CBA
Congresso Brasileiro de Automática
João Pessoa - PB

Recognition and Tracking of Vehicles in Highways using Deep Learning

Ludwin Lope Cala

Profa. Dra. Roseli Aparecida Francelin Romero

USP – ICMC - 2018

INTRODUCTION

Motivation

- Experience images taken from a drone's perspective and recognize cars.
- Simulate the tracking of a vehicle from a video captured by the drone.
- The work can give origin to an intelligent drone capable of track down suspicious vehicles on the highways.

INTRODUCTION

Objective

To build a system to simulate the tracking of a vehicle with real videos taken from a drone, using visual attention and deep learning algorithms.

STATE OF THE ART

We divide the related works into two areas:

- Detection and Classification of Objects.
- Moving Object Tracking.

STATE OF THE ART

Detection and Classification of Objects.

- Object detection and classification is an active area of research. Nowadays, one of the most important competitions is ImageNet Large Scale Visual Recognition Challenge (ILSVRC).

STATE OF THE ART

These works have different dataset image car, and achieved a different accuracy:

- R. Montanari (Montanari, 2015). Recognition the car with bag-of-feature techniques achieved an accuracy 79.82%.
- In (Huttunen et al., 2016), it was proposed a system for recognizing 4 types of vehicles, with own arquitecture CNN technique achieved an accuracy 97%.
- In (Riveros et al., 2017), wiht LeNet-5(CNN) achieved an accuracy 95.6%.

STATE OF THE ART

Moving Object Tracking.

- In its simplest form, tracking can be defined as the problem of estimating the trajectory of an object in an environment or around a scene.

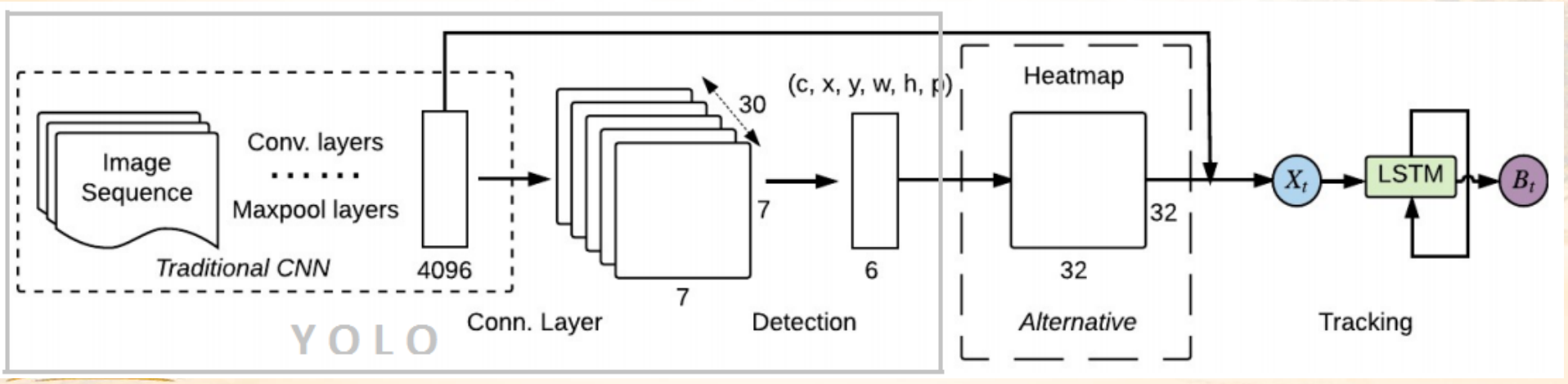
STATE OF THE ART

Almost all tracking algorithms require the detection of objects.

- R. Montanari (Montanari, 2015), it was utilized Camshift technique (Bradski, 1998) with Kalman filter (Welch and Bishop, 1995).
- Riveros et al. (Riveros et al., 2017) used Camshift with correlation filter tracker (Danelljan et al., 2014).
- ROLO (Ning et al., 2016), a tracking framework, a robust object tracking that requires knowledge and understanding of the object being tracked (Gordon et al., 2017).

