural networks (CNNs), weakly supervised semantic segmentation (WSSS).**

I. Introduction

AURORA borealis and aurora australis, often called the northern lights and southern lights, are spectacular phenomena that appear around the high latitude area of the earth. The light is emitted by atmospheric atoms and molecules that have been excited by collisions with electrons and protons that precipitate into the atmosphere from the outer space [1]. As an optically thin projection screen reflecting the solar activities and changes in the earth's magnetosphere, aurora is an important way to monitor and investigate the physical processes in the near-earth space for geosciences. Of various facilities for aurora observation, the ground-based optical all-sky imager (ASI) captures 2-D morphological information with satisfactory spatial and temporal resolutions [2]. However, with the dramatically and ceaselessly increasing amount of ASI images, how to efficiently analyze such huge data set faces enormous challenges. The traditional analysis of ASI images via human visual inspection is usually performed on a small number of images, and the corresponding analysis results are difficult to reproduce due to the tedious work burden.

Since the morphological types of aurora have turned out to be correlated with specific magnetospheric regimes and dynamic activity [3] and influenced by the solar wind parameters [4], many computer vision and machine learning techniques have been developed to assist aurora research over the past few decades, such as, ASI aurora image retrieval [1], [5], [6], ASI aurora image classification [7]–[11], and ASI aurora image segmentation [12], [13]. Particularly, an automatic classification of ASI aurora images on a large data set can help scientists to study the relationship between morphological types and physical processes of aurora. In return, scientists can construct specific model to forecast solar activities by automatically analyzing the morphological types of aurora images, and thus, some disastrous space weather caused by strong disturbance in the magnetosphere (e.g., the magnetospheric substorm which seriously interferes the communication, electricity supply, aviation, and global positioning system) can be avoided [6].

In this paper, two essential problems in an ASI aurora image analysis are considered: classification of the global morphology of aurora and localization of the key local structure (KLS) in the ASI aurora images. Specifically, the aurora classification corresponds to a basic image classification problem, which assigns a predefined aurora type to an image. In general, the existing aurora classification methods predict the aurora types from the whole image features [6], [10], [11]. However, the complex, nonrigid deformable spatial structure, and fast temporal morphological evolution of aurora make the classification of aurora image a challenging task, especially for the fine-grained classification, e.g., further classifying the corona type [14] into the subtypes of drapery, radial, and hot spot (HS) [2]. As shown in Fig. 1, drapery, radial, and hotspot aurora share some similar ray structures. The differences among these ray structures mainly lie in two aspects: local

Manuscript received September 1, 2017; revised January 7, 2018 and April 11, 2018; accepted May 24, 2018. Date of publication July 12, 2018; date of current version November 22, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61571353 and Grant 41504115 and in part by the Natural Science Basic Research Plan in Shaanxi Province of China under Grant 2015JZ019 and Grant 2015JQ6223. *(Corresponding author: Jimin Liang.)*

C. Niu and J. Liang are with the School of Life Science and Technology, Xidian University, Xi'an 710071, China (e-mail: niuchuang@stu.xidian.edu.cn; jimleung@mail.xidian.edu.cn).

J. Zhang is with the Department of Radiology, Duke University, Durham, NC 27705 USA (e-mail: xdzhangjun@gmail.com).

Q. Wang is with the School of Telecommunication and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710121, China (e-mail: xinrzhsh24@126.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2018.2848725

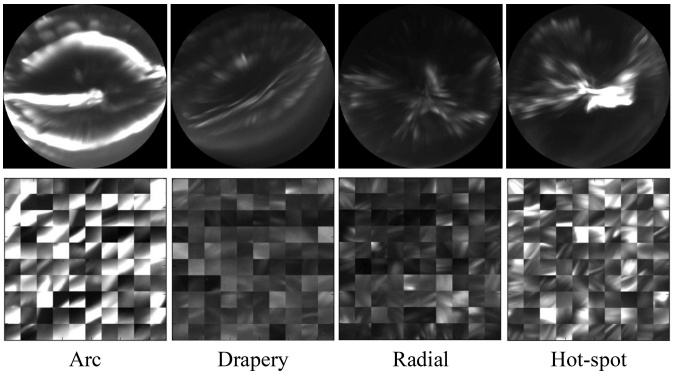


Fig. 1. Typical types of ASI aurora images. (Top row) From left to right: arc, drapery, radial, and HS. (Bottom row) Corresponding small patches of each type randomly selected from the labeled data sets.

details and overall arrangements, which should be analyzed with both the small and large size of field of view (FOV).

Actually, when aurora experts annotate the aurora images, they tend to first localize the KLSs of particular aurora type in the different sizes of FOV and then classify the aurora image based on these KLSs. Motivated by this observation, we put our focus on the joint KLS localization and classification of the aurora image. We refer the mask that determines a certain type of aurora as the KLS of this aurora type. It should be noted that the KLS localization differs from the aurora image segmentation [13] that aims to segment all aurora pixels from the dark sky, while the KLS localization aims to identify all the pixels belonging to each type. The KLS localization can not only help identify the aurora image types but also obtain the scale information of type-specific structures to calculate the proportion of the aurora region to sky, which is a significant cue for describing the scale of aurora [13]. Moreover, it can also provide the position information to statistically analyze both the spatial and the morphological evolution of aurora [2], [15].

For the purpose of the KLS localization, every pixel of the ASI aurora image will be classified into predefined aurora types. This is the same as the semantic segmentation task in the computer vision. Particularly, the state-of-the-art semantic segmentation methods based on fully convolutional networks (FCNs) [16] require massive pixel-level annotations. However, it is of great difficulty to obtain large amounts of pixel-level annotated aurora images, since the aurora is transparent and the shape boundaries of aurora structures are difficult to deal with [13]. In addition, the fully supervised semantic segmentation methods hardly scale to more morphological types of aurora. To tackle this problem, we explore the weakly supervised semantic segmentation (WSSS) method that only requires easily obtained image-level tags. In the state-of-the-art WSSS methods, there exist two main components.

1) Discovering the local semantic regions or their latent information related to the image-level labels. Various of methods were proposed, including cross-image contextual analysis [17], saliency object detection models [18], sparse learning models [19], conditioned random field models [20], and convolutional neural network (CNN)-based methods [21], [22].
2) Training the semantic segmentation models, such as FCN, using the above-obtained information.

However, most of the existing WSSS methods can hardly apply to aurora images due to the unique characteristics of aurora morphology. First, the spatial size of the emission area is a key factor of aurora morphology. The large-scale variation of aurora structures will change the morphological type. For example, the large scale of bright bands is the main component of arc-type aurora, while the small scale of bright bands usually appears in other types of aurora images. However, the existing methods for a natural image analysis usually assume that the objects in different images sharing a similar appearance belong to the same category irrespective of their scale changes. Second, the differences among various types of aurora are subtle, especially for the ray structures. Without a lot of accurate pixel-level annotations in the WSSS setting, the FCN is inferior to distinguish such subtle differences (see Section IV-E).

Considering the unique characteristics of aurora, we develop a patch-scale model (PSM) and a region-scale model (RSM) for analyzing the low-level detail features and high-level overall arrangement features, respectively. The PSM and RSM together achieve joint KLS localization and aurora image classification. According to the scale characteristic of aurora, the PSM is designed to estimate whether a fixed-size patch is specific to a given aurora type, which agrees to the top–down processing in recognition [23]. Specifically, we assume that the fixed-size patches, which represent the small-scale structure details, are discriminable among different types to some extent (see Fig. 1). On one hand, it is of great difficulty to classify these small patches into different morphological types due to the existence of common small-scale structures in different types of aurora images. On the other hand, if the type of an aurora image is known, we can determine the specific small-scale structures for each type and the common structures between one type and the rest. Thus, the PSM can discover the type-specific local regions of each training image with a label. Subsequently, an RSM is trained with the semantic regions generated by the PSM to detect each type of regions using bounding boxes from different sizes of FOV, which is regarded as the coarse localization of KLS. The aurora image is classified into the type with the maximum area of coarsely detected KLS. Note that in this paper, we classify an aurora image to a single type only as in [10], but it is possible to extend the proposed method to the multilabel classification. Given the predicted type of an aurora image, the PSM is further used to refine the KLS in the pixel level by identifying the type-specific small-scale structure details. Extensive experiments demonstrate the effectiveness of the proposed method in terms of both classification and segmentation.

The main contributions of this paper are summarized as follows.

1) We propose a novel WSSS framework for joint pixel-level KLS localization and classification of the ASI aurora images according to the unique characteristics of the aurora, which can help to analyze the huge aurora image data sets.

2) We design a PSM to discover the type-specific local regions of aurora images with the image-level labels which is a key component in WSSS methods.
3) Motivated by the image annotation process of experts, the proposed method classifies the aurora images based on the KLSs from multiple sizes of FOV, which capture both local structures and overall shape of the aurora. The classification accuracy of the aurora image is obviously improved.
4) We propose a coarse-to-fine process for KLS localization, which can distinguish the subtle differences among various types of aurora. More specifically, the coarse localization is determined by the overall arrangements from multiple sizes of FOV with a bottom–up processing, and the fine localization is achieved by identifying the type-specific small-scale structure details in a top–down manner.
5) Extensive experiments are carried out to validate the effectiveness of the proposed methodology compared with the existing approaches, which suggests the potential application value of our method to the automatic analysis of large-scale aurora images.

