Semantic Segmentation of High-resolution Aerial Images

Alena Tesařová (xtesar36)

Pavel Kohout (xkohou15)

Petr Kohout (xkohou14)

Our task 1: A Dual-Path Convolutional Neural Network for High-Resolution Aerial Image Segmentation

- Motivation
    - more and more high resolution aerial images (urban planning, environment monitoring etc.)
- Challenges
    - intra and inter variability
    - computational efficiency
- Proposed solution
    - dual path DCNN
        - **Intra-class heterogeneity** - **spatial** path to use information from global context
        - **Inter-class homogeneity** - **edge** path to create semantic boundaries to detect changes between similar objects
    - further modification of **MobileNetV2** to optimize performance

Article

A Dual-Path and Lightweight Convolutional Neural Network for High-Resolution Aerial Image Segmentation

Gang Zhang [1,2], Tao Lei [1,*], Yi Cui [1] and Ping Jiang [1]

1 Institute of Optics and Electronics, Chinese Academy of Sciences, P.O. Box 350, No.1 Guangdian Avenue, Chengdu 610209, China; zhanggang@ioe.ac.cn (G.Z.); cuiyi@ioe.ac.cn (Y.C.); jiangping@ioe.ac.cn (P.J.)
2 University of Chinese Academy of Sciences, No. 19 (A) Yuquan Road, Beijing 100049, China
* Correspondence: taoleiyan@ioe.ac.cn

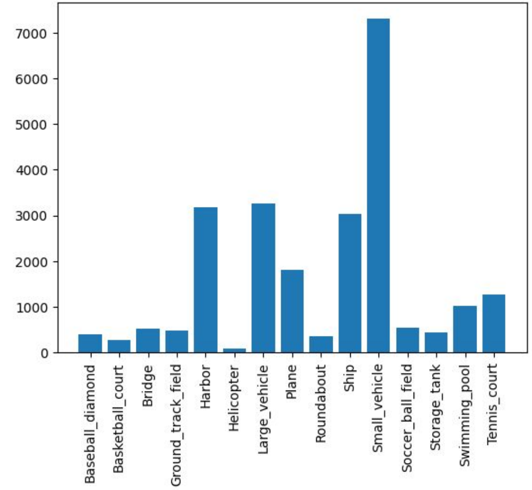
Received: 7 October 2019; Accepted: 9 December 2019; Published: 12 December 2019

check for updates

Abstract: Semantic segmentation on high-resolution aerial images plays a significant role in many remote sensing applications. Although the Deep Convolutional Neural Network (DCNN) has shown great performance in this task, it still faces the following two challenges: intra-class heterogeneity and inter-class homogeneity. To overcome these two problems, a novel dual-path DCNN, which contains a spatial path and an edge path, is proposed for high-resolution aerial image segmentation. The spatial path, which combines the multi-level and global context features to encode the local and global information, is used to address the intra-class heterogeneity challenge. For inter-class homogeneity problem, a Holistically-nested Edge Detection (HED)-like edge path is employed to detect the semantic boundaries for the guidance of feature learning. Furthermore, we improve the computational efficiency of the network by employing the backbone of MobileNetV2. We enhance the performance of MobileNetV2 with two modifications: (1) replacing the standard convolution in the last four Bottleneck Residual Blocks (BRBs) with atrous convolution; and (2) removing the convolution stride of 2 in the first layer of BRBs 4 and 6. Experimental results on the ISPRS Vaihingen and Potsdam 2D labeling dataset show that the proposed DCNN achieved real-time inference speed on a single GPU card with better performance, compared with the state-of-the-art baselines.

Our task 2: iSAID dataset

- **total of 15 categories**: Plane, Ship, Storage tank, Baseball diamond, Tennis court, Basketball court, Ground track field, Harbor, Bridge, Large vehicle, Small vehicle, Helicopter, Roundabout, Swimming pool, Swimming pool
- **2806** high resolution images
- **uneven** class distribution
- must be preprocessed



