UNIVERSIDAD
NACIONAL
DE COLOMBIA

CLASIFICACIÓN Y RECONOCIMIENTO DE PATRONES

Jorge E. Espinosa

Profesor

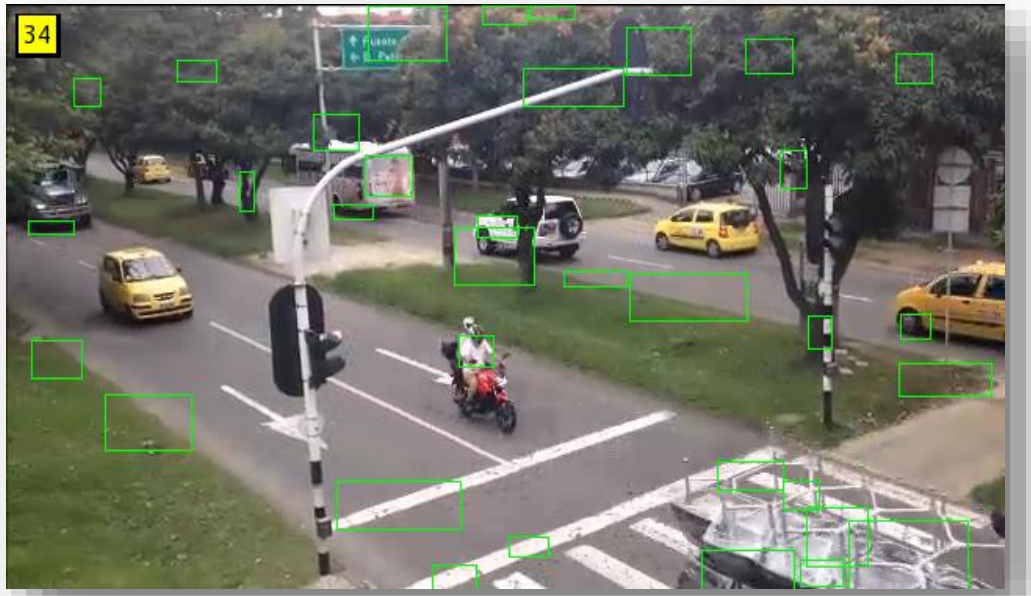
Departamento de Ciencias de la Computación y de la Decisión

Investigador del Grupo de I+D en Inteligencia Artificial – GIDIA

jeespinosao@unal.edu.co

CLASIFICACIÓN Y RECONOCIMIENTO DE PATRONES

Reconocimiento de Objetos en tiempo real Mediante técnicas Basadas en Deep Learning



Experimentation Image

Experimentation Image

PhD Jorge E. Espinosa

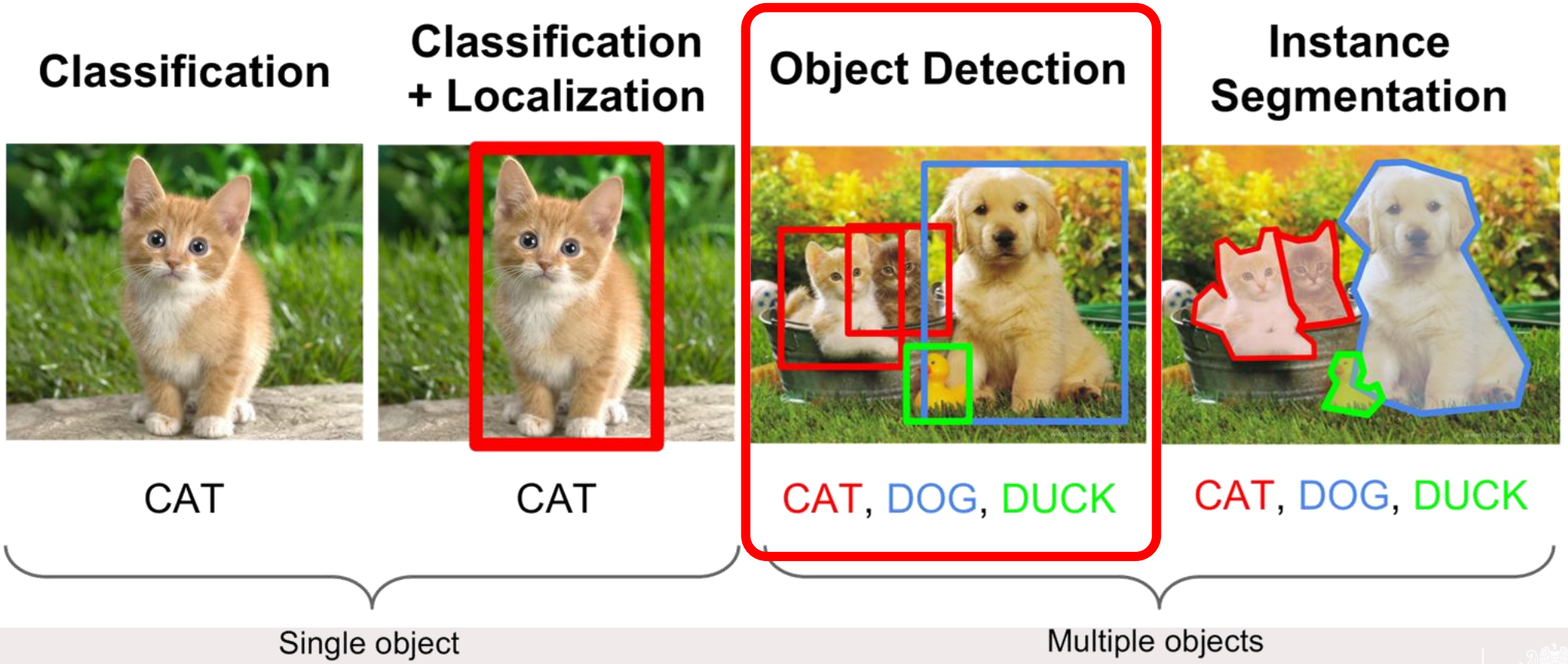
POLITÉCNICO COLOMBIANO
JAIME ISAZA CADAVID
FACULTAD DE INGENIERÍAS

Gidia
Grupo de I+D
en Inteligencia Artificial

UNIVERSIDAD
NACIONAL
DE COLOMBIA

Motivación

Las Tareas de la Visión por Computador



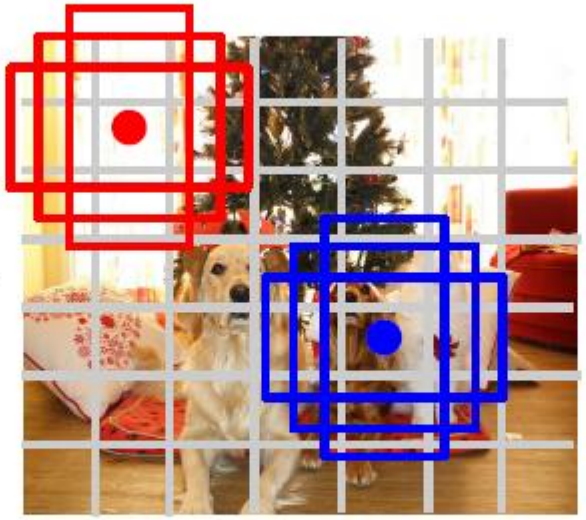
DEMO
YOLO
You Look Only Once


YOLOv2 COCO
car
car
car
car
car
person
YOLOv3 COCO
Mask RCNN COCO
person 99%
car 97%
DeepLab V3 xception_cityscapes_trainfine (GTX980M) INPUT_SIZE=1539
Prediction time: 776ms (1.3 fps) AVG: 776ms (1.3 fps)
DeepLab xception

YOLO
You Look Only Once
Detection without proposals

Go from input image to tensor of scores with one big convolutional network!

