Skeleton-Aware Network for Aircraft Landmark Detection

Yuntong Ye[1,2], Yi Chang[2], Yi Li[1], and Luxin Yan[1(✉)]

[1] National Key Laboratory of Science and Technology on Multi-spectral Information Processing, School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China
{yuntongye,li_yi,yanluxin}@hust.edu.cn
[2] AI Center, Pengcheng Lab, Shenzhen, China
yichang@hust.edu.cn

Abstract. The landmark detection has been widely investigated for the human pose with rapid progress in recent years. In this work, we aim at dealing with a new problem: *aircraft landmark detection in the wild*. We have a key observation: the aircraft is a rigid object with global structural relationships between local landmarks. This motivates us to progressively learn the global geometrical structure and local landmark localization in a coarse-to-fine guidance manner. In this paper, we propose a simple yet effective skeleton-aware landmark detection (SALD) network, including one stream for exploiting the coarse global skeleton structure and one stream for the precise local landmarks localization. The global skeleton structure models the aircraft "images" into skeleton "lines", in which the multiple skeletons of the holistic aircraft and the parts are explicitly extracted to serve as the geometrical structure constraints for landmarks. Then, the local landmark localization precisely detects the key "points" with the guidance of skeleton "lines". Consequently, the progressive strategy of "extracting lines from images, detecting points with lines" significantly eases the landmark detection task by decomposing the task into the simpler coarse-to-fine sub-tasks, thus further improving the detection performance. Extensive experimental results show the superiority of proposed method compared to state-of-the-arts.

Keywords: Aircraft · Landmark detection · Convolutional neural network · Skeleton

1 Introduction

Landmark detection refers to the task of locating keypoints in the given images. In aircraft landmark detection these keypoints are predefined at aircraft endpoints and joints such as head, tip and stabilizer, as shown in Fig. 1 (a). The aircraft landmark detection serves as an important prior work for applications like aircraft fine-grained classification [1–3], and aircraft detection [4,5]. In this paper, we focus on the problem of single aircraft landmark detection.

Y. Peng et al. (Eds.): ICIG 2021, LNCS 12888, pp. 185–197, 2021.
https://doi.org/10.1007/978-3-030-87355-4_16

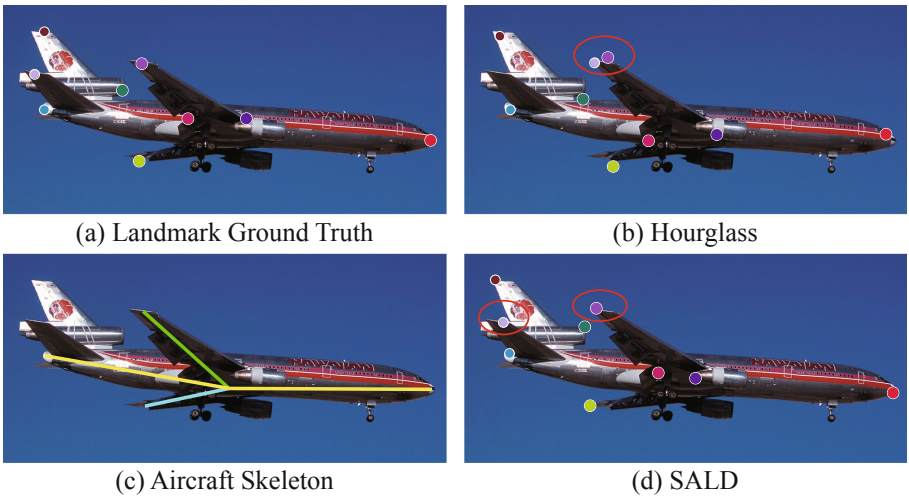

Fig. 1. Illustration of the proposed skeleton-aware landmark detection (SALD). (a) Landmark ground truth. (b) Hourglass [16] wrongly detects the right tip on the right stabilizer due to the similarity of local features in aircraft. (c) The skeletons model the coarse global geometrical structure of the aircraft. (d) With the guidance of the aircraft structure, SALD predicts the precise landmark locations in a progressive coarse-to-fine manner. The right stabilizer landmark is located near the end point of the stabilizer part, with a correct semantic and geometrical relationship with the structure.