The rest of this paper is organized as follows. Section II gives a brief review of the related work. Section III presents the proposed WSSS method for joint key local structure localization and aurora image classification. The experimental results and discussion are presented in Section IV. Finally, Section V concludes this paper.

II. Related Work

In this section, we first review the automatic methods for ASI aurora images, including retrieval, classification, and segmentation, and then introduce the most related CNN-based WSSS methods using image-level annotations. It is noted that there are many other methods, such as auroral oval segmentation [24]–[28] and aurora event detection [29], are developed for the satellite-based aurora images captured by the ultraviolet imager. Considering the relevance to the proposed algorithm, this section focuses on the methods for ASI images only.

A. Automatic Analysis Methods for ASI Aurora Images

1) Retrieval: Image retrieval is to search a set of images that have most similar appearance to a given query image from a large-scale database, which is a basic technique for analyzing the morphological characteristics of ASI aurora images. Syrjäsuo *et al.* [1], [5], [9] have used the shape information to represent aurora images and developed the first search engine for ASI aurora image data sets. Since then, researchers have been designing representation methods for ASI aurora images to improve the retrieval performance. Considering that similar patterns may have different shapes and not all aurora structures have typical extractable contours (e.g., corona aurora), the shape information alone is insufficient for representation. Accordingly, some texture description methods were proposed to represent aurora images. Specifically, in [9], the gray-level aura matrices (GLAM) [30] was used to extract the texture features of aurora. However, the GLAM only provides the global information without much local cues, and is sensitive to spatial scale, orientation, and intensity variation. In the past years, the local binary pattern (LBP) [31] and its variants have shown a great ability to describe the local textures for ASI aurora images. Wang *et al.* [10] applied an LBP descriptor to aurora images combined with a delicately designed block partition scheme and achieved both global shape and local texture representations. According to the characteristic of ASI, Yang *et al.* [6] presented a polar embedding method by combining the scale-invariant feature transform (SIFT) [32] and deep LBP features to represent an aurora image, and a large-scale aurora image retrieval system was developed based on the bag-of-visual-words (BoVW) model.

2) Classification: Classification aims to assign a predefined aurora type to each ASI image in the interesting data sets. Similarly, all representation methods for ASI aurora image retrieval can be applied to the classification problem by training a classifier, such as support vector machines [9] and k-nearest neighbors (kNNs) [10]. In addition, Syrjäsuo and Partamies [11] evaluated the selection of numeric image features, including simple intensity features, texture features, and brightness-invariant features, for the task of aurora detection, which can be regarded as a binary classification problem. They found that the local methods perform better than global ones, and simple intensity features, such as mean, minimum, and maximum intensity, are most accurate when the training and testing sets have the same brightness range for determining the existence of aurora in an ASI image. In recent years, CNN-based methods have become the de facto technique to the pattern recognition problems in the computer vision due to its great representation learning ability. Han *et al.* [33] have trained the CNN models to classify the ASI aurora images. Comparing with the traditional CNN models (e.g., AlexNet [34]), the classification accuracy of their method is improved on two aspects: pretraining the first layer of a multisize kernels CNN with eye movement annotations and fine-tuning a three-stream CNN with image-level labels to capture different receptive fields. However, obtaining the eye movement annotation is extremely expensive and it is hard to expand to large data sets or many other aurora types. Apart from the static ASI image classification, Yang *et al.* [35] and Zhang *et al.* [36] also explored the ASI aurora image sequence classification by taking dynamic information into consideration.

3) Segmentation: ASI aurora image segmentation is to segment all aurora structures from dark sky, which is a pixel-level binary classification problem. In order to segment the ray structures, Fu *et al.* [12] proposed an adaptive LBP (ALBP) descriptor to extract the ASI image features and a block threshold strategy to estimate the aurora region. Furthermore, Gao *et al.* [13] proposed an ASI image segmentation algorithm with two parts: a texture part based on the ALBP features to segment ray structures and a patch part based on the modified Otsu method to segment bright patch structures. They assumed that the aurora image is coarsely composed of patch and texture parts.

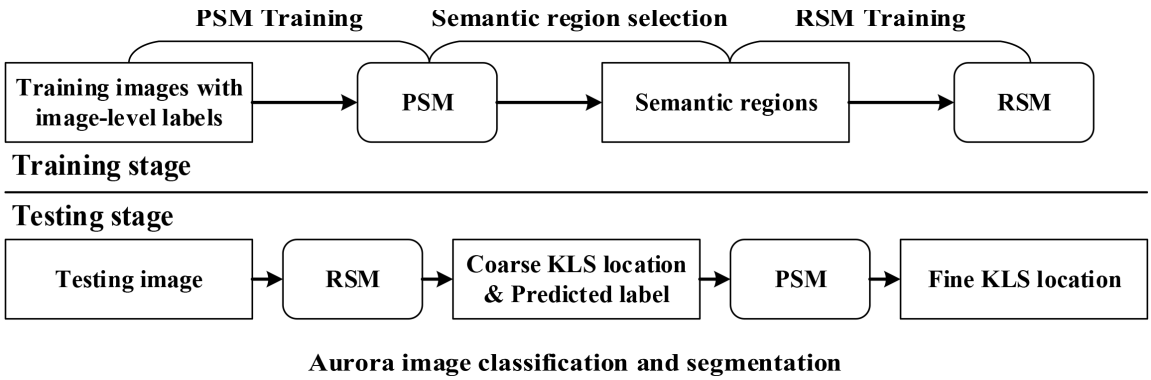


Fig. 2. Pipeline of the proposed method.

Motivated by the analysis process of aurora experts for ASI images, our proposed algorithm is to achieve the joint KLS localization and classification of an aurora image. This challenging task can be regarded as the combination of classification and segmentation, but not the simple cascading process that classification followed by segmentation. The proposed algorithm predicts the aurora type of an ASI image by coarsely localizing KLSs. It significantly improves the classification accuracy compared with the strong baseline of CNN models (Section IV-D2). On the other hand, KLS localization is a pixel-level multiclass classification problem, and the KLS of a particular aurora type is localized as the union of pixels belonging to this type. Thus, the KLS provides not only morphological information but also spatial location of aurora forms, which further improves the automatic analysis ability compared with the existing methods. In this paper, we formulate the KLS localization as the WSSS problem discussed next.

B. Weakly Supervised Semantic Segmentation Methods

WSSS aims to classify every pixel into predefined classes using weak annotations, such as image-level labels in this paper. Zhou *et al.* [21] proposed a class activation map (CAM) technique for joint classification and discriminative localization by training a classification model. However, the CAM has a low resolution and only localizes the most discriminative parts instead of complete objects. Recently, Kwak *et al.* [22] developed a classification model, named superpixel pooling network (SPN), for WSSS task. In the SPN, the resulting superpixel-pooled CAM (SP-CAM) can localize the local regions to each class, which has shown a better performance than the CAM. In addition, some WSSS methods [17]–[19] first find the corresponding local semantic regions to each image-level label and then train an FCN. Unfortunately, these methods cannot directly apply to the aurora data with only image-level labels due to the unique morphological characteristics of aurora.

III. Method

A. Overview of the Proposed Method

Given the ASI training images with only image-level annotations, the objective of this paper is to achieve the joint pixel-level KLS localization and image-level classification of aurora images. The proposed WSSS method consists of two main modules: a PSM and a RSM for analyzing small-scale structures and complex overall arrangements, respectively. The pipeline of the training and testing stage is shown in Fig. 2. The training stage contains three components: training PSM, semantic region selection, and training RSM. Specifically, the PSM is trained to identify the type-specific patches using the fixed-size patches densely detected from the training set with image-level labels. Based on the PSM, the semantic region selection component then selects the type-specific local regions of the training images. Finally, the RSM is trained with the selected semantic regions to coarsely localize the KLS using bounding boxes. At the testing stage, the RSM first outputs the coarse KLS location of each type and the aurora image type is predicted based on the KLSs. Then, given the predicted label, the PSM further refines the KLS location in a pixel level.

In Sections III-B–III-E, we detail each component of the proposed method.

B. Patch-Scale Model

In this section, we design a PSM[1] to estimate whether a small-scale structure is specific to a given aurora type. Specifically, we assume that the small-scale structures provide some morphological information as shown in Fig. 1, while it is hard to classify a patch into a definite aurora type due to the existence of common small-scale structures in different types of aurora images. Instead, assuming that the image type is known, we can determine whether a patch within the image is specific to the given type, which is known as the top–down processing [23]. However, the premise is that we have the knowledge of what kinds of small-scale structures are specific for each type and what are common between one type and other types.

To discover the specificity of small-scale structures, we construct a bag of semantic visual words (SVWs) for each aurora type, and all types of SVWs form the semantic codebook. The diagram of the semantic codebook construction is shown in Fig. 3. First, small patches are detected. We extract patches

[1]The patch scale is based on the BoVW method. To better understand the PSM, please refer [37] for more details about the BoVW method.

