A framework to derive geospatial attributes for aircraft type recognition in large-scale remote sensing images

Rajeshreddy Datla *,†, Chalavadi Vishnu †, C Krishna Mohan †

*Advanced Data Processing Research Institute (ADRIN), Secunderabad, India

†Department of Computer Science and Engineering, Indian Institute of Technology Hyderabad, India

ABSTRACT

Aircraft type recognition remains challenging, due to their tiny sizes and geometric distortions in large-scale panchromatic satellite images. This paper proposes a framework for aircraft type recognition by focusing on shape preservation, spatial transformations, and geospatial attributes derivation. First, we construct an aircraft segmentation model to obtain masks representing the shape of aircrafts by employing a learnable shape-preserved and deformable network in the mask RCNN architecture. Then, the orientation of the segmented aircrafts is determined by estimating the symmetrical axes using their gradient information. Besides template matching, we derive the length and width of aircrafts using the geotagged information of images to further categorize the types of aircrafts. Also, we present an effective inferencing mechanism to overcome the issue of partial detection or missing aircrafts in large-scale images. The efficacy of the proposed framework is demonstrated on large-scale panchromatic images with ground sampling distances of 0.65m (C2S).

Keywords: Geospatial attributes, aircraft type, remote sensing images, instance segmentation, partial objects.

1. INTRODUCTION

Remote sensing images with high spatial resolution are able to capture the objects with their structural properties at finer level often covering a large area on the earth. The utilization of these images is greatly promoted in the development of both civil and strategic applications. In general, the analysis of tiny objects like aircrafts in large-scale high resolution panchromatic images is performed through detection [1]. Undoubtedly, finer categorization or new findings of such objects will help in improving the level of interpretation and decision making process. Aircraft type recognition is one such task that identifies the aircrafts along with their type. However, the lack of spectral information in the panchromatic images exhibits difficulty in separating the objects from their background. In addition, some issues such as shadow of aircraft, aero bridge attachment, and oil spills on the ground pose challenges in deriving the precise shape of aircrafts. Also, the presence of inherent geometric distortions in the remote sensing images makes the type recognition task further complex.

A typical C2S [2] image with a ground sampling distance (GSD) of 0.65m covering a length of 28km with 10.2km swath will result in 790 mega pixels $\approx$1.5 GB. These kind of large-scale images are not viable for inferencing directly due to imperceptibility of tiny objects and limitation on the computation capability. Apparently, tiles obtained by partitioning the large-scale image as input for inferencing would manifest partial objects at the borders of the tiles. The issues, such as partial detection or complete object missing, besides complex background necessitate an effective mechanism for inferencing the large-scale images.

Existing methods for aircraft type recognition in remote sensing images have explored the use of image features [3, 4, 5], template matching [6, 7, 8], and deep learning concepts [9, 10]. The convolutional neural networks (CNN) in most of the existing instance segmentation methods [11, 12, 13, 13, 15, 16] are not effective in handling the geometric distortions and influenced by shadows, aero bridges, etc. To overcome these issues, the proposed framework focuses on handling distortions using spatial transformations, preserving shape of masks with fusion blocks, and deriving geospatial attributes by combining geotagged information. We construct an aircraft segmentation model by employing deformable convolutions in mask R-CNN framework [12] to exploit aircraft shape information thereby to improve localization accuracy at mask-level. The predicted aircraft masks are used to determine the type of an aircraft by a simple template matching process. We demonstrate the efficacy of the proposed framework on C2S [2] panchromatic images. We also present an inferencing mechanism to recognize the types of aircrafts effectively in large-scale remote sensing images.

The main contributions of this paper are:

- We construct an aircraft instance segmentation model on C2S panchromatic images by employing deformable convolutions.

- Developed a method to generate a shapefile by geo-tagging the segmented aircrafts with the help of the metadata of image for better interoperability.

- A mechanism to derive spatial attributes of the aircrafts such as their type, orientation, length, and width is presented.

