Aircraft Type Recognition in Remote Sensing Images using Deep CNN

S.Dialine[1] K.P.Ajitha Gladis[2]
[1]PG Student [2]Professor
[1,2]Department of Computer Science and Engineering
[1,2]C.S.I Institute of Technology, Tamil Nadu, India

Abstract— Aircraft type recognition is a meaningful task in remote sensing images. It remains challenging due to the difficulty of obtaining appropriate representation of aircrafts for recognition. To tackle these problems, a novel and robust aircraft type recognition framework based on Convolutional Neural Networks (CNN)s has been proposed in this paper. First, the proposed system focus on obtaining more detailed aircraft segmentation results without using refined annotations in the training stage. A convolutional encoder decoder network is designed to capture coarse segmentations. Then, a conditional random field (CRF) is used to refine the segmentation results. Second, to acquire more accurate direction estimations, the direction estimation is transformed to a keypoints' detection task, and a convolutional regression network is built to locate the positions of aircraft's keypoints. Besides, a multirotation refinement (MRR) method is proposed to further improve the precision of keypoints' positions. At last, the template matching procedure is carefully designed to recognize aircrafts based on the direction estimations and segmentation results. The similarity between segmentation results and templates is evaluated by adopting the Intersection Over Union (IOU) measure. The proposed framework takes advantage of both shape and scale information of aircrafts for recognition. Experimental results show that the proposed method outperforms the state-of-the-art methods and can achieve 95.6% accuracy on the challenging data set.

Key words: Aircraft Type Recognition, Deep Convolutional Neural Networks (CNNs), Image Segmentation, Keypoints Detection

I. INTRODUCTION

AIRCRAFT type recognition is one among the meaningful tasks in remote sensing image interpretation. However, since some types of aircrafts are similar to each other different types can only be distinguished by slight differences in shape and scale, which makes aircraft recognition a challenging task. Besides, complex backgrounds, shadows, and illumination changes in images make this problem more difficult. Many effective methods have been proposed to handle the aircraft type recognition problem. Several methods classified aircrafts directly based on image features [11], [12], [3]. Diao et al. [11] proposed a method based on deep belief networks for classification. Fang et al. [12] built an aircraft recognition method based on moment invariants and back propagation neural networks. However, training these models requires a large amount of data in each type. Besides, these methods are sensitive to the distribution of data, and the imbalanced data distribution will lead to bad recognition accuracy. However, existing methods still suffer from two main limitations:

1) The rough shape representations of these methods lose lots of details, which are essential to distinguish aircrafts.
2) These methods usually need direction estimation to align aircrafts or segmentations, but limited by the capacity of the methods, the predicted directions are not accurate enough.

Inspired by the great success of deep convolutional neural networks (CNNs) in image classification tasks [4], [8],recently the researchers solve the image segmentation problem by converting existing CNN architectures to pixel wise labelling models, such as a fully convolutional network (FCN) [6], DeepLab [1], and DeconvNet [7]. Although these methods achieve outstanding segmentation performance, they have to be trained on large-scale data set with refined pixel wise annotations, which are quite expensive to be acquired. To tackle these problems, a novel and robust aircraft type recognition framework based on CNNs is presented in this paper. First of all, we focus on obtaining more detailed aircraft segmentation results without using refined annotations in the training stage. A convolutional encoder–decoder network is designed to capture coarse segmentations. Then, we adopt a conditional random field (CRF) [5] to refine the segmentation results. Second, to acquire more accurate direction estimations, we transform direction estimation to a keypoints' detection task, and build a convolutional regression network to locate the positions of aircraft's keypoints.

Besides, a multirotation refinement (MRR) method is proposed to further improve the precision of keypoints' positions. At last, we carefully design the template matching procedure to recognize aircrafts based on the direction estimations and segmentation results. Experiments demonstrate the effectiveness of the proposed framework.

The main advantages of the proposed framework are as follows.

1) The proposed method is not sensitive to the imbalanced type distribution of data, and it only requires keypoints' annotations to train both segmentation and keypoints' detection models. Neither pixel wise segmentation annotations nor aircraft type labels are demanded.
2) It has strong expansibility. We do not need to retrain the model if we want to recognize a new type of aircraft.
3) It provides more detailed shape and scale information for recognition, which improves the recognition accuracy significantly.

