Automatic keying algorithm for multi-category target recognition

Liping Mao*
Nanchang Institute of Technology, Nanchang, 330044, China

*Corresponding author: maoliping_nclg@nut.edu.cn

Abstract. In this paper, through an in-depth study of the automatic keying algorithm for target recognition and using multi-class algorithms for its analysis, a saliency detection model based on the hypercomplex Fourier transform is proposed, which can quickly search for information related to the current task requirements. The problem of sample imbalance in deep neural network training exists, the module is used many times to fuse multi-scale features, the loss function uses weighted cross-entropy loss, and the weights are determined according to the proportion of samples in the training sample, which can solve the problem of the model tends to fit the category with more samples. After analysis, the proposed hyperspectral image fast feature enhancement algorithm based on guided filtering can effectively solve the problem of "the same object, different spectrum", and the classification accuracy of small sample high-dimensional data is improved greatly. At the same time, the complexity of processing high-dimensional data such as hyperspectral remote sensing images is greatly reduced. The experimental results show that the processing time of the proposed fast feature enhancement process for hyperspectral remote sensing images in this paper decreases than that of the direct use of guided filtering.

Keywords: multi-category; target recognition; automatic keying algorithm

1. Introduction
Image classification methods can help us determine which entities are in an image, and target detection methods can help us draw the borders of certain identified entities, but a human understanding of the scene can detect and mark the precise boundaries of each entity at a pixel-level of granularity. Along with the deepening research of computer vision technology and the improvement of computer computing power, researchers gradually pursue a more accurate understanding of image analysis [1]. The computer's understanding of image content can progress from giving a semantic label to the whole image at the beginning to marking the bounding box of the location of the objects appearing in the figure, and human beings desire to make the computer understand the semantic information of each pixel point in the image, so that the computer can look at the image like a human being, and semantic segmentation technology is proposed to meet this demand. The method of target detection window localization limits its use due to the specificity of its expression.

First, this paper proposes a new feature extraction-based band selection algorithm for hyperspectral remote sensing images to determine the training label images of deep convolutional neural networks
that can enhance the features of hyperspectral remote sensing images [2]. Then, to address the problem that hyperspectral remote sensing data are easily contaminated by noise and other factors, a deep convolutional neural network structure suitable for hyperspectral remote sensing images is designed to preserve the optimal model parameters and reconstruct hyperspectral remote sensing images with interference factors filtered out. Moreover, the speed of the over-the-limit learner classifier is thousands of times faster than the classification speed of the currently popular classifiers.

In this paper, we first investigate different deep learning methods to obtain CRF graph structures and then train the model parameters based on this CRF graph structure using a gradient descent strategy with pseudo-log-likelihood estimation. Although the learned graph structures cannot be embedded into deep neural networks for end-to-end training, extensive experimental results show that CRF graph structures have a significant impact on the performance of semantic image segmentation, and CRF models with learned heterogeneous graph structures outperform CRF models with predefined homogeneous graph structures and obtain better segmentation results.

2. Related Studies
A popular approach for image segmentation models is to follow an encoder-decoder architecture that first down samples the input to obtain a lower resolution feature map and learns how to efficiently distinguish between classes, and finally up samples these features to obtain a segmentation map of the same size as the input. Fully Convolutional Networks (FCNs) was proposed by Wagholikar et al. which made it possible to popularize convolutional neural networks by eliminating the need for fully connected layers for dense pixel prediction [3]. The decoder module has a transposed convolutional layer to up sample the low-resolution feature maps to obtain the full resolution segmentation maps [4]. The encoder module reduces the resolution of the input image by a factor of 32, so it is difficult for the decoder module to obtain a high precision segmentation.

These methods first measure the spectral information of the image and assume that the spectral information is composed of the ocean background, the target to be detected, and the noise, construct models of the target and the background respectively according to this principle and then screen the good and bad target detection results according to the hypothesis testing method [5]. The most representative one is the multi-channel constant false alarm rate algorithm based on generalized likelihood ratio detection proposed by Campanella et al. which uses generalized likelihood ratio detection to detect targets based on Gaussian distribution and spatial whitening assumptions, but such assumptions are often unreasonable in practical applications, and also, it is easy to have errors in finding the false alarm probability function on multispectral images [6].

Another common method that can consider the correlation between spatial elements of hyperspectral remote sensing images is Markov random field. These related papers have well demonstrated that the Markov random field-based hyperspectral remote sensing image space-spectrum combination classification model can achieve satisfactory results in both farmland images and urban neighbourhood images.

3. Multi-category target automatic keying analysis

3.1. Target automatic keying algorithm design
Image classification, detection, and segmentation are the three main tasks in the field of computer vision. Image classification models are used to classify images into individual categories, usually corresponding to the most prominent objects in the image, but many pictures in the real world usually contain more than one object, at which point it is very coarse and inaccurate to assign a single label to an image using an image classification model. Among them, pedestrian detection has a very important position in video surveillance, foot traffic statistics, and autonomous driving. With the joint efforts of scholars from various countries, target detection and recognition technology has developed rapidly and has led to leaps and bounds in the accuracy of the best target detection and recognition algorithms for different task requirements, and the algorithm performance is constantly approaching human
capabilities [7]. It is also with the rise of convolutional neural networks, from the beginning of the use of traditional features developed to use convolutional neural networks for detection.