- An inferencing mechanism is established to derive the spatial attributes of the segmented aircrafts effectively in large-scale remote sensing images.

The rest of the paper is organized as follows. Sections 2 presents the proposed framework for aircraft type recognition in large-scale panchromatic remote sensing images. The experimental results and their analysis are provided in Section 3. Finally, Section 4 concludes this paper.

2.   PROPOSED FRAMEWORK FOR AIRCRAFT TYPE RECOGNITION

Fig. 1 shows the proposed framework for aircraft type recognition in large-scale remote sensing images. It consists of two phases, namely, training and inference phase. In training phase, we predict the masks representing the shape of aircrafts using aircraft instance segmentation model, which is described in Section 2.1. We also devise an inferencing mechanism to ensure no partial detections or objects missing due to partitioning of large-scale images. Finally, the predicted masks and the derived geo-spatial attributes such as length and width of aircrafts are used for type recognition.
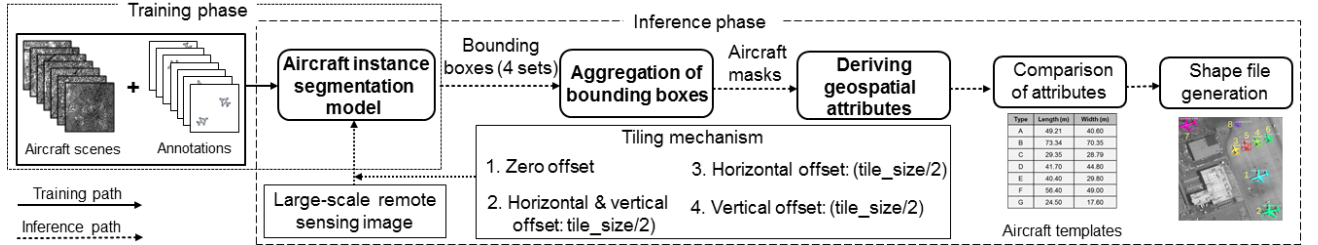

Figure 1. Block diagram of the proposed framework for aircraft type recognition using geospatial attributes.

2.1 Aircraft segmentation model

We employ deformable convolutions with fusion blocks to improve the mask head in the construction of an effective aircraft instance segmentation model, which is illustrated in Fig. 2.

2.1.1 Deformable convolutions

Assume that w_l and z_l represent the weight and offset for l^{th} location, respectively, for a convolutional kernel of L sampling locations. A 3×3 convolutional kernel with dilation 1 is defined $z_l \in \{(-1,1),(-1,0),\cdots,(1,1)\}$ and $L = 9$. Let $f(z)$ and $g(z)$ denote the features at location z from the input feature maps f and output feature maps g, respectively. The deformable convolution [17] can be formulated as