The rest of this paper is structured as follows; section II describes the related works and the existing systems. Section III describes the proposed system and discusses its different components in detail. Implementation and results are provided in section IV. Section V deals with the performance evaluation. A conclusion is drawn in section VI.

II. RELATED WORKS

This section deals with analyzing the existing system. It is the process of gathering information and diagnosing the problems in the existing system, then suggesting an idea for the improvement of the existing system.

In [9] D. Wang et al. had suggested a model to handle the aircraft type recognition problem. In [12] A. Zhao et al. adopted template matching based methods for aircraft type recognition. These methods usually get aircrafts' binary segmentations or shape representations at first, and compare them with standard templates in certain ways. In [10] Wu et al. had suggested a model jigsaw reconstruction method to obtain aircrafts' shape extractions. In [12] Zhao et al. transformed recognition to a keypoints' detection problem, and adopted keypoints template matching to recognize aircrafts. In [4] K. He et al. had suggested a model inspired by the great success of deep convolutional neural networks (CNNs) in image classification tasks. In [8] K. Simonyan et al. researchers solve the image segmentation problem by converting existing CNN architectures to pixel wise labelling models. In [6] J. Long et al. had suggested a model a fully convolutional network (FCN), In [1] L. C. Chen at al. had suggested a DeepLab, and in [7] H. Noh at al. had suggested a model a DeconvNet. Although these methods achieve outstanding segmentation performance, they have to be trained on large-scale data set with refined pixel wise annotations, which are quite expensive to be acquired. In [5] P. Krahenbuhl at al. adopts a conditional random field (CRF) to refine the segmentation results.

III. PROPOSED SYSTEM

This section describes the problem, the proposed system. Also the different components in the proposed system are discussed here.

A. Problem Description

Aircraft type recognition is one of the meaningful tasks in remote sensing image interpretation. Since some types of aircrafts are similar to each other, different types can only be distinguished by slight differences in shape and scale, which makes aircraft recognition a challenging task. Besides, complex backgrounds, shadows, and illumination changes in images make this problem more difficult. Many effective methods have been adopted to handle the aircraft type recognition problem. Several methods classified aircrafts based on image features, or moment invariants and back propagation neural networks. However, training these models requires a large amount of data in each type, which are quite expensive to be acquired. Besides, these methods are sensitive to the distribution of data, and the imbalanced data distribution will lead to bad recognition accuracy. Hence it is required to develop a robust and cost effective framework for aircraft type recognition.

B. Proposed Framework

The proposed framework consists of three parts: an aircraft segmentation model, direction estimation and bounding box localization model, and a template matching model. To recognize an aircraft, the aircraft's binary segmentation is got from the image at first. Before segmentation initially the output image is pre-processed to remove the noise if any; the proposed system uses median filter and LADCT filter to denoise the image. Then, the segmentation result will be rotated upright and placed in the middle of a fixed-size background image. At last, the similarity between the segmentation results with the standard aircraft templates is estimated to identify the aircraft's type.

