Extracting and Distilling Direction-Adaptive Knowledge for Lightweight Object Detection in Remote Sensing Images

Zhanchao Huang, Wei Li, Ran Tao Beijing Institute of Technology

E-mail: zhanchao.h@outlook.com

Paper ID: #8521

icassp 2022 Singapore

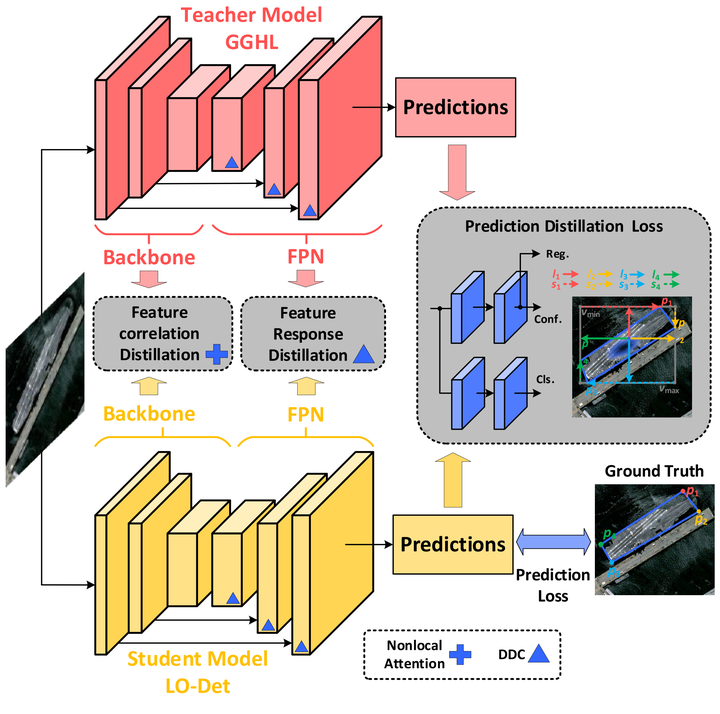
Knowledge Distilling for Object Detection in Remote Sensing Images

◆ For Spaceborne or Airborne Online Remote Sensing Object Detection (RSOD) Tasks:

- Limited Computing Resources → Lightweight CNN Models
- Lightweight CNN Models → Performance Drop
- Performance Compensation → Knowledge Distillation

◆ Our Contributions:

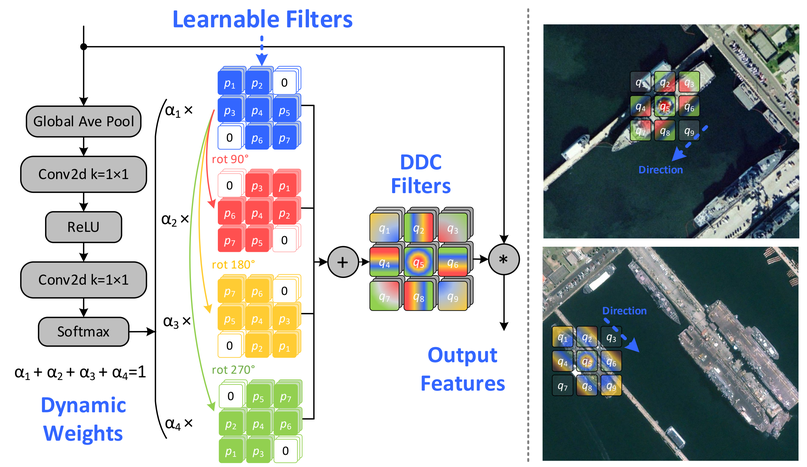
- **Propose a direction-adaptive knowledge extraction and distillation (DKED) method that compensates for the performance penalty of lightweight RSOD models;**
- **The proposed DKED distills arbitrary-oriented feature responses & correlations & predictions of objects in remote sensing images.**



Knowledge Distillation of DKED

◆ Dynamic Directional Convolution (DDC)

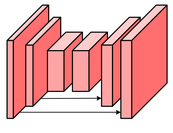
Because of the top-down perspective of remote sensing images, extracting arbitrary oriented features adaptively is important.