$$g(z) = \sum_{l=1}^{L} w_l \cdot f(z + z_l + \Delta z_l). \tag{1}$$

Here, Δz_l is the learnable offset for the lth location. These deformable convolutions capture the spatial transformations in the objects by learning the sampling grid along with location offsets. However, the earlier feature maps along with additional convolutional layers determine these offsets. Hence, the deformation can be adaptively conditioned locally and densely on input features.
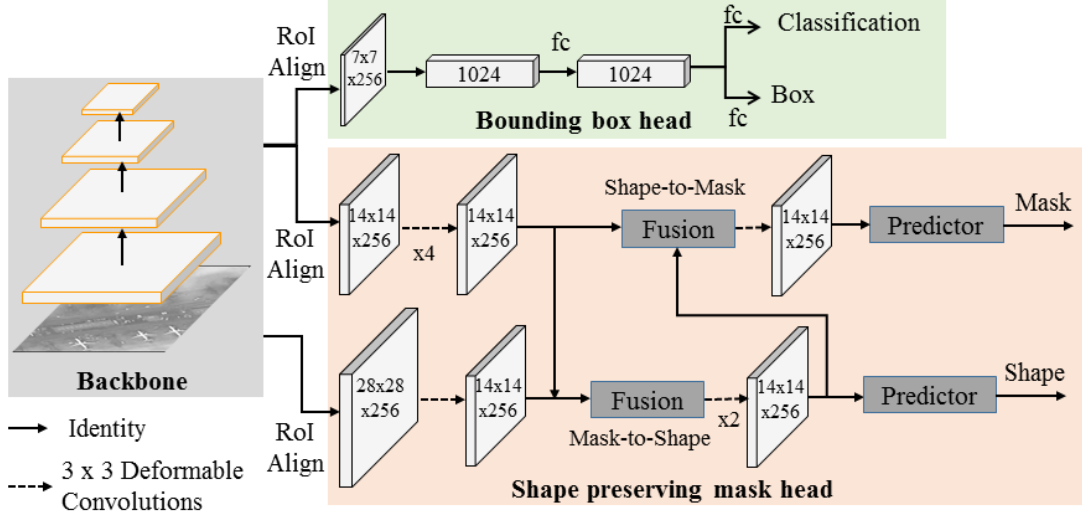
Figure 2. Aircraft instance segmentation model by employing deformable convolutions for predicting precise masks.

2.1.2 Shape preserving mask head

Let R_s and R_m be the region of interest features for shape prediction and mask prediction, respectively. We perform RoIAlign by setting the larger resolution for R_s than R_m to have better spatial information for shape prediction. Then, the obtained $\tilde{R}_s$ (same resolution as R_m) after downsampling R_s with a strided 3×3 deformable convolution is used in feature fusion as shown in Fig. 2. The R_m features are fed to four consecutive 3×3 deformable convolutions to get mask features F_m. Subsequently, shape features $\tilde{R}_s$ are fused with F_m and fed to 2 consecutive 3×3 deformable convolutions.

a) Mask-to-shape fusion: F_m containing pixel-wise object information is integrated with $\tilde{R}_s$ to predict object shapes. The fusion block is represented as

$$F_s = \varphi(F_s) + \tilde{R}_s, \tag{2}$$

where, F_s and φ denote the output shape features and 1×1 convolution, respectively.

b) Shape-to-mask fusion: The fusion of final shape features with mask features enriches the mask features and guides for better mask prediction. This block is identical to mask-to-shape fusion.

c) Shape loss function: We formulate the shape loss function Ψ_s as

$$\Psi_s(q_s, r_s) = \Psi_{Dice}(q_s, r_s) + \lambda \Psi_{BCE}(q_s, r_s), \tag{3}$$

Where $q_s \in \mathbb{R}^{\alpha \times \beta}$ and $r_s \in \mathbb{R}^{\alpha \times \beta}$ represent the predicted shape and ground truth of an aircraft, respectively. And α and β denote height and width of predicted shape, respectively. The hyper-parameter λ denotes weight of dice loss (we use $\lambda = 1$ in experiments). Here, Ψ_{BCE} denotes the binary cross-entropy loss and Dice loss is expressed as

$$\Psi_{Dice}(q_s, r_s) = 1 - \frac{\sum_k^{\alpha \times \beta} q_s^k r_s^k + \epsilon}{\sum_k^{\alpha \times \beta} \left(q_s^k\right)^2 + \sum_k^{\alpha \times \beta} \left(q_s^k\right)^2 + \epsilon}, \tag{4}$$

where k denotes the kth pixel & ϵ (unity term) is set to 1. To further enhance the features for both mask and shape predictions, we employ joint loss as

$$\Psi = \Psi_{class} + \Psi_{bbox} + \Psi_{mask} + \Psi_s \tag{5}$$

Here, Ψ_{class}, Ψ_{bbox}, Ψ_{mask} represent classification, regression and mask losses which are taken from mask R-CNN. And, Ψ_s denotes shape loss given in Equation (3).