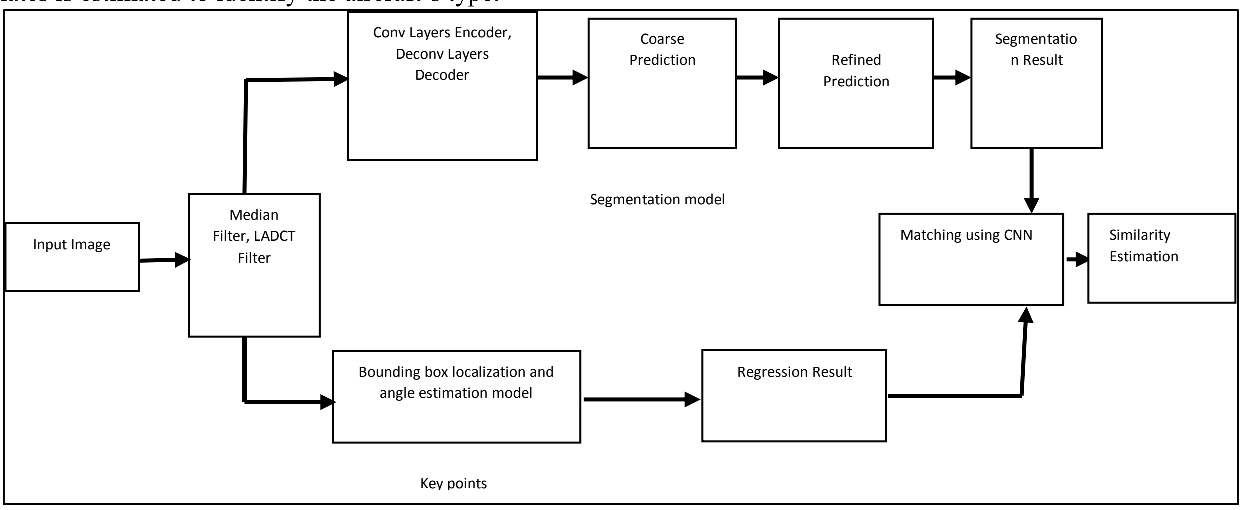

Fig. 1: Architecture of the aircraft recognition framework

The architecture of the aircraft recognition framework is depicted in Fig.1, showing the various components involved in the proposed aircraft recognition system.

Image is taken as an input. The input image is being pre-processed i.e. Image resizing and filtering. The filtering is done by using LADCT filter and median filter. Then the image is processed via the segmentation model. The segmentation model consists of encoder–decoder network to predict coarse segmentation results. Then, a fully connected CRF is adopted to refine the segmentation results. Meanwhile the filtered image is inputted into bounding box localization and angle estimation to detect the keypoints. Both the image i.e. Segmentation & regression result are combined & matched by using CNN. Then the type of aircraft is detected. We carefully design methods to improve the performance of each step. The proposed framework is elaborately described in the following subsections.

1) Pre-Processing

The pre-processing of image aims at selectively removing the redundancy present in captured images without affecting the details that play a key role in the overall process. This involves the following basic step:

a) Filtering

1) Median Filter

Median filtering is a nonlinear method used to remove noise from images. It is widely used as it is very effective for removing noise while preserving edges. Particularly it is effective at removing 'salt and pepper' type noise. The median filter works by moving through the image pixel by pixel, replacing each value with the median value of neighbouring pixels. The pattern of neighbours is called the "window", which slides, pixel by pixel over the entire image 2 pixels, over the entire image. The median is calculated by first sorting all the pixel values from the window in numerical order, and then replacing the pixel being considered with the middle (median) pixel value.

2) LADCT

An improved version of the Locally Adaptive Discrete Cosine Transform (LADCT) filter is used for filtering in the proposed system. The LADCT filter operates on sliding blocks (local action filter), which could offer more information about the local effect of noise on the image in a better fashion. The LADCT filter for a type of noise that contaminates the signal through a multiplicative rule. They use a sliding block window to obtain de-noised estimates of neighbouring and overlapping blocks. The multiple estimates are then averaged to suppress artifacts caused by undershoots and overshoots around the highly textured regions. The threshold for each block depends on the local mean of the block and the local noise variance.

2) Aircraft Segmentation

Different from objects in natural images, aircrafts in remote sensing images have inerratic shapes and discriminative colours compared with backgrounds, which provide important information for accurate segmentation. Therefore, we take advantage of these characteristics to design the segmentation method and avoid the requirement of refined annotations. First of all an encoder–decoder network is designed to predict coarse segmentation results. Then, a fully connected CRF is adopted to refine the segmentation results.

a) Coarse Segmentation Network

The data set used to train the network is only annotated with eight aircraft keypoints; choose these keypoints, because they can represent the basic shape of an aircraft. The segmentation ground truths by generating polygons using eight key points. All pixels in the polygons are regarded as the aircraft class, while other pixels are regarded as the background class. Build the aircraft segmentation network based on the architecture of DeconvNet. The network consists of two parts: a convolutional encoder network and a deconvolutional decoder network. The encoder network aims to extract features from the input image. On the contrary, the decoder network takes the features as input, and outputs the segmentation score map, which has the same size as the input image. The score map indicates the probability of each pixel that belongs to the aircraft or background, and it is followed by a softmax layer to decide the final class of each pixel.