There are few researchers devoted to aircraft landmark detection in the wild despite its importance. In remote sensing images, the landmarks of aircraft aerial view images are detected by effective convolutional neural networks (CNN) and utilized by the following aircraft type recognition and detection tasks. The focus is on the landmark utilization rather than the detection. For instance, Zhao *et al.* [2] proposed a six-layer model based on vanilla CNN [25] to regress the aircraft landmarks and perform landmark template matching to recognize the aircraft type. Zhou *et al.* [13] predicted aircraft keypoints via convolutions and designed attention mechanism on the keypoints to enhance the features for detection. In this paper, detecting the aircraft landmarks in the wild is more challenging due to the vastly different appearances in variable viewpoints.

Related to the aircraft landmark detection problem, the 2D human landmark detection has achieved rapid progress in recent years [7,9,11,20,22,24]. To alleviate the problems of the occlusion and variable viewpoints in human body, the landmark relationships are studied to guide landmark detection [7,12,15,17], which improves the robustness and accuracy. However, the landmark relationships in the deformable human body mostly lie in the local body parts, and are modelled in an implicit manner. Chen *et al.* [12] trained a landmark distribution discriminator in an adversarial to make the predicted landmarks distribute naturally like the real ones. Ke *et al.* [15] connected landmarks in the same body part and designed a structure-aware loss to preserve the structure layout. Tang

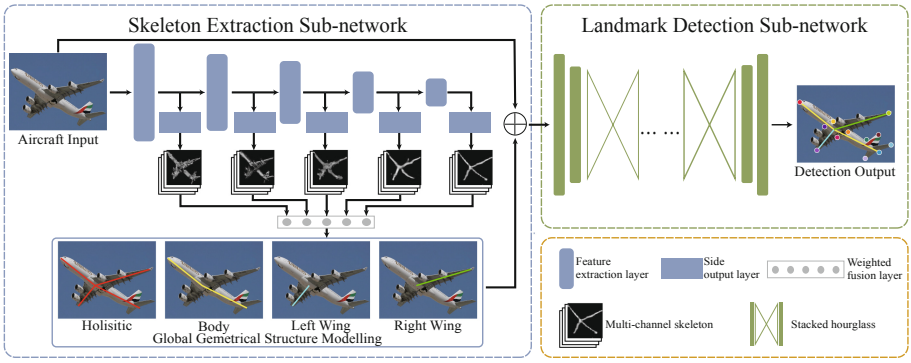

Fig. 2. Overview of the proposed skeleton-aware landmark detection (SALD) method. Our network consists of a skeleton extraction sub-network for modelling the coarse global aircraft structure and a landmark detection sub-network for precisely locating landmarks with the structure guidance. The framework decompose the landmark detection into the two simpler sub-tasks in a progressive coarse-to-fine manner, which eases the learning procedure and improves the detection performance.

et al. [7] classified the landmarks into five groups and jointly learned the shared features for the landmarks without explicit constraints. In this paper, we focus on the landmark detection for the aircraft with the rigid property, in which the landmark relationships are not only lie in the local parts, but also exist in the global geometrical distributions for all the landmarks. This motivates us to explicitly model the global aircraft structure, which serves as the coarse guidance to provide the structure cues and ease the precise local landmark detection task. Consequently, the aircraft landmark detection task is decomposed into the progressive coarse-to-fine learning of two simpler sub-tasks: extracting the coarse global structure from the image, and locating the precise local landmarks with the guidance of structure. As illustrated in Fig. 1, compared with Hourglass [16] which wrongly detects the right stabilizer on the similar right tip (Fig. 1 (b)), the aircraft skeletons serve as the global structure (Fig. 1 (c)), and guide the correct local landmarks localization on the right stabilizer (Fig. 1 (d)).

To this end, we propose a skeleton-aware landmark detection (SALD) network consisting of a structure extraction stream for explicitly modelling the geometrical structure via the hierarchical aircraft skeletons, and a landmark detection stream for locating the landmarks with the guidance of aircraft skeletons, as shown in Fig. 2. Specifically, the hierarchical skeletons have multiple channels including the holistic skeleton channels for all the landmarks, and part skeleton channels for the aircraft parts, such as the wings and aircraft body. The skeletons provide the global relationships among the landmarks in geometrics and semantics. With the guidance of the global understanding for aircraft structure, the precise local landmark localization is performed to achieve the coherency between landmark distributions and the structure. Consequently, the two streams in SALD progressively focus on the sub-task of extracting skeleton

"lines" from the aircraft "images", and detecting the key "points" with the skeleton "lines", which are simpler than directly detecting landmarks in the aircraft images.