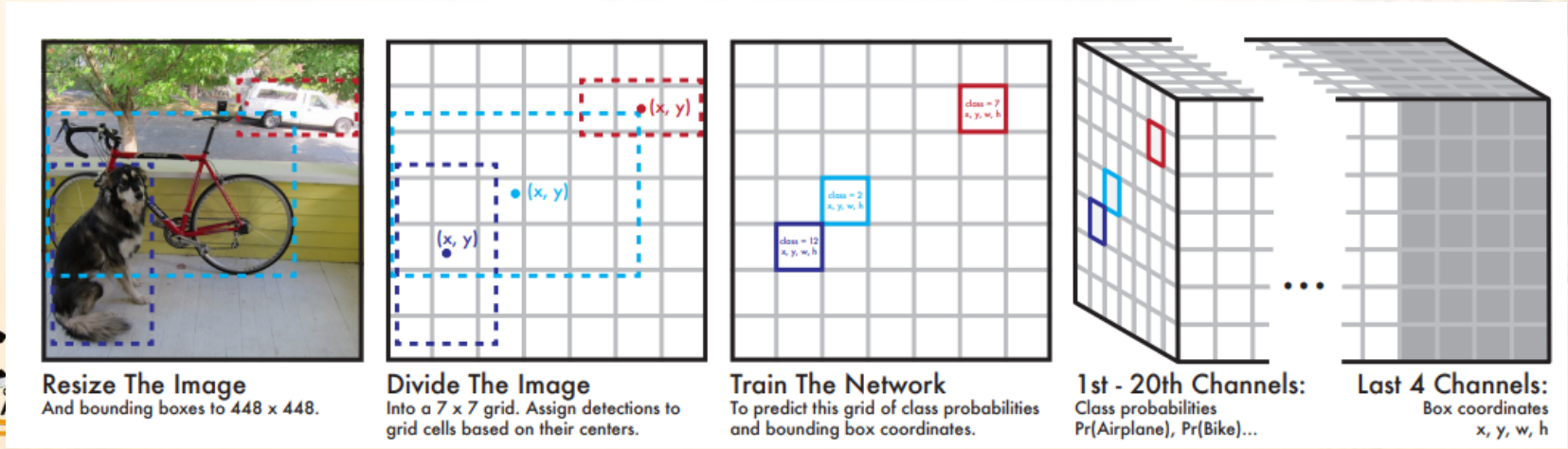
STATE OF THE ART

- ROLO (Ning. et. Al.. 2016): They used YOLO (*Redmon et. al., 2016)* to collect rich and robust visual features as well as preliminary location inferences; and then they used LSTM in the next stage as it is spatially deep and appropriate for sequence processing (see Figure).



STATE OF THE ART

- YOLO (*Redmon et. al., 2016)* : It is a unified pipeline for object detection. The system divides the input image into a 7×7 grid. If the center of an object falls into a grid cell, that grid cell is responsible for detecting that object. Each grid cell predicts a bounding box and class probabilities associated with that bounding box (see Figure).



THE PROPOSED SYSTEM

Proposed methods

- Saliency : VOCUS2 (Frintrop et al., 2015).
- Recognition: our own CNN arquitecture, with some parameters obtained from (Huttunen et al., 2016).
- Tracking : LSTM from ROLO (Ning et al., 2016), with less number LSTM cell.

