
# Detection and Tracking of Motorbikes/People using Deep Learning Techniques

Prof. Sergio Velastín -FIET CEng SMIEEE - CORTEXICA

PhD(c) - Jorge Espinosa - PCJIC – UNAL

Prof. John William Branch, UNAL

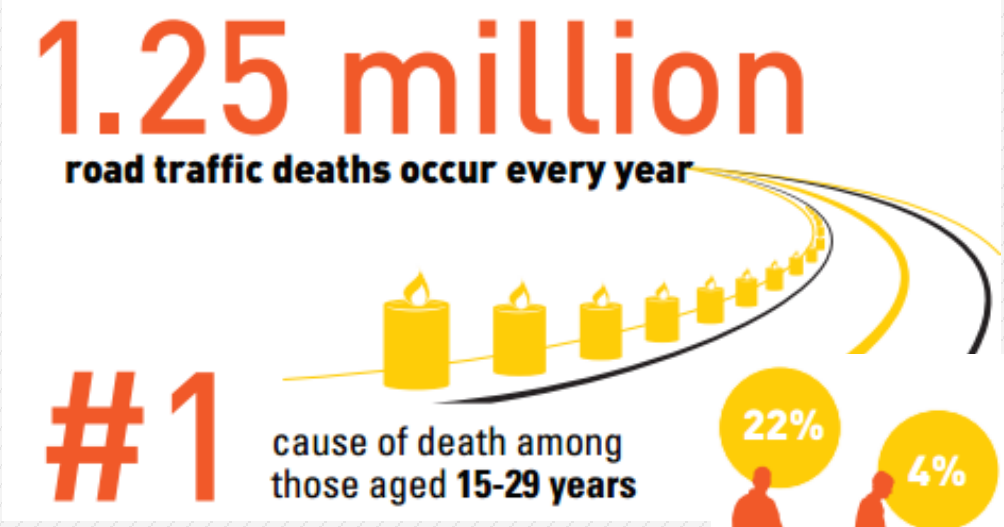
Prof. Rodrigo Fernandez, U los Andes, Chile

# Content

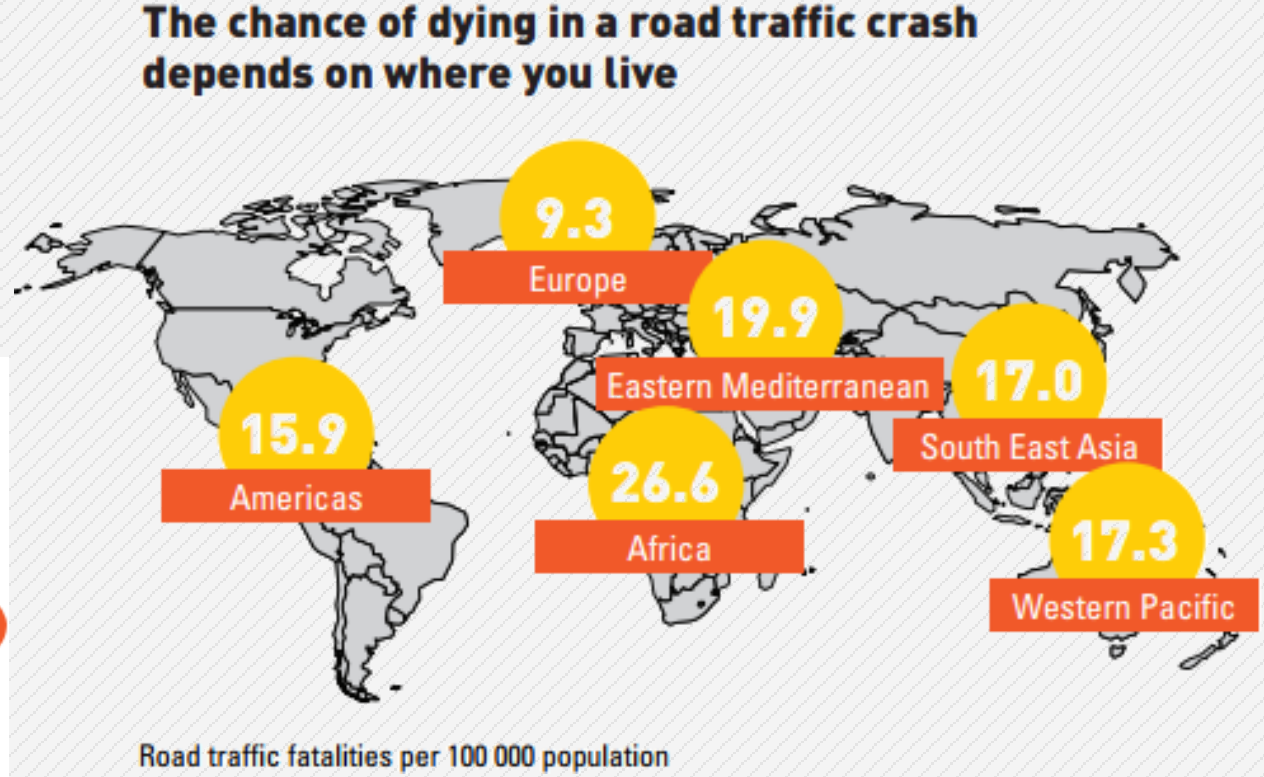
1. Motivation
2. Deep Learning capabilities and architectures
3. A new public Urban Motorcycle Dataset (UMD)
4. Motorcycle detection and classification on UMD and for a traditional traffic CCTV systemTracking
5. People detection
6. Conclusions and Future Work

# Motivation

## Why Motorcycles and Pedestrians??








Global status report on road safety 2015 - WHO

# Motivation

## Why Motorcycles in Colombia??

13.245.856
Total Vehicles Registered 2Q-2017
Motorcycles = 7.470.663 (56,4%)

Source: Registro Único Nacional de Tránsito, RUNT



Source FENALCO 2016

## High Accidentally Rate (2Q -2017)


Sources: Secretarías de Movilidad y Medicina Legal

# Motivation (more numbers)

## Pollution & Enviromental Issues ☹

COLOMBIA

Motorcycles = 7.740.838 (57%)
Internal Combustion Engine (ICE)
[Runt 2018]

COLOMBIA



Motorcycles (76,64%)
111 – 135 c.c.
Street Sport Segment
[Fenalco % ANDI 2017]

MEDELLÍN

P.M.5 (59%)
Produced by Land Transportation
Motorcycles (40%)
[Área Metropolitana 2017]