We summarize the main contributions as follows:

- We study the new problem of aircraft landmark detection in the wild. In this task, by taking the full advantage of the rigid property, we not only utilize local features of the landmarks, but also exploit the global geometrical relationships between aircraft landmarks, which consequently achieve the consistency between the landmark layout and aircraft structure.
- We propose a skeleton-aware landmark detection (SALD) network consisting of two streams, including one stream for extracting the coarse global structure from the image, and one stream for locating the precise local landmarks with the guidance of structure. The framework decomposes the landmark detection task into two simpler sub-tasks in a progressive coarse-to-fine manner, which eases the learning procedure, thus further improves the performance of landmark detection.
- Extensive quantitative and qualitative evaluations on the aircraft datasets show that SALD performs favorably against the state-of-the-art methods, which demonstrates the effectiveness of the skeleton guidance for aircraft landmark detection.

2 Skeleton-Aware Landmark Detection Network

2.1 Global Geometrical Structure

Skeleton Structure. To model the global geometrical structure of the rigid aircraft, we intuitively resort to the aircraft skeleton as the representations which is an important graphics description with intrinsic relationships to the landmark, and possess strong constraints with the landmarks in both geometrics and semantics. For instance, the aircraft tip should distribute on the wing, and the precise location should be near the endpoint of the wing skeletons. To fully exploit the skeletons as the global geometrical structure, we propose the hierarchical multi-channel skeletons to present the structure of both holistic aircraft and the parts, including the including holistic channel, body channel, left wing channel and right wing channel as shown in Fig. 2. Specifically, the holistic channel encodes the geometrical relationships for all the landmarks from a global viewpoint, while the skeletons of aircraft parts in the other channels provide explicit guidances for locating landmarks near the corresponding parts. We obtain the skeleton labels by connecting the head, tail and wing tip landmarks with aircraft center point which is calculated as the average coordinate of the leading and trailing edge flaps. The label generating procedure dose not require additional manual work.

Skeleton Extraction. To extract the aircraft skeletons, we introduce a hierarchical deep-supervised network for the skeleton extraction sub-network which generates high quality skeletons by fully utilizing the multi-scale spatial information in aircraft images. Specifically, the multi-channel skeletons are side outputted at different scales of the network, which are deeply supervised during training by the ground truth skeletons connected by the landmarks, as shown in Fig. 2. Then the multi-scale side outputs are fused together through a weighted fuse layer to generate the final multi-channel skeletons. Denoting X as the given aircraft image, we extract the multi-channel skeletons $\hat{L}$ by the skeleton extraction sub-network G, which is formulated as:

$$\hat{L} = G(X; W, w_G), \tag{1}$$

where w_G, W denote the parameters of the side output layers and feature extraction layers. For each side output of the skeletons $\hat{L}^c$, we impose cross-entropy loss function, which is defined as:

$$\begin{aligned} Loss^c_{ske} = &- \sum_{j \in L^c_+} \log \Pr(\hat{L}^c_j = 1 | X; W, w^c_G) \\ &- \alpha_c \sum_{j \in L^c_-} \log \Pr(\hat{L}^c_j = 0 | X; w^c_G), \end{aligned} \tag{2}$$

where the function $\Pr(\cdot)$ is computed by the output of sigmoid function on activation value at pixel j. w^c_G denotes the parameters of c-th side output layer. α_c denotes the balance weights corresponding the ratio of skeleton pixels. The final loss for the skeleton extraction sub-network is defined as:

$$Loss_{ske} = \sum_{c}^{C} \gamma_c Loss^c_{ske}, \tag{3}$$

where C is the total number of the side output layers, while γ_c denotes the balance weights. The extracted hierarchical skeletons represent the coarse global aircraft structures are then utilized for precise localization.

