RDR-KD: A Knowledge Distillation Detection Framework for Drone Scenes

Jinxiang Huang, Hong Chang, Xin Yang, Yong Liu, Shuqi Liu, and Yong Song

***Abstract*— Drone object detection (DOD) with real-time deployment is a research hotspot. On the one hand, the performance of tiny object detection is closely related to the ground detection capability of the drone platform. Existing methods are keen on designing complex networks to enhance the accuracy of tiny objects, which significantly increases computational costs. On the other hand, the limited drone hardware resources urgently require lightweight models for deployment. To address the dilemma of balancing detection accuracy and computational efficiency, we propose a regenerated-decoupled-responsive knowledge distillation (RDR-KD) framework specifically for drone scenes. First, we design the Regenerated Distillation and the Decoupled Distillation to fully transfer the tiny object feature information from the teacher model to the student model. Meanwhile, we devise the logit-based Responsive Distillation based on focal loss and efficient intersection over union (EIoU) to alleviate class imbalance. Finally, we conduct extensive experiments on the VisDrone2019 dataset. The experimental results demonstrate that the proposed RDR-KD framework improves AP and AP_S of the student model by 3.3% and 2.9% respectively, which outperforms other state-of-the-art distillation frameworks.**

***Index Terms*— Drone object detection (DOD), feature distillation, knowledge distillation (KD), responsive distillation.**

I. Introduction

DRONE object detection (DOD) is a crucial technology for obtaining ground information in the field of low-altitude remote sensing. In recent years, the high efficiency and flexibility of this technology have led to a wide range of applications including fire monitoring [1], infectious disease control [2], precision agriculture [3], and so on. However, the sharp scale changes and cluttered backgrounds in drone scenes usually result in poor feature information about tiny objects. Meanwhile, the models deployed on most drone platforms are relatively lightweight due to hardware resource constraints, which makes the DOD task more challenging. Current DOD research primarily focuses on modifying network structure [4] and designing a tiling detection framework [5] to fully extract the features of tiny objects. However, these methods change the structure of existing models and reduce computational efficiency, making actual deployment more inconvenient.

Generally speaking, knowledge distillation (KD) [6] is a learning method that transfers information from a large teacher network to a compact student network. It can remarkably improve the student performance without incurring additional costs. So far, it can be roughly divided into two categories depending on the location of distillation.

The first category is logit-based KD [6], which was first proposed in the field of image classification but rarely used in object detection tasks. Specifically, the student model learns the logit distribution of the teacher model at the prediction layer, and then the distillation loss is computed by applying the Kullback–Leibler (KL) divergence [7] between the logits of the student and teacher models under a softmax with temperature. Zhou et al. [8] analyzed the role of soft labels in terms of bias-variance tradeoff and redesigned weighted soft labels for distilling. Zhao et al. [9] introduced a novel logit decoupled distillation to divide the classification KD loss into target classes and nontarget classes. Yang et al. [10] normalized the nontarget logits, which can enhance knowledge transfer from teacher to student. Despite the above logit-based KD methods performing well in the image classification task, they are still weak in the face of the challenge of extremely unbalanced categories in the DOD task.

Another category is feature-based KD, which is extensively used in object detection. Romero et al. [11] first proposed to let the student model mimic the feature map of the teacher model at the middle layer. Shu et al. [12] proposed a channel-wise KD (CWD) method for dense prediction, which simply minimizes the loss of the probability map between the teacher and student networks. Yang et al. [13] improved the performance of the student by forcing it to reconstruct random pixels, but the reconstructed regions are not necessarily critical. Cao et al. [14] proposed to employ the Pearson correlation coefficient to mimic the corresponding features, thereby focusing on relational information from the teacher. However, existing feature-based KD methods are unable to focus on the tiny objects, resulting in the feature information of such objects being overshadowed during the distillation process.