Input image
3 x H x W



Divide image into grid
7 x 7

Image a set of **base boxes**

Within each grid cell:

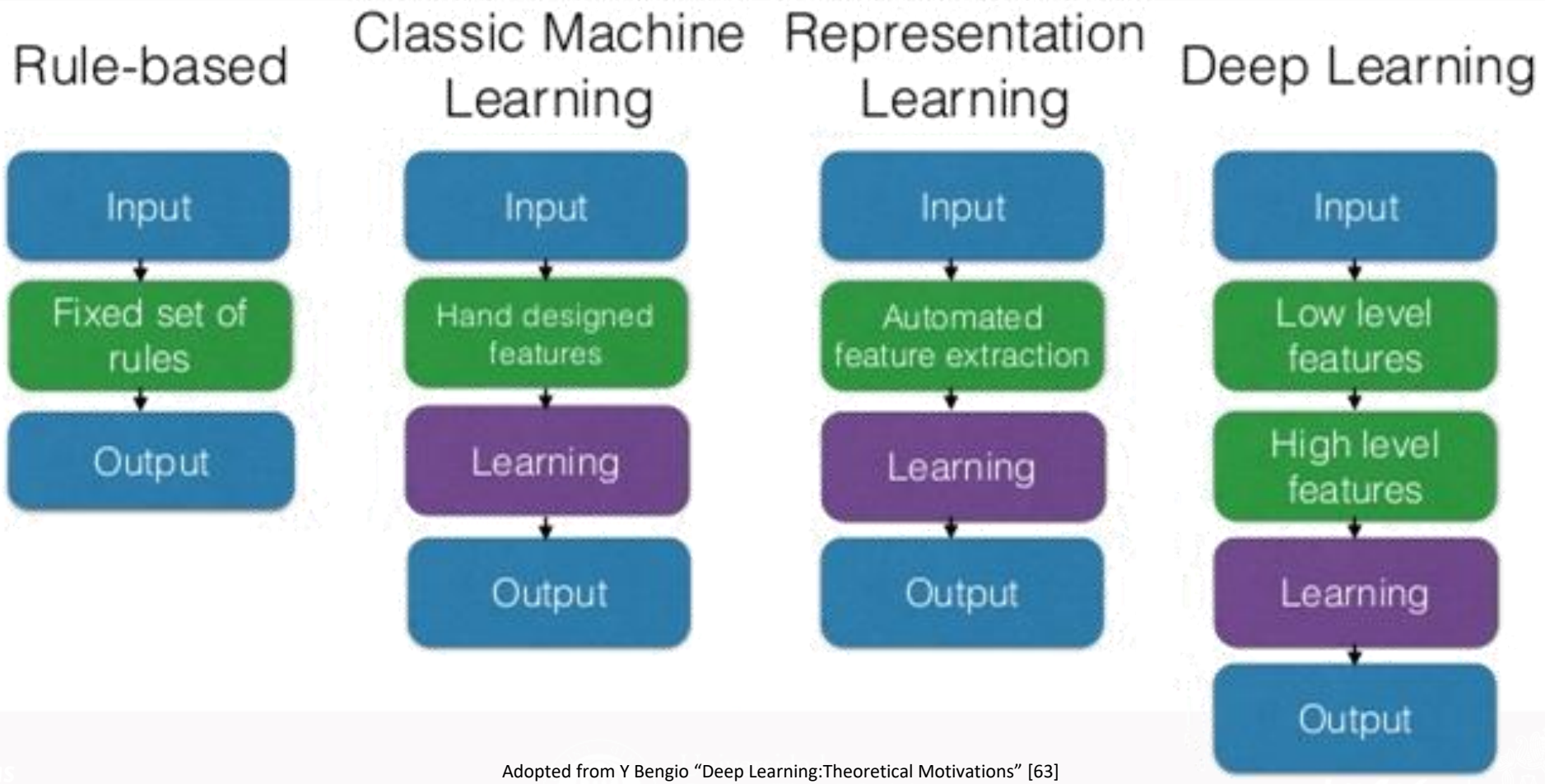
- Regress from each of the B base boxes to a final box with 5 numbers:
  (dx, dy, dh, dw, confidence)
- Predict scores for each of C classes (including background as a class)

Output:
7 x 7 x (5 * B + C)

[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).