2.2 Inference on large-scale remote sensing images

To avoid memory overheads, inference is performed on the tiles by partitioning the large-scale image into the tiles of size used in training. However, an effective inferencing mechanism is required to avoid the partial object detection/missing objects at the borders of the tiles as shown in Fig. 3. We describe the partitioning scheme (with no offsets) for a given image I of width W and height H during inference as:

$$T_1 = \left\{ \cup_{i=0}^{N_X} \left\{ \cup_{j=0}^{N_Y} \{I(i \times t_s, j \times t_s)\} \right\} \right\}, \tag{6}$$

where $\cup$ represents union operation, i and j refer indices to locate cells along X and Y direction, respectively. And, $I(i, j)$ denotes a tile of size $(t_s \times t_s)$ that starts at the cell (i, j). The values N_X and N_Y represent the number of tiles along X and Y direction, which are computed as $N_X = \frac{W}{t_s}$ and $N_Y = \frac{H}{t_s}$, respectively. Similarly, other three partitioning schemes are performed by offsetting $\left(\frac{t_s}{2}\right)$ horizontally, vertically, and in both the directions to ensure sufficient overlaps across the tiles.

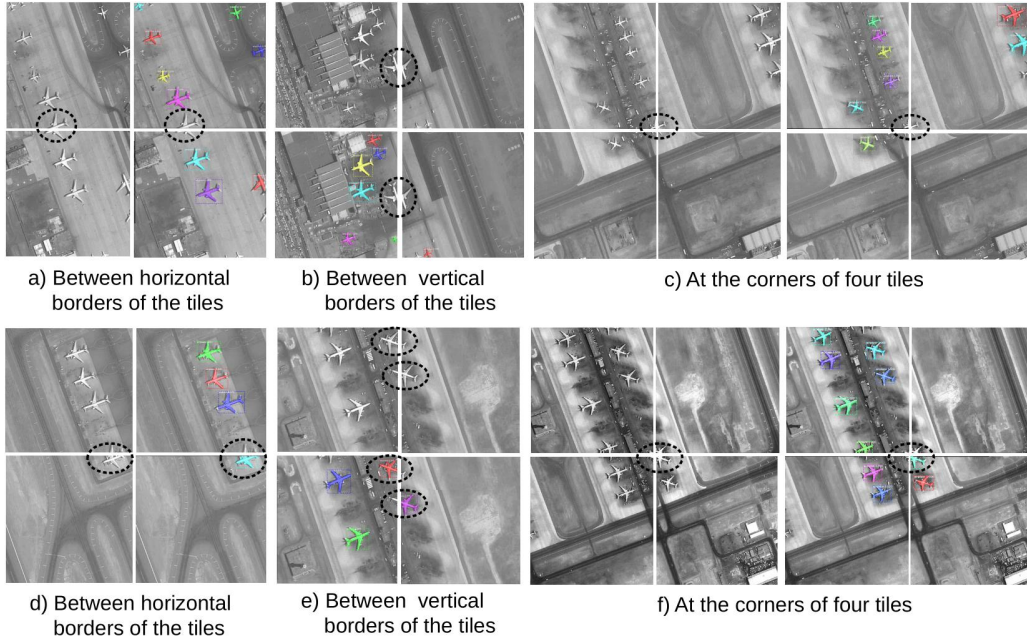| a) Between horizontal borders of the tiles | b) Between vertical borders of the tiles | c) At the corners of four tiles |
| d) Between horizontal borders of the tiles | e) Between vertical borders of the tiles | f) At the corners of four tiles |

Figure 3. Issues with large-scale C2S [2] images. (a-c): complete object missing. (d-f): partial detection.