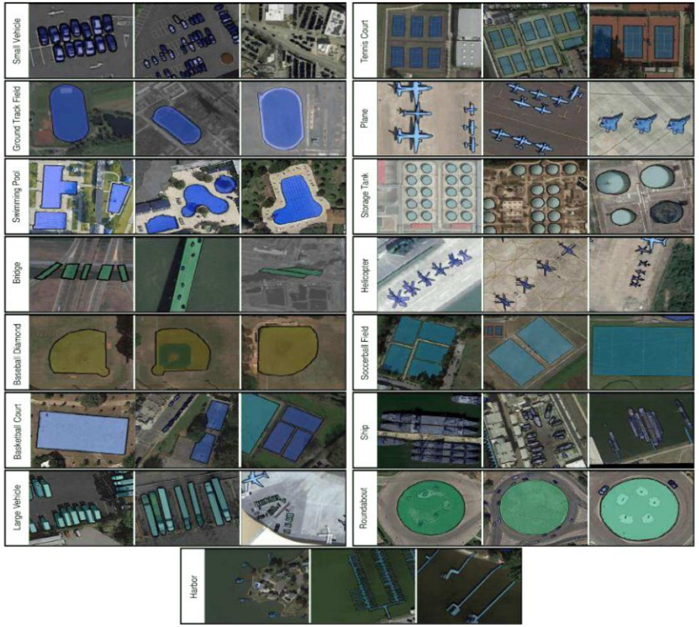
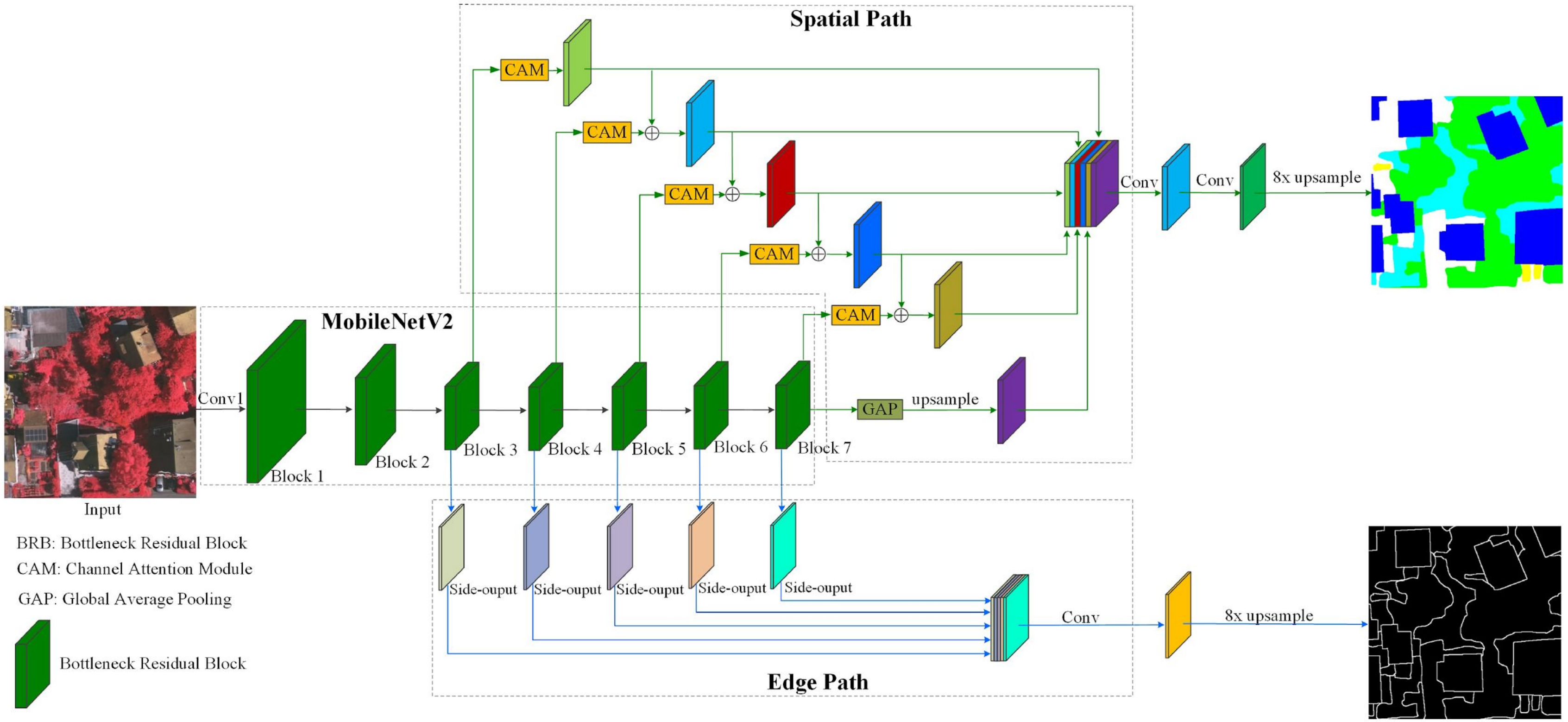


Figure 8: Samples of annotated images in iSAID.