IA- Machine Learning

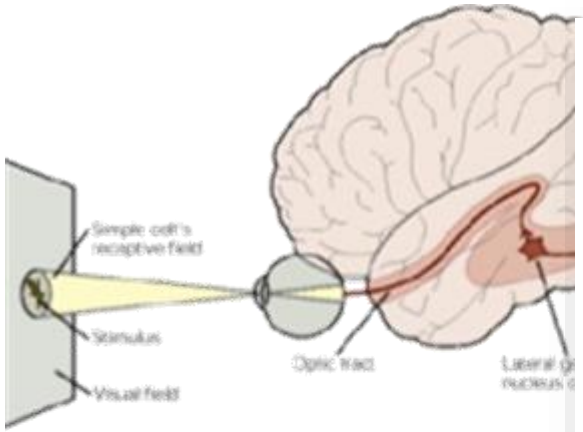
Evolution of Machine Learning Methods

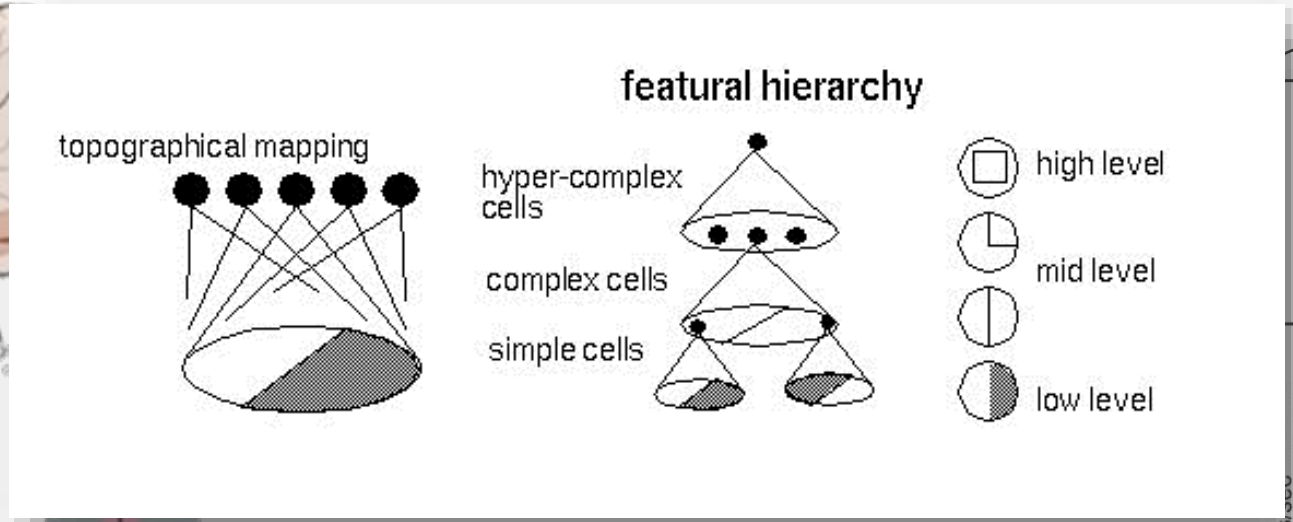


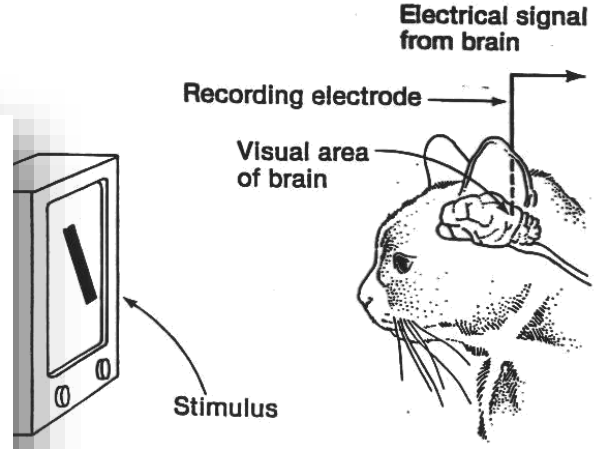
Adopted from Y Bengio "Deep Learning:Theoretical Motivations" [63]

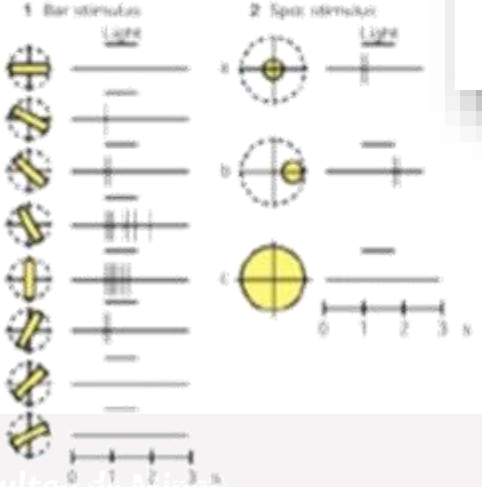
Convolutional Neural Networks

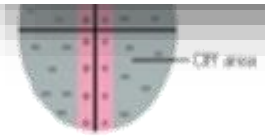
A bit of History

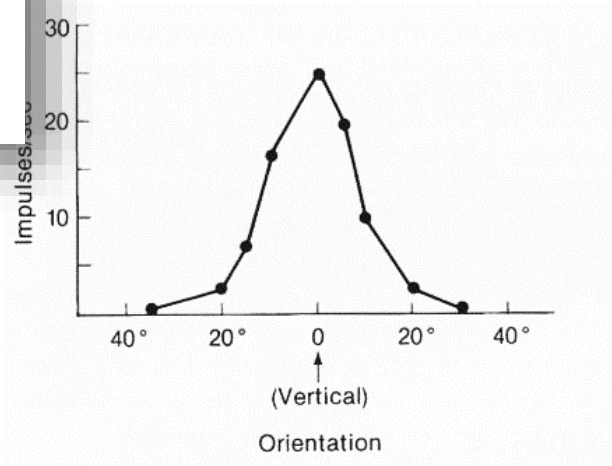








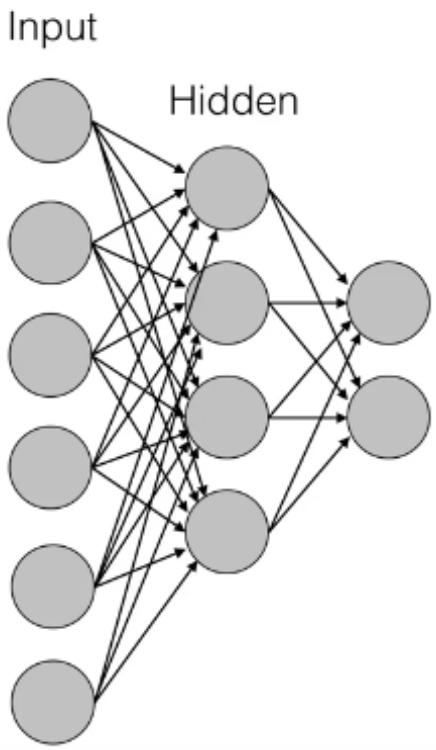




Hubel & Wiesel, 1959, 1962, 1965,1968

Convolutional Neural Networks

Computational Implications

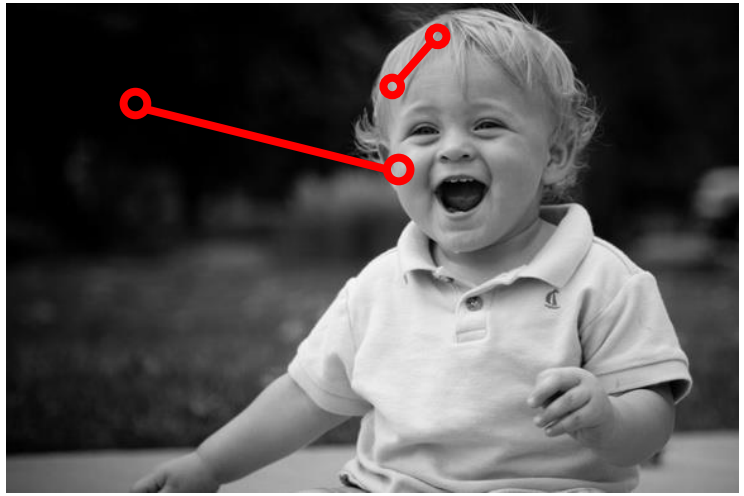


NN MLP Tradtional Architecture

MNIST dataset: 28 x28 pixels (784 pixels)
First layer weights: ~78k parameters

Typical Image: 256 x 256 (56,000 pixels)
First layer weights: 560k parameters !

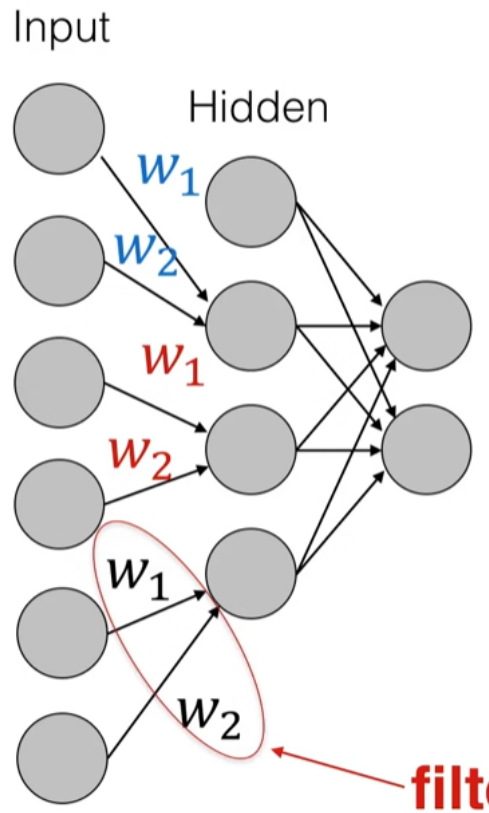
Too many parameters!!