2.3 Aggregation of bounding boxes

Each of the above four partitioning schemes provides a collection of bounding boxes and corresponding masks of the aircrafts separately over the same image. We ignore storing both partial and complete bounding boxes of the aircrafts on either side of vertical or horizontal borders of the tiles with a threshold of $\left(\frac{t_s}{2}\right)$. However, these ignored bounding boxes of one partitioning scheme, by virtue of their overlaps, are available in anyone of the other partitioning schemes. This ensures the inclusion of bounding boxes of the complete objects from the four sets and exclusion of the partial bounding boxes as well at the borders.

2.4 Deriving geospatial attributes of aircraft

The predicted mask of the aircraft and their gradient information help in determining its symmetry [18]. The vector describing the gradient of a surface $K[x, y]$ is defined by

$$K_x = \frac{\partial K(x,y)}{\partial x}, K_y = \frac{\partial K(x,y)}{\partial y} . \tag{7}$$

And, its orientation (ϕ) with domain $[0, 2\pi)$ is given by $arctan\left(\frac{K_y}{K_x}\right)$. We use Fourier transform to determine the orientation of symmetrical axis by searching the convolution peaks of the gradient orientation histogram. Subsequently, its position is determined based on the center of gravity of the aircraft mask. The accuracy of the orientation of the symmetrical axis is influenced by the noise in the image. Hence, the predicted mask of the aircraft alone is considered by suppressing the outside pixels uniformly. The symmetrical axis and its orientation information of the aircraft are used to rotate the mask upright. The length and width of aircraft are obtained by multiplying the number of pixels in Y-axis and X-axis with GSD of the image, respectively.

2.5 Aircraft type recognition

We determine the type of an aircraft based on IoU value between the predicted aircraft mask and the reference templates given in Fig. 4. Additionally, the length and width are used to further categorize the aircrafts of the same type. The derivation of geospatial attributes eliminates the need of separate templates for the same type of aircrafts with different length and width. Further, we use metadata to geotag both bounding boxes and the masks. This metadata consists of georeferenced corner coordinates of the large-scale input image that helps to store the identified aircrafts instance-wise while generating a shapefile.

	A	B	C	D	E	F	G
Row 3	49.21	73.34	29.35	41.70	40.40	56.40	24.50
Row 4	40.60	70.35	28.79	44.80	29.80	49.00	17.60

Figure 4. Reference templates of aircrafts. Row 1 shows 7 types of sample aircrafts from C2S PAN images. Row 2 depicts corresponding binary templates. Row 3 & Row 4 indicate length & width of aircrafts. Row 5 indicates aircraft type.

3. EXPERIMENTAL RESULTS

This section presents both quantitative and qualitative analysis of the proposed framework to demonstrate its efficacy on large-scale panchromatic remote sensing images.

3.1 Datasets and implementation details

We consider 5600 C2S [2] panchromatic aircraft scenes of size 512×512 with ground sampling distance (GSD) of 0.65m for all the experiments. A ratio of 80% - 20% is considered for training and validation. We perform data augmentation in the training process in order to accommodate the characteristics of remote sensing images in learning the various features. In particular, we explore several data augmentation techniques which include horizontal flipping, upside down flipping, rotation from -90° to 90°, scaling from 0.5 to 1.5, and shear range -16 to 16. To have fair comparison, we conducted all the experiments with ResNet-101 backbone architecture by setting the hyper-parameters: number of iterations = 1.2k; learning rate = 0.01; batch size = 1 image; weight attenuation = 0.0001; optimizer as SGD with a momentum of 0.9. The performance of aircraft segmentation model is validated on 1400 aircrafts from 1120 scenes.

