Weekly Report on Road analytics

Faculty: Mehul S. Raval

Date : 17/2/22 - 23/2/22

Member: Yagnik Bhavsar

En. No: AU2149006

Outline of peformed task :

- Literature survey
- List out different approaches for vehicle detection and classification tasks

Literature Survey:

[1] AU-AIR: A Multi-modal Unmanned Aerial Vehicle Dataset for Low Altitude Traffic Surveillance

AU-AIR a multi-purpose aerial dataset that has multi-modal sensor data (i.e., visual, time, location, altitude, IMU, velocity) collected in real-world outdoor environments. Comparision with other UAV datasets are given below,

TABLE I
COMPARISON WITH EXISTING UAV DATASETS.

Dataset	Environment	Data type	Visual data	Object annotations	Time	GPS	Altitude	Velocity	IMU data
VisDrone [7]	outdoor	real	yes	yes	no	no	no	no	no
UAVDT [8]	outdoor	real	yes	yes	no	no	partial	no	no
CARPK [9]	outdoor	real	yes	yes	no	no	no	no	no
Stanford [10]	outdoor	real	yes	yes	no	no	no	no	no
UAV123 [11]	outdoor	synthetic	yes	yes	no	no	no	no	no
VIVID [12]	outdoor	real	yes	yes	no	no	no	no	no
highD [13]	outdoor	real	yes	yes	no	no	no	no	no
Mid-Air [20]	outdoor	synthetic	yes	no	yes	yes	yes	yes	yes
Blackbird [21]	indoor	real	yes	no	yes	yes	yes	yes	yes
EuRoC MAV [22]	indoor	real	yes	no	yes	yes	yes	yes	yes
Zurich Urban MAV [23]	outdoor	real	yes	no	yes	yes	yes	yes	yes
UPenn Fast Flight [24]	outdoor	real	yes	no	yes	yes	yes	yes	yes
AU-AIR	**outdoor**	**real**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**

UAV Platform and data recording style :

- Parrot Bebop 2 quadrotor
- Video Resolution of 1920 × 1080 pixels @30 fps
- Sensor data have been recorded for every 20 milliseconds
- Flight altitude -between 10 m - 30 m
- Camera angle- between 45 degrees - 90 degrees

ML Models and other specifications:

- YOLOv3-Tiny
- MobileNetv2-SSDLite
- Batch size of 32 and Adam optimizer default parameters (alpha= 0.001, beta1=0.9, beta2=0.999)

End results with different model are shown below,

TABLE III

CATEGORY-WISE AVERAGE PRECISION VALUES OF THE BASELINE NETWORKS.

Model	Training Dataset	Human	Car	Truck	Van	Motorbike	Bicycle	Bus	Trailer	mAP
YOLOV3-Tiny	AU-AIR	34.05	36.30	47.13	41.47	4.80	12.34	51.78	13.95	30.22
MobileNetV2-SSDLite	AU-AIR	22.86	19.65	34.74	25.73	0.01	0.01	39.63	13.38	19.50
YOLOV3-Tiny	COCO	0.01	0	0	n/a	0	0	0	n/a	n/a
MobileNetV2-SSDLite	COCO	0	0	0	n/a	0	0	0	n/a	n/a

[2] VAID: An Aerial Image Dataset for Vehicle Detection and Classification

It contains about 6000 images captured under different traffic conditions, and annotated with 7 common vehicle categories for network training and testing. Comparision with other UAV datasets are given below,

Dataset	Number of Images	Image Resolution	Pixel Scale	Vehicle Size
VEDAI	1,250	512 × 512 1,024 × 1,024	25cm 12.5cm	10 × 20 20 × 40
COWC	53	2,000 × 2,000 19,000 × 19,000	15cm	24 × 48
DLR-MVDA	20	5,616 × 3,744	13cm	20 × 40
KIT-AIS	241	300 – 1,800	12.5cm – 18cm	15 × 25 20 × 40
VAID	5,985	1137 × 640	12.5cm	20 × 40

UAV Platform and data recording style :

- DJI Mavic Pro
- Video Resolution of 2720 × 1530 pixels @23.98 fps
- Flight altitude -between 90 m - 95 m

ML Models and other specifications:

- Faster R-CNN
- Modified Faster R-CNN (softplus)
- Modified Faster R-CNN (ELU)
- Modified Faster R-CNN (ReLU)
- YOLOv4
- MobileNetv3
- RefineDet
- U-Net