Space Matters!!

Convolutional Neural Networks

Convolutional Neural Networks
Why Convolution?



- Edges are filtered – Pixel with high contrast
- This operation is perfomed all around the image

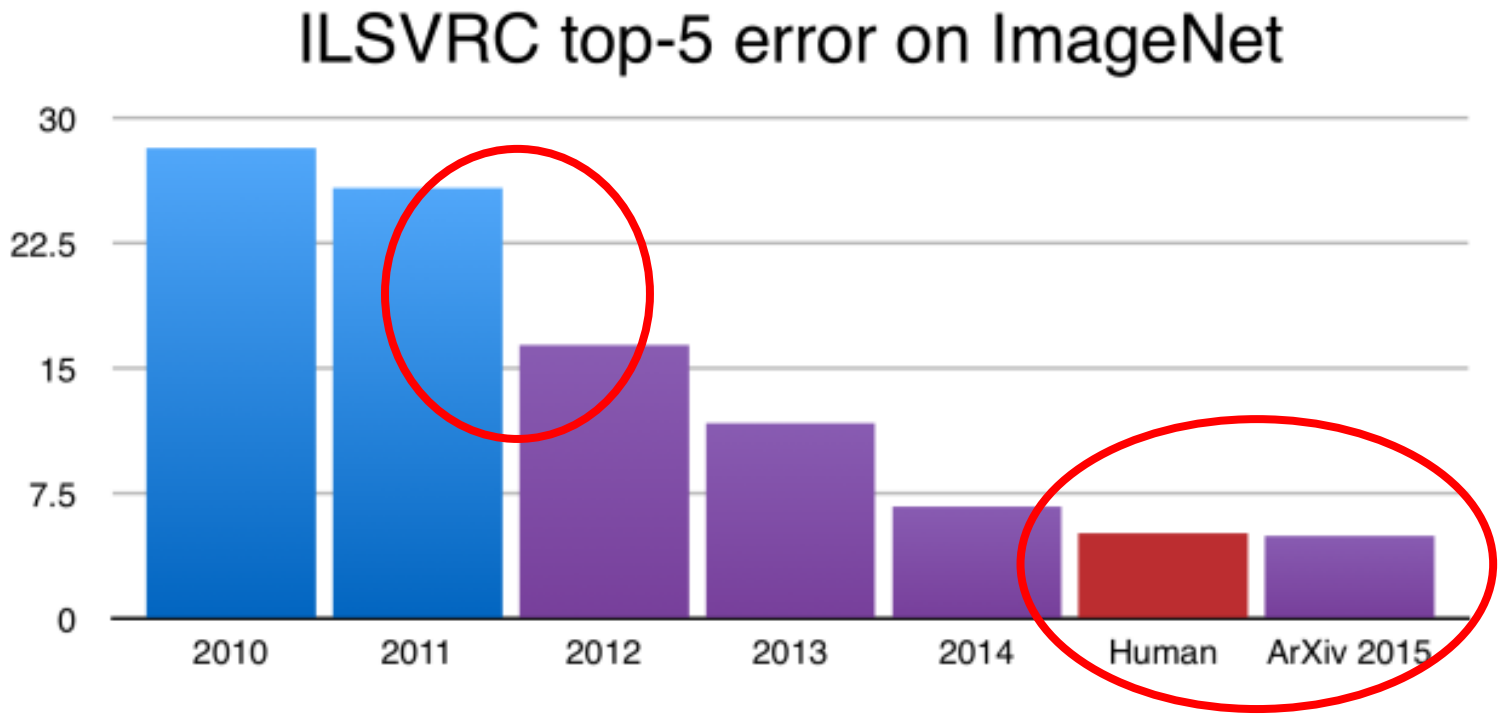
$$y = w_1 x_1 + w_2 x_2$$

If $(w_1, w_2) = (1, -1)$: $y = x_1 - x_2$

y maximal when $(x_1, x_2) = (1, 0)$

Convolutional Neural Networks

IMAGENET **Image Large Scale Visual Recognition Challenge**

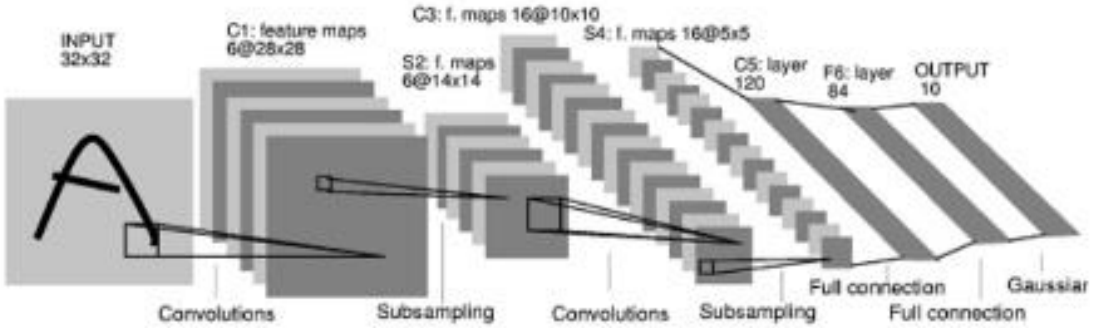


Source: http://image-net.org/

Convolutional Neural Networks

1998

LeCun et al.



# of transistors

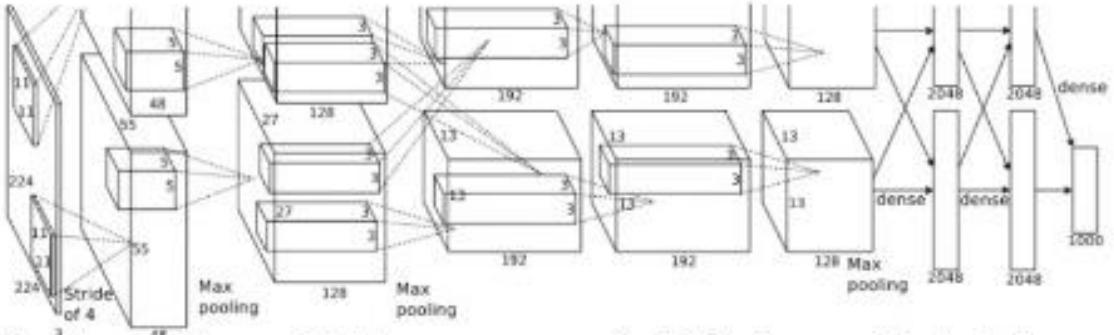
10^6

# of pixels used in training

10^7 NIST

2012

Krizhevsky et al.