3.2 Quantitative analysis of aircraft segmentation model

In this section, we compare the performance of our proposed aircraft segmentation model with the state-of-the art instance segmentation models. The evaluation is based on average precision (AP) metric, specifically for the threshold of 0.5 & 0.75, to indicate a predicted bounding box if its Inter-section over Union (IoU) is greater than 0.5 or 0.75. Also, we use mean length error (L_{error}) and mean width error (W_{error}) to evaluate the proposed segmentation. Table 1 provides the performance comparison of the proposed method with state-of-the-art instance segmentation methods. It can be observed from Table 1 that our proposed model outperforms existing state-of-the art approaches by a margin of 6%. Also,

it can be observed that the proposed method is able to bring down the mean length and width errors by 5 to 6 pixels in comparison to state-of-the-art-methods. These small errors are accounted for the fragmented nose/tail/wings. The results in Table 1 signify the robustness of our proposed model and its ability to incorporate the characteristics of even more harder examples in the segmentation process. Fig. 5 shows the confusion matrix among the 7 types of aircrafts from C2S scenes from our approach. As it can be observed from Fig. 5, the type-wise recognition is near to the overall average precision. This ensures that our model is able to predict the shapes of all 7 types of aircraft equally well. This also indicates that our proposed method effectively preserves the shape of the aircrafts even along their nose and wings. Thus the derived geospatial attributes from our approach are effective for aircraft type recognition.

Table 1. Performance comparison of the proposed method with state-of-the-art approaches on C2S scenes.

	Average Precision (AP) (%)	AP@50 (%)	AP@75 (%)	$L_{error}(m)$	$W_{error}(m)$
FCIS [11]	79.42	88.26	76.49	12.64	12.03
Mask-RCNN [12]	82.07	90.27	78.93	11.25	10.56
MS RCNN [14]	83.66	91.54	81.14	10.11	9.98
PA-Net [13]	85.49	92.03	82.35	9.54	8.87
HTC [15]	89.24	92.76	85.63	8.92	8.04
Proposed method	**95.43**	**97.34**	**90.72**	**5.23**	**4.57**
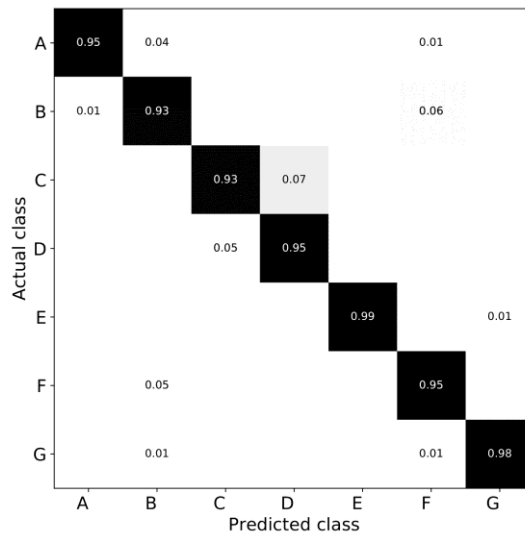
Figure 5. Confusion matrix from the proposed approach on 7 types of aircrafts over C2S panchromatic remote sensing images.

3.3 Qualitative results

Fig. 6 provides the aircraft segmentation output from Mask-RCNN and our proposed method on a C2S scene. It can be observed that our proposed method effectively preserves the shape of the aircrafts even along their nose and wings. Thus the derived geospatial attributes from our approach are effective for aircraft type recognition.