2.2 Local Landmark Localization

After obtaining the coarse global aircraft structure represented by the hierarchical skeletons, we perform the precise landmark localization under the explicit guidance of the aircraft geometrics and landmark relationships. Specifically, we feed the concatenation of the aircraft image and the hierarchical skeletons to stacked Hourglass [16], in which up and down sampling processes are repeated with intermediate conjunction and supervision to learn features across all the scales. During training, the skeleton channels indicate the corresponding localization of the aircraft parts. Taking an example, the left wing channel skeleton explicitly represents the structure of the wing. With the guidance of the left wing channel skeleton, the left leading and trailing edge flap should be detected

Table 1. Effectiveness of the global skeleton structure. L and $\hat{L}$ refer to the ground truth and extracted skeletons. Y and $\hat{Y}$ refer to the ground-truth and predicted landmarks. $-$ refers to the average Euler distances. The small deviation in the second and third columns justify the accuracy of the skeleton extraction, while the last two columns show that with the skeleton guidance, the landmarks are more related to the skeleton structures. The detection performances are improved in terms of the PCKh, which demonstrates the effectiveness of the skeleton guidance, and further illustrate the accuracy of the skeleton extraction.

Landmarks	$L-Y$	$\hat{L}-Y$	$\hat{L}-\hat{Y}$ w/o Guidance	$\hat{L}-\hat{Y}$ w/ Guidance
Head	0	1.23	2.42	1.91
Wing tip	0	4.46	11.02	9.74
Leading edge flap	6.39	6.55	8.52	6.72
Trailing edge flap	6.07	6.96	11.36	8.20
PCKh	100	100	86.17	87.57

near the endpoint of the left wing skeleton, while the left tip should be located near the other endpoint, which is possible to shift onto the right wing or the stabilizers due to the similar local features without the global skeleton guidance. Learning both global geometrical structure from the skeletons and local features from the aircraft appearances in the images, the detection sub-network regresses landmark heatmaps $\hat{Y}$ in which the location of highest value is determined as the final prediction. MSE loss is imposed on the heatmaps, which is defined by:

$$Loss_{land} = \sum_{i}^{N} \|\hat{Y}^i - Y^i\|_2, \tag{4}$$

where $\hat{Y}^i$ denotes the heatmap for the i-th landmark. Denoting β as the balance weight, the full objective is given by:

$$Loss = Loss_{ske} + \beta Loss_{land}. \tag{5}$$

2.3 Effectiveness of the Global Skeleton Structure

Accuracy of the Skeleton Extraction. The extracted skeleton should possess the strong relationships with the aircraft landmarks in both geometrics and semantics. Wrong skeletons will do harm to the landmark detection, even though the structure extraction task is coarse compared with the precise localization. To illustrate the accuracy of the skeleton extraction, we calculate the average Euler distances between landmarks and skeletons in the 256×256 aircraft images in the dataset. The small distances indicate the strong geometrical relationships between the skeletons and landmarks. The first three columns in Table 1 show that compared to the ground-truth skeletons which have the strongest relationships with the landmarks, the extracted skeletons by SALD deviate from the

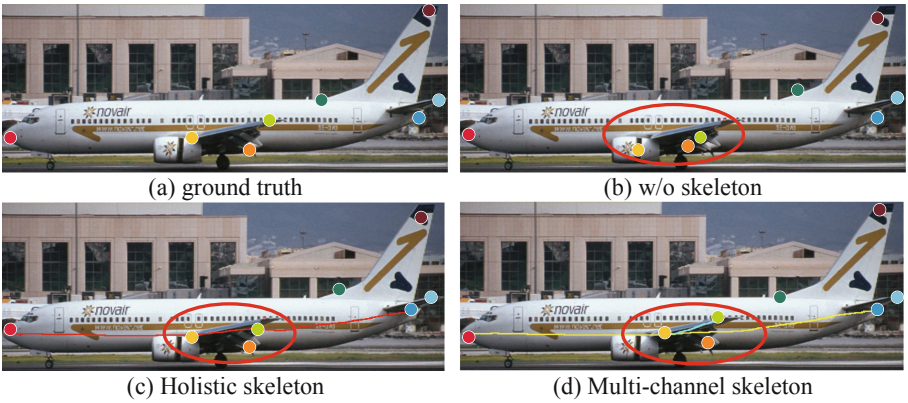

Fig. 3. The superiority of multi-channel skeleton guidance. (a) The landmark ground truth. (b) Without skeleton guidance, the detected left tip shifts onto the wing. (c) The holistic skeleton provides a provides a rough structure information of the aircraft, which corrects the location of the left edge flap landmarks. (d) The multi-channel skeleton models the structure more explicitly, in which the left wing structure is further represented, thus obtaining better performance.

ground-truth in a small extent, but still holds the strong relationships with the ground-truth landmarks in terms of the small distance, which can help the landmark detection sub-network to locate the landmarks with the coarse global understanding.