**Emission from combustion engines is one of the most important causes of pollution in Colombia**

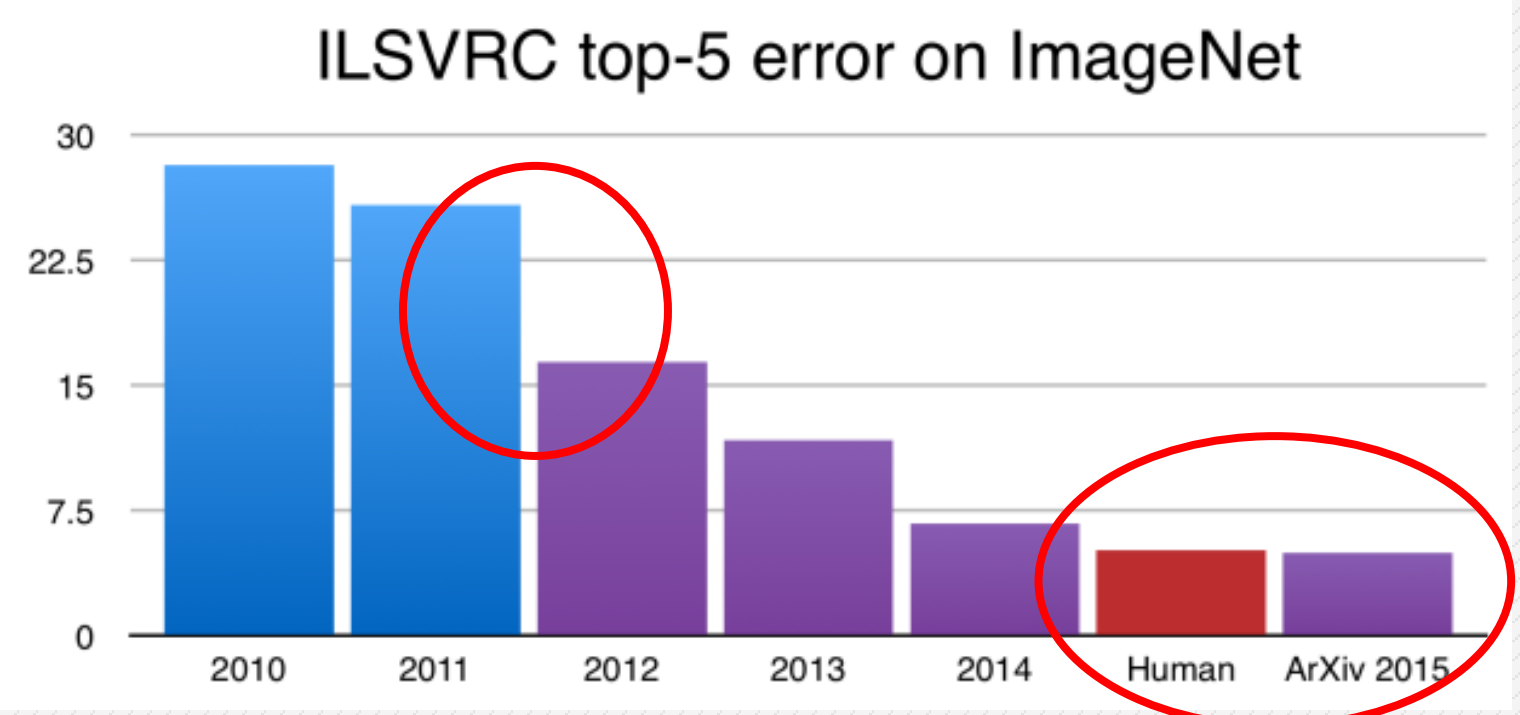
MOYCOT
Movilidad y Control del Tráfico

# Why Motorcycles?? (Jakarta ☹)

# Motivation

# Deep Learning Breakthrough

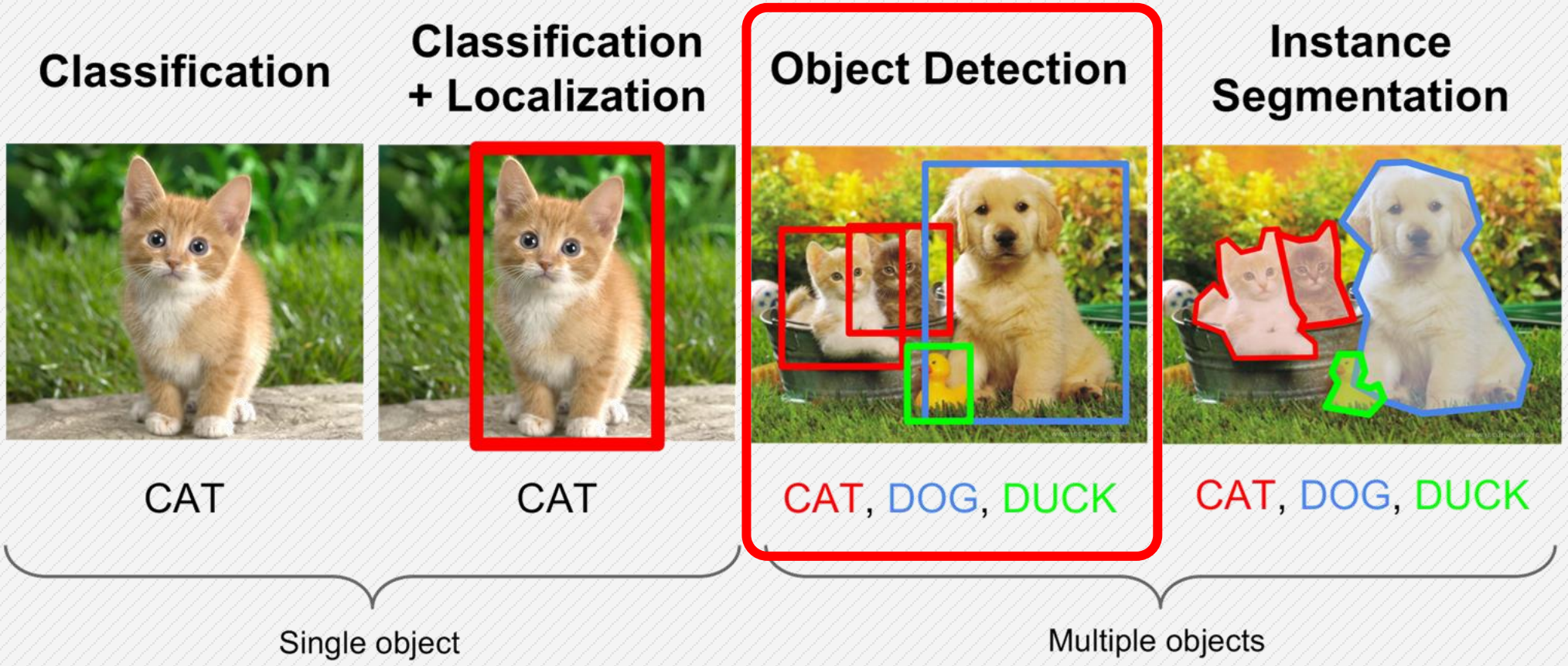
IMAGENET **Image Large Scale Visual Recognition Challenge**