End results with different model are shown below,

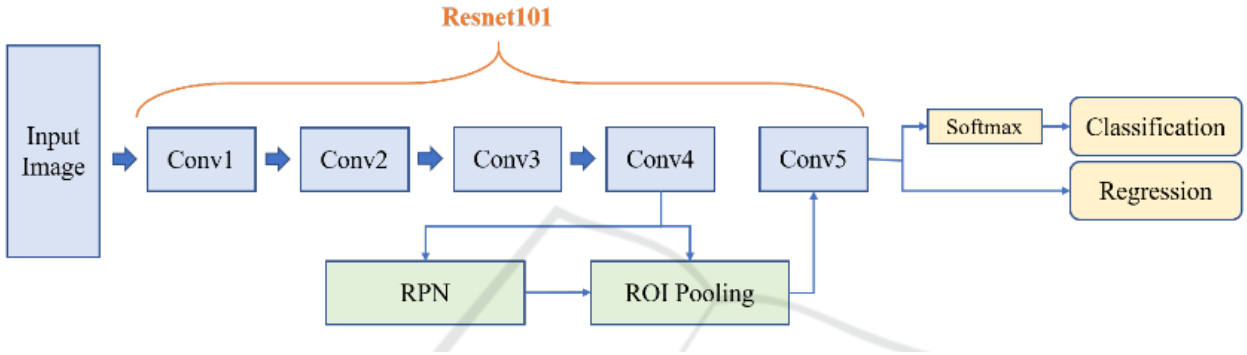
	Original Faster R-CNN	Modified Faster R-CNN	Modified Faster R-CNN (softplus)	Modified Faster R-CNN (ELU)	Modified Faster R-CNN (ReLU)
Sedan	90.0%	90.2%	90.2%	90.2%	**90.2%**
Minibus	95.0%	**97.9%**	92.2%	96.6%	95.6%
Truck	83.4%	84.9%	**87.5%**	87.1%	86.8%
Pickup Truck	76.8%	78.7%	79.3%	78.6%	**79.6%**
Bus	88.9%	89.4%	89.6%	89.8%	**90.3%**
Cement Truck	94.2%	97.6%	93.8%	97.6%	**98.1%**
Trailer	**85.6%**	83.8%	81.5%	84.7%	84.5%
Average	87.7%	88.9%	87.7%	89.2%	**89.3%**

	Modified Faster R-CNN	YOLOv4	MobileNetv3	RefineDet	U-Net
Sedan	90.22%	**98.49%**	70.46%	89.08%	67.20%
Minibu	90.80%	**96.04%**	89.02%	90.14%	94.36%
Truck	89.34%	**96.44%**	64.92%	82.21%	83.46%
Pickup Truck	**88.93%**	57.25%	75.73%	84.59%	82.80%
Bus	90.87%	**97.03%**	87.67%	90.46%	97.84%
Cement Truck	90.96%	69.94%	90.30%	80.68%	**91.24%**
Trailer	89.75%	**95.45%**	78.14%	86.64%	80.74%
mAP	90.12%	**96.91%**	73.9%	86.26%	85.38%
Precision	0.041	**0.94**	0.2818	0.2420	0.9099
Recall	**0.9755**	0.97	0.8807	0.9640	0.9016
F1 score	0.0731	**0.96**	0.4178	0.3739	0.9057

[3] Vehicle Detection and Classification in Aerial Images using Convolutional Neural Networks

ML Model : Modified Faster R-CNN
Dataset: VAID
Feature Learning model: ResNet101

Proposed modified Faster R-CNN architecture



The results from our dataset and the modified Faster R-CNN with ReLU,

[4] Vehicle Detection and Type Classification Based on CNN-SVM

ML Models : YOLOv2-tiny (vehicle detection)
CNN(AlexNet)+SVM (vehicle classification)
Dataset: BIT Datasets, UA-DETRAC dataset

Here, two-step approach is proposed for vehicle detection and classification.

For vehicle detection,

- target extraction is done using the Yolov2-tiny
- network parameter adjustment through K-means clustering during network training