Effectiveness of the Skeleton Guidance. The last two columns in Table 1 illustrates the effectiveness of skeleton guidance. Compared with the results in which no global geometrical constraint is imposed as shown in the fourth column, our extracted skeleton guides the detector to locate the landmarks near the corresponding part, enforcing the landmark layout to be consistent with the structure. The distances between landmarks and the aircraft structure in the fifth column become smaller when the skeleton constraint is imposed, and consequently improve the accuracy of landmark detection in terms of the quantitative evaluation method PCKh, thus justifying the effectiveness of skeleton guidance and further illustrating the accuracy of the skeleton extraction.

To further show the superiority of the multi-channel skeleton guidance, we perform the comparison between the landmark detection results with no skeleton guidance, with only holistic skeleton guidance, and with multi-channel skeleton guidance in Fig. 3. Without the skeleton guidance, the local features of left wings are not discriminative enough, the landmarks on the left wing shift from the correct location. The holistic skeleton provides a rough structure of the aircraft, which guides the detector to locate the leading and trailing edge flaps at the joints of left wing and aircraft body. The multi-channel skeletons model both the holistic and part structures as more explicit cues to distinguish the left

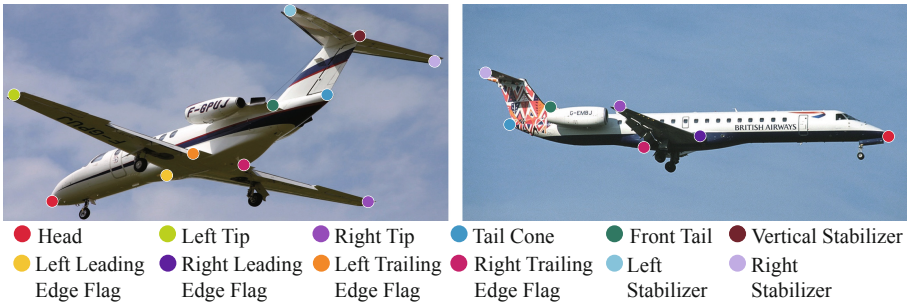

Fig. 4. Illustration of the landmark labelling on the ALDW dataset.

Table 2. Quantitative Comparisons with state-of-the-art methods on FGVC dataset. SALD outperforms the competing methods, especially in landmarks on the wings whose structure is explicitly modelled by the multi-channel skeleton.

Methods	Head & tail cone	Leading edge flap	Wing tip	Trailing edge flap	Horizontal stabilizer	Vertical stabilizer	PCKh
Hourglass [16]	96.60	79.55	82.30	72.93	83.13	87.95	86.17
DAN [8]	97.46	80.52	80.82	79.49	83.80	88.22	86.78
PoseAtten [21]	97.27	81.87	77.05	78.54	**86.40**	89.29	87.16
PyraNet [19]	97.15	82.69	77.59	81.14	85.06	88.69	87.17
CU-Net [18]	97.54	80.43	78.29	80.75	84.38	88.28	87.14
SALD	**97.98**	**82.89**	**83.47**	**82.55**	85.74	**89.37**	**87.57**

wing and the body via additional body and left wing channel, and consequently further improving the detection performance on left tip.

2.4 Implementation Detail

SALD is implemented using Tensorflow framework on a RTX 2080Ti GPU. The input images are resized to 256×256 and random flip is applied for augmentation. The skeleton balance weights α_c are 186, 113, 45 and 51 for skeletons of holistic, body, left wing and right wing, which are the average ratio of skeleton pixels in ground truth. γ_c and β are set as 1. We respectively train the two sub-networks for 200 epochs with initial learning rate 0.001, decay rate 0.99 and decay step 5000. Then we fine-tune the two sub-networks together with learning rate 0.0001 for 50 epochs. For the optimizer, the RMSprop is adopted with batch size 6.