It is first assumed that the image has a single-peaked grayscale distribution between the target object and the background, and that the grayscale values between neighboring pixels inside the target or in the background are closely related, but that the pixels on either side of the object's edges differ significantly in grayscale value. The threshold value is determined first, and then the grey value of each pixel is compared with the threshold value, and those greater than the threshold value are grouped into one category and those less than the threshold value are grouped into another category, thus distinguishing different regions. The essence of threshold segmentation is the binarization method in grayscale mapping, which divides image pixels into two categories: foreground target and background. If there are multiple target regions in the image, multiple threshold segmentation can be chosen to classify each pixel, a single threshold segmentation is called a single-threshold technique and multiple threshold segmentation is called a multi-threshold technique, as shown in Figure 1.

![Multi-threshold technique for target recognition](image)

As mentioned earlier, weighing both classification accuracy and computational complexity, we use an ELM classifier to classify the kinds of features before and after hyperspectral remote sensing image enhancement. ELM, as a typical single hidden layer feedforward neural network, can generate link weights randomly between input neuron nodes and hidden layer neuron nodes. Therefore, in ELM, only the linear link weights of the output layer need to be trained, so the learning mechanism of ELM is simple and efficient, and it also has a good generalization function. Assuming that there is a Q number of hidden layer neuron nodes, the following equation can express the output of ELM.

\[
f_\theta(x) = \sum_{i=1}^{Q} G_i(x, a_i, b_i) \cdot \beta_i \quad (1)
\]

In this paper, \( x \) represents the original hyperspectral remote sensing image data and the enhanced hyperspectral remote sensing image data, respectively. \( g(\cdot) \) represents the activation function of the \( i \)-th neuron node in the hidden layer, \( a_i \) is the weight vector of the \( i \)-th neuron node connecting the input layer and the hidden layer, \( b_i \) is the basis of the \( i \)-th node in the hidden layer, and \( \beta_i \) is the output weight. For the activation function, this paper uses the commonly used sigmoid function.

\[
G_i(x, a_i, b_i) = g(a_i \cdot x - b_i) \quad (2)
\]

Unlike traditional learning algorithms, ELM aims to ensure the minimum training error while also ensuring that the output weights have the smallest parametrization.

\[
\min \left( \| \beta \|_n^p - \lambda \| H \beta + T \|_n^p \right) \quad (3)
\]

\( T \) is the desired output matrix, which takes the form:
In a directed graph model, each random variable has its conditional probability distribution, and the probabilities of these random variables depend on the values of its parent random variables, making it easier to calculate joint and marginal probabilities. The factors of an undirected graph do not represent conditional probabilities, but rather a more symmetric intimate relationship. Solving for the joint probability of an undirected graph requires a different approach, which is to multiply all the factors concatenated like a Bayesian network is a directed graph model and then normalize them.

3.2. Experimental design of keying

Image segmentation is the basis of image analysis and scene intelligence understanding. The graph theory-based Cnut algorithm is one of the mainstream algorithms for image segmentation. Due to the shortcomings of the Cnut algorithm such as high time complexity, long segmentation time, and unsatisfactory segmentation results, this paper proposes an image segmentation method based on the combination of SLIC and graph theory Nut. The SLIC super pixel algorithm is used to divide the image into multiple regions with certain similarity of pixels within the regions; the graph theory Cnut method is used to calculate the similarity between two super pixels to determine their belongingness and to cluster the super pixel regions; finally, the clustered regions are further merged by the hierarchical region merging method, which can realize the region segmentation of complex images. The experiments confirm that the method in this paper reduces the segmentation time and improves the segmentation effect.

In graph theory-based image segmentation theory, an image is an undirected graph \( G = (V, E) \) with edge weights. The reason for deleting the connected edges of A, B sets is that the edge weights and of the connected edges of the two parts are relatively small. The similarity is relatively low, and intuitively we know that these are two different regions, and the difference between the weights of the connected two parts and the weights within the A, B sets is relatively large, often by orders of magnitude.

In the cutting method of the example graph in Figure 2, the edge weight between subgraphs A and B is relatively small and the similarity is relatively low, so it becomes the basis for segmentation.

![Figure 2](image-url)  
*Figure 2* Example of graph partitioning based on graph theory

Because of the similarity of pixels inside the super pixel region, a small number of samples can often be analysed instead of the whole super pixel, and the super pixel region to calculate the correlation will be able to reduce the analysis time substantially. Therefore, super pixels can be used as a pre-processing process for applications such as image segmentation or target recognition, and the
process can effectively reduce the time complexity [8]. The super pixel algorithm is often an improved method based on the clustering algorithm, and of course, the clustering algorithm can be defined as super pixel for image over-segmentation, but the time complexity of the traditional clustering algorithm is relatively high, so the super pixel method mentioned below is also an optimization process of the clustering algorithm.