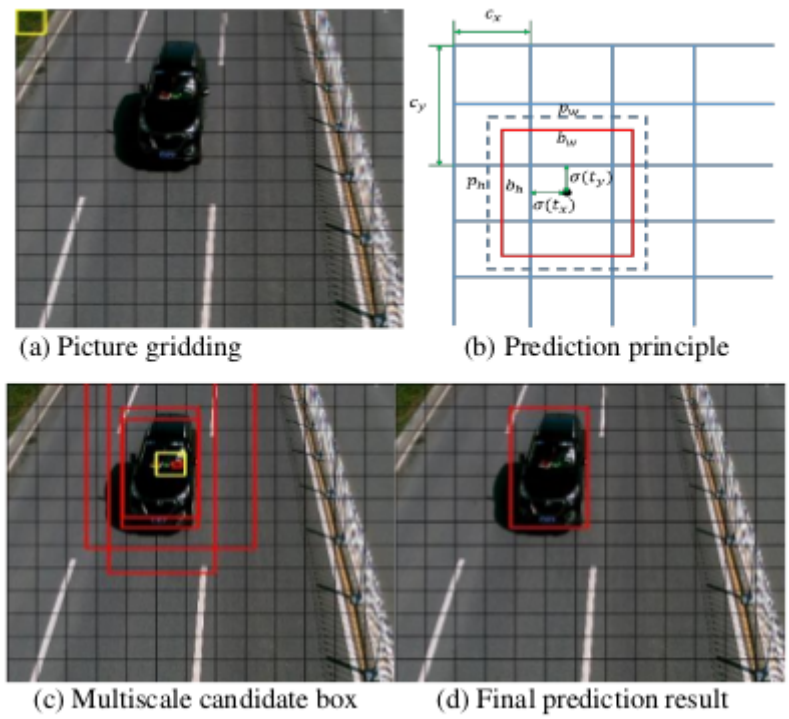


Fig. Yolov2-tiny detection procedure

For vehicle classification,

- improved CNN network is used for feature extraction to overcome poor generalization of manual feature extraction
- Modified network based on AlexNet is used for feature extraction
- SPP layer is added to solve the problem of low classification accuracy caused by image resizing and rescaling
- Secondary training done on the SVM which helps to reduce the overfitting of the network, enhances the generalization ability of the model, and further improves the accuracy of the network

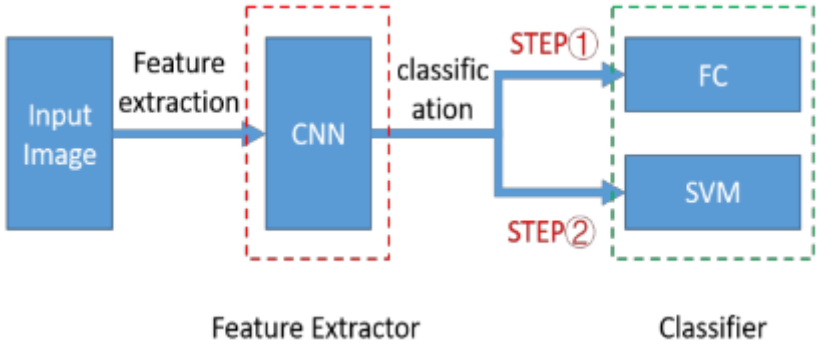


Fig. Composite Network

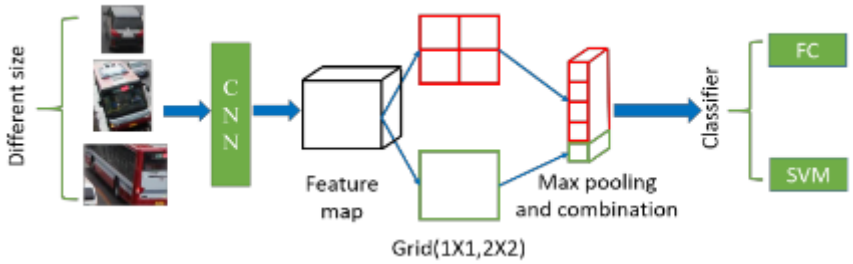


Fig. SPP layer

TABLE II: THE CNN STRUCTURE OF IMPROVED ALEXNET

Name	Kernel	Stride	Activation	Output
Conv1	5×5×96	1	Relu	?×?×96
Pool1	3×3×96	2		?×?×96
LRN				
Conv2	3×3×128	1	Relu	?×?×128
Pool2	3×3×128	2		?×?×128
LRN				
Conv3	3×3×128	1	Relu	?×?×128
Conv4	3×3×100	1	Relu	?×?×100
SPP(1×1,2×2)				500

Fig. Improved AlexNet

[? is unknown value]

Specification(w.r.t original AlexNet) of Improved AlexNet,

- Uses SPP instead of Pool3 layer
  - SPP layer will normalized features instead of resizing images at the beginning, which avoids the loss of accuracy caused by image distortion.

- Fewer layers
  - results in smaller sized model

- Smaller kernel size
  - Instead of 11×11 kernel of the original Conv1 layer this uses 5×5 kernel, and other layer's kernels size are 3×3

- Fewer feature maps, fewer model parameters, and faster recognition
  - Max number of feature maps of the proposed AlexNet is 128, originally it is 384

Model accuracy for different classes are shown below ,

TABLE III: TESTING RESULT ON EPOCH 10

Methods	Car	Bus	Van	Others
AlexNet	98.17%	91.24%	68.44%	49.35%
Improved AlexNet	98.37%	93.81%	76.07%	71.43%
Improved AlexNet+SVM	98.87%	95.02%	76.50%	74.03%

Tentive list of tasks for next session :

- Understand deep learning models (specifically convolution neural network)