3 Experiments

Dataset. Since there exist few datasets for the task, we apply a new dataset for aircraft landmark detection in the wild (ALWD), in which we annotate 7819 (6245 for training and 1574 for testing) aircraft images from the FGVC [26]

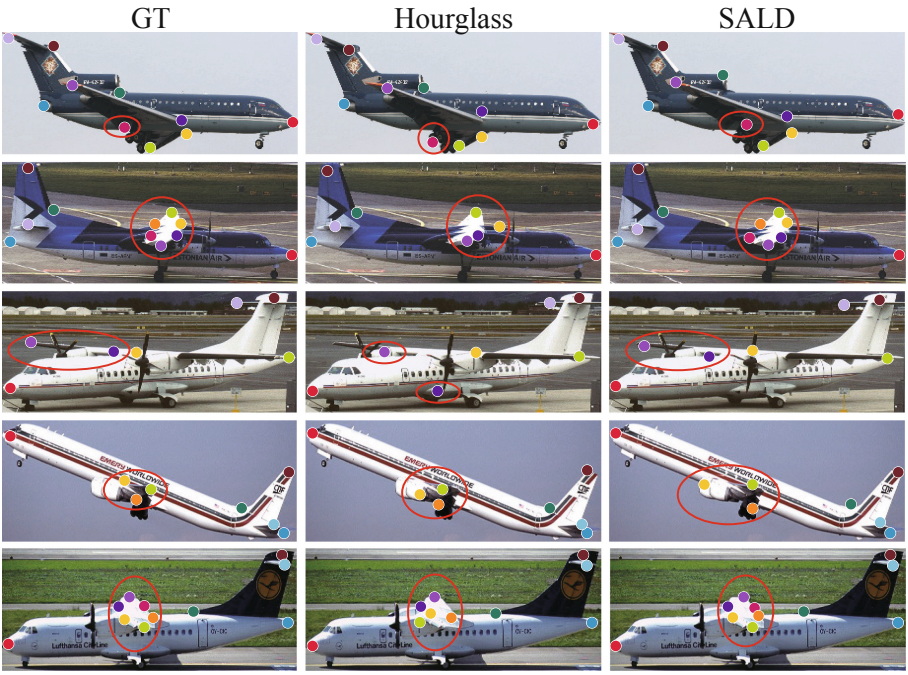

Fig. 5. Visualization of landmark detection results on ALDW dataset. Under the global geometrical skeleton guidance, SALD predicts more accurate landmarks whose layout is consistent with the aircraft structure.

dataset and Google search. For each aircraft, we annotate the location and visibility of 12 aircraft landmarks, whose locations are defined on the joints or endpoints of the head, wings and stabilizers as illustrated in Fig. 4.

Experiment Settings. We select five state-of-the-art human landmark detection methods: Hourglass [16], DAN [8], PoseAtten [21], PyraNet [19] and CU-Net [18] and fine-tune them on ALDW for comparison, in which PCKh [7,21,24] is utilized for quantitative assessment. The codes of our methods will be released in the homepage of the author.

Evaluation on ALDW Dataset. The results in Table 2 show that SALD outperforms the competing methods. Especially, SALD achieves significantly better performance in landmarks on the wings whose structures are explicitly modelled by the multi-channel skeletons, demonstrating the effectiveness of the skeleton guidance. The qualitative comparison is shown in Fig. 5. The feature extracted by Hourglass is not discriminative enough due to similarity of the aircraft parts, resulting in the landmark shifts. SALD generates structure-consistent results which distribute near the aircraft skeletons in a reasonable layout.

Fig. 6. Visualization results of landmark detection in real scene. The SALD still obtains accurate predictions.

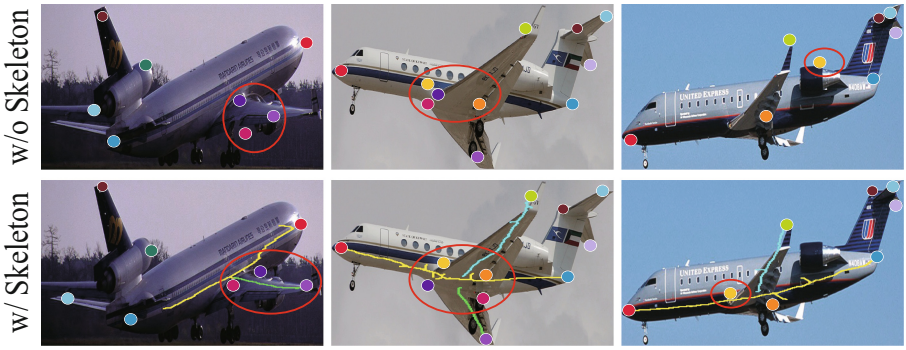

Fig. 7. Effectiveness of the skeleton guidance. The skeletons provide geometrical constraints for the consistency between landmark layout and aircraft structure.