Implemented solution



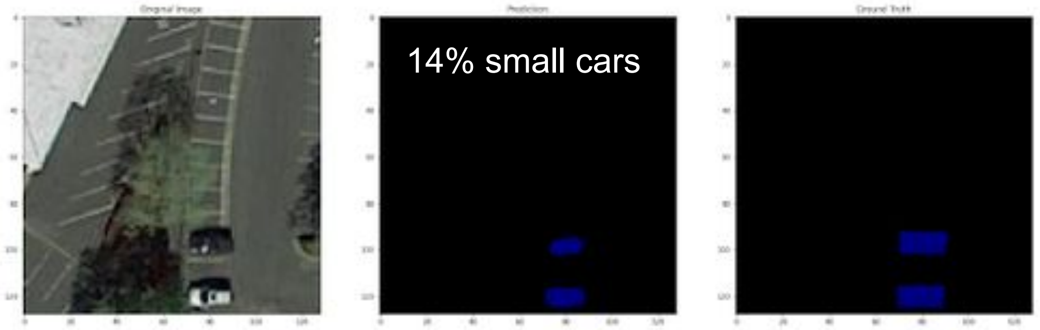
Experiment settings

- Image sizes 512x512 and 128x128
- Adam optimizer, decay weight factor, learning rate 0.001 (adaptive)
- Loss: Cross-entropy, Mean square error, Dice loss
- Epochs 50 - 75, batch sizes 2 - 10
- Final loss function composed of 2 separated loss functions
    - Spatial path loss, edge path loss, coefficient setup

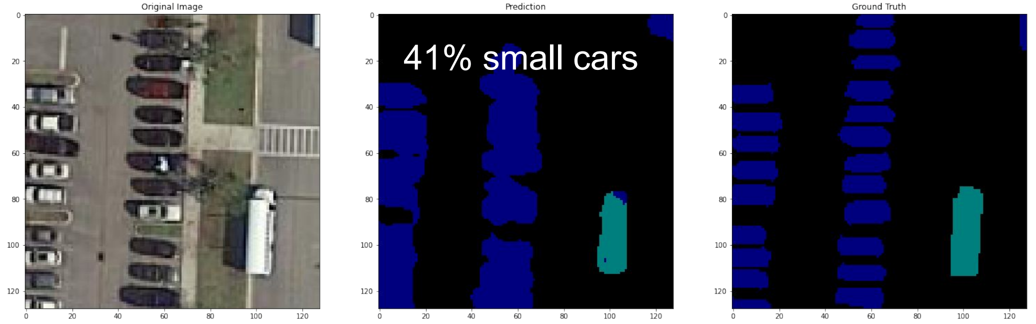
$$L_{total} = \alpha \times L_{spatial} + \beta \times L_{edge}$$

Our results

Our implementation of Dual-path



Baseline solution FPN (Feature Pyramid Network)



Experiments a results

Method	AP	Plane	BD	Bridge	GTF	SV	LV	Ship	TC	BC	ST	SBF	RA	Harbor	SP	HC
Mask R-CNN	25.7	37.7	42.5	13.0	23.6	6.9	7.4	26.6	54.9	34.6	28.8	20.8	35.9	22.5	25.1	5.3
Mask R-CNN+	33.4	41.7	39.6	15.2	25.9	16.9	30.4	48.8	72.9	43.1	32.0	26.7	36.0	29.6	36.7	5.6
PANet	34.2	39.2	45.5	15.1	29.3	15.0	28.8	45.9	64.1	47.4	29.6	33.9	36.9	26.3	36.1	9.5
PANet++	40.0	48.7	50.3	18.9	32.5	20.4	34.4	56.5	78.4	52.3	35.4	38.8	40.2	35.8	42.5	13.7
FPN	6.4	0.0	0.0	0.0	0.0	41.1	28.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
SpatialPath	4.7	0.0	0.0	0.0	0.0	12.5	5.6	0.03	0.0	0	5.2	0.0	0.0	28.3	0.0	0.0
SpatialEdgePath	3.6	7.4	0.0	0.0	0.0	14.0	15.8	1.88	0.0	0	0.01	0.0	0.0	0.28	0.0	0.0

Table 1: Class-wise instance segmentation results on iSAID test set. Note that short names are used to define categories: BD-Baseball diamond, GTF-Ground field track, SV-Small vehicle, LV-Large vehicle TC-Tennis court, BC-Basketball court, ST-Storage tank, SBF-Soccer-ball field, RA-Roundabout, SP-Swimming pool, and HC-Helicopter

Thank you for your attention