To build the encoder network, employ the first 13 convolutional layers of VGG-16 net and remove all of the fully connected layers. Train the encoder network and decoder network end-to-end by minimizing cross entropy loss. Since an aircraft only takes up a small proportion of pixels in an image, the imbalanced distribution of data increases the difficulty of training the network. Thus, a weighted loss function is designed to handle this problem. The weighted loss function L can be formulated as in equation (1)

$$L = \frac{1}{N}\sum_{n=1}^{N} \omega_0 p_n \log \hat{p}_n + (1 - \hat{p}_n)\log(1 - \hat{p}_n) \qquad (1)$$

Where N is the amount of pixels in this batch and we set batch size to 4 in experiments. Pins the ground truth and ˆpn is the prediction of the nth pixel. In the training data set, only 5.16% pixels belong to the aircraft class. So, we set ω0 = 0.95 and ω1 = 0.05 in experiments.

b) CRF Refinement

To acquire accurate aircraft segmentation results, a fully connected CRF is applied to refine the predictions of the segmentation network, since it has shown outstanding ability to acquire detailed boundaries. We use the score map generated by the segmentation network to build the energy function of the CRF can be formulated as in equation (2)

$$E(x) = \sum_i \theta_i(x_i) + \sum_{ij} \theta_{ij}(x_i, x_j) \qquad (2)$$

All of the pixels' classes in the input images are reassigned based on the energy function. The CRF makes full use of the colour information to classify each pixel. A lot of shape details are presented after adopting CRF.

3) Direction Estimation and Bounding Box Localization

Transform the aircraft direction estimation and bounding box localization task to a keypoints' detection task. By locating the positions of aircraft's four keypoints, i.e., nose, tail, and left and right wingtips we can easily obtain the direction and bounding box at the same time. Aircraft's large direction variations increase the difficulty of locating keypoints accurately. To tackle this problem and further improve the precision of results, an MRR method is proposed in this paper. At test time, we will get aircraft's keypoints following the procedure given below.

1) Rotate and flip the image to get multiple images.
2) Input these images into the model separately, and the network will output the keypoints' prediction for each image.
3) Rotate or flip the predictions back based on the corresponding augmentation method used before.
4) Merge the predictions to get the final result. The way to merge the predictions is as follows.

Given the good capacity of the model, there are only a few outliers which are far away from the ground truth, while most of predictions are just around the real keypoint. Thus, instead of simply calculating the mean value of predictions, an iterative cluster-based algorithm is proposed to merge the results and eliminate the bad predictions' influence. Specifically, for each aircraft keypoint, we estimate whether the diameter of the minimal circle which covers all predictions is less than is the desired error tolerance. If so, the position of this keypoint is the mean value of these predictions.

Otherwise, K-means algorithm will be used to cluster these predictions into two groups. The cluster that contains fewer predictions will be regarded as an outlier group and removed. If the two clusters have the same amount of predictions, the one which has larger will be removed. We will keep removing outliers in this way until all of remaining predictions are in the desired circle or there are only two predictions left. The final result is the mean value of the remaining predictions.

4) Aircraft Type Recognition

The CNN is used to recognition the aircraft type. For the convenience of comparing the segmentation result with the standard templates, pre-process the segmentation result at first. Specifically, calculate the aircraft's angle referring to the horizontal line with the predicted positions of nose and tail, and rotate the segmentation result upright based on the angle. Then, use all of the four predicted keypoints to obtain the minimum bounding box of the rotated segmentation result. All of the pixels outside the bounding box will be abandoned. Like the templates, the segmentation result in the bounding box will be placed in the middle of an image, which has a pure black background. Finally, the result is compare to the database to get the accurate output.

IV. IMPLEMENTATION AND RESULTS

The proposed system is implemented by using MAT Lab. A brief description of implementation is given below. First, an image is taken as the input. Then the image is pre-processed by filtering. Filtering is done by two methods LADCT and median filter. Second the filtered image is processed by segmentation model. In segmentation model, image is coarse predicted and refined predicted, and then the segmentation result is given out. Third the deep CNN is created, and the images are trained. Then the network for particular input image is simulated. Finally by matching process the type of the aircraft is recognized.