# of transistors GPUs

10^9

# of pixels used in training

10^{14} IMAGENET

Convolutional Neural Networks

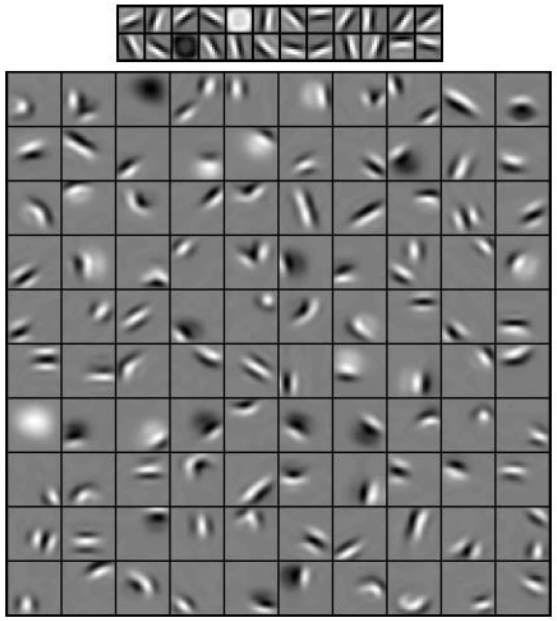
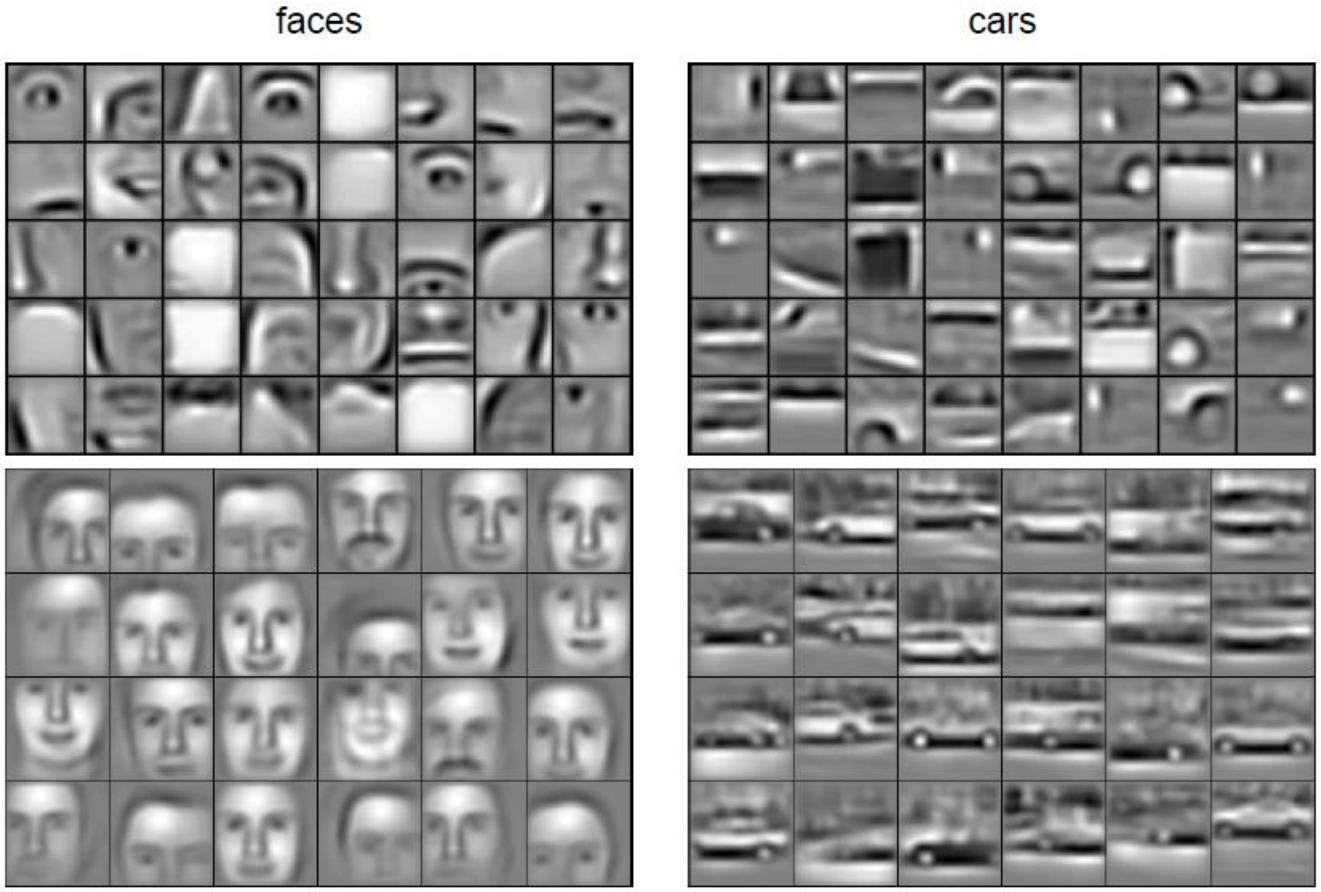


Figure 2. The first layer bases (top) and the second layer bases (bottom) learned from natural images. Each second layer basis (filter) was visualized as a weighted linear combination of the first layer bases.