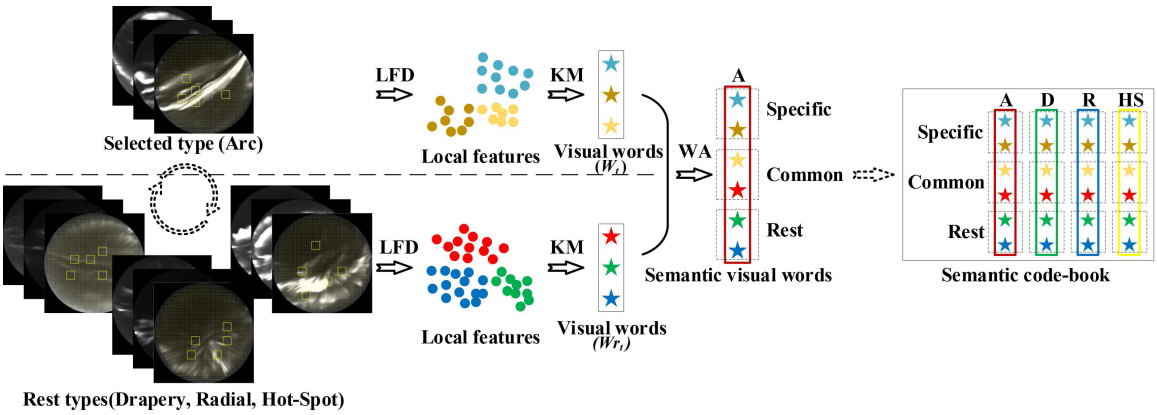


Fig. 3. Diagram of the semantic codebook construction. For a particular aurora type (e.g., arc in this figure), a bag of SVWs is constructed as follows. In step 1, small patches are generated through uniform grid and the yellow boxes on ASI images represent some of the generated patches. In step 2, all patches are mapped to feature space by an LFD. In step 3, clustering method is used to calculate the visual words. Steps 1–3 are carried out separately for the given type and the rest types of training images. In step 4, the obtained two BoVW are merged into one bag of SVWs by the proposed WA method. All types of SVWs [arc (A), drapery (D), radial (R), and HS (HS)] are obtained by conducting the above process for each type repeatedly (shown as dashed arrows), and they together form the final semantic codebook.

of size $S \times S$ evenly by a step of 10 pixels in each training image. Second, the patches are represented by some kinds of local feature descriptor (LFD) which is a key step in constructing a BoVW model. Third, a clustering algorithm is used to generate the visual words; hereafter, the simple k-means method is used. In our approach, two BoVWs are generated for each aurora type: one is from the given type images and the other from the rest types. Finally, we propose a word analysis (WA) method to merge the two bags of each type into one bag of SVWs which contains three categories: *specific*, *common*, and *rest*. The basic idea is that the *common* words are very "close" ones from the two bags and the *specific* and *rest* words are relative "far" ones from the given type and rest types of visual words, respectively.

The PSM can be formally described as follows. We denote the two BoVWs of type t by W_t and W_{r_t}. The visual words $W_t = \{w_1^t, \ldots, w_{V_t}^t\}$ are generated from the images of type t, and $W_{r_t} = \{w_1^{r_t}, \ldots, w_{V_{r_t}}^{r_t}\}$ are generated from the rest types (all predefined types, excluding type t) images, and V_t and V_{r_t} are the word numbers. Then, the interdistance matrix $D^{tr} = \{d_{ij}^{tr}\}_{V_t \times V_{r_t}}$ and intradistance matrices $D^{tt} = \{d_{ij}^{tt}\}_{V_t \times V_t}$ and $D^{rr} = \{d_{ij}^{rr}\}_{V_{r_t} \times V_{r_t}}$ are calculated using the Euclidean distance, where d_{ij}^{tr} denotes the distance between word vectors w_i^t and $w_j^{r_t}$, and d_{ij}^{tt} and d_{ij}^{rr} have the similar form. A closeness measurement is defined as

$$\alpha_{\text{closeness}} = \max(\min\_L(D^{tt}), \quad \min\_L(D^{rr})) \tag{1}$$

where min_L is to find the Lth minimum item in the intra-matrix except for the diagonals.

In order to merge the two BoVWs W_t and W_{r_t}, all the words are categorized using the following rule. If $d_{ij}^{rt} < \alpha_{\text{closeness}}$, then w_i^t and $w_j^{r_t}$ are categorized as the *common* words. If $d_{ij}^{rt} \geq \alpha_{\text{closeness}}$, then w_i^t and $w_j^{r_t}$ are labeled the *specific* and *rest* words for type t, respectively. The visual words W_t and W_{r_t} and their category labels $C_t = \{c_1^t, \ldots, c_{V_t}^t\}$, $C_{r_t} = \{c_1^{r_t}, \ldots, c_{V_t^r}^{r_t}\} (c^t, c^{r_t} \in \{specific, common, rest\})$ together form the SVWs of type t. All types of the SVWs construct the semantic codebook.

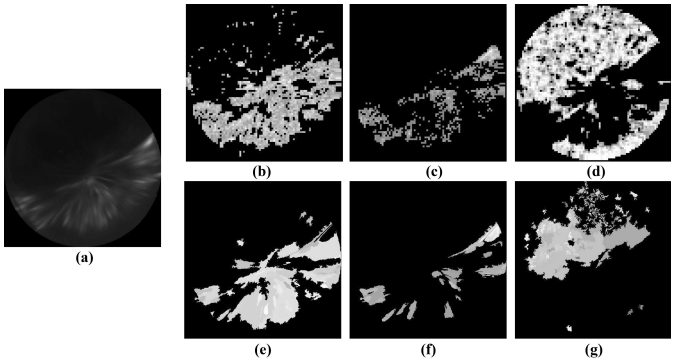


Fig. 4. Example of KLSs localized by the PSM. (a) Radial-type aurora image. (b)–(d) Heat maps of *specific*, *rest*, and *common* to the given image, respectively. The heat maps are generated by estimating p_o^s, p_o^c, and p_o^{re} of each sliding patch with a step size of 5 pixels and $k = 19$ in kNN. (e)–(f) Heat maps generated using a superpixel as the layout, and the probability of each superpixel is estimated by (3).

Based on the semantic codebook, the probability distribution of a small patch o with type t over *specific*, *common*, and *rest* category, denoted as p_o^s, p_o^c, and p_o^{re}, respectively, can be estimated by a kNN density estimator [38]. For example, given a radial-type aurora image as shown in Fig. 4(a), the SVWs of the radial type is selected to estimate p_o^s, p_o^c, and p_o^{re} using the kNN density estimator and generate the corresponding *specific*, *common*, and *rest* heat maps to the radial-type image [Fig. 4(b)–(d)]. Fig. 4 shows that the PSM can distinguish the subtle difference of small-scale structures. However, by estimating each sliding patch independently without considering its context, the generated heat maps [Fig. 4(b)–(d)] contain many noisy points. To solve this problem, we further use the superpixel as the basic layout, in which all patches are assigned to the same category. The probability of each superpixel is computed by (3) (see Section III-D). Thereafter, the noisy points are effectively removed, as shown in Fig. 4(e)–(g).

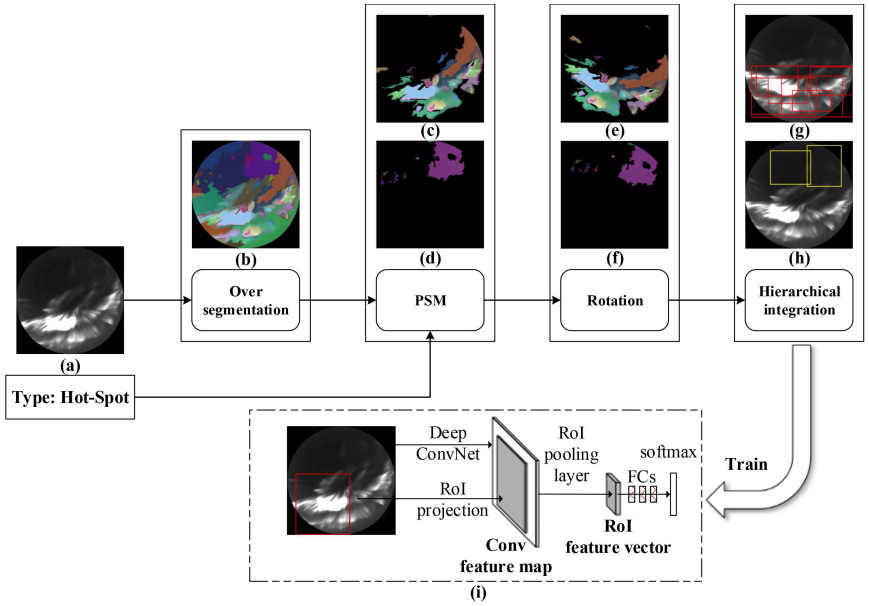


Fig. 5. Semantic region selection and RSM training. (a) Training image with a type label is (b) oversegmented to get superpixels. Superpixels are processed by the PSM to find (c) specific regions and (d) common regions. (e) and (f) Specific and common regions are rotated, respectively, to reduce the coupling between different categories in a bounding box. (g) and (h) Different sizes of semantic regions and the corresponding bounding boxes are generated by the hierarchical integration. (i) All the selected semantic regions and their type labels are used to train the RSM.