Independent logit-based or feature-based KD frameworks may prevent the student model from learning comprehensive knowledge. Meanwhile, the existing framework that integrates two distillation types [15] is mainly designed for a two-stage detection algorithm, which is unsuitable for deployment on lightweight drone platforms. In this letter, we propose a novel regenerated-decoupled-responsive KD (RDR-KD) framework

Manuscript received 20 January 2024; revised 13 April 2024; accepted 4 May 2024. Date of publication 8 May 2024; date of current version 17 May 2024. This work was supported in part by the National Natural Science Foundation of China General Program under Grant 82272130, in part by the National Natural Science Foundation of China Key Program under Grant U22A20103, and in part by the Aeronautical Science Foundation under Grant 2023Z019072001. *(Corresponding authors: Yong Song; Xin Yang.)*

Jinxiang Huang, Xin Yang, Shuqi Liu, and Yong Song are with the School of Optics and Photonics, Beijing Institute of Technology, Beijing 100081, China (e-mail: huangjinxiang@bit.edu.cn; xinyang@bit.edu.cn; liushuqi@bit.edu.cn; yongsong@bit.edu.cn).

Hong Chang and Yong Liu are with the Beijing Aerospace Institute for Metrology and Measurement, Beijing 100076, China (e-mail: sunshine8299@126.com; bcst@sohu.com).

Digital Object Identifier 10.1109/LGRS.2024.3398140

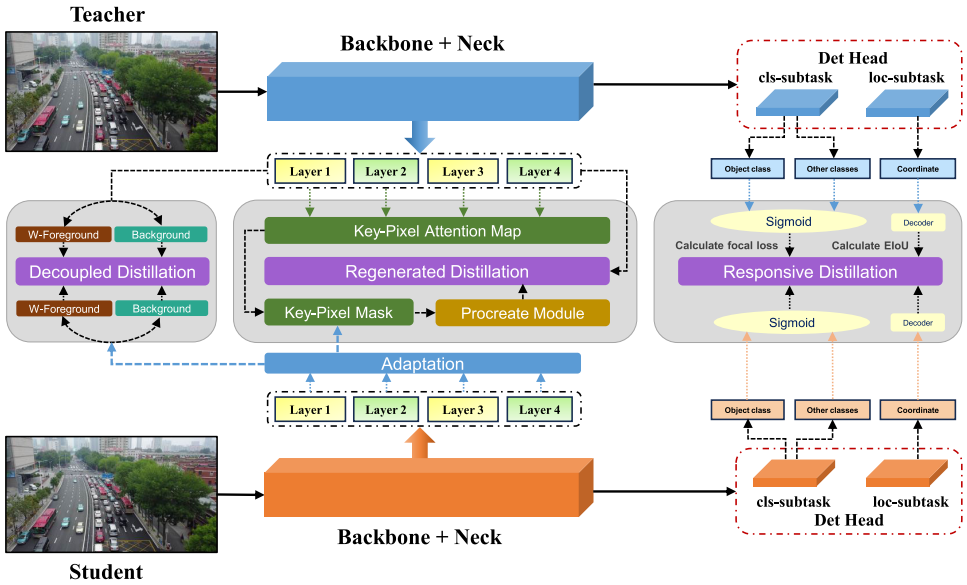

Fig. 1. Overall framework of the proposed RDR-KD. It consists of three distillation modules regenerated distillation, decoupled distillation, and responsive distillation. Blue and orange represent the feature information and logit predictions of the teacher and student models, respectively.

specifically for drone scenes to tackle the aforementioned weaknesses. The objective is to integrate logit-based and feature-based KD to maximize the DOD accuracy of the lightweight student, especially for tiny objects. In a nutshell, our contributions are listed as follows.

1) To encourage the student model to reconstruct features according to the teacher model, a Regenerated Distillation is designed using a key-pixel mask.
2) A Decoupled Distillation employing area weighting is established for forcing the student to pay more attention to the tiny object regions, thereby reducing the influence from other regions when mimicking the teacher features.
3) We develop a Responsive Distillation based on both the classification and localization prediction, which can help the student alleviate class imbalance.