Source: http://image-net.org/

# Toward localisation

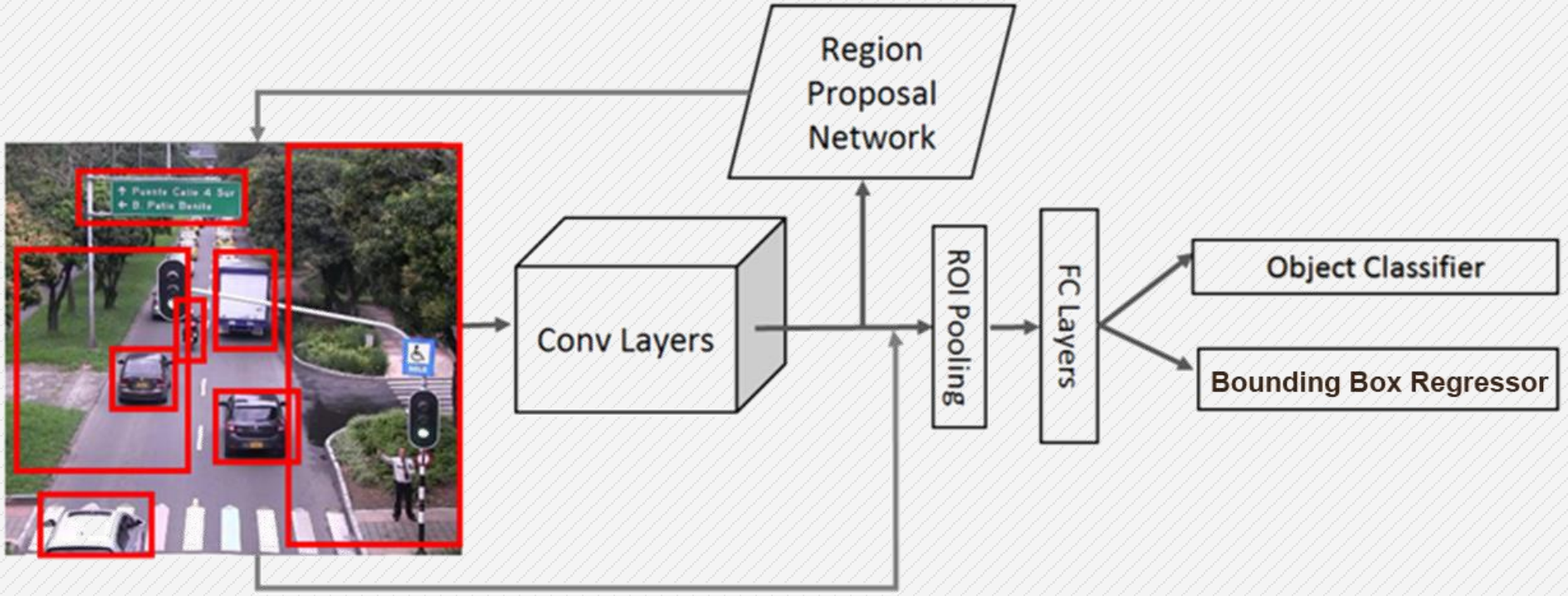
## Computer Vision Tasks




*CS231n Winter 2016: Lecture 8: Localization and Detection*

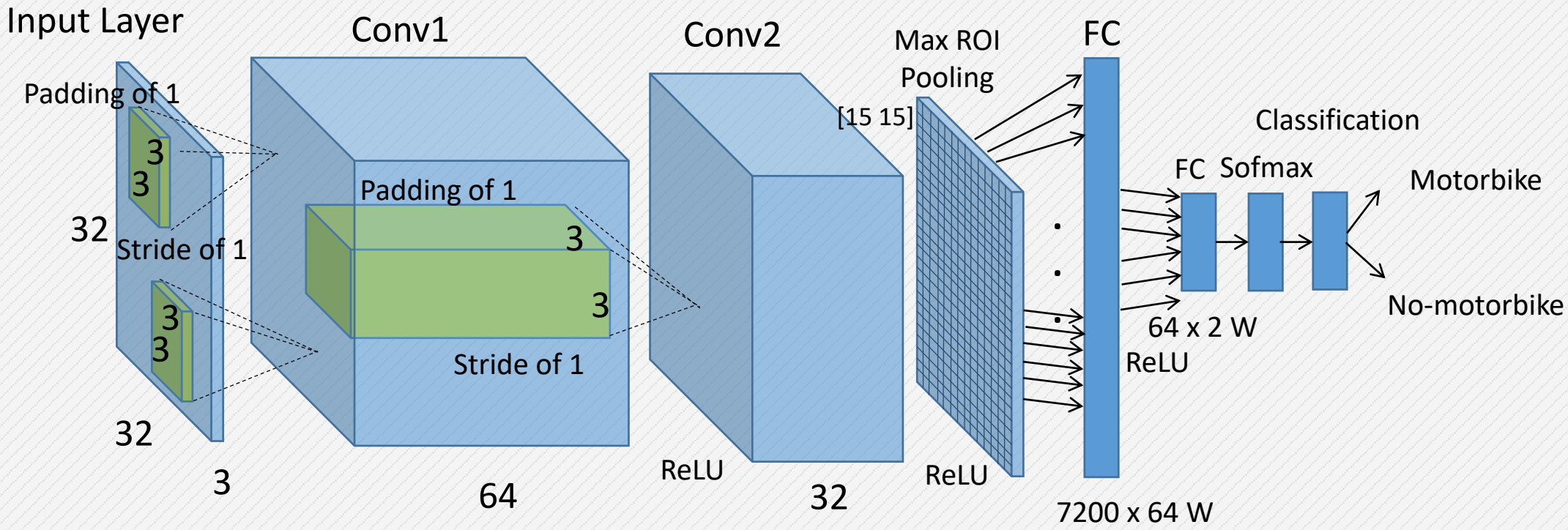
# CNN Architectures used

## Faster R-CNN




Region Proposal: attention director (where to look), where objects are more likely to be.
Classifier: what those objects are likely to be