By this time, the PSM construction is finished. Then, the PSM is applied to select the semantic region and refine KLS which are described in Sections III-D and III-E.

C. Region-Scale Model

Most of the recent WSSS models use the FCN for semantic segmentation. However, the high performance of the FCN depends on a vast amount of accurate pixel-level annotations, which is absent in the WSSS setting. Although many methods are proposed to compensate for this missing information, it is hard to obtain results as accurate as the human annotations. However, the coarse location (bounding box region) is much easier to obtain and can be easily localized by the bottom–up deep detection model. Therefore, in this paper, we decompose the KLS localization of an aurora image into two procedures. First, an RSM is developed to localize each type of KLSs coarsely using bounding boxes. Then, the final pixel-level KLSs are obtained by refining the coarse KLSs with the PSM presented in Section III-B.

The RSM is a modified version of fast region-based CNN (Fast R-CNN) for object detection [39]. The region of interest layer in the Fast R-CNN represents a different size of regions with a fixed-dimension vector, and thus, it can avoid the requirement of the fixed-size input regions. This property is exactly fit for the multiple sizes of an FOV analysis of aurora images, since the resized region will probably change its original morphological type.

The RSM takes as input an entire aurora image and a set of semantic regions generated by the selective search method [40], and outputs a discrete probability distribution $p = (p_0, \ldots, p_T)$ over $T + 1$ ($+1$ for the *common*) types for each region. The original loss function in the Fast R-CNN consists of two components: a classification loss and an object bounding box regression loss. However, the RSM does not need output the object bounding boxes, and thus, the loss function of the RSM is modified as the log loss of the true type t

$$L(p, t) = -\log p_t. \tag{2}$$

More details can be obtained in [39].

D. Semantic Region Selection

In order to train the RSM, different sizes of *specific* and *common* regions of each type are selected by the PSM based on the selective search method [40]. The procedure of the semantic region selection is shown in Fig. 5. For a training image I with a label t, the graph-based segmentation method [41] oversegments the image into a set of superpixels $I = \{r_1, \ldots, r_M\}$, where M is the number of superpixels.

The category of a superpixel can be determined by its internal patches, because the patches in the same superpixel have a similar appearance and are discriminable among different categories ($specific, common$, and $rest$) in the top–down processing. Thus, the fixed-size patches are randomly sampled within each superpixel, $r_m = \{o_m^k\}, k = 1, \ldots, K_m$, where o_m is a fixed-size image patch and K_m is the number of patches set as 10% of the number of region pixels. The probability of superpixel r_m belonging to each category is defined as the average probability of its internal patches

$$P_{r_m}^s = \frac{1}{K_m} \sum_{k=1}^{K_m} p_{o_m^k}^s. \tag{3}$$

$P_{r_m}^c$ and $P_{r_m}^{re}$ have the similar form and meaning. The *specific* regions satisfying $P_{r_m}^s > \max\{P_{r_m}^c, P_{r_m}^{re}\}$ are assigned a label t, and the *common* regions ($P_{r_m}^c > \max\{P_{r_m}^s, P_{r_m}^{re}\}$) are regarded

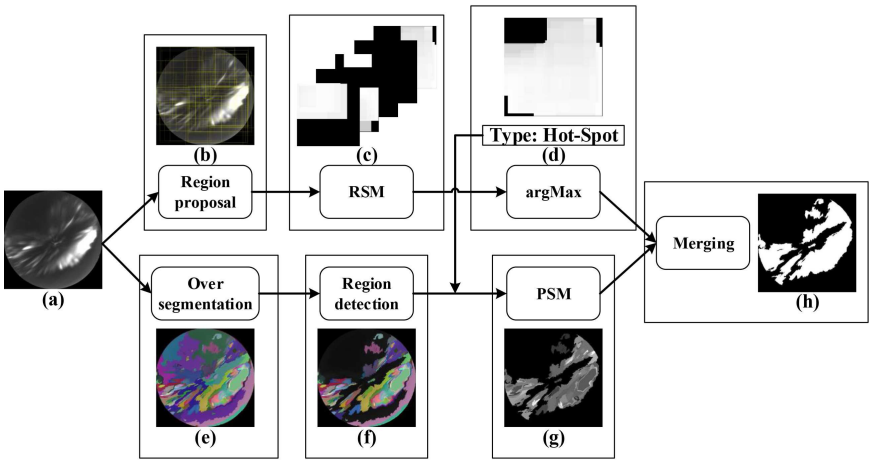


Fig. 6. Joint classification and KLS localization. Inference includes two procedures. (Bottom–up) Given (a) test image, (b) region proposal method generates a set of bounding boxes. (c) RSM calculates the coarse heat maps of each aurora type. (d) Based on these heat maps, the image type and coarse KLS location are predicted by the argMax operation. (Top–down) Set of superpixels of the image are obtained by (e) oversegmentation method. (f) Superpixel containing aurora structures are detected and processed by (g) PSM to calculate the pixel-level heat map conditioned by the predicted type. (h) Localization mask is obtained using the proposed merging strategy.

as background since we find that most of them do not contain typical aurora structures.

In order to reduce the coupling among different categories in a bounding box, we rotate the image according to the principle direction using the principal component analysis method so that its horizontal direction is at the maximum variance axis of the *specific* and *common* regions. To capture multiple sizes of FOV, we hierarchically integrate the regions with a similar appearance (measurements are the same as in [40]) for each category until all regions have been integrated into one. Finally, the bounding boxes fitting all sizes of semantic regions with the assigned labels are used as the training samples. In this process, many redundant bounding boxes may be generated. To reduce the redundancy, we repeatedly keep the biggest bounding box first and remove the boxes overlapped with the biggest one more than 90%.

Although some superpixels may be assigned with false categories, they can hardly affect the training of RSM. This is because the small boxes (e.g., the size is less than 10 000 pixels, as discussed in Section IV-D1) that contain the isolated noisy superpixels will not be included in the RSM training data [see Fig. 5(g)]. In addition, the effect of small noisy superpixels can be ignored with respect to the relative large training boxes, in which most of structures are consistent with the assigned labels. It is one of the main reasons that RSM has a superior performance over weak image-level annotations.

E. Aurora Image Classification and KLS Localization

The proposed framework of the joint KLS localization and aurora image classification is shown in Fig. 6. There exist two procedures. The first procedure performs the coarse KLS localization and image classification in a bottom–up manner. Given an entire ASI image as an input, a set of regions are generated by the selective search method [40] and forwarded to the RSM to calculate the corresponding scores of $T+1$ types. Then, these scoring semantic regions are merged into $T+1$ coarse heat maps $\{h_t\}$ in which the probability of each pixel is calculated by averaging the scores ($\geq$0.8) of all the bounding boxes containing the pixel. The predicted type l of the whole aurora image is decided by the coarse heat map with a maximum area

$$l = \arg\max_t \{\text{area}(h_t > 0.8)\}. \tag{4}$$

The coarse heat map h_l can be used to coarsely localize the KLS of the predicted image type l.

The second procedure is to refine the KLS in a pixel level, given the predicted type l and coarse heat-map h_l. First, the input image is oversegmented to obtain a set of superpixels. Then, a region detection (RD) method is conducted to detect regions containing aurora structures before further processing, as motivated by [11] and [13]. The basic idea is that the dark superpixels with the mean intensity lower than a certain threshold of dark region should be filtered out, because they do not contain KLS. As the size of bright region increases, the intensity threshold of the dark region should be increased to eliminate artifacts generated by large bright regions. Thus, the threshold is adaptively set by a bounded linear function

$$\text{th} = \min(25 + 0.05 S_{180}, 80) \tag{5}$$

where S_{180} is the number of pixels with the value greater than 180 in the aurora image (the image pixel value is from 0 to 255). S_{180} represents the size of bright region, and 25 and 80 represent the minimum and maximum values of the dark region threshold. These values were set empirically by analyzing the ASI image histograms motivated by [13]. Afterward, given the predicted image type, the PSM generates a fine-grained heat map by identifying the *specific* regions. Finally, the fine-grained and the coarse heat map are averaged and thresholded to suppress the false positives generated by PSM. The KLS is determined as the pixels with the average heat-map values larger than 0.5.

IV. Experiments

A. Aurora Image Data

The aurora data used in this paper were observed by the ASI system installed in the Chinese Arctic Yellow River Station

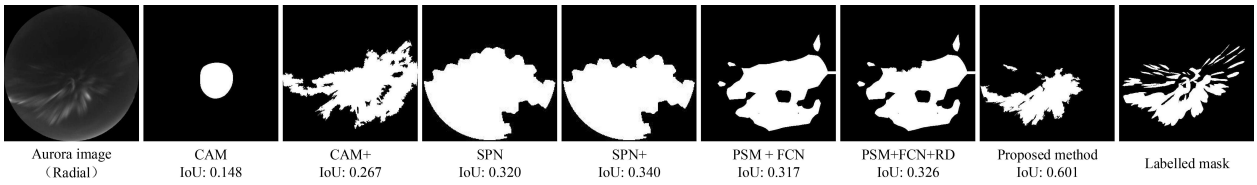


Fig. 7. KLS localization results of a radial-type aurora image by various WSSS methods.