Evaluation in Real Scene. We also test SALD on the aircraft images collected from Google. The qualitative results in Fig. 6 show that we still obtain accurate predictions. By recognizing the main structure of the aircraft in the real scene as the guidance for the precise landmark detection, SALD achieves the coherency between landmarks and structure, thus performing well in the real scene.

Ablation Study. We perform ablation study on the effectiveness of the multi-channel skeleton guidance quantitatively and qualitatively. Figure 7 shows that with the geometrical constraints of the skeletons, the detector achieves the consistency with aircraft structure, and consequently obtains significant improvement of detection accuracy. Table 3 further illustrates the effect of each channel. Compared with the first row, the other rows show the improvements brought by each skeleton channel. Especially, compared with the second row, the fourth and fifth rows show the wing channel brings more improvement in the landmarks

Table 3. Ablation study on the effectiveness of different skeleton channel. * refers to guiding detection with ground truth skeletons.

Holistic	Body	Left Wing	Right wing	Head	Wing	Stabilizer	PCKh
				97.95	80.23	84.32	86.17
✓				97.94	82.25	84.44	86.42
✓	✓			97.95	82.38	84.43	86.63
✓		✓		97.97	82.87	84.43	86.61
✓			✓	97.97	82.53	85.46	86.72
✓	✓	✓	✓	**97.98**	**83.03**	**86.13**	**87.57**
✓*	✓*	✓*	✓*	98.12	85.37	88.77	88.84

on the wing, demonstrating the specific contribution of the wing channel skeleton in guiding their corresponding landmarks. The landmark detection with all skeleton channel guidance achieves the best performance. In the last row, the performance of landmark detection is the best with the guidance of ground truth skeleton, which further demonstrates the effectiveness of the skeleton guidance.

4 Conclusions

In this paper, we have studied the new problem of aircraft landmark detection in the wild by utilizing the rigid property to progressively learn global structure extraction and local landmark localization in a coarse-to-fine manner. Specifically, we propose a skeleton-aware aircraft landmark (SALD) method consisting of two streams, including one stream for modelling the coarse aircraft structure by extracting the hierarchical skeletons, and one stream for detecting the precise landmark localization with the guidance of the global skeleton structures. Consequently, the landmark detection is decomposed into two simpler sub-tasks. By the global guidance for local landmark detection, SALD achieves the consistency between landmark layout and aircraft structure, which improves the accuracy and the robustness of aircraft landmark detection. Extensive experiments show that SALD outperforms state-of-the-art landmark detection methods.

Acknowledgement. This work was supported by This work was supported by National Natural Science Foundation of China under Grant No. 61971460, China Postdoctoral Science Foundation under Grant 2020M672748, National Postdoctoral Program for Innovative Talents BX20200173, the Open Research Fund of the National Key Laboratory of Science and Technology on Multispectral Information Processing under Grants 6142113200304 and Industrial Technology Development Program grant JCKY2018204B068.