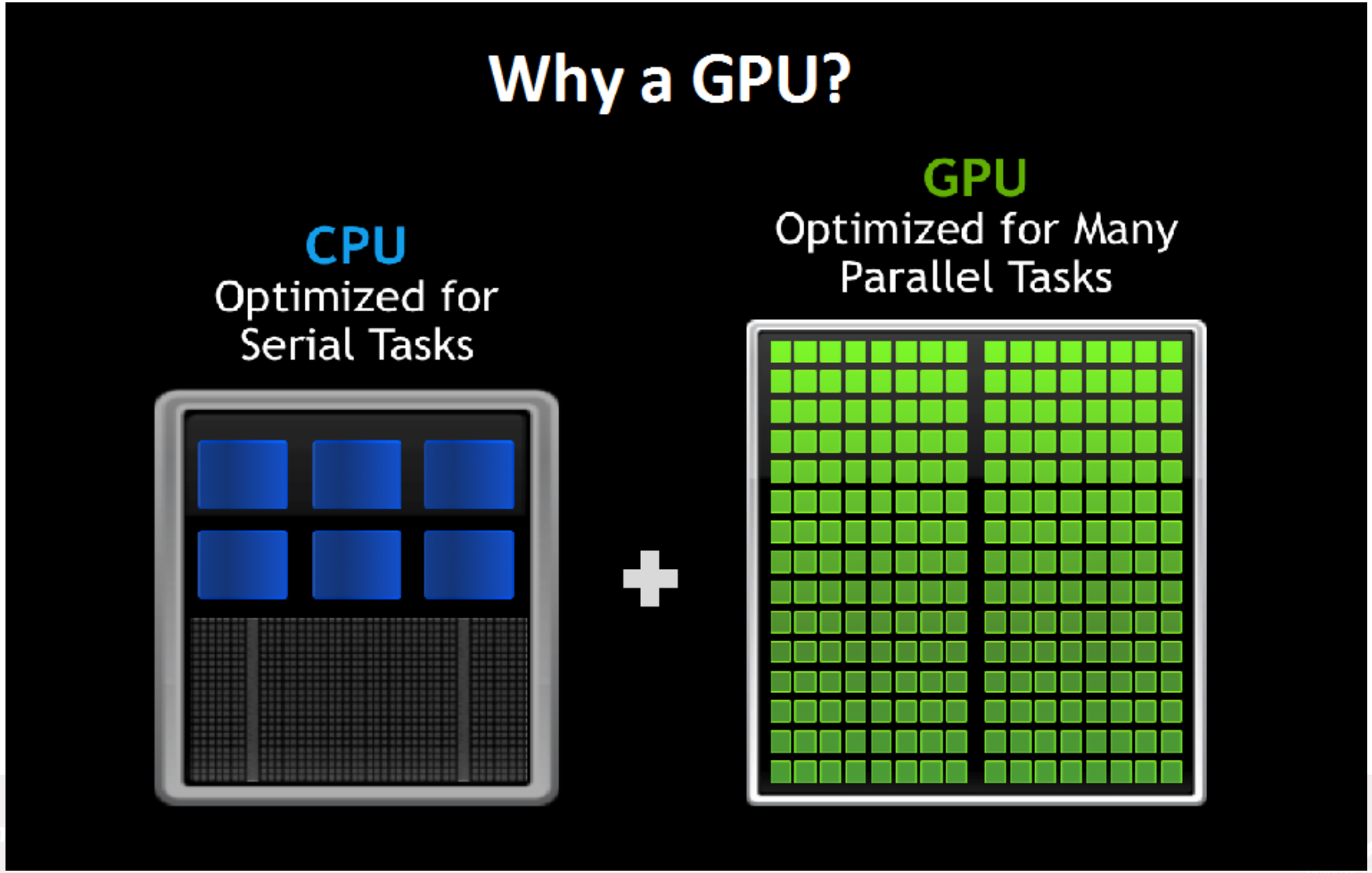


[8] H. Lee, R. Grosse, R. Ranganath, y A. Y. Ng,

Todo esto se puede hacer gracias a...

NVIDIA GRANT

Por qué GPU´s?


Why a GPU?
CPU
Optimized for
Serial Tasks
GPU
Optimized for Many
Parallel Tasks

En que lo hemos aplicado?

Why Motorcycles?? (Yakarta ☹)

Source: Escrito en el futuro. (2017, November 5). Retrieved 25 November 2017, from http://www.hectorabad.com/escrito-en-el-futuro/

CNN used as Feature extraction

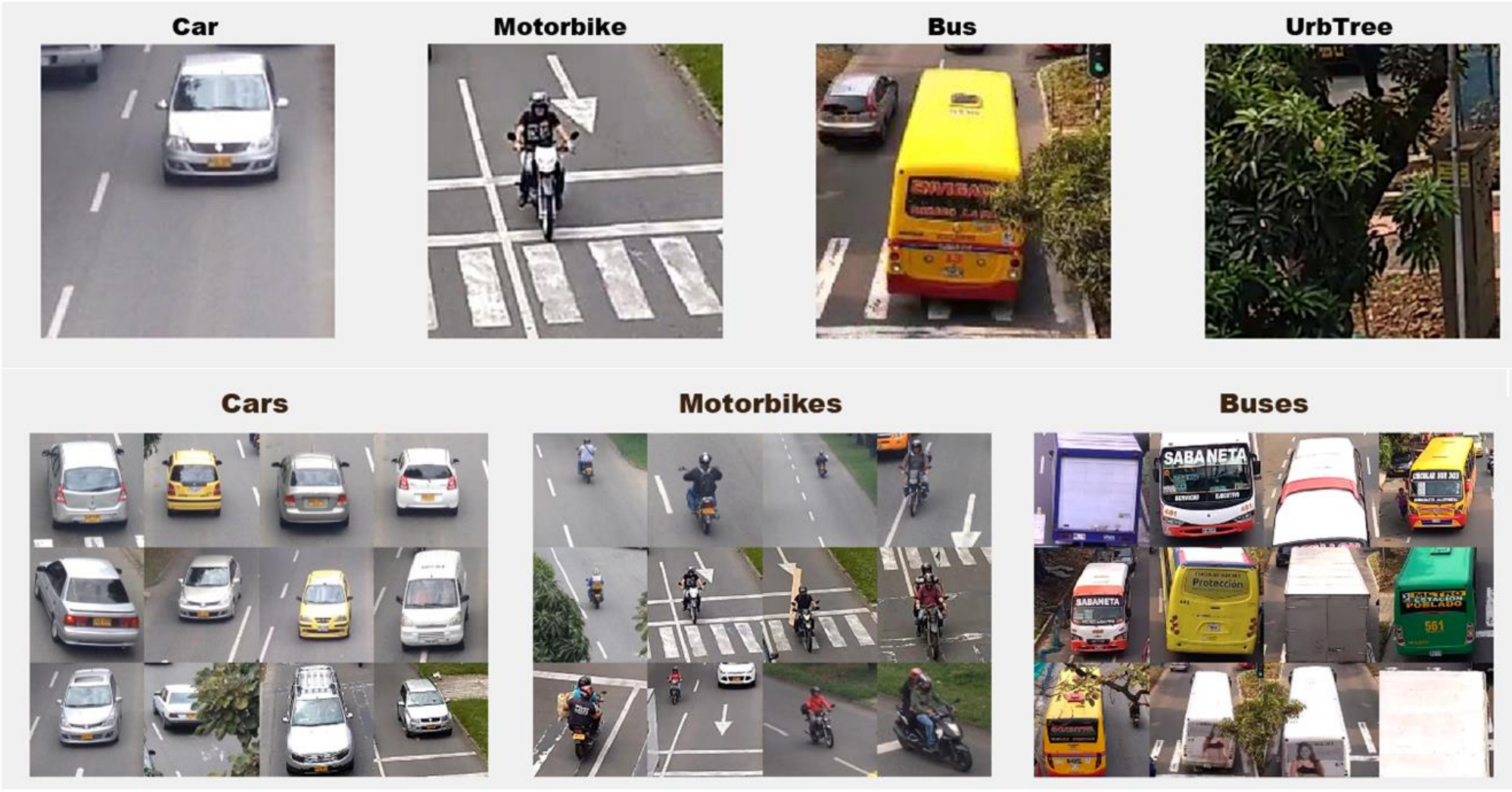
AlexNet as a Pretraining CNN

- AlexNet (ImageNet [2]) took almost a **entire week** for training , running in two GPUs GTX 580 3GB, **ImagineNet** dataset contains more than 15 millions of high res images, labelled and classified on 1000 categories.

- CNN can be pretrainned, for two purposes:
  - **Feature extraction:** Feature extraction: where a CNN is used to extract features from data (in this case images) and then use the learned features to train a different classifier, e.g., a support vector machine (SVM).
  - **Transfer learning:** Where a network already trained on a big dataset is retrained in the last few layers on a more compact data set.