(YRS), Ny-Ålesund, Svalbard. YRS is located on geographic coordinates 78.92° N, 11.93° E. Three ASIs were installed in the optical system to measure the multiple wavelengths of photoemissions at 427.8, 557.7, and 630.0 nm since December 2003 [2]. The optical instruments at YRS can provide 24-h surveys of aurora emissions with a temporal resolution of 10 s in the winter season from October to March of the following year. In this paper, we concentrate on the dayside (03∼15 UT/06∼18 magnetic local time) aurora at 557.7 nm from December 2003 to February 2009 the same as in [10]. To better focus on the study of aurora classification and KLS localization algorithm, the images that do not contain aurora structures or are captured under bad weather conditions (e.g., aurora structures are severely covered by clouds or containing the moonlight) are eliminated by the human visual inspection. In a practical automatic aurora image analysis system, the human preprocessing can be replaced by the aurora detection methods, such as [11], which can be regarded as a front-end processor.

The same as in the previous works [2], [10], all ASI aurora images have been preprocessed before human visual inspection and automatic algorithms as follows.

1) *Subtracting Dark Current and Rescaling:* Dark current is deemed as system noise caused by equipment. It is removed from images before further operation. Every image is then stretched with a cutoff value of 4000 and rescaled from 16 to 8 bits. The advantage of using image stretching is that the relative intensities of the pixels are preserved, while the image contrast is enhanced. After stretching, the images are more easily categorized and labeled by human visual inspection in the following studies.
2) Masking and Cropping: A circle mask with a radius of 220 pixels is applied to cut off the outer regions where a significant wide-angle distortion happens and may contain YRS lights. Then, the original image size of 512 × 512 is cropped to 440 × 440 pixels.

In this paper, we assume that there are four types of aurora[2]: arc (A), drapery (D), radial (R), and HS. Based on the previous image-level annotated data for classification [10], we manually construct several data sets according to the image observation time to train and evaluate the proposed WSSS model.

SetG38K contains 38 044 images with the image-level annotations from December 2003 to January 2004 the same as in [10], which is used to pretrain the RSM.

SetG2K contains 2000 images in which each type has a balanced number of 500 images selected from *SetG38K*. In this data set, each image only includes a single typical type of aurora, which is used to mine the knowledge of semantic codebook and fine-tune the RSM.

SetGcls2K contains 2000 images with the image-level annotations in which each type has a balanced number of 500 images selected from December 2004 to January 2009. It is used to evaluate the classification accuracy of our proposed WSSS method.

SetGseg200 contains 200 images with the pixel-level annotations selected from *SetGcls2K*, which is mainly used to evaluate the segmentation effectiveness of the proposed WSSS method.

B. Implementation Details

Motivated by previous works [6], [10], [11], [13], we evaluate LBP [31], SIFT [42], and intensity histogram local descriptors for describing small patches when training PSM. As suggested in [31], the "uniform" LBP incorporated by different spatial resolutions and different angular resolutions [$(P = 8, R = 1)$, $(P = 16, R = 2)$, and $(P = 24, R = 3)$] is used. The intensity histogram is divided into 64 bins for the 8-bit gray-scale aurora image. We take $k = 1$ of k-NN estimator for efficiency in this paper, since we find that k has a little impact on the final results. This phenomenon can be explained by (3) in which the probability of a superpixel has already considered as many SVWs as the number of the internal patches even for $k = 1$.

We use the *VGG_CNN_M_1024* net [43] as the base network, which is pretrained for classification on the SetG38K. Then, the RSM is fine-tuned from the pretrained network on SetG2K. During the fine-tuning, we keep the weights in the first layer freezing and tune all other layers. The network is trained by backpropagation and stochastic gradient decent. Each mini-batch consists of 64 semantic regions, including 10 background regions and 54 predefined types of regions randomly. We use an initial learning rate of 0.005 and decrease it by a factor of 0.1 every 2000 iterations. We use a momentum of 0.9 and a weight decay of 0.0005. Our method is implemented with Caffe [44].

C. Existing WSSS Methods for Comparison

We compare the proposed method with two state-of-the-art WSSS methods [21], [22] for the joint localization of KLS and classification of aurora images. Since many WSSS methods directly find the local regions related to the image-level labels and then use these local regions as pixel-level supervision to train the segmentation model, for the purpose of comparison, we also train an FCN using the regions selected by the PSM for semantic segmentation of aurora images.

[2]In this paper, the proposed algorithm can only recognize the four types of aurora: arc, drapery, radial, and HS. In addition, it can be extended to analyze other types of aurora by constructing the corresponding data sets.

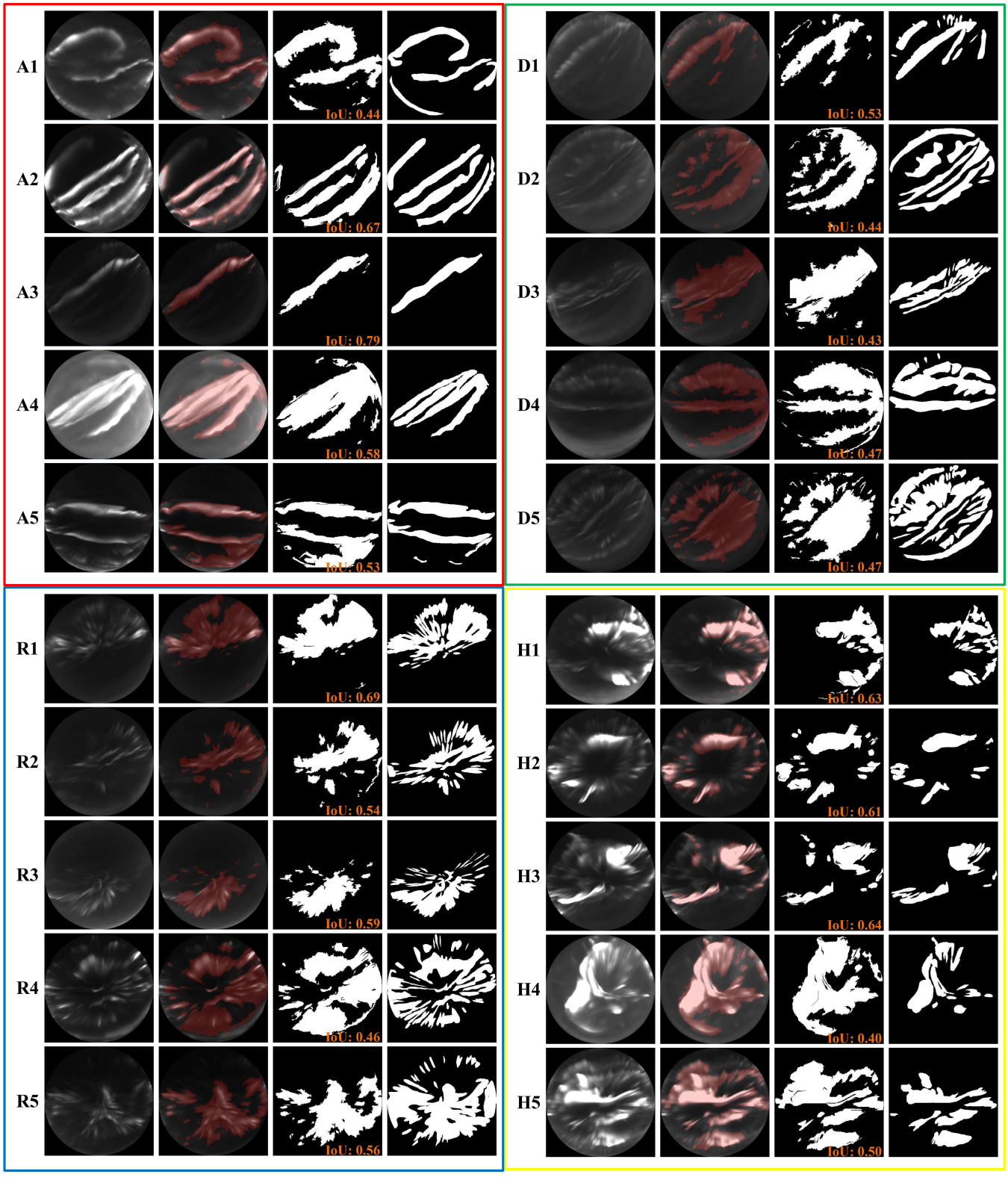


Fig. 8. Examples of classification and KLS localization for arc (red box), drapery (green box), radial (blue box), and HS (yellow box) aurora. In each box, each row presents (Left to right) the original image, original image covered by predicted mask, predicted mask, and human labeled mask, respectively. Different rows present different examples.