References

1. Fu, K., Dai, W., Zhang, Y., Wang, Z., Yan, M., Sun, X.: MultiCAM: multiple class activation mapping for aircraft recognition in remote sensing images. Remote Sens. **11**(5), 544–553 (2019)
2. Zhao, A., et al.: Aircraft recognition based on landmark detection in remote sensing images. IEEE Geosci. Remote Sens. Lett. **14**(8), 1413–1417 (2017)
3. Zuo, J., Xu, G., Fu, K., Sun, X., Sun, H.: Aircraft type recognition based on segmentation with deep convolutional neural networks. IEEE Geosci. Remote Sens. Lett. **15**(2), 282–286 (2018)
4. Yang, Y., Zhang, Y., Bi, F., Shi, H., Xie, Y.: M-FCN: effective fully convolutional network-based airplane detection framework. IEEE Geosci. Remote Sens. Lett. **14**(8), 1293–1297 (2017)
5. Qiu, S., Wen, G., Deng, Z., Fan, Y., Hui, B.: Automatic and Fast PCM generation for occluded object detection in high-resolution remote sensing images. IEEE Geosci. Remote Sens. Lett. **14**(10), 1730–1734 (2017)
6. Zou, X., Zhong, S., Yan, L., Zhao, X., Zhou, J., Wu, Y.: Learning robust facial landmark detection via hierarchical structured ensemble. In: ICCV, pp. 141–150 (2019)
7. Tang, W., Wu, Y.: Does learning specific features for related parts help human pose estimation? In: CVPR, pp. 1107–1116 (2019)
8. Kowalski, M., Naruniec, J., Trzcinski, T.: Deep alignment network: a convolutional neural network for robust face alignment. In: CVPR Workshop, pp. 88–97 (2017)
9. Qiu, Z., Qiu, K., Fu, J., Fu, D.: Learning recurrent structure-guided attention network for multi-person pose estimation. In: ICME, pp. 418–423 (2019)
10. Zhou, L., Chen, Y., Wang, J., Tang, M., Lu, H.: Bi-directional message passing based ScaNet for human pose estimation. In: ICME, pp. 1048–1053 (2019)
11. Zhu, M., Shi, D.: Deep geometry embedding networks for robust facial landmark detection. In: ICME, pp. 1222–1227 (2019)
12. Chen, Y., Shen, C., Wei, X., Liu, L., Yang, J.: Adversarial PoseNet: a structure-aware convolutional network for human pose estimation. In: ICCV, pp. 1221–1230 (2017)
13. Zhou, M., Zou, Z., Shi, Z., Zeng, W., Gui, J.: Local attention networks for occluded airplane detection in remote sensing images. IEEE Geosci. Remote Sens. Lett. **3**(17), 381–385 (2020)
14. Liu, Z., Yan, S., Luo, P., Wang, X., Tang, X.: Fashion landmark detection in the wild. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9906, pp. 229–245. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46475-6_15
15. Ke, L., Chang, M.-C., Qi, H., Lyu, S.: Multi-scale structure-aware network for human pose estimation. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11206, pp. 731–746. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01216-8_44
16. Newell, A., Yang, K., Deng, J.: Stacked hourglass networks for human pose estimation. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9912, pp. 483–499. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46484-8_29
17. Tang, W., Yu, P., Wu, Y.: Deeply learned compositional models for human pose estimation. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11207, pp. 197–214. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01219-9_12

18. Tang, Z., Peng, X., Geng, S., Wu, L., Zhang, S., Metaxas, D.: Quantized densely connected u-nets for efficient landmark localization. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11207, pp. 348–364. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01219-9_21
19. Yang, W., Li, S., Ouyang, W., Li, H., Wang, X.: Learning feature pyramids for human pose estimation. In: ICCV, pp. 1281–1290 (2017)
20. Chandran, P., Bradley, D., Gross, M., Beeler, T.: Attention-driven cropping for very high resolution facial landmark detection. In: ICCV, pp. 5861–5870 (2020)
21. Chu, X., Yang, W., Ouyang, W., Ma, C., Yuille, A., Wang, X.: Multi-context attention for human pose estimation. In: CVPR, pp. 1831–1840 (2017)
22. Zhang F., Zhu, X., Dai H., Ye, M., Ce, Z.: Multi-context attention for human pose estimation. In: CVPR, pp. 7093–7102 (2020)
23. Xie, S., Tu, Z.: Holistically-nested edge detection. In: CVPR, pp. 1395–1403 (2015)
24. Ke, W., Chen, J., Jiao, J., Zhao, G., Ye, Q.: SRN: side-output residual network for object symmetry detection in the wild. In: CVPR, pp. 1068–1076 (2017)
25. Wu, Y., Hassner, T., Kim, K., Medioni, G., Natarajan, P.: Facial landmark detection with tweaked convolutional neural networks. IEEE Trans. Pattern Anal. Mach. Intell. **40**(12), 3067–3074 (2017)
26. Maji, S., Rahtu, E., Kannala, J., Blaschko, M., Vedaldi, A.: Fine-grained visual classification of aircraft. arXiv preprint arXiv:1306.5151 (2013)