(a) Outputs from Mask R-CNN

(b) Outputs from the proposed aircraft segmentation model

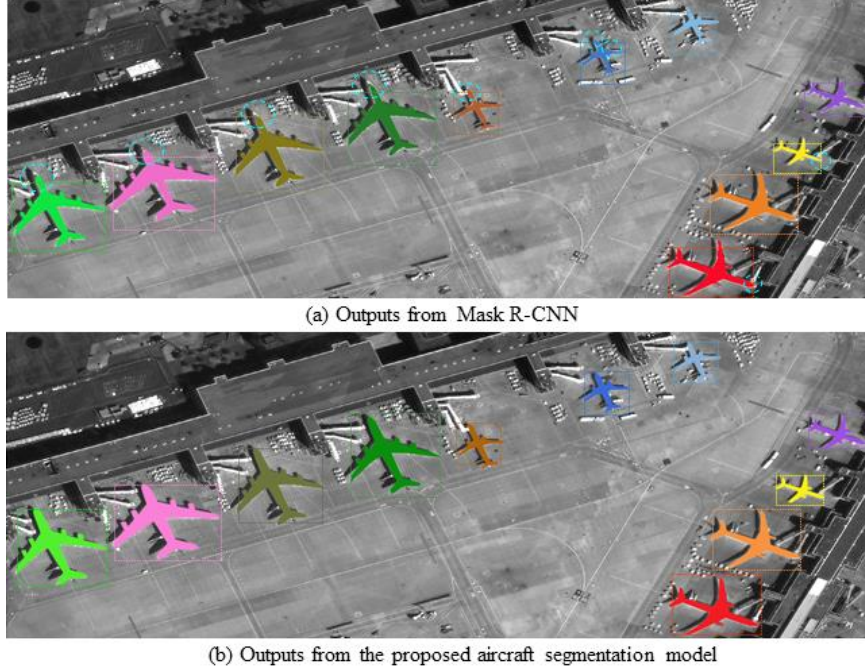Figure 6. Aircraft segmentation outputs from C2S panchromatic remote sensing images.

3.4 Evaluation of inference mechanism

Table 2 presents the evaluation of four partitioning schemes and their aggregation on fourteen large-scale C2S panchromatic images. The aggregate method is able to overcome the issues, such as the inclusion of partial objects and exclusion of the complete objects at the borders of the tiles. We considered two cases for evaluation, in which one case (indicated with *) does not consider partial bounding boxes (PBB) as they are not useful for the recognition process. The second case includes both PBB and Full bounding box (FBB). It is evident from the average values of precision, recall, and F1- scores that the aggregate method is able to detect the aircrafts completely, without leaving a possibility for partial/missing objects along the borders of tiles. Thus, the description of objects is improved relatively by minimum 4% to maximum 9% with the aggregation of four schemes. This shows the efficacy of the proposed framework in identifying the aircrafts and deriving their spatial attributes in large-scale panchromatic remote sensing images.

Table 2. Performance evaluation of our inference mechanism on large-scale C2S panchromatic images.

Partitioning Schemes	False Positive	False Negative	True Positive (FBB+PBB)	Precision*	Recall*	F1 Score*	Precision	Recall	F1 score
Scheme1	28	14	180 (156+24)	0.847	0.917	0.881	0.865	0.927	0.895
Scheme2	41	18	176 (157+19)	0.792	0.897	0.841	0.811	0.907	0.856
Scheme3	36	23	171 (146+25)	0.802	0.863	0.831	0.826	0.881	0.852
Scheme4	30	21	173 (153+20)	0.836	0.879	0.857	0.852	0.891	0.871
Proposed method	**28**	**4**	**190 (190+0)**	**0.871**	**0.979**	**0.922**	**0.871**	**0.979**	**0.922**

4. CONCLUSION

In this paper, we proposed a novel framework for effective aircraft type recognition in large-scale remote sensing images. Initially, we construct a learnable shape-preserved and deformable aircraft segmentation model to handle

geometric distortions in remote sensing images. Specifically, we employ deformable convolutions in two fusion blocks to improve the quality of the predicted masks. The metadata of images is used to derive the length and width in addition to geotagged masks that represent aircraft shapes. Finally, these geospatial attributes are used to recognize aircraft type by comparing with the reference templates. We also presented an effective inferencing mechanism to avoid the partial detections/missing of the aircrafts at the borders of tiles in the partitioning process. The efficacy of the proposed method is demonstrated on large-scale C2S panchromatic images. This approach would also be helpful to catalog the segmented aircrafts with their geospatial attributes, if they are not found in reference templates.

REFERENCES

[1] Z. Han, H. Zhang, J. Zhang, and X. Hu, "Fast aircraft detection based on region locating network in largescale remote sensing images," in IEEE International Conference on Image Processing (ICIP), pp. 2294–2298, (2017).