II. Methodology

The overall structure of RDR-KD is shown in Fig. 1. It comprises three components: Regenerated Distillation, Decoupled Distillation, and Responsive Distillation. The first two are feature-based distillation and the last one is logit-based distillation. Based on the same feature pyramid network (FPN) structure and input image resolution, the teacher and student have corresponding multiscale feature layers F^t and F^s. To fully obtain the feature knowledge of objects at different scales, we use the feature maps of Layer 1/2/3/4, and their downsampling rates are 4/8/16/32, respectively. Referring to [13], Adaptation is utilized for aligning the dimension of the feature map in student to teacher. In Regenerated Distillation, we obtain a key-pixel attention map from the feature map of the teacher and generate a spatial and channel mask to cover the key-pixel in the student feature map. Then, a more robust feature representation is regenerated with the assistance of the procreate module (PM). In Decoupled Distillation, we decouple the foreground and background regions while weighting the foreground examples according to the object area. In Responsive Distillation, we employ focal loss and efficient intersection over union (EIoU) loss for the classification subtask and localization subtask respectively, ensuring abundant transfer of the teacher prediction.

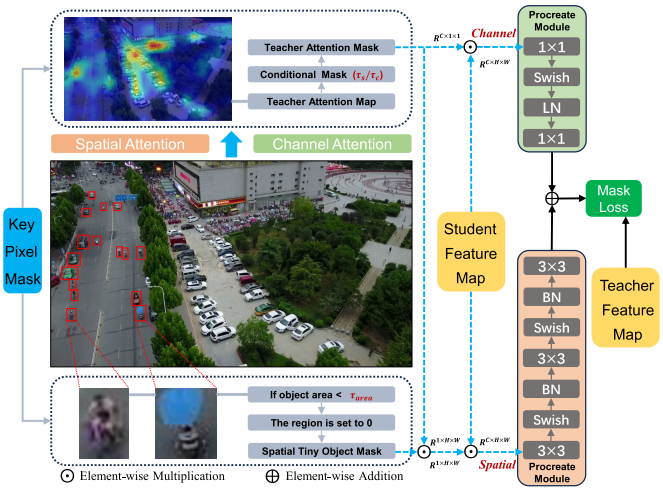

Fig. 2. Illustration of the proposed regenerated distillation.

A. Masked Distillation Based on Key-Pixel Regeneration

As illustrated in Fig. 2, we use a dual attention mechanism mask and a tiny object area mask to jointly build up the key-pixel mask. First, the teacher attention maps are acquired by spatial attention and channel attention, which can be formulated as follows:

$$A^s = H \cdot W \cdot \text{softmax}\left(\frac{\sum_{k=1}^{C}\left|F_k^t\right|}{TC}\right) \tag{1}$$

$$A^c = C \cdot \text{softmax}\left(\frac{\sum_{i=1}^{H}\sum_{j=1}^{W}\left|F_{i,j}^t\right|}{THW}\right) \tag{2}$$

where $A^s \in \mathbb{R}^{1\times H\times W}$ and $A^c \in \mathbb{R}^{C\times 1\times 1}$ denote the spatial and channel attention maps, respectively. H, W, C are the shape

of F^t. T is the temperature to adjust the distribution. Meanwhile, we observe that the spatial attention values for most tiny objects are relatively low, so they are easily overlooked when regenerating feature maps. To force the student to focus more on the representation of tiny objects during the reconstruction process, we introduce an area judgment condition when generating the spatial attention mask. Therefore, the mask used in the student feature map can be represented in two parts

$$M^s_{i,j} = \begin{cases} 0, & \text{if } A^s_{i,j} > \tau_s \text{ or } S_{\text{bbox}} < \tau_{\text{area}} \\ 1, & \text{Otherwise} \end{cases} \tag{3}$$

$$M^c_k = \begin{cases} 0, & \text{if } A^c_k > \tau_c \\ 1, & \text{Otherwise} \end{cases} \tag{4}$$