# EspiNet2: New CNN model inspired in Faster R-CNN (2 layers=fast)



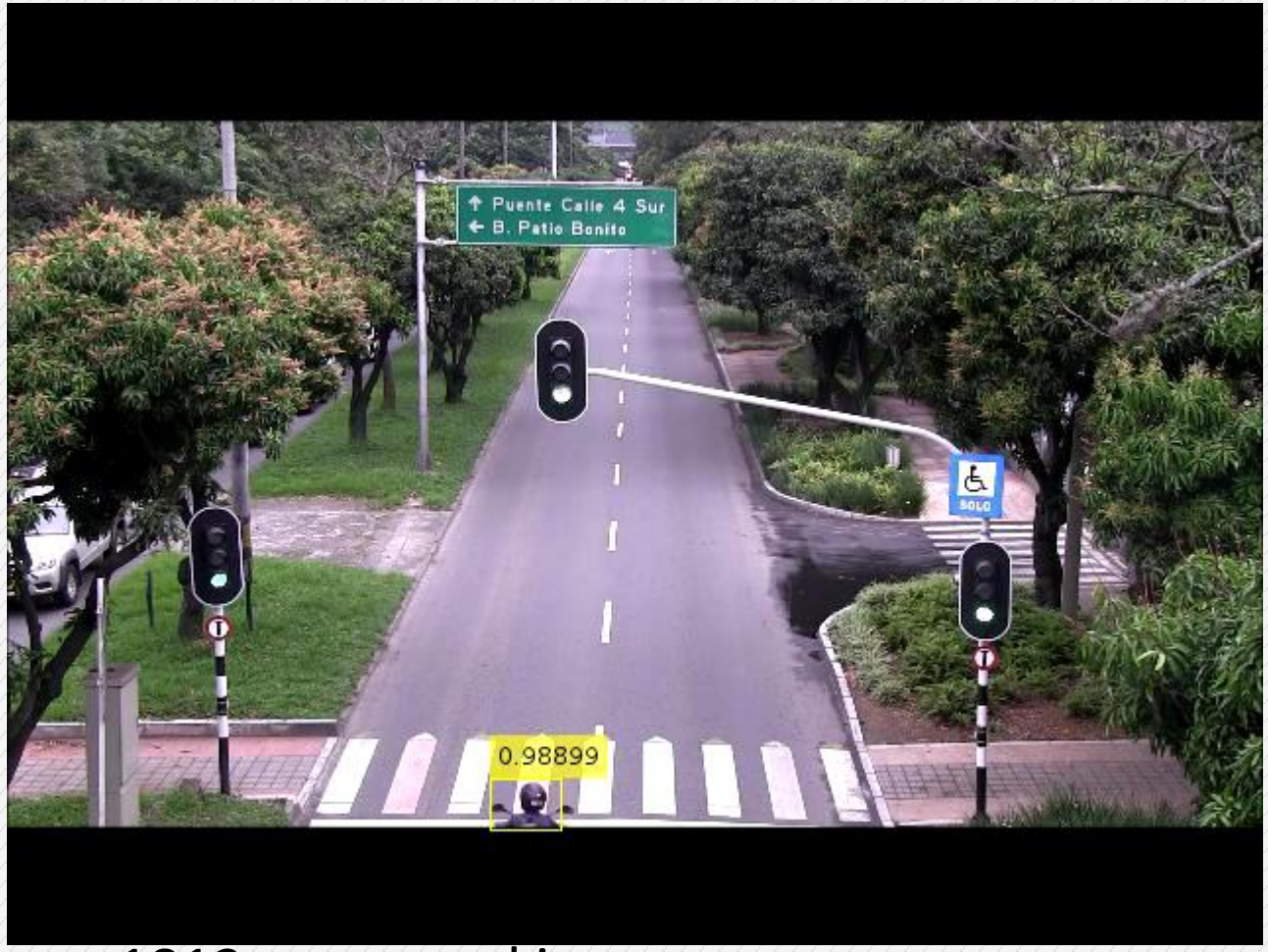

$$n_{out} = \left\lfloor \frac{n_{in} + 2p - k}{s} \right\rfloor + 1$$

- Optimization Algorithm for training:
  Stochastic Gradient Descent with momentum (SGDM)

$$\theta_{\ell+1} = \theta_{\ell} - \alpha \nabla E(\theta_{\ell}) + \gamma(\theta_{\ell} - \theta_{\ell-1})$$

- Took 32 hours for training the dataset
  (50% Training – 30%Validating – 20%Testing)