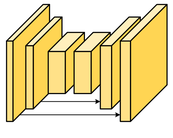
- **DDC dynamically generates convolutional filters that match the optimal orientation based on the input.**

◆ Direction-Adaptive Knowledge Extraction and Distillation (DKED) Framework

1) Teacher & Student Models for Distilling



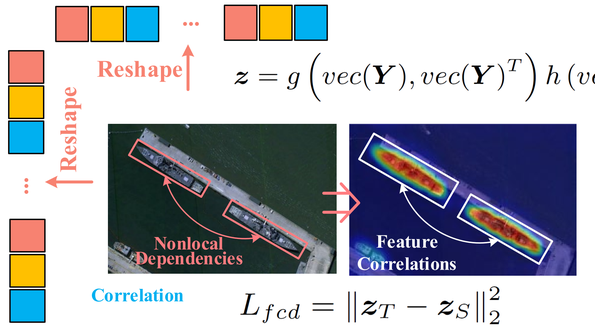
Teacher Model: GGHL (Huang et al., TIP 2022)



Student Model: LO-Det (Huang et al., TGRS 2022)

2) Feature Correlation Distillation

Distill the non-local correlation features extracted by the backbone CNNs.



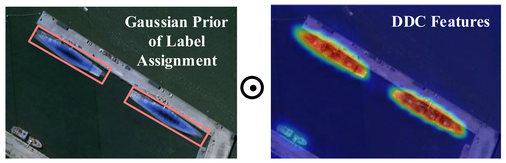
$$z = g\left(vec(\boldsymbol{Y}), vec(\boldsymbol{Y})^T\right) h\left(vec(\boldsymbol{Y})\right)$$

$$L_{fcd} = \|\boldsymbol{z}_T - \boldsymbol{z}_S\|_2^2$$

Used in the auxiliary module of DKED framework without introducing additional inference costs

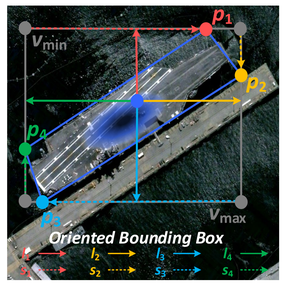
3) Feature Response Distillation

$$L_{drfd} = \left\|\boldsymbol{G} \odot \left(\boldsymbol{Y}_T' - \boldsymbol{Y}_S'\right)\right\|_2^2$$



Distill the direction-adaptive feature responses in the FPN of CNNs.

4) Prediction Distillation



Distill the predictions of detection heads in CNN models.

$$L_{obb}\left(\boldsymbol{r}_T, \boldsymbol{r}_S, \boldsymbol{r}_{GT}\right) = \begin{cases} JOL_{obb}\left(\boldsymbol{r}_S, \boldsymbol{r}_{GT}\right), \\ \text{if } JOL_{obb}\left(\boldsymbol{r}_S, \boldsymbol{r}_{GT}\right) + \delta \\ > JOL_{obb}\left(\boldsymbol{r}_T, \boldsymbol{r}_{GT}\right) \\ 0, \quad otherwise \end{cases}$$

$$L_{loc} = JOL_{obb}\left(\boldsymbol{r}_S, \boldsymbol{r}_{GT}\right) + \lambda \times L_{obb}\left(\boldsymbol{r}_T, \boldsymbol{r}_S, \boldsymbol{r}_{GT}\right)$$

$$L_{cls} = \lambda \times JOL_{cls}\left(\boldsymbol{c}_S, \boldsymbol{c}_{GT}\right) + (1-\lambda) \times JOL_{cls}\left(\boldsymbol{c}_T, \boldsymbol{c}_S\right)$$

Use the predictions of teacher model as soft labels.

The Representation of Predictions: GGHL (Huang et al., TIP 2022)

Experimental Results

◆ Comparative Results w/wo Using DKED

Methods	mAP @HRSC	mAP @DOTA	Speed (fps) @RTX3090	Speed (fps) @TX2	Speed (fps) @AGX Xavier
LO-Det	80.05	66.17	62.12	7.34	23.51
DKED + LO-Det	85.17(+5.12)	71.39(+5.22)	62.16	7.38	23.51
GGHL	87.30	76.95	42.39	4.67	14.72
DKED + GGHL	93.45(+6.15)	77.43(+0.48)	42.39	4.69	14.73

- **DKED improves the detection performance of lightweight model without introducing any inference costs;**
- **DKED can also be used for self-distillation of large models to further improve detection performance.**

◆ Detection Results on HRSC & DOTA Datasets