1) Class Activation Mapping: The CAM method was introduced in [21]. It performs global average pooling on the convolutional feature maps and use those as features for a fully connected layer to produce the desired output (categories). When training the classification architecture, only image-level annotations are needed and the trained model is used

to classify aurora images. Given this simple connectivity structure, the CAMs representing the importance of image regions can be generated by projecting back the weights of the output layer onto the convolutional feature maps. Simply upsampling the CAMs to the size of the input image, the image regions most relevant to the particular category are identified. Nevertheless, the resolution of this localization method is low and it only localizes the most discriminative parts instead of the complete structures, as demonstrated by the example shown in Fig. 7. To solve this problem, we oversegment the aurora image into a set of superpixels, as described in Section III-D. Then, the RD method (Section III-E) is used to detect the aurora regions. For propagating the identified most discriminative parts to complete regions, the detected regions containing at least one point with its corresponding score in the CAM larger than 0.5 are defined as the final localized KLS. The modified CAM method is denoted as CAM+.

2) Superpixel Pooling Network: The SPN was proposed in [22]. The SPN takes two inputs for inference: an image and its superpixel map. Given an input image, the network extracts high-resolution feature maps using a CNN encoder followed by several upsampling layers, and the superpixel pooling layer aggregates features inside each superpixel by exploiting an input superpixel map as the pooling layout. The aggregated features are forwarded to fully connected layer to output classes. An additional branch of global average pooling is added for regularization, which prevents undesirable training noises introduced by superpixels. The trained SPN can be directly applied to classify aurora images. To achieve semantic segmentation, the feature vector to each superpixel is generated through the superpixel pooling layer and fed to the fully connected classification layer to output the class scores of each superpixel. As a result, the activation map for the associated class is obtained, which is called the SP-CAM. The segmentation mask is set as the superpixels with the score larger than 50% of the maximum of SP-CAM. This is equivalent to that the score map values are first normalized into [0, 1] by the maximum for the probabilistic interpretation and the segmentation mask is the union of pixels with probability larger than 0.5. We further improve the segmentation results for aurora images by filtering out superpixels not containing aurora using the RD method. The modified SPN method is denoted as SPN+.

3) PSM + FCN for KLS Localization: We evaluate another WSSS strategy for semantic segmentation of aurora images that is training a bottom–up FCN model [16] using the selected semantic regions by the proposed PSM. It should be noted that the selected semantic regions for training FCN do not fit bounding boxes as for training the RSM. Given an image of type t, the trained FCN directly generates $T + 1$ types of heat maps. The final segmentation mask is set as the pixels in the heat map of type t whose value is greater than that of pixels in all other heat maps. The segmentation results can also be improved by the RD method.

D. Classification

1) Parameter Analysis: One of the key strategies in our proposed WSSS framework is analyzing aurora images from multiple sizes of FOV. Larger regions observe more complete aurora forms but have weak fine-grained localization ability, while smaller regions emphasize more details but cannot provide distinct semantics of predefined morphological types. To explore the effect of FOV, we train and test RSM for classification using a different size range of regions (in pixels), from 10 000 (about 1/20 of the image size) to 48 400 (half image size) denoted by *S*, from 48 400 to 193 600 (full image size) denoted by *L*, and from 10 000 to 193 600 denoted by *F*, where the image size is 440 × 440. The RSM is pretrained on data set SetG38K, fine-tuned on SetG2K, and tested on SetGcls2K. We use an LBP descriptor, a patch size of 16 × 16, and the words number of 500 to construct the semantic codebook for semantic region selection. Table I shows the classification accuracy of each aurora type and its mean value by RSM trained and tested on a different size range of regions. The RSM trained on *F* and tested with *S* has the best classification results. This means that a half size of the image is sufficient to distinguish different aurora morphologies. In addition, there is an interesting phenomenon that a large size of FOV has strong discriminability for arc-type and radial-type auroras but weak for drapery-type and HS-type auroras. Actually, the phenomenon coincides with the facts that the arc-type and radial-type auroras usually cover a very large size of FOV, while the drapery-type and HS-type auroras are mainly distinguished by KLS, as shown in Fig. 8.

TABLE I
RSM CLASSIFICATION ACCURACY FROM DIFFERENT SIZES OF FOV

Train	Test	Arc	Drapery	Radial	Hot-spot	Mean
S	S	0.944	**0.932**	0.710	0.900	0.8715
S	L	0.946	0.926	0.646	0.844	0.8405
S	F	0.942	0.934	0.644	0.850	0.8425
L	S	0.944	0.896	0.678	0.820	0.8345
L	L	**0.954**	0.826	**0.770**	0.878	0.8570
L	F	0.946	0.832	0.762	0.872	0.8530
F	S	0.950	0.928	0.682	**0.934**	**0.8735**
F	L	0.940	0.900	0.656	0.894	0.8475
F	F	0.940	0.914	0.652	0.902	0.8520

2) Compared With Existing Methods: In order to demonstrate the effectiveness of multiple sizes of FOV, we compare the proposed RSM with a traditional CNN classifier which is trained and tested directly using the whole image under the same settings. Table II shows the classification results of the traditional CNN classifier. The best RSM classification accuracy is 5.5% higher than the traditional CNN classifier on *SetGcls2K*. It is noted that the main deficiency of the traditional CNN is the relative low discriminability of the drapery-type and the HS-type aurora compared with the RSM. We interpret this result to the classification mechanism that these aurora morphologies are regarded as the subcategories of *patchy aurora* [13]. Just as our proposed method, the subcategories need KLS for fine-grained classification. In Table II, we also compare the RSM with the state-of-the-art WSSS methods [21], [22] in terms of classification. Since both CAM and SPN use global pooling which significantly damages the details, they have weak ability to distinguish these

TABLE II
COMPARED WITH EXISTING METHODS FOR CLASSIFICATION

Model	Arc	Drapery	Radial	Hot-spot	Mean	Time
CNN [43]	0.992	0.712	0.750	0.820	0.8185	0.0374
SPN [22]	0.996	0.570	0.814	0.650	0.7575	0.0821
CAM [21]	0.868	0.570	0.682	0.812	0.7330	0.1525
Our method	0.950	0.928	0.682	0.934	**0.8735**	1.2338

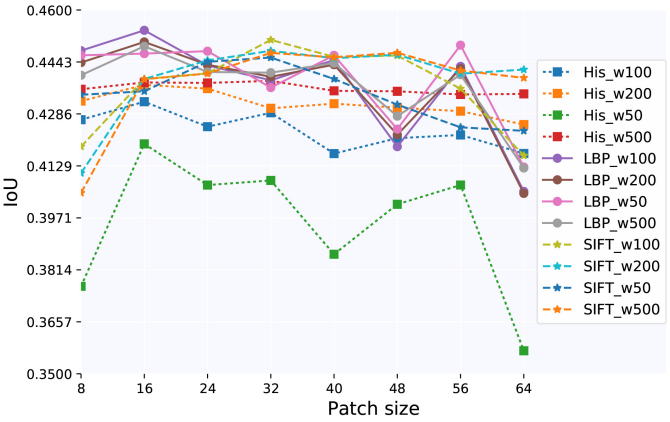


Fig. 9. Segmentation results over different patch sizes. In the legends, His, LBP, and SIFT denote the local patch descriptors and the numbers represent the numbers of visual words.

TABLE III
SEGMENTATION ACCURACY FOR DIFFERENT COMPONENTS IN TERMS OF IOU

Method	Arc	Drapery	Radial	Hot-spot	Mean
RD	0.493	0.270	0.388	0.401	0.388
PSM	0.481	0.351	0.457	0.376	0.416
PSM+RD	0.548	0.350	0.443	0.468	0.452
PSM+RD+Merge	0.555	0.342	0.442	0.476	0.454

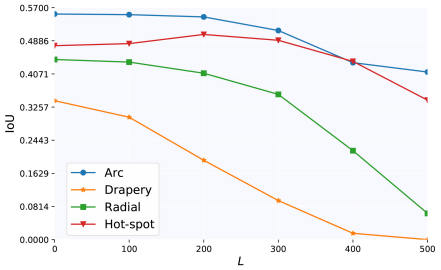


Fig. 10. Segmentation results for different values of L with 100 word numbers.

subcategories. In addition, the proposed method has weak discriminability for radial-type aurora compared with other types due to two aspects. First, we select the RSM trained on F and tested with S as the best classification model in terms of the mean accuracy while testing with S will hurt the radial-type classification accuracy as discussed in Section IV-D1. Second, the radial-type aurora structures are usually contained in all other types, so that it is easy to confuse the radial-type aurora with other types.

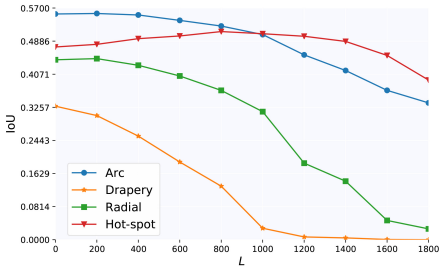


Fig. 11. Segmentation results for different values of L with 200 word numbers.

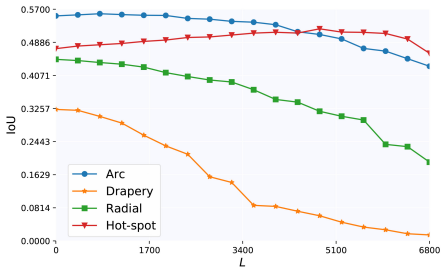


Fig. 12. Segmentation results for different values of L with 500 word numbers.