where $A^s_{i,j}$ represents the spatial attention value at coordinates (i, j), and A^c_k represents the attention value of the kth channel. $M^s \in \mathbb{R}^{1\times H\times W}$ and $M^c \in \mathbb{R}^{C\times 1\times 1}$ denote the spatial and channel masks, respectively. Three hyperparameters are employed simultaneously to control the mask proportions: spatial attention threshold τ_s, channel attention threshold τ_c, and area threshold τ_{area}. Specifically, when the area of a certain object S_{bbox} is less than τ_{area}, its corresponding region on the spatial mask is set to 0 so that it will participate in feature regeneration. Then, we overlay the student feature map F^s with M^s and M^c, and the student's masked feature is regenerated using spatial PM Ω^s_{PM} and channel PM Ω^c_{PM}, which can be formulated as follows:

$$F^s_{\text{new}} = \Omega^s_{PM}\left(F^s \odot M^s\right) + \Omega^c_{PM}\left(F^s \odot M^c\right) \tag{5}$$

where F^S_{new} denotes the student feature regenerated after spatial PM and channel PM. We refer to the generation block in MGD [13] and redesign our PM module, the specific flow is shown in Fig. 2. Finally, the mask loss function is as follows:

$$L^{\text{dis}}_{\text{mask}} = \sum_{h=1}^{H}\sum_{w=1}^{W}\sum_{c=1}^{C}\left(F^t(h,w,c) - F^s_{\text{new}}(h,w,c)\right)^2. \tag{6}$$

B. Decoupled Distillation Based on Foreground Weighting

In Regenerated Distillation, the mask of the student feature leads to the lack of original feature information. Therefore, the feature-based mimicry distillation should be appropriately retained. As is well known, the background occupies a substantial proportion of the pixels in drone images. However, the knowledge contained in the background is not as vital as in the foreground. If both types of knowledge are taught to students equally, this will affect the final learning outcome.

To compensate for the missing original information in Regenerated Distillation and distinguish the importance of various foreground knowledge, we propose a decoupled distillation based on foreground weighting. As shown in Fig. 3, we first generate a binary mask according to the ground-truth boxes and use them to decouple the foreground and background regions, obtaining the foreground feature maps $F^t_{\text{fore}}/F^s_{\text{fore}}$ and the background feature maps $F^t_{\text{back}}/F^s_{\text{back}}$. Next, we design an **Area-softmax** function with temperature $T_s = 20$ and bias $b = 0.005$ to establish an inverse proportionality between the object area and the corresponding distillation

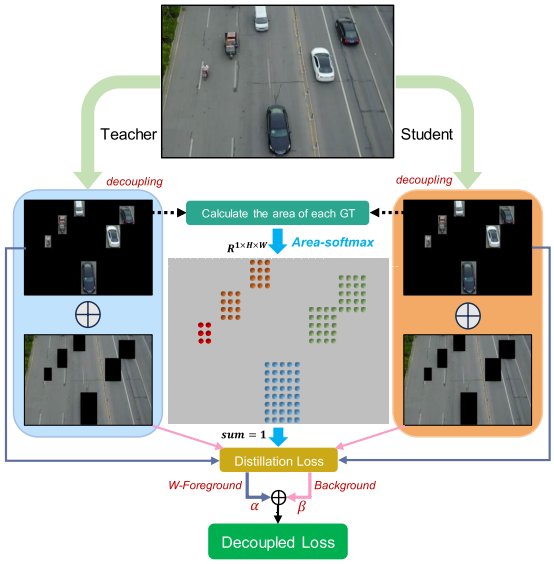

Fig. 3. Illustration of the proposed decoupled distillation.

loss weights

$$\text{softmax}(x_i, T_s, b) = \frac{e^{(1-x_i-b)/T_s}}{\sum_j e^{(1-x_j-b)/T_s}}. \tag{7}$$