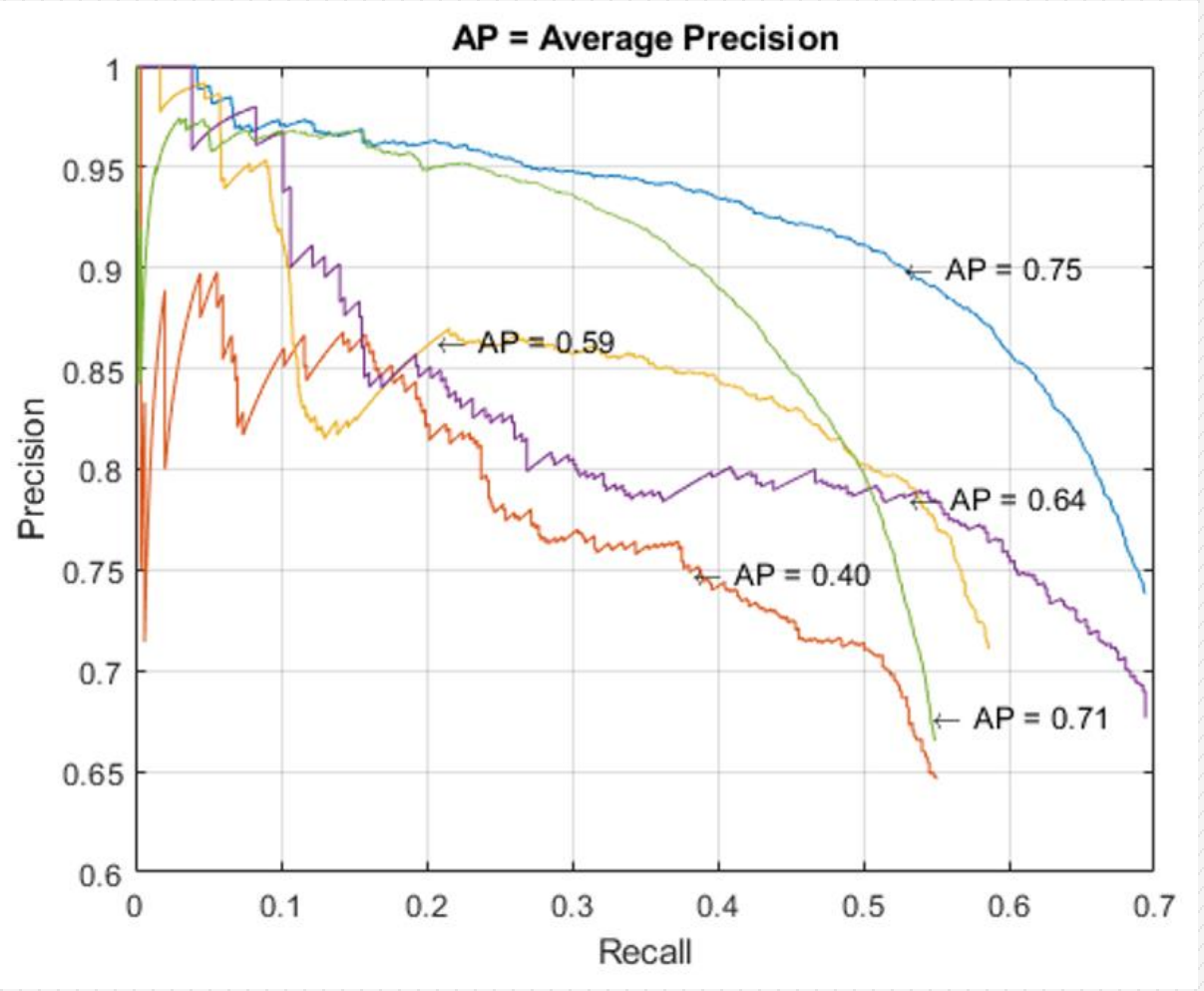
# Initial Experiments




- 1812 annotated images
- 640 x 480
- No significant occlusions

**AP= 92 % (EspiNet2)**

# New Urban Motorbike Dataset (UMD)

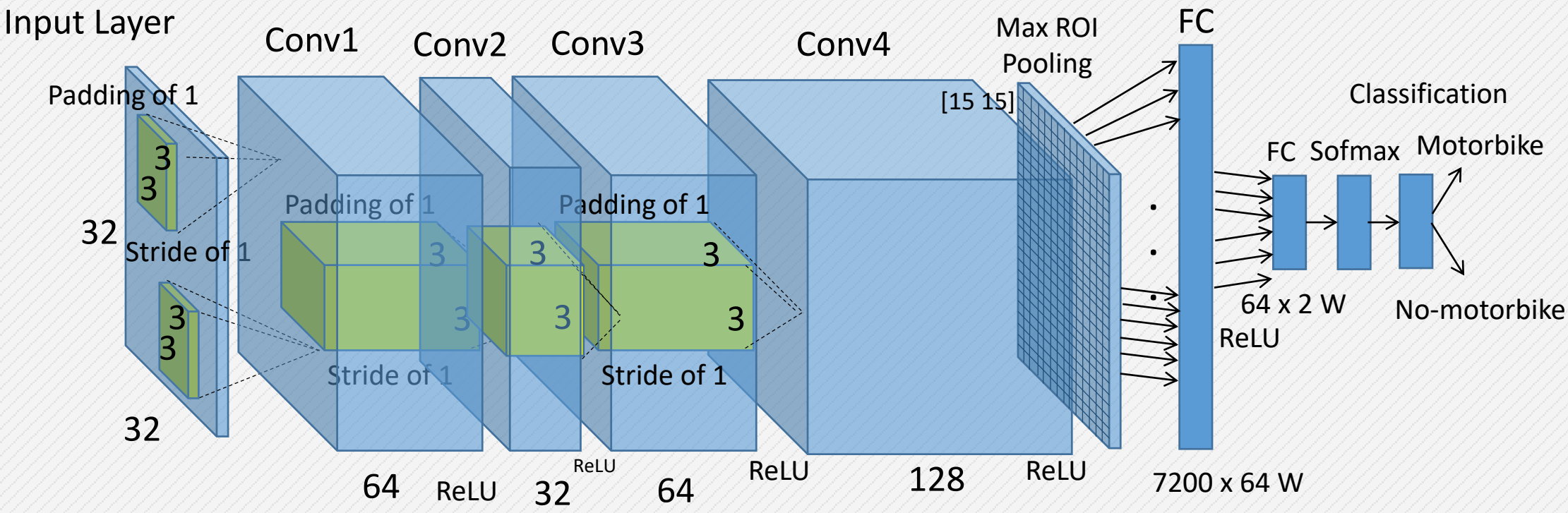



- 7,500/10,000 annotated images
- 220/317 motorcycles on urban traffic.
- 41,040/56,795 ROI annotated objects
- **60% Annotated object are occluded**