A few screenshots of the recognition results obtained are shown in Fig.2 and Fig.3. Fig.2 shows the recognition result using median filter. Fig.3 shows the recognition result using LADCT filter.

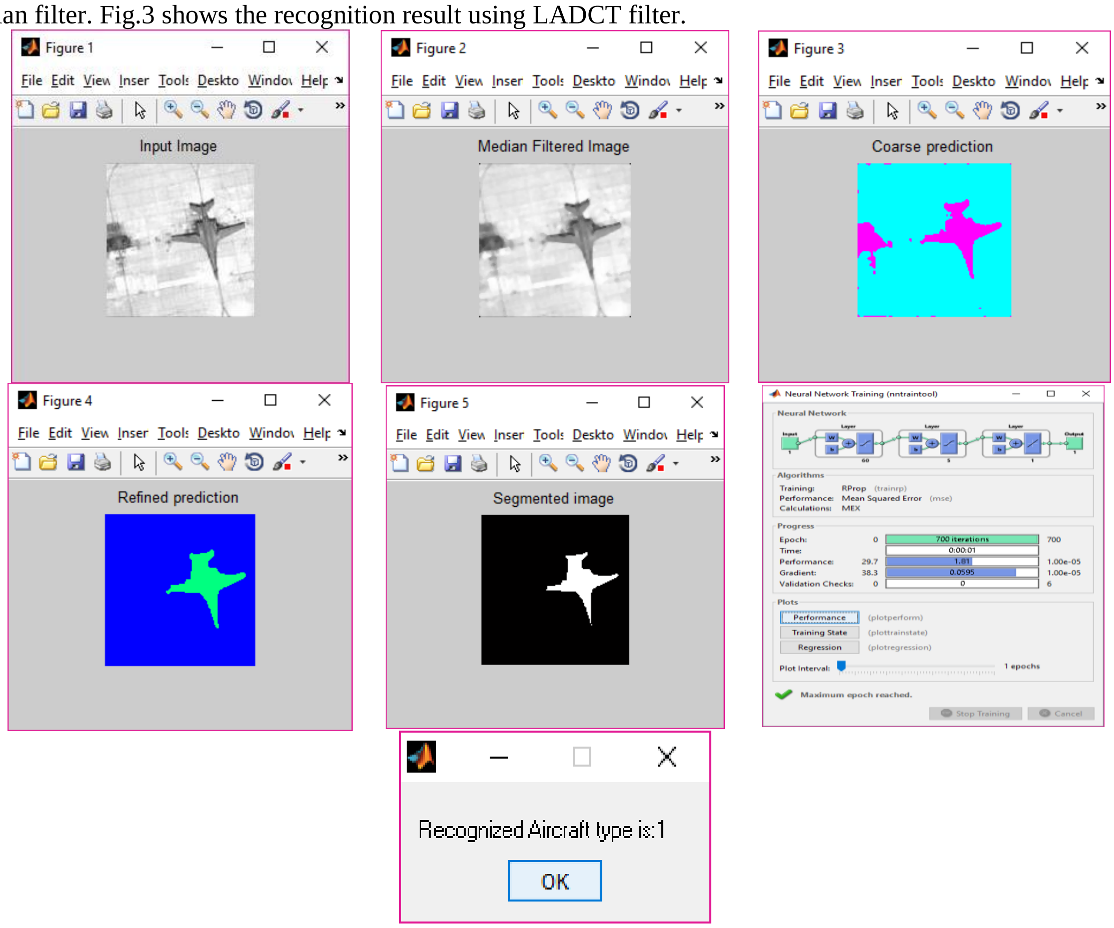

Fig. 2: The recognition result using median filter.