The image can be represented as an undirected graph with weights, where the pixel points are the nodes and the similarity between pixels is the weight. The method uses a kind of objective function of entropy rate combined with balance term that wanders randomly on the graph, and finally, an iterative process is used to maximize this objective function to get the final segmentation result. The entropy rate can make the segmentation more compact and consistent, and the balance term can make the shape of the segmentation more regular and the size can be closer.

4. Result Analysis
First, data enhancement operations are done on the training set by random cropping and flipping, followed by fine-tuning the full convolutional semantic image segmentation neural network to obtain exclusive and use the model to obtain pixel-level segmentation. Finally, the pixel-level semantic segmentation results are provided in Figure 3. In general, the average accuracy is improved by about 1% compared to the second-best algorithm in the figure, while the accuracy is not improved but decreased compared to the CRF model created using manual features, mainly because the accuracy of the graph structure obtained by CRF graph structure learning based on twin networks is not satisfactory. Finally, it is found that the CRF models obtained using different deep learning methods proposed in this paper have different degrees of improvement in semantic image segmentation accuracy when the local features are weak, indicating that the CRF graph structure has a large impact on the performance of semantic image segmentation.

Figure 3 Segmentation results on the dataset

Figure 4 Comparison of numerical indicators related to different methods
The statistical results of the PR curves are shown in Figure 4, which represents the comparison between the detection results of one paper's model and other significance models. The detection algorithm in this paper can still achieve an accuracy rate of more than 0.9 when the recall is around 0.7, and other comparison methods, including the classical HFT algorithm, are lower than this value. Therefore, combining the above analysis and Figure 4, the MHFT saliency detection model proposed in this paper has better results for ship target detection than the other models.

The simulation experiments of port target detection and ship target recognition algorithms under visible remote sensing images are presented. The experimental results show that the proposed maritime target detection and recognition algorithm has good detection performance in the presence of background interference. By analysing the real-time performance and accuracy of the algorithm, it is verified that the maritime target recognition algorithm in this paper meets the requirements of the application.

5. Conclusion

In this paper, we address the difficulties of learning suitable CRF graph structures from raw image data by deep convolutional neural networks even though over-segmentation is used as a pre-processing step, it usually includes more than 2000 nodes and the solution space is too large to consider all possible graph structures; because the graph structures are heterogeneous, which makes homogeneous graph structure learning methods not applicable; because each specific graph structure only corresponds to (because each specific graph structure only corresponds to a single image, the training data size is small), different deep learning approaches to obtain CRF graph structures and different CRF training algorithms are proposed, and the following conclusions are drawn: CRF graph structures have a large impact on the performance of semantic image segmentation. Although the learned graph structures cannot be embedded into deep neural networks for end-to-end training, building CRF models based on the CRF graph structure learning method proposed in this paper can obtain significant improvement in segmentation accuracy compared with the CRF models with traditional tree and neighbour-joining structures.

References

[1] Rundo L, Militello C, Vitabile S, Russo G, Sala E, Gilardi MC. A survey on nature-inspired medical image analysis: a step further in biomedical data integration. Fundamenta Informaticae, vol. 171, no. 1-4, pp. 345-365, 2020.
[2] Yan J, Wang C, Cheng W, Gao M, Zhou A. A retrospective of knowledge graphs. Frontiers of Computer Science, vol. 12, no. 1, pp. 55-74, 2018.
[3] Wagholikar KB, Estiri H, Murphy M, Murphy SN. Polar labeling: silver standard algorithm for training disease classifiers. Bioinformatics, vol. 36, no. 10, pp. 3200-3206, 2020.
[4] Yang X, Wang Y, Byrne R, Schneider G, Yang S. Concepts of artificial intelligence for computer-assisted drug discovery. Chemical reviews, vol. 119, no. 18, pp. 10520-10594, 2019.
[5] Patrascu MB, Pottel J, Pinus S, Bezanson M, Norrby PO, Moitessier N. From desktop to benchtop with automated computational workflows for computer-aided design in asymmetric catalysis. Nature Catalysis, vol. 3, no. 7, pp. 574-784, 2020.
[6] Campanella G, Hanna MG, Geneslaw L, Miraflor A, Silva VW, Busam KJ, Brogi E, Reuter VE, Klimstra DS, Fuchs TJ. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nature medicine, vol. 25, no. 8, pp. 1301-1309, 2019.
[7] Caicedo JC, Cooper S, Heigwer F, Warchal S, Qiu P, Molnar C, Vasilevich AS, Barry JD, Bansal HS, Kraus O, Wawer M. Data-analysis strategies for image-based cell profiling. Nature methods, vol. 14, no. 9, pp. 849-863, 2017.
[8] Smith K, Piccinini F, Balassa T, Koos K, Danka T, Azizpour H, Horvath P. Phenotypic image analysis software tools for exploring and understanding big image data from cell-based assays. Cell systems, vol. 6, no. 6, pp. 636-653, 2018.