**AP=75% (harder!)**

Available at: http://videodatasets.org

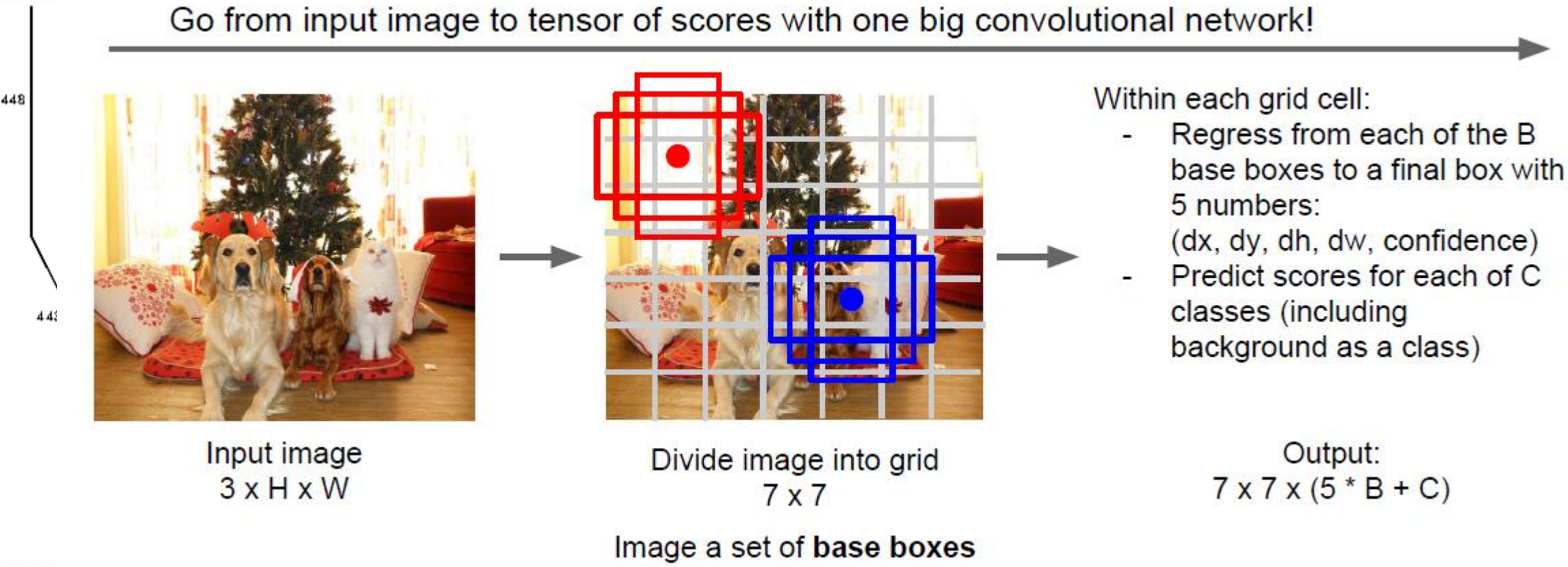
# EspiNet4 - A Deeper model




- Took 62 hours for training the dataset
  (90% Training – 5%Validating – 5%Testing)

- **AP=89,3% on UMD10K (Improves EspiNet2)**

# YOLO
# You Look Only Once
## Detection without proposals (fast!)

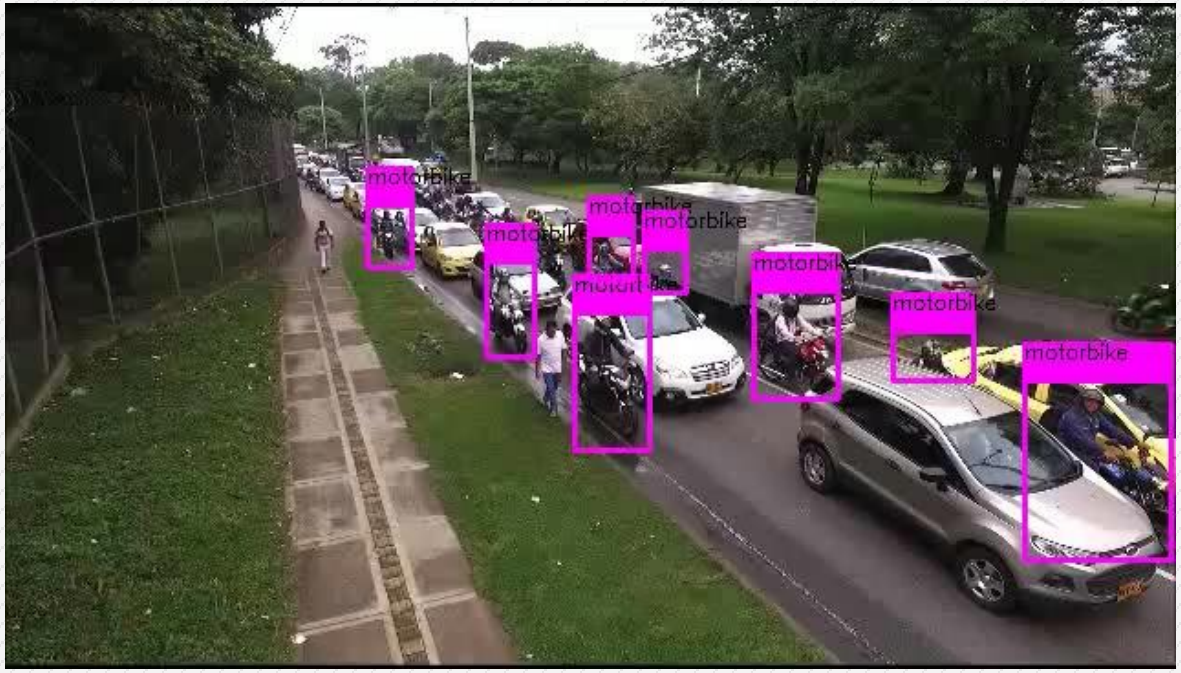



[55] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

MOYCOT
Movilidad y Control del Tráfico

# YOLO vs EspiNet (UMD 10K)

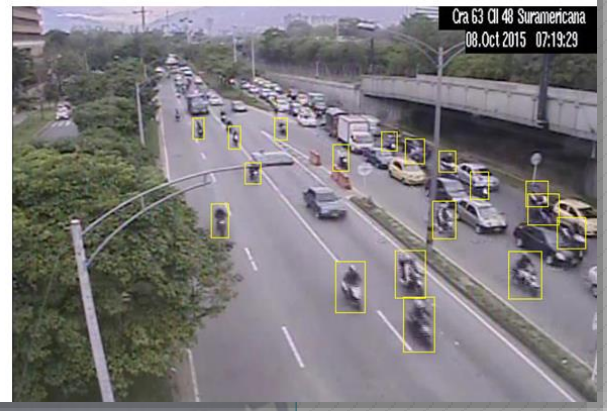
## YOLO V3 AP = 80%




## EspiNet AP = 89,3%






# New "Secretaría de Movilidad" Dataset

- 5000 annotated images (6 different cameras)
- 827 motorcycles tracks on urban traffic
- 704 x 480 (low resolution)
- **21,625 ROI** annotated motorbike objects
  Minimum H size 25 px
- **40 % Annotated object are occluded**