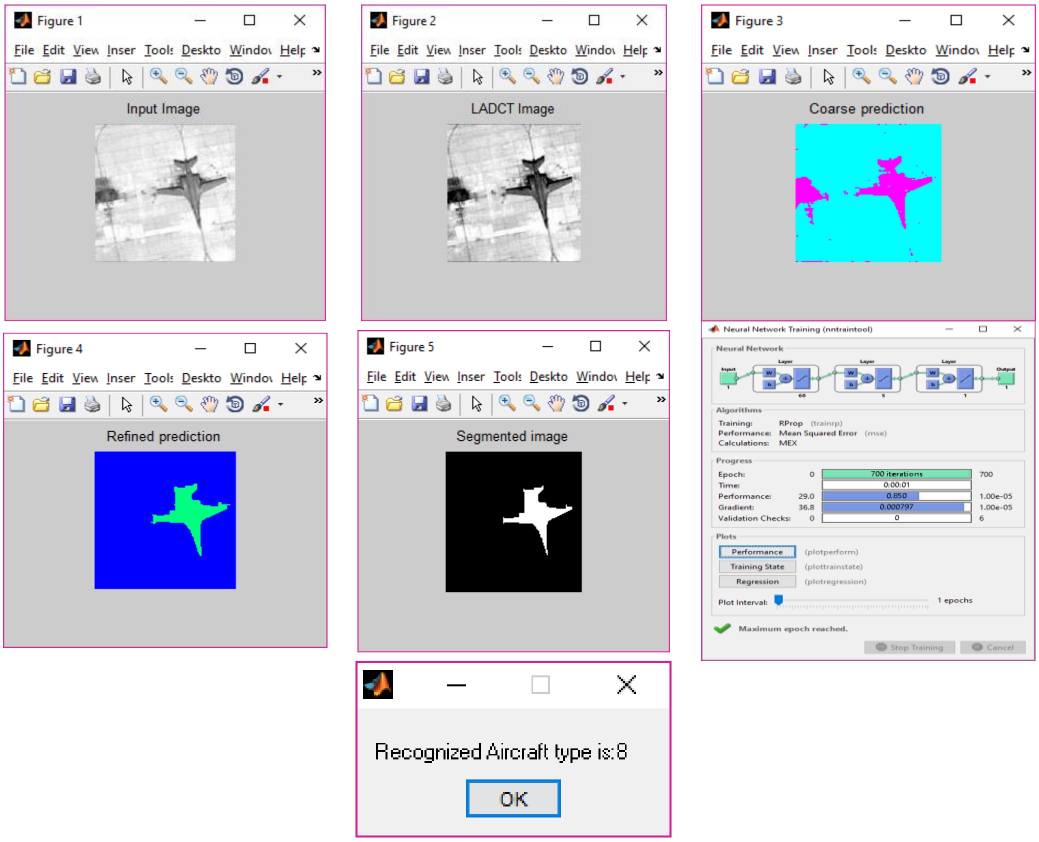

Fig. 3: The recognition result using LADCT filter.

We evaluate the performance of the proposed aircraft recognition method using the IOU parameter. We calculate the mean IOU between the segmentation predictions and the standard templates.

Intersection over Union

We adopt intersection over union (IOU) for the measurement of similarity. The IOU can effectively distinguish different aircrafts. Also IOU considers every pixel; it can classify aircrafts correctly even if the segmentation is partially fragmentary or the direction is slightly slanting. For each pixel in the predicted segmentation result, we evaluate whether it is classified correctly by comparing it with templates. The IOU is defined as

$$\boldsymbol{IOU} = \frac{\boldsymbol{True\ Positive}}{\boldsymbol{True\ Positive + False\ Positive + False\ Negative}} \tag{3}$$

- True positives (TP): These refer to the positive tuples that were correctly labelled by the classifier.
- True negatives (TN): These are the negative tuples that were correctly labelled by the classifier..
- False positives (FP): These are the negative tuples that were incorrectly labelled as positive (e.g., tuples of class buys computer = no for which the classifier predicted buys computer= yes).
- False negatives (FN): These are the positive tuples that were mislabelled as negative (e.g., tuples of class buys computer = yes for which the classifier predicted buys computer= no).