For foreground distillation loss, we calculate a weighting matrix $W^{\text{area}} = \text{softmax}(S, T_s, b)$, where S indicates the area of each object. Each object region receives a corresponding weight after Area-softmax. Finally, the decoupled loss function is as follows:

$$\begin{aligned} L^{\text{dis}}_{\text{decouple}} = \alpha \cdot & \sum_{h=1}^{H}\sum_{W=1}^{W}\sum_{c=1}^{C}\cdot\left(W^{\text{area}}_q / S_q\right) \\ & \cdot\left(F^t_{\text{fore}}(h,w,c) - F^s_{\text{fore}}(h,w,c)\right)^2 \\ & + \beta \cdot \sum_{h=1}^{H}\sum_{W=1}^{W}\sum_{c=1}^{C}\left(F^t_{\text{back}}(h,w,c) - F^s_{\text{back}}(h,w,c)\right)^2 \end{aligned} \tag{8}$$

where α and β are the hyperparameters to balance the loss between foreground and background. W^{area}_q and S_q denote the qth object region weight and the qth object area, respectively. W^{area}_q/S_q denotes the weight at each point of the qth object region and $\sum_{q=1}^{Q}((W^{\text{area}}_q/S_q)\cdot S_q) = 1$. Q is the number of object regions in an image.

C. Responsive Distillation Based on Focal Loss and EIoU

In general, KD in object detection has primarily focused on feature-based distillation, which ignores the knowledge embedded in the probability distribution of the model outputs. To maximize the prediction information from the teacher, inspired by Bridging Cross-task Protocol Inconsistency for Knowledge Distillation (BCKD) [16], we further propose a Responsive Distillation based on focal loss and EIoU. Due to the extreme imbalance of foreground categories in drone scenes, we treat classification logit outputs as multiple binary classification outputs during distillation. Specifically, l^s and l^t are the logits of the student and teacher, respectively. P is the number of anchors or points and N is the number of foreground categories. We calculate $l^{s'} = \text{sigmoid}(l^s)$ and

$l^{t\prime} = \text{sigmoid}(l^t)$. $l^{s\prime}$ and $l^{t\prime}$ are binary classification scores of the size $P \times N$. On this basis, a binary classification is performed for each object category, and the distillation loss for the classification subtask is calculated by using focal loss [17]

$$L_{\text{cls}}^{\text{dis}} = \sum_{i=1}^{P} \sum_{j=1}^{N} L_{\text{focal}}\left(l_{i,j}^{s\prime}, l_{i,j}^{t\prime}\right). \tag{9}$$

In the localization subtask, we leverage the intersection over union (IoU) between the teacher and the student. Specifically, the predicted bounding boxes of the teacher and student model are obtained as $\text{bbox}_i^t = \text{Decoder}(\mathbb{A}_i, c_i^t)$ and $\text{bbox}_i^s = \text{Decoder}(\mathbb{A}_i, c_i^s)$, respectively. $\mathbb{A}_i$ denotes the ith anchor. c_i^t and c_i^s represent the coordinate predictions at the ith position from the teacher and the student, respectively. Then, EIoU [18] between bbox_i^t and bbox_i^s is calculated as follows:

$$L_{\text{loc}}^{\text{dis}} = \sum_{i=1}^{P} \text{PEIoU}\left(\text{bbox}_i^t, \text{bbox}_i^s\right). \tag{10}$$

Therefore, the responsive distillation loss can be computed as $L_{\text{responsive}}^{\text{dis}} = L_{\text{cls}}^{\text{dis}} + L_{\text{loc}}^{\text{dis}}$. In summary, we train the student with the holistic loss as follows:

$$L_{\text{holistic}} = L^{\text{ori}} + L_{\text{mask}}^{\text{dis}} + L_{\text{decoupled}}^{\text{dis}} + L_{\text{responsive}}^{\text{dis}} \tag{11}$$

where L^{ori} is the original training loss.