E. KLS Localization

1) Parameter Analysis: In this section, we evaluate the effectiveness of the KLS localization in terms of intersection over union (IoU) on the challenging data set *SetGseg200*. First, we analyze the effects of a local patch descriptor, a patch size, and a word number on the mean IoU. The results are shown in Fig. 9. Using the patch size of 16×16 and 32×32 can achieve the best segmentation accuracy for LBP and SIFT, respectively, both on the 100 visual words. The number of visual words is more important for the intensity histogram feature. More visual words will have a better segmentation result, but they also require more computation and memory costs. Therefore, we select the LBP descriptor, a patch size of 16×16, and 100 visual words as the optimal parameters for further analysis.

2) Ablation Experiments on Region Detection and Merging Strategies: To investigate the importance of RD and merging strategies, we conduct four KLS localization experiments by using different combinations of components: RD-only, PSM-only, RD and PSM (PSM + RD), and RD and PSM followed by merging strategy (PSM + RD + Merge). The results are shown in Table III. RD is actually an adaptive thresholding segmentation method constrained by superpixels. Despite the simplicity, the RD-only achieves good segmentation results. The PSM-only has a better accuracy than the RD-only. By combining both RD capturing the intensity feature and PSM capturing the texture feature, the segmentation accuracy is significantly improved. The merging strategy has very little improvement (0.2%) in segmentation results, since we mainly focus on ASI images containing a single type of aurora. The merging strategy is necessary in the multilabel classification problem.

3) "Keyness" of KLS: In fact, the key local structure is vaguely defined, especially for HS type of aurora whose

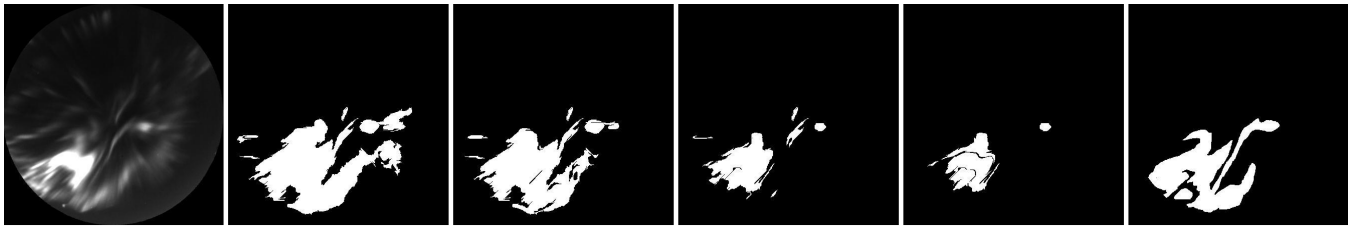


Fig. 13. Segmentation results of different values of L. (Left to right) Original image, KLS of L=1, 300, 500, 700, and labeled mask, respectively, with 100 words.

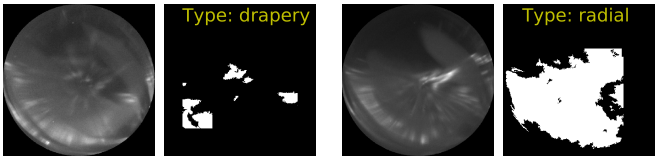


Fig. 14. Failed examples under nonideal conditions. (a) Classification error: *radial* aurora is wrongly classified as a *drapery* type. (b) Segmentation error: clouds are wrongly localized as *radial* KLSs.

morphologies are very complex and usually consist of irregular bright patch and rays. The problem is that a ray structure is also the main characteristic of drapery aurora and radial aurora. In our proposed method, the parameter L in the WA method (Section III-B) can control the "keyness" of KLS. We test a different value of L for segmentation on a different size of semantic codebook using the LBP descriptor. The results are shown in Figs. 10–12. With the increase in the L value, only the segmentation accuracy curve of the HS type aurora first rises and then drops, while others directly decline. According to (1), with the increase in the L value, the closeness measurement α will increase and the algorithm will keep more-specific structures. Particularly, L to be 0 means that the closeness measurement α is set as the minimum of intradistance matrix introduced in Section III-B. Thus, we can infer that the HS aurora has more common structures with other types. Fig. 13 shows an example for visual inspection about the relation between the "keyness" and L. It is noted that the value of L depends on the test data set, so L is set to 0 for each type of aurora to give a fair comparison with other methods and the value of L can be changed accordingly for future applications.

4) Compared With Existing Methods: In this section, the KLS localization effectiveness of the proposed method is evaluated qualitatively and quantitatively based on aurora experts labeled masks. It is noted that the labeled masks are not used for training the proposed model, while they are the ground truth for evaluating the KLS localization performance of automatic methods. In Fig. 7, an example image is presented for visual comparison of the proposed method with other WSSS methods. More KLS localization examples by the proposed method are presented in Fig. 8. Trained with only image-level labels, the proposed method can jointly classify the aurora image and localize the KLS in a pixel level with satisfying results. As shown in Fig. 8, the proposed method cannot only localize various types of arc (A1–A5) and irregular patch structures (H1–H5) but also various ray structures (D1–D5 and R1–R5). Even with the unlabeled structure interference, the proposed method can still localize the KLS. For instance, the example A2 is partially covered by cloud and the proposed method only localizes the key arc structures and ignores the cloud, which agrees with the human understanding. The drapery-type aurora image often includes diffuse aurora as shown in D2, D3, and D4. The proposed method can still localize the key draperylike structures and is not affected too much by the diffuse structure.

To make a comprehensive evaluation, we also show two typical failed examples from the proposed algorithm under nonideal conditions that the aurora structures are severely covered by clouds (see Fig. 14). These ASI images are not contained in our interesting data sets. In addition, the errors caused by nonideal conditions can be mitigated by the front-end processing of aurora detection manually or automatically [11].

The proposed method was quantitatively evaluated on data set *SetGseg200*. The results presented in Table IV indicate that our method achieves the best semantic segmentation performance for each type of aurora images. The main deficiency of the existing methods is mainly caused by drapery and radial images which consist of similar transparent rays with a varying brightness and similar appearance as shown in Fig. 8. By directly using the bottom–up processing for semantic segmentation of aurora images, the existing methods have great difficulties to segment the detail structures. We believe that the superiority of the proposed method is resulted from the combination of both bottom–up and top–down processing.

F. Runtime

Since the training stage is performed offline, users are more concerned about the runtime in the testing stage. The runtime[3] of the proposed method as well as the compared methods for the classification and KLS localization of a 440×440 ASI aurora image is reported in Tables II and IV. The proposed method needs about 1.2 s for classification and about 1.7 s to localize the KLS. The process of generating multisize FOV is the most time-consuming component (about 1.1 s), so that the proposed method takes more time than the existing methods, especially for the classification task. However, it can still achieve the "nowcasting" application for a 10-s imaging cadence now. For the community to make use of the discoveries and further research, the source code is released at *https://github.com/niuchuangnn/Aurora-ASI-KLS*.

[3]The reported runtime is tested on a generic PC (Intel i7 CPU, 8 GB memory) with an NVIDIA GeForce GTX 750Ti.

TABLE IV
SEGMENTATION ACCURACY COMPARISON WITH THE EXISTING METHODS IN TERMS OF IOU

Method	Arc	Drapery	Radial	Hot-spot	Mean	Time
PSM + FCN [16]	0.550	0.168	0.342	0.422	0.370	0.5709
PSM + FCN [16] + RD	0.548	0.170	0.322	0.475	0.379	0.7133
CAM [21]	0.163	0.149	0.172	0.118	0.151	0.1730
CAM+	0.376	0.205	0.292	0.280	0.288	0.4199
SPN [22]	0.308	0.208	0.267	0.171	0.239	4.1997
SPN+	0.440	0.227	0.384	0.401	0.363	4.2335
Proposed method	0.555	0.342	0.442	0.476	0.454	1.7553

G. Discussion

From both the quantitative and qualitative results, we can see that proposed WSSS method performs better than the state-of-the-art methods for joint KLS localization and aurora image classification. The improvement is due to two aspects: bottom–up analysis of an aurora image from multiple sizes of FOV and top–down localization of the KLS using the specificity of small-scale structures. By combining a large size of FOV that captures the overall arrangements and a small size of FOV that emphasizes the local structure, the classification accuracy has been significantly improved, especially for the subcategories. Using the obtained image type and the specificity of small-scale structures, the top–down processing can distinguish the subtle differences among different type of aurora rays. However, the proposed method has several drawbacks: 1) the hand-crafted local descriptors may be suboptimal for describing the small patches and 2) the training of the PSM for more types of aurora by clustering algorithm will be a time-consuming process. How to integrate the learning processes of small-scale structure specificity and high-level features into one end-to-end network is a problem needed to be solved in the feature.

V. CONCLUSION

This paper proposed a WSSS method for joint KLS localization and aurora image classification using image-level annotations only. By analyzing the aurora images from multiple sizes of FOVs with the deep convolutional network, the designed RSM has significantly improved the classification accuracy. To accurately localize the KLS of aurora images in the pixel level, a from-coarse-to-fine procedure was developed by combining both the RSM and the PSM. Extensive experiments were conducted on the experts' labeled data sets, and the results have demonstrated that the proposed method achieves higher accuracy in terms of both classification and segmentation for aurora images compared with other methods.