[2] ISRO, "https://www.isro.gov.in/spacecraft/cartosat-2-series-satellite-2," Tech. Rep., ISRO, (2018).

[3] J. Hsieh, J. Chen, C. Chuang, and K. Fan, "Aircraft type recognition in satellite images," IEE Proceedings - Vision, Image and Signal Processing, vol. 152, no. 3, pp. 307–315, (2005).

[4] D. Wang, X. He, W. Zhonghui, and H. Yu, "A method of aircraft image target recognition based on modified PCA features and SVM," in 9th International Conference on Electronic Measurement Instruments, pp. 177–181, (2009).

[5] Wenhui Diao, Xian Sun, Fangzheng Dou, Menglong Yan, Hongqi Wang, and Kun Fu, "Object recognition in remote sensing images using sparse deep belief networks," Remote Sensing Letters, vol. 6, pp. 745–754, (2015).

[6] Q. Wu, H. Sun, X. Sun, D. Zhang, K. Fu, and H. Wang, "Aircraft recognition in high-resolution optical satellite remote sensing images," IEEE Geoscience and Remote Sensing Letters, vol. 12, no. 1, pp. 112–116, (2015).

[7] A. Zhao, K. Fu, S. Wang, J. Zuo, Y. Zhang, Y. Hu, and H. Wang, "Aircraft recognition based on landmark detection in remote sensing images," IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 8, pp. 1413–1417, (2017).

[8] J. Zuo, G. Xu, K. Fu, X. Sun, and H. Sun, "Aircraft type recognition based on segmentation with deep convolutional neural networks," IEEE Geoscience and Remote Sensing Letters, vol. 15, no. 2, pp. 282–286, (2018).

[9] H. Li, B.Guo, T. Gao, and H. Chen, "A transfer learning method for aircrafts recognition," International Conference on Artificial Intelligence for Communications and Networks, vol. 286, pp. 175–185, (2019).

[10] K. Fu, W. Dai, Y. Zhang, Z. Wang, M. Yan, and X. Sun, "MultiCAM: Multiple class activation mapping for aircraft recognition in remote sensing images," Remote Sensing, vol. 11, no. 5, (2019).

[11] Y Li, H Qi, J Dai, X Ji, and YWei, "Fully convolutional instance-aware semantic segmentation," in Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 4438–4446, (2017).

[12] K. He, G. Gkioxari, P. Dol´ar, and R. Girshick, "Mask RCNN," in Proc. IEEE Conference on Computer Vision, pp. 2980–2988, (2017).

[13] S. Liu, L. Qi, H. Qin, J. Shi and J. Jia, "Path Aggregation Network for Instance Segmentation," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8759-8768, (2018).

[14] Z Huang, L Huang, Y Gong, C Huang, and X Wang, "Mask scoring R-CNN," in Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp.6409–6418, (2019).

[15] Chen, Kai, Jiangmiao Pang, Jiaqi Wang, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng et al. "Hybrid task cascade for instance segmentation." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4974-4983, (2019).

[16] X. Liu and X. Di, "Global context parallel attention for anchor-free instance segmentation in remote sensing images," IEEE Geoscience and Remote Sensing Letters, pp. 1–5, (2020).

[17] Dai, Jifeng, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. "Deformable convolutional networks." In Proceedings of the IEEE international conference on computer vision, pp. 764-773. (2017).

[18] C.Sun and D.Sib, "Fast reflectional symmetry detection using orientation histograms," Real-Time Imaging, vol. 5, pp. 63–74, (1999).

AUTHORS' BACKGROUND

Your Name	Title*	Research Field	Personal website
Rajeshreddy Datla	Scientist-SE/ External PhD candidate	Remote sensing imagery analysis, computer vision, machine learning	
Chalavadi Vishnu	PhD candidate	Aerial and drone imagery analysis, machine learning	
C Krishna Mohan	Full Professor	Pattern Recognition, machine learning	https://www.iith.ac.in/~ckm/