Available Soon at: http://videodatasets.org

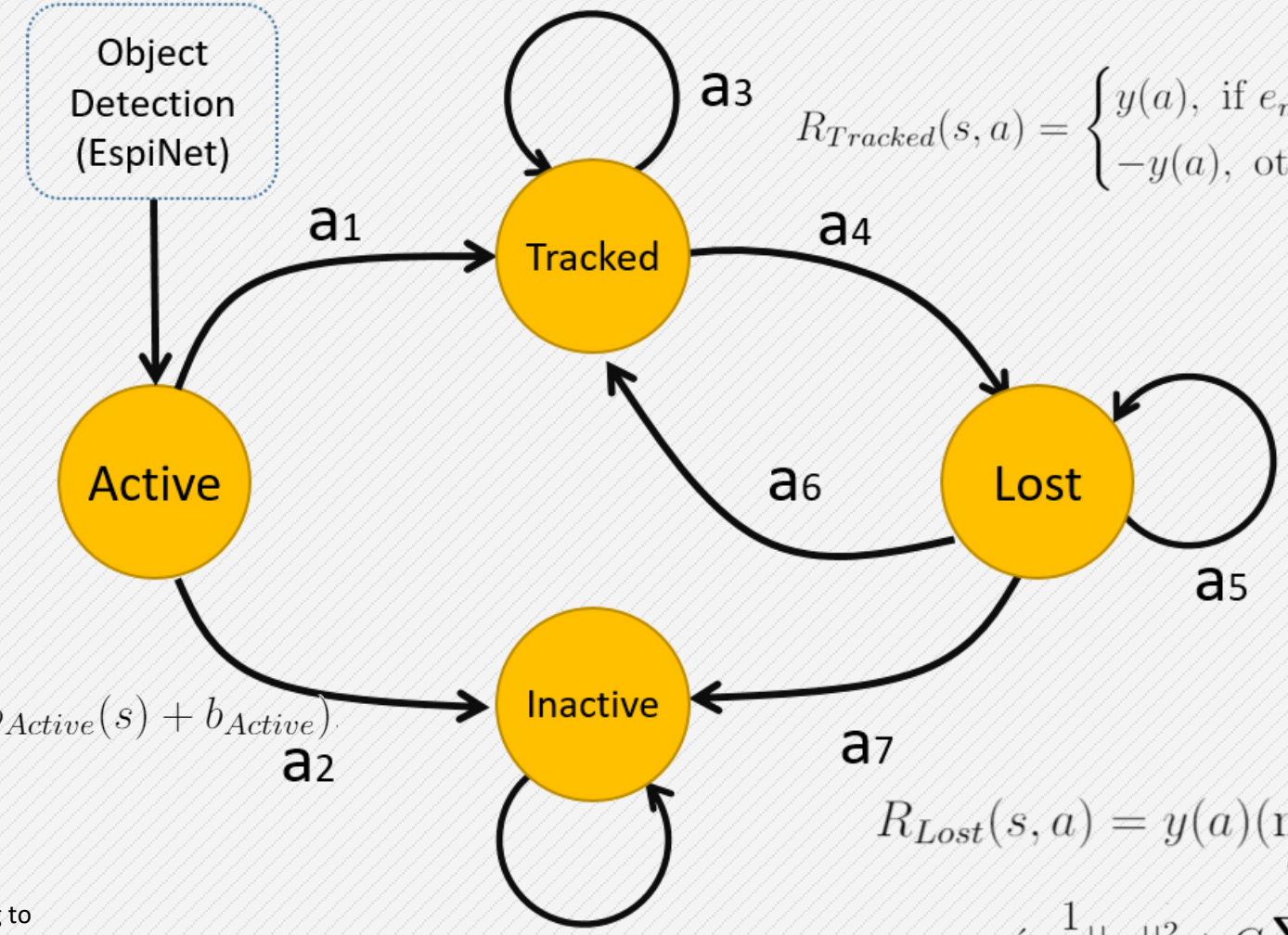
MOYCOT
Movilidad y Control del Tráfico

# YOLO vs EspiNet (Sec5k)

## YOLO V3 AP = 77%


## EspiNet = 80%

# Tracking by Detection




The MDP of the algorithm

$$R_{Tracked}(s,a) = \begin{cases} y(a), & \text{if } e_{medFB} < e_0 \text{ and } o_{mean} > o_0 \\ -y(a), & \text{otherwise,} \end{cases}$$

$$R_{Active}(s,a) = y(a)(W_{Active}^T \phi_{Active}(s) + b_{Active})$$

$$R_{Lost}(s,a) = y(a)(\max_{k=1}^{M}(\mathrm{w}^T \phi(t,d_k) + b))$$

$$\min_{\mathrm{w},b,\xi} \frac{1}{2}||\mathrm{w}||^2 + C\sum_{k=1}^{M} \xi k$$

$$s.t. \quad y_k(\mathrm{w}^T \phi(t_k,d_k) + b) \geq 1 - \xi k, \xi k \geq 0, \forall k$$

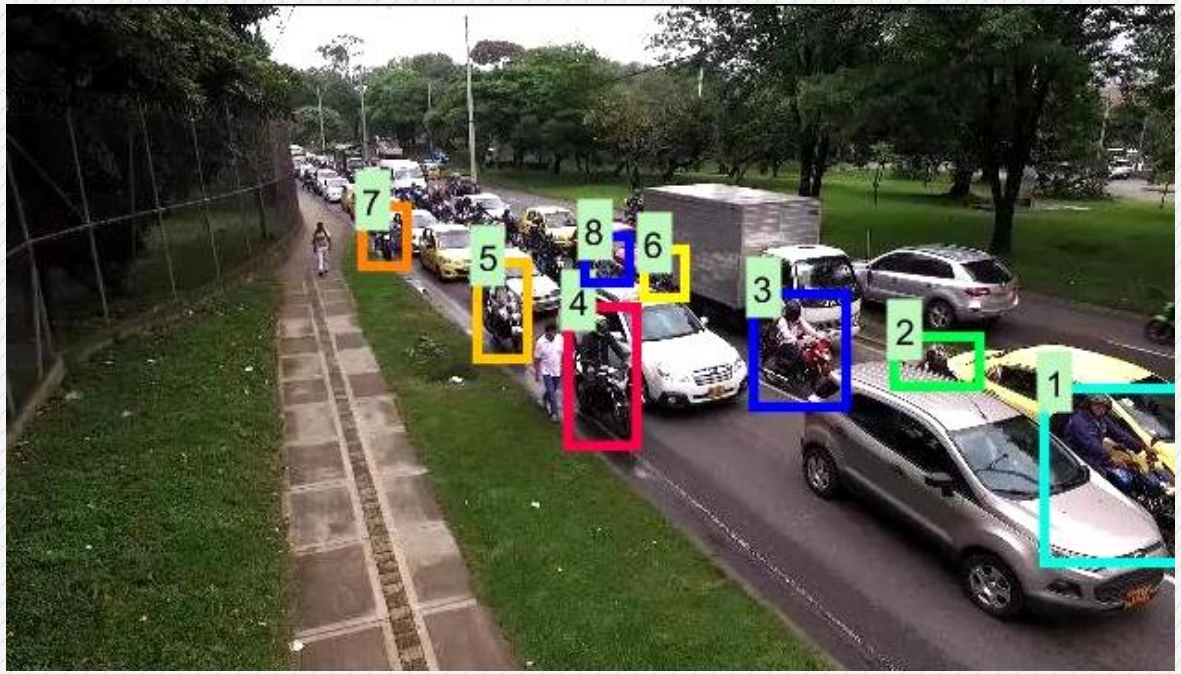

[35] Y. Xiang, A. Alahi, y S. Savarese, “Learning to track: Online multi-object tracking by decision making”, en *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 4705–4713.