In the future, we will solve the time-consuming problem by integrating the PSM and the RSM into one end-to-end deep network. In addition, both better overall appearance representation and local detail representation methods will be taken into account to further improve the classification and key local structure localization performance. Since the local structure analysis has recently attracted attention in aurora research, e.g., throat aurora [45], we will apply our proposed joint classification and KLS localization framework for analyzing more types of local structures automatically.

ACKNOWLEDGMENT

The authors would like to thank D. Han and Z. Hu for discussion and the Polar Research Institute of China for providing the aurora data sets.

REFERENCES

[1] M. T. Syrjäsuo, E. F. Donovan, and L. L. Cogger, "Content-based retrieval of auroral images—Thousands of irregular shapes," in *Proc. IASTED Int. Conf. Vis., Imag., Image Process.*, 2004, pp. 224–228.

[2] 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. Solar-Terrestrial Phys.*, vol. 71, nos. 8–9, pp. 794–804, 2009.

[3] Y. I. Feldstein and R. D. Elphinstone, "Aurorae and the large-scale structure of the magnetosphere," *J. Geomagn. Geoelectr.*, vol. 44, no. 12, pp. 1159–1174, 1992.

[4] A. Kullen, M. Brittnacher, J. A. Cumnock, and L. G. Blomberg, "Solar wind dependence of the occurrence and motion of polar auroral arcs: A statistical study," *J. Geophys. Res. Atmos.*, vol. 107, no. A11, pp. 13-1–13-23, 2002.

[5] M. T. Syrjäsuo and E. F. Donovan, "Using relevance feedback in retrieving auroral images," in *Proc. IASTED Int. Conf. Comput. Intell.*, 2005, pp. 420–425.

[6] 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.

[7] M. T. Syrjäsuo and T. I. Pulkkinen, "Determining the skeletons of the auroras," in *Proc. Int. Conf. Image Anal. Process.*, Sep. 1999, pp. 1063–1066.

[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, 2004.

[9] M. T. Syrjäsuo, E. F. Donovan, X. Qin, and Y.-H. Yang, "Automatic classification of auroral images in substorm studies," in *Proc. Int. Conf. Substorms*, 2006, pp. 309–313.

[10] Q. Wang *et al.*, "Spatial texture based automatic classification of dayside aurora in all-sky images," *J. Atmos. Solar-Terrestrial Phys.*, vol. 72, nos. 5–6, pp. 498–508, 2010.

[11] M. T. Syrjäsuo and N. Partamies, "Numeric image features for detection of aurora," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 2, pp. 176–179, Mar. 2012.

[12] R. Fu, X. Gao, and Y. Jian, "Patchy aurora image segmentation based on ALBP and block threshold," in *Proc. Int. Conf. Pattern Recognit.*, vol. 1, Aug. 2010, pp. 3380–3383.

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

[14] H. Yang *et al.*, "Synoptic observations of auroras along the postnoon oval: A survey with all-sky TV observations at Zhongshan, Antarctica," *J. Atmos. Solar-Terrestrial Phys.*, vol. 62, no. 9, pp. 787–797, Jun. 2000.

[15] K. Axelsson *et al.*, "Spatial characteristics of wave-like structures in diffuse aurora obtained using optical observations," *Ann. Geophys.*, vol. 30, no. 12, pp. 1693–1701, 2012.

[16] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3431–3440.

[17] Y. Wei *et al.*, "Learning to segment with image-level annotations," *Pattern Recognit.*, vol. 59, pp. 234–244, Nov. 2015.
[18] Y. Wei *et al.*, "STC: A simple to complex framework for weakly-supervised semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 11, pp. 2314–2320, Nov. 2017.
[19] Z. W. Lu, Z. Fu, T. Xiang, P. Han, L. Wang, and X. Gao, "Learning from weak and noisy labels for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 3, pp. 486–500, Mar. 2017.
[20] Y. Li, Y. Guo, Y. Kao, and R. He, "Image piece learning for weakly supervised semantic segmentation," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 4, pp. 648–659, Apr. 2017.
[21] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2921–2929.
[22] S. Kwak, S. Hong, and B. Han, "Weakly supervised semantic segmentation using superpixel pooling network," in *Proc. AAAI*, 2017, pp. 4111–4117.
[23] M. Bar *et al.*, "Top-down facilitation of visual recognition," *Proc. Nat. Acad. Sci. USA*, vol. 103, no. 2, pp. 449–454, 2006.
[24] G. A. Germany *et al.*, "Analysis of auroral morphology: Substorm precursor and onset on January 10, 1997," *Geophys. Res. Lett.*, vol. 25, no. 15, pp. 3043–3046, Aug. 1998.
[25] X. Li *et al.*, "Comparing different thresholding algorithms for segmenting auroras," in *Proc. Int. Conf. Inf. Technol Coding Comput.*, vol. 2, Apr. 2004, pp. 594–601.
[26] C. Cao, T. S. Newman, and G. A. Germany, "New shape-based auroral oval segmentation driven by LLS-RHT," *Pattern Recognit.*, vol. 42, no. 5, pp. 607–618, May 2009.
[27] 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.
[28] J. Shi, Y. Lei, J. Bai, and J. Wu, "Gradually evolved fuzzy active contour model for auroral oval segmentation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2017, pp. 3771–3774.
[29] 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.
[30] D. A. Clausi and H. Deng, "Design-based texture feature fusion using Gabor filters and co-occurrence probabilities," *IEEE Trans. Image Process.*, vol. 14, no. 7, pp. 925–936, Jul. 2005.
[31] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
[32] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[33] B. Han, F. Chu, X. Gao, and Y. Yan, "A multi-size kernels CNN with eye movement guided task-specific initialization for aurora image classification," in *Proc. Chin. Conf. Comput. Vis.*, 2017, pp. 533–544.
[34] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[35] 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.
[36] J. Zhang, Q. Wang, Z. Hu, and M. Liu, "Auroral event representation based on the n-ary fusion of multiple oriented energies," *Neurocomputing*, vol. 253, pp. 42–48, Aug. 2017.
[37] C.-F. Tsai, "Bag-of-words representation in image annotation: A review," *ISRN Artif. Intell.*, vol. 2012, no. 1, 2012, Art. no. 376804.
[38] C. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2007.
[39] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Jun. 2015, pp. 1440–1448.
[40] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," *Int. J. Comput. Vis.*, vol. 104, no. 2, pp. 154–171, Apr. 2013.
[41] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 167–181, Sep. 2004.
[42] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[43] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," in *Proc. Brit. Mach. Vis. Conf.*, 2014, pp. 1–12.
[44] Y. Jia *et al.* (2014). "Caffe: Convolutional architecture for fast feature embedding." [Online]. Available: https://arxiv.org/abs/1408.5093
[45] D.-S. Han *et al.*, "Observational properties of dayside throat aurora and implications on the possible generation mechanisms," *J. Geophys. Res. Space Phys.*, vol. 122, no. 2, pp. 1853–1870, Feb. 2017.

Chuang Niu received the B.S. degree from Xidian University, Xi'an, China, in 2015, where he is currently pursuing the Ph.D. degree.

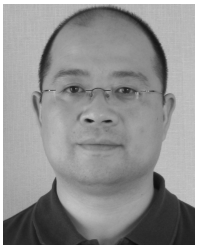
His research interests include machine learning, pattern recognition, and computer vision.

Jun Zhang received the B.S. and Ph.D. degrees from Xidian University, Xi'an, China, in 2009 and 2014, respectively.

From 2014 to 2017, he was a Post-Doctoral Research Associate with the Biomedical Research Imaging Center, Department of Radiology, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. He is currently a Post-Doctoral Research Associate with the Department of Radiology, Duke University, Durham, NC, USA. His research interests include image processing, machine learning, pattern recognition, and medical image analysis.

Qian Wang received the Ph.D. degree in pattern recognition and intelligence system from Xidian University, Xian, China, in 2011.

From 2012 to 2015, she was a Post-Doctoral Research Fellow with the SOA Key Laboratory for Polar Science, Polar Research Institute of China, Shanghai, China. From 2017 to 2018, she was a Visiting Scholar with the Biomedical Research Imaging Center, Department of Radiology, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. She is currently with the Xian University of Posts and Telecommunications, Xi'an, where she is also an Associate Professor with the School of Telecommunication and Information Engineering. Her research interests include image/video processing, computer vision, and multimedia information retrieval.



Jimin Liang (M'09) received the B.E., M.E., and Ph.D. degrees in electronic engineering from Xidian University, Xi'an, China, in 1992, 1995, and 2000, respectively.

He joined Xidian University in 1995, where he is currently a Professor with the School of Life Science and Technology. In 2002, he was a Research Associate Professor with the Electrical and Computer Engineering Department, The University of Tennessee, Knoxville, TN, USA. His research interests include biomedical imaging, image analysis, and pattern recognition. He has authored or co-authored over 100 journal papers and holds 30 granted patents on those topics.

Dr. Liang is an Executive Member of the Shaanxi Society of Biomedical Engineering, a Standing Member of the Molecular Imaging Society of China, and a member of the Society of Photographic Instrumentation Engineers. He was a recipient of the 2010 Progress Award in Science and Technology of the Ministry of Education of China and the 2012 National Invention Award of China.