Input Image	Median filter	LADCT
1	255	46
2	232	1
3	173	2
4	102	3
5	236	1
6	138	1
7	255	1
8	255	1
9	142	1
10	255	35

Table 1: Intersection over union

The above Table 1 shows the intersection over union values obtained for various test images. In the testing stage, we calculate the IOU between the segmentation result and the standard templates. The testing image which has the highest IOU with the segmentation result will be classified into the same type as the template. Fig.4 shows the graphical representation of IOU value.

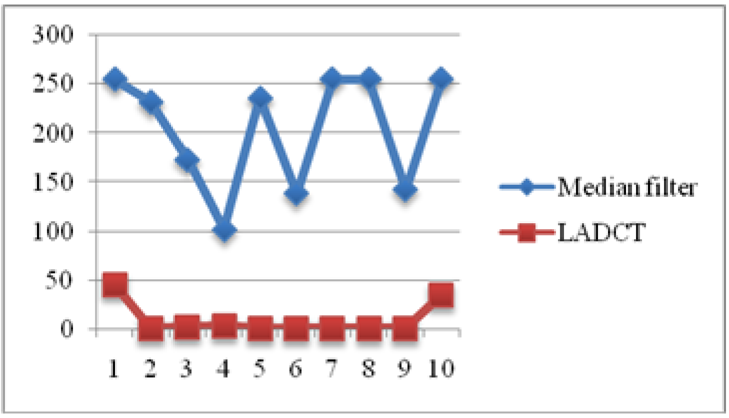

Fig. 4: Graphical representation of IOU value

V. CONCLUSION

In this paper, a novel aircraft type recognition framework is proposed to identify aircrafts precisely. An aircraft segmentation network is proposed and a keypoints' detection network is obtained accurately and detailed representations of aircrafts. Then, template matching is adopted to recognize aircrafts' types. The proposed method only requires keypoints' annotations in the training stage. Besides, it has good expansibility, since it does not need to retrain the segmentation and keypoints' detection model if a new type of aircraft needs to be recognized. Designing methods are done carefully to improve the performance at each step. Experimental results demonstrated the effectiveness of the proposed method.

Future researches may be directed to investigate the classifier types for further improvement on the aircraft identification process; applying the Neuro-fuzzy techniques may reduces the training time. The automatic identification of aircraft by soft-computing techniques, such as fuzzy systems, will give better identification results than when performed by subjective manual approaches.

REFERENCES

[1] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," IEEE Trans. Pattern Anal. Mach. Intell., to be published.
[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2016, pp. 770 -778.
[3] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected CRFs with Gaussian edge potentials," in Proc. Adv. Neural Inf. Process. Syst., 2011, pp. 109 -117.
[4] 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.
[5] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in Proc. IEEE Int. Conf. Comput. Vis., Dec. 2015, pp. 1520–1528.
[6] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," unpublished paper, 2014. [Online]. Available: https://arxiv.org/abs/1409.1556
[7] D. Wang, X. He, Z. Wei, and H. Yu, "A method of aircraft image target recognition based on modified PCA features and SVM," in Proc. Int. Conf. Electron. Meas. Instrum., Aug. 2009, pp. 4-177–4-181.
[8] Q. Wu, H. Sun, X. Sun, D. Zhang, K. Fu, and H. Wang, "Aircraft recognition in high-resolution optical satellite remote sensing images," IEEE Geosci. Remote Sens. Lett., vol. 12, no. 1, pp. 112-116, Jan. 2015.
[9] C. Xu and H. Duan, "Artificial bee colony (ABC) optimized edge potential function (EPF) approach to target recognition for low-altitude aircraft," Pattern Recognit. Lett., vol. 31, no. 13, pp. 1759-1772, 2010.
[10] Zhao et al., "Aircraft recognition based on landmark detection in remote sensing images," IEEE Geosci. Remote Sens. Lett., vol. 14, no. 8, pp. 1413–1417, Aug. 2017.
[11] W. Diao, X. Sun, F. Dou, M. Yan, H. Wang, and K. Fu, "Object recognition in remote sensing images using sparse deep belief networks,"Remote Sens. Lett., vol. 6, no. 10, pp. 745–754, 2015.
[12] Z. Fang, G. Yao, and Y. Zhang, "Target recognition of aircraft based on moment invariants and BP neural network," in Proc. World Autom. Congr., Jun. 2012, pp. 1–5.