# Tracking by Detection


| Rcll | Prcn | FAR | GT | MT | ML | IDs | MOTA | MOTP |
|---|---|---|---|---|---|---|---|---|
| 66.5 | 67.5 | 1.53 | 44 | 24 | 20 | 44 | 33.9 | 74.8 |

Detector Faster R-CNN

Detection base on AlexNet + GMM




| Rcll | Prcn | FAR | GT | MT | ML | IDs | MOTA | MOTP |
|---|---|---|---|---|---|---|---|---|
| 86.5 | 87.5 | 0.75 | 128 | 126 | 2 | 128 | **93.52** | **96.8** |

Detection base on EspiNet

# Tracking by Detection

## Secretaría de Medellín Dataset




| Rcll | Prcn | FAR | GT | MT | ML | IDs | MOTA | MOTP |
|---|---|---|---|---|---|---|---|---|
| 83,3 | 56,3 | 2,70 | 816 | 411 | 81 | 503 | 16,3 | 67,2 |

MOYCOT
Movilidad y Control del Tráfico

# Detection of People Boarding/Alighting a Metropolitan Train

## PAMELA-UANDES Dataset

# Detection of People Boarding/Alighting a Metropolitan Train

- 8 videos people Alighting
- 7 videos of people waiting and then Boarding
- 352 x 288 resolution
- **43,751 positive examples**
- Took two months to complete annotation

## PAMELA-UANDES Dataset

MOYCOT
Movilidad y Control del Tráfico

# Detection of People Boarding/Alighting a Metropolitan Train

Boarding

**EspiNet AP= 82,14%**

Alighting

**PAMELA-UANDES Dataset**

MOYCOT
Movilidad y Control del Tráfico

# Conclusions

- Deep Learning is a promising avenue for vision-based traffic analysis
- Urban traffic still represents a challenge for image detection classification and tracking for high levels of occlusion (greater than 80 %) e.g. Jakarta
- The method is comparable (better?) than manual observation.
- The model is able to deal with levels of occlusion ~60%, achieving a precision of 75%-82% tested with new realistic datasets and comparable with the state-of-the-art.
- As typical of deep learning, the model gives better results by using a large number of training examples (5k – 10k), implying long training times even for GPUs.
- We are looking at a number of improvements to increase detection and tracking performance

sergio.velastin@ieee.org
jeespinosa@elpoli.edu.co
jwbranch@unal.edu.co



# Gracias !!!

# Referencias


- [1] G. E. Moore, *Cramming more components onto integrated circuits*. McGraw-Hill, 1965.
- [2] A. Yilmaz, O. Javed, and M. Shah, “Object tracking: A survey,” *ACM Comput Surv*, vol. 38, no. 4, Dec. 2006.
- [3] S. J. Dickinson, “Object representation and recognition,” *What Cogn. Sci.*, vol. 7, pp. 172–207, 1999.
- [4] L. M. Fuentes and S. A. Velastin, “Advanced Surveillance: From Tracking To Event Detection,” *Lat. Am. Trans. IEEE Rev. IEEE Am. Lat.*, vol. 2, no. 3, pp. 206–211, 2004.
- [5] W. Hu, T. Tan, L. Wang, and S. Maybank, “A survey on visual surveillance of object motion and behaviors,” *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 34, no. 3, pp. 334–352, 2004.
- [6] M. Barnard and J. Odobez, “Sports Event Recognition Using Layered HMMS,” in *IEEE International Conference on Multimedia and Expo, 2005. ICME 2005*, 2005, pp. 1150–1153.
- [7] N. Buch, S. A. Velastin, and J. Orwell, “A Review of Computer Vision Techniques for the Analysis of Urban Traffic,” *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 3, pp. 920–939, Sep. 2011.

# Referencias (Cont.)


- [8] O. Masoud and N. P. Papanikolopoulos, “Using geometric primitives to calibrate traffic scenes,” *Transp. Res. Part C Emerg. Technol.*, vol. 15, no. 6, pp. 361–379, 2007.
- [9] N. Cornelis, B. Leibe, K. Cornelis, and L. Van Gool, “3d urban scene modeling integrating recognition and reconstruction,” *Int. J. Comput. Vis.*, vol. 78, no. 2–3, pp. 121–141, 2008.
- [10] K. Park, D. Lee, and Y. Park, “Video-based detection of street-parking violation.,” in *IPCV*, 2007, pp. 152–156.
- [11] B. Morris and M. Trivedi, “Robust classification and tracking of vehicles in traffic video streams,” in *IEEE Intelligent Transportation Systems Conference, 2006. ITSC ’06*, 2006, pp. 1078–1083.
- [12] K. Muller, A. Smolic, M. Drose, P. Voigt, and T. Wiegand, “3-D reconstruction of a dynamic environment with a fully calibrated background for traffic scenes,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 4, pp. 538–549, Apr. 2005.
- [13] N. Buch, J. Orwell, and S. A. Velastin, “3d extended histogram of oriented gradients (3dhog) for classification of road users in urban scenes,” 2009.
- [14] S. Messelodi, C. M. Modena, and M. Zanin, “A computer vision system for the detection and classification of vehicles at urban road intersections,” *Pattern Anal. Appl.*, vol. 8, no. 1–2, pp. 17–31, 2005.

# Referencias (Cont.)


- [15] X. Chen and C. Zhang, “Vehicle classification from traffic surveillance videos at a finer granularity,” in *Advances in Multimedia Modeling*, Springer, 2006, pp. 772–781.
- [16] J. Lou, T. Tan, W. Hu, H. Yang, and S. J. Maybank, “3-D model-based vehicle tracking,” *IEEE Trans. Image Process.*, vol. 14, no. 10, pp. 1561–1569, Oct. 2005.
- [17] X. Song and R. Nevatia, “Detection and Tracking of Moving Vehicles in Crowded Scenes,” in IEEE Workshop on Motion and Video Computing, 2007. WMVC ’07, 2007, pp. 4–4.
- [18] Y. Guo, C. Rao, S. Samarasekera, J. Kim, R. Kumar, and H. Sawhney, “Matching vehicles under large pose transformations using approximate 3D models and piecewise MRF model,” in *IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008*, 2008, pp. 1–8.
- [19] Ali Ghodsi, *Dimensionality Reduction A Short Tutorial.* Department of Statistics and Actuarial Science University of Waterloo, Waterloo, Ontario, Canada, 2006
- [20] “Imagery Library for Intelligent Detection Systems - Detailed guidance - GOV.UK.” [Online]. Available: https://www.gov.uk/imagery-library-for-intelligent-detection-systems. [Accessed: 02-Dec-2014].
- [21] “Imagery Library for Intelligent Detection Systems - Detailed guidance - GOV.UK.” [Online]. Available: https://www.gov.uk/imagery-library-for-intelligent-detection-systems. [Accessed: 02-Dec-2014].