III. Experiments

A. Experimental Setup

1) Dataset and Metrics: We evaluate the proposed RDR-KD on VisDrone2019. It comprises 10 209 images (6471 for training, 548 for validation, 3190 for testing) with 10 object categories. The presence of diverse angles, scales, and dense tiny objects makes high-precision detection exceptionally challenging. We employ the COCO evaluation protocol to assess the detection performance, which includes different IoU thresholds and object scales. Specifically, AP is computed by averaging over all categories, and AP_{50} and AP_{75} are computed at a single IoU threshold of 0.5 and 0.75, respectively. Furthermore, we employ AP_S, AP_M, and AP_L to evaluate the detection performance on tiny, medium, and large objects.

2) Implementation Details: All relevant distillation experiments are implemented using the Ultralytics framework with PyTorch. We train all the models with stochastic gradient descent (SGD) optimizer, where the momentum is 0.937 and the weight decay is 0.0005. There are five hyperparameters in the distillation process, and we adopt $\{\tau_s = 1.0, \tau_c = 1.0, \tau_{\text{area}} = 500, \alpha = 0.5, \beta = 0.5\}$ for the experimented model. Meanwhile, to balance three distillation losses, we use 1×10^{-5}, 1×10^{-1}, and 1×10^{-5}, respectively, to weight $L_{\text{mask}}^{\text{dis}}$, $L_{\text{decoupled}}^{\text{dis}}$, and $L_{\text{responsive}}^{\text{dis}}$. All experiments are performed on an NVIDIA RTX 3070 GPU.

B. Comparisons With State-of-the-Art KD Methods

We compare RDR-KD with recent state-of-the-art KD methods on VisDrone2019, including feature-based methods, logit-based methods, and the combination. As shown in Table I, RDR-KD almost achieves the best results. The advantages of RDR-KD against the most powerful feature-based AWD and logit-based CrossKD are evidenced by respective further performance gains of 1.0% and 0.7%. To validate the effectiveness of RDR-KD in enhancing the detection accuracy of tiny objects, we focus on AP_S. It can be observed that our framework improves AP_S from 9.7% to 12.6%. This proves that the student can learn more tiny object feature information from the teacher compared to other frameworks. Meanwhile, due to Decoupled Distillation, increasing the distillation loss weights for tiny objects results in decreased weights for large objects. The improvement in AP_L is not significant. Overall, the above results demonstrate the superiority of our RDR-KD among all compared distillation frameworks.

TABLE I
RESULTS OF DIFFERENT DISTILLATION METHODS ON VisDrone2019

Type	Method	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
-	*Baseline*	18.8	32.0	19.0	9.7	29.7	36.7
Feature	CWD [12]	19.4(+0.6)	33.3	19.1	9.7	30.1	44.7
	MGD [13]	19.7(+0.9)	33.5	19.7	10.0	30.8	44.9
	FGD [19]	20.0(+1.2)	33.9	20.0	10.5	31.0	46.4
	PKD [14]	20.2(+1.4)	34.5	20.2	10.5	31.4	**48.5**
	AWD [20]	21.1(+2.3)	35.9	21.1	11.5	32.2	42.8
Logit	WSLD [8]	20.1(+1.3)	34.4	20.0	10.3	31.1	47.4
	LD [21]	20.5(+1.7)	34.7	20.8	11.5	31.0	45.4
	DKD [9]	20.6(+1.8)	35.1	20.8	11.6	31.1	45.5
	NKD [10]	21.0(+2.2)	35.8	21.3	12.4	31.1	40.8
	CrossKD [22]	21.4(+2.6)	36.3	21.3	11.7	32.5	43.0
Feature + Logit	CWD + WSLD	21.2(+2.4)	36.2	21.1	11.9	31.7	44.5
	MGD + WSLD	21.4(+2.6)	36.4	21.4	11.7	32.6	44.9
	MGD + DKD	21.6(+2.8)	36.7	21.8	12.1	32.5	45.6
	PKD + NKD	21.6(+2.8)	36.7	21.6	11.9	32.7	42.0
	PKD + DKD	21.7(+2.9)	36.7	21.8	12.4	32.6	46.1
	RDR-KD	**22.1(+3.3)**	**37.3**	**22.3**	**12.6**	**33.1**	45.6