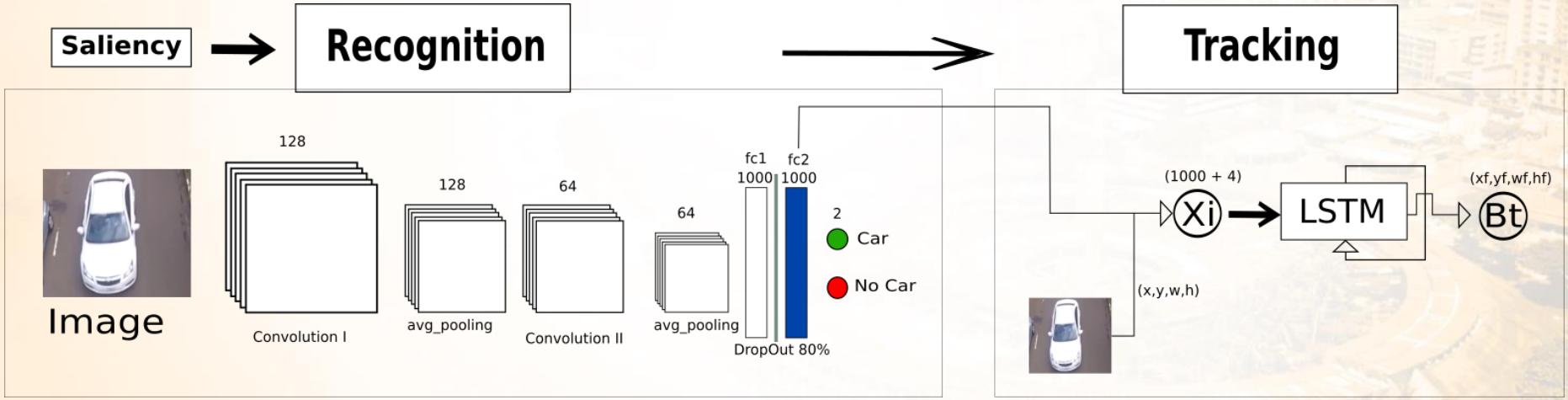
THE PROPOSED SYSTEM

Parameter CNN architecture:

Number of training epoch.	5000
Number of convolutional layers	2
Number of layers fully connected	2(1000, 1000)
Input image size	96x96
Number of feature maps	{128,64}
Kernel size of the convolutional layers	5x5
Pooling kernel size	2x2 (avg)
Cost function	Mean Squared Sum
Optimizer	AdagradOptimizer
Learning Rate	0.001643
Dropouts	80%

THE PROPOSED SYSTEM

Architecture proposed for classification and tracking:



EXPERIMENTAL RESULTS

Data Collection

- For collecting videos, a drone took images from inside our university (USP - Campus I and II – São Carlos).
- This dataset was increased up 13705 images (between car, not-car).

EXPERIMENTAL RESULTS

Training for Vehicle Recognition

- The dataset was trained with different architectures of networks, with 2 architectures proposed and Inception-v2 (Szegedy et al., 2015).

EXPERIMENTAL RESULTS

Training for Vehicle Recognition

- Table of errors in each Cross-Validation Folds (Arch-2).

Fold	Error %
I	14.75
II	12.08
III	9.27
IV	1.73
Avg. Error	**9.45**

EXPERIMENTAL RESULTS

Test for Vehicle Recognition

- Average of accuracy and precision of architectures 1, 2 and Inception-v2.

	Arch. 1	Arch. 2	Inception-v2
Avg. accuracy	0.4158	0.4192	0.5775
Avg. precision	0.9313	0.9572	0.8546

EXPERIMENTAL RESULTS

Training for Vehicles Tracking

- In Table, are shown the number of epochs used to train the LSTM recurrent network and the average of Intersection-over-Union (IoU).

No. Epochs	IoU
Video#1	
200	0.37077
1000	0.70266
Video#2	
200	0.75823
500	0.79602

EXPERIMENTAL RESULTS

Test for Vehicles Tracking

- Comparison between Correlation Filter Tracker (CFT) and LSTM tracker.

	CFT	**Our Tracker**
IoU of Video#1	0.3154	0.7027
IoU of Video#2	0.9163	0.7960
Avg. IoU	**0.6159**	**0.7494**

CONCLUSION

- The results obtained showed 90.55 % of average accuracy.
- From the comparative results, in terms of accuracy, Inception-v2 network was better than the architecture proposed, but it holds to note that for the classification Inception-v2 network took longer time. On the other hand, in terms of precision, the proposed architecture had a better performance.
- The obtained result showed that the proposed system got an IoU mean of approximately 75% compared to 62% obtained for the correlation filter tracker.

Limitation and Future Works

- It is useful for images took to daylight, but it does not at night light.
- We intend to increase the dataset and to use transfer learning with the proposed architecture, for training in real time for a specific car that is being tracked.

Referencias:

1 - Ning, G., Zhang, Z., Huang, C., He, Z., Ren, X. and Wang, H. (2016). Spatially supervised recurrent convolutional neural networks for visual object tracking, CoRR abs/1607.05781. (ROLO)

2 – Huttunen, H., Yancheshmeh, F. S. and Chen, K. (2016). Car type recognition with deep neural networks, IEEE Intelligent Vehicles Symposium (IV) .

3 - *Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi*; "You Only Look Once: Unified, Real-Time Object Detection", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788 (YOLO)