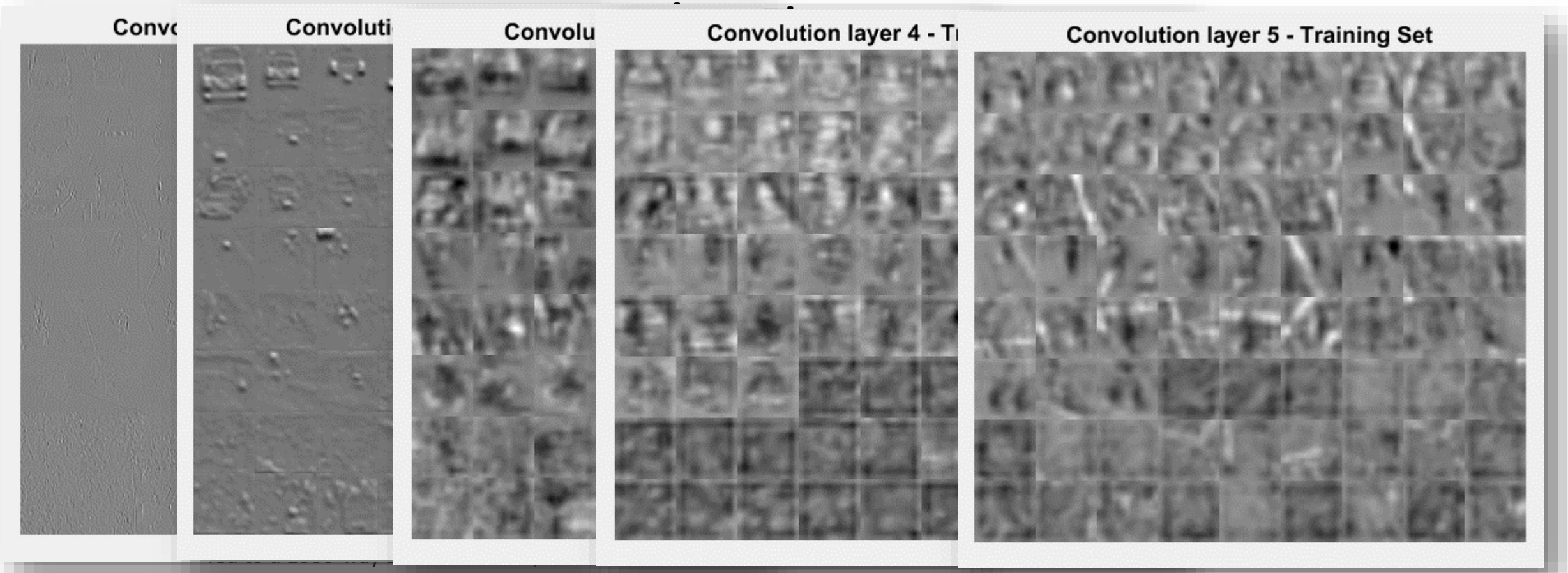
  This can be verified in Razavian et al. [10], where a generic descriptor is generated from a CNN and then it is used in the net OverFed[11] to perform task of object recognition and classification.

AlexNet for Vehicle Classification



The four categories Dataset created for classification used in **AlexNet model**
Only 80 Examples per Category = 320 Total

AlexNet for Vehicle Classification

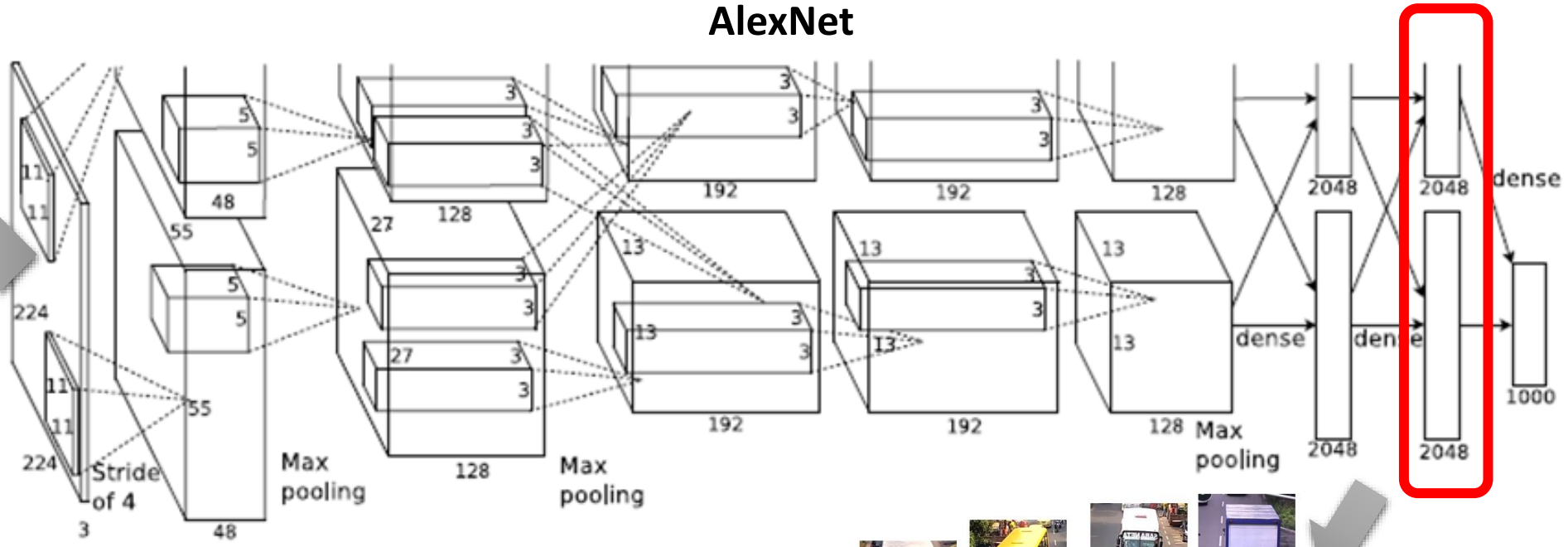


The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers with a final 1000-way softmax

AlexNet - A. Krizhevsky, I. Sutskever, y G. E. Hinton, [2]

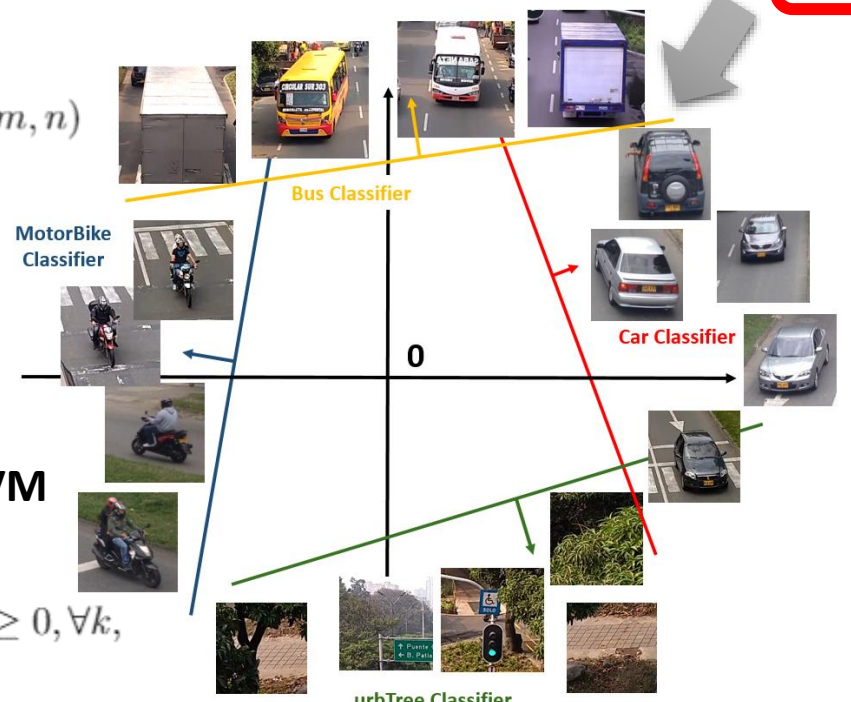
AlexNet for Vehicle Classification

AlexNet



$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(i+m, j+n) K(m,n)$$

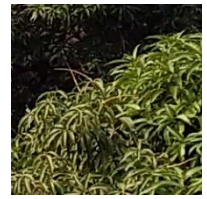
$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right)$$