TABLE II
ABLATION STUDY ON VisDrone2019

Regenerated (Feature)	Decoupled (Feature)	Responsive (Logit)	Teacher: YOLOv8l Student: YOLOv8n AP	AP_S	AP_L
-	-	-	18.8	9.7	36.7
✓	-	-	20.9	12.3	35.7
-	✓	-	21.0	12.0	41.4
✓	✓	-	21.7	12.2	43.3
-	-	✓	21.2	11.8	42.5
✓	✓	✓	**22.1**	**12.6**	**45.6**

C. Ablation Studies

To assess the effectiveness of the proposed three distillation modules, we conduct ablation experiments using YOLOv8 to systematically evaluate the impact of each individual module. The results are presented in Table II. We primarily discuss two aspects: whether Regenerated Distillation and Decoupled Distillation improve the detection performance of tiny objects, and whether Responsive Distillation is worth adopting. First, we individually retain only one distillation module in the framework. The results indicate that Regenerated Distillation and Decoupled Distillation are indeed effective in improving tiny object detection accuracy, and AP_S increases by 2.6% and 2.3%, respectively. Moreover, the combination of two distillation modules can produce superior performance and exceed the aforementioned SOTA methods based on feature and logit. Furthermore, the proposed logit-based distillation achieves excellent scores by observing the ablation study results of Responsive Distillation. Even when compared with other logit-based SOTA methods, Responsive Distillation is

TABLE III
QUANTITATIVE COMPARISON ON THE SELF-CAPTURED DATASET

-	Baseline	PKD + DKD	**RDR-KD**
AP	19.1	19.7	**24.9**
AP_{50}	28.8	31.2	**37.7**
AP_{75}	21.2	22.7	**29.3**

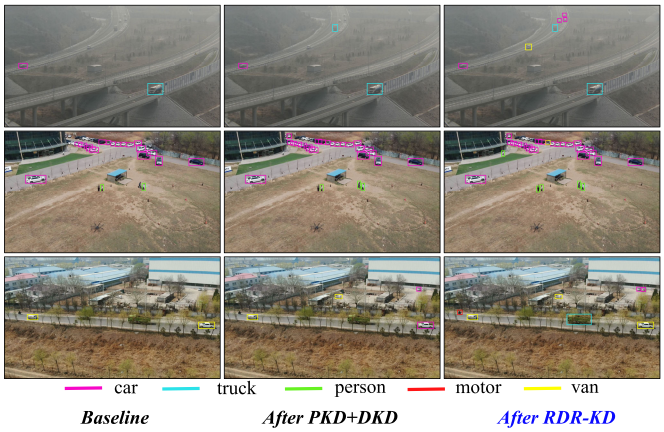

Fig. 4. Qualitative comparison on the self-captured dataset.

still extremely competitive. Finally, the cooperation of the three distillation modules can further facilitate the student to learn more features and logit information, resulting in optimal performance.

D. Dual Experiments on the Self-Captured Dataset

To more directly demonstrate the effectiveness of RDR-KD in drone scenes, we use a DJI-M300-RTK drone to capture a compact DOD dataset consisting of 79 images in the suburbs of Beijing. The dataset was captured at altitudes ranging from 50 to 120 m, and the image resolution is 1920 × 1080. We adopt the same category settings as VisDrone2019 and use YOLOv8 as the detection algorithm. The results on the self-captured dataset are shown in Table III. Some of the visualization results are shown in Fig. 4. Both quantitative and qualitative experimental results from the self-captured dataset reveal that our proposed framework exhibits a significant advantage in handling drone images.

IV. CONCLUSION

In this letter, we propose a novel KD framework for object detection, which is specifically applied to drone scenes. Our RDR-KD comprises three distillation modules and integrates both feature-based and logit-based distillation. First, we design Regenerated Distillation based on key pixels to mask and reconstruct the feature map of the student. Then, Decoupled Distillation with area weighting is further introduced to enhance the learning capability of the student for the tiny objects and compensate for the lost feature information in the preceding module. Finally, we devise Responsive Distillation using focal loss and EIoU to efficiently deliver classification and localization prediction information. The experimental results demonstrate the effectiveness of the proposed RDR-KD framework, particularly in improving tiny object detection performance in drone scenes. Therefore, we believe that our distillation framework can stimulate further research to tackle the DOD problem and the proposed method is also effective for other aerial detection platforms.

REFERENCES

[1] Z. Qadir, K. Le, V. Nguyen Quoc Bao, and V. W. Y. Tam, "Deep learning-based intelligent post-bushfire detection using UAVs," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, pp. 1–5, 2024.
[2] Z. Shao, G. Cheng, J. Ma, Z. Wang, J. Wang, and D. Li, "Real-time and accurate UAV pedestrian detection for social distancing monitoring in COVID-19 pandemic," *IEEE Trans. Multimedia*, vol. 24, pp. 2069–2083, 2021.
[3] H. Huang et al., "Object-based attention mechanism for color calibration of UAV remote sensing images in precision agriculture," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4416013.
[4] Y. Zhang, C. Wu, T. Zhang, Y. Liu, and Y. Zheng, "Self-attention guidance and multiscale feature fusion-based UAV image object detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.
[5] X. Yang et al., "An efficient detection framework for aerial imagery based on uniform slicing window," *Remote Sens.*, vol. 15, no. 17, p. 4122, Aug. 2023.
[6] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," 2015, *arXiv:1503.02531*.
[7] J. Guo et al., "Distilling object detectors via decoupled features," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2154–2164.
[8] H. Zhou et al., "Rethinking soft labels for knowledge distillation: A bias-variance tradeoff perspective," 2021, *arXiv:2102.00650*.
[9] B. Zhao, Q. Cui, R. Song, Y. Qiu, and J. Liang, "Decoupled knowledge distillation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 11953–11962.
[10] Z. Yang, A. Zeng, Z. Li, T. Zhang, C. Yuan, and Y. Li, "From knowledge distillation to self-knowledge distillation: A unified approach with normalized loss and customized soft labels," 2023, *arXiv:2303.13005*.
[11] A. Romero, N. Ballas, S. Ebrahimi Kahou, A. Chassang, C. Gatta, and Y. Bengio, "FitNets: Hints for thin deep nets," 2014, *arXiv:1412.6550*.
[12] C. Shu, Y. Liu, J. Gao, Z. Yan, and C. Shen, "Channel-wise knowledge distillation for dense prediction," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 5311–5320.
[13] Z. Yang, Z. Li, M. Shao, D. Shi, Z. Yuan, and C. Yuan, "Masked generative distillation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2022, pp. 53–69.
[14] W. Cao, Y. Zhang, J. Gao, A. Cheng, K. Cheng, and J. Cheng, "Pkd: General distillation framework for object detectors via Pearson correlation coefficient," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 35, 2022, pp. 15394–15406.
[15] G. Chen, W. Choi, X. Yu, T. Han, and M. Chandraker, "Learning efficient object detection models with knowledge distillation," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 742–751.
[16] L. Yang et al., "Bridging cross-task protocol inconsistency for distillation in dense object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 17175–17184.
[17] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.
[18] Y.-F. Zhang, W. Ren, Z. Zhang, Z. Jia, L. Wang, and T. Tan, "Focal and efficient IOU loss for accurate bounding box regression," 2021, *arXiv:2101.08158*.
[19] Z. Yang et al., "Focal and global knowledge distillation for detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 4643–4652.
[20] G. Yang, Y. Tang, J. Li, J. Xu, and X. Wan, "AMD: Adaptive masked distillation for object detection," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jun. 2023, pp. 1–8.
[21] Z. Zheng et al., "Localization distillation for dense object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 9407–9416.
[22] J. Wang, Y. Chen, Z. Zheng, X. Li, M.-M. Cheng, and Q. Hou, "CrossKD: Cross-head knowledge distillation for object detection," 2023, *arXiv:2306.11369*.