Car

Bus

MotorBike



UrbTree

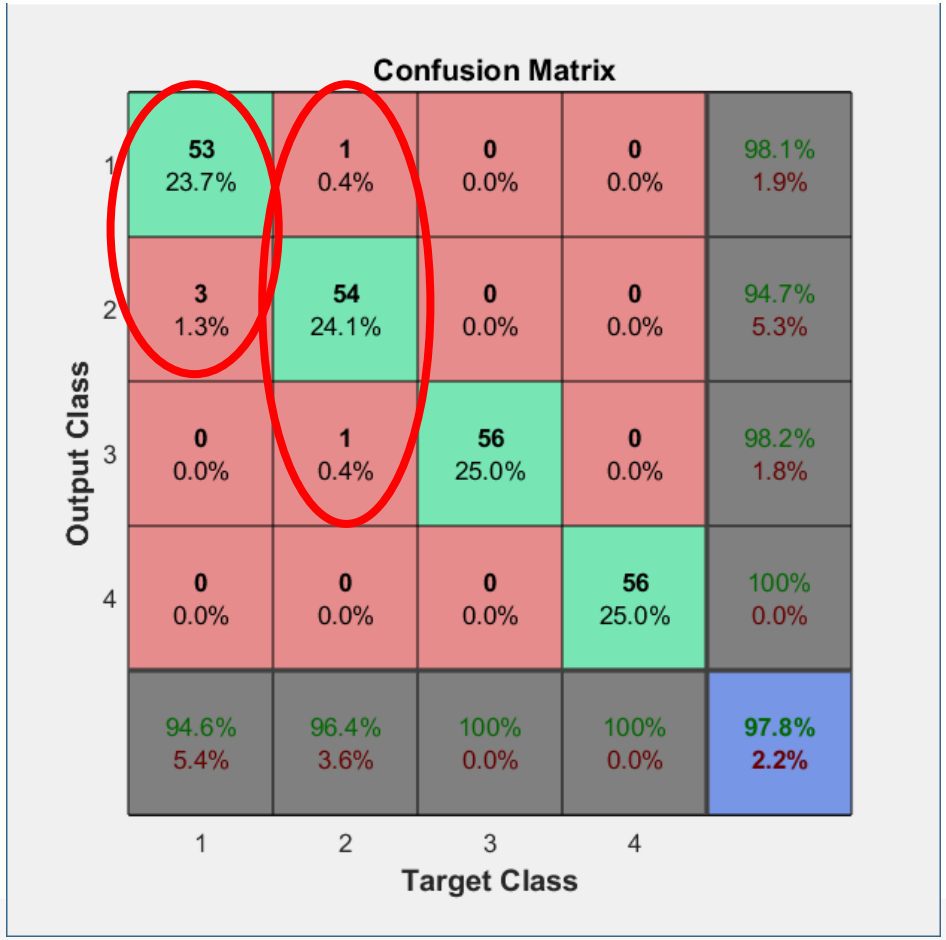
Linear SVM

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

$$s.t. \quad y_k(w^T \phi(x_i, y) + b) \geq 1 - \xi k, \xi k \geq 0, \forall k,$$

AlexNet for Vehicle Classification

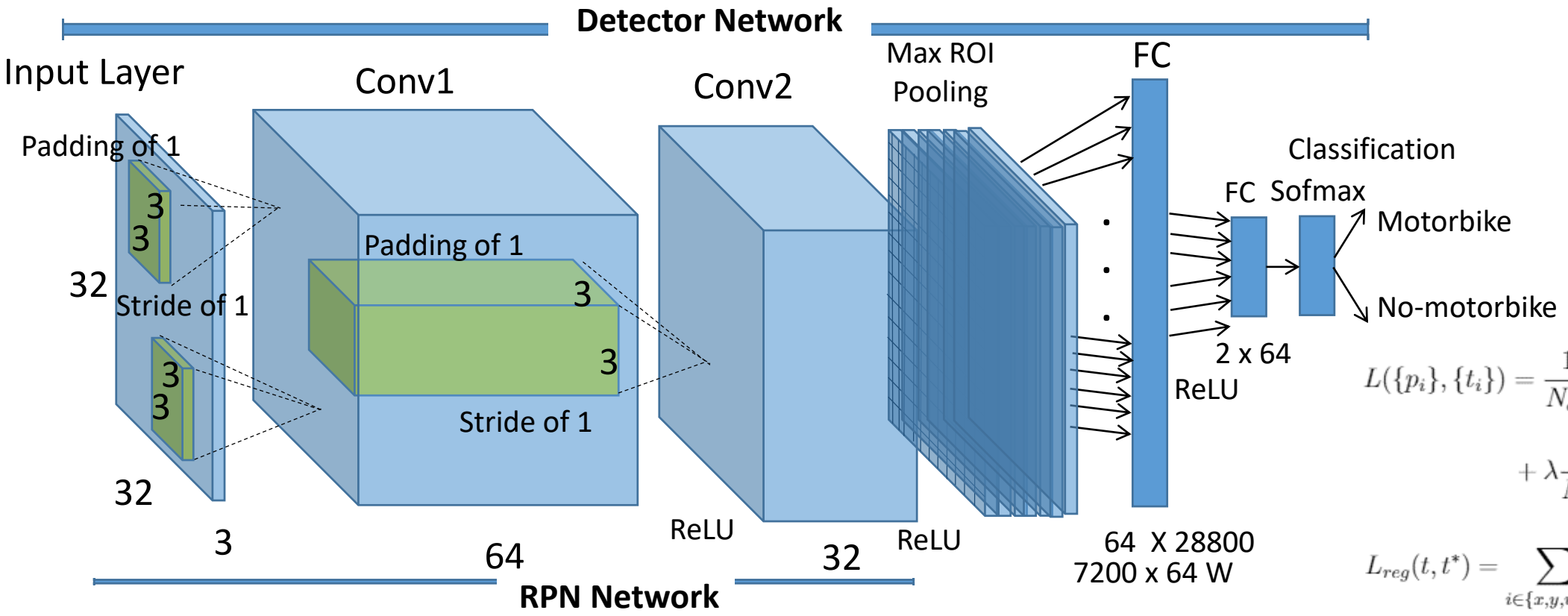
Classification Results



Confusion Matrix of the experiments.
(Class 1: Buses 2: Cars 3: Motorcycles 4: urbTree)

- Mean Accuracy: 97,80%
  (Training 30% – Test 70 %)
- Cross Validated Mean Accuracy : 100%
  (k=10, Training 90% – Test 10 %)
- Cross Validated Mean Accuracy : 99,31%
  (k=10, Training 10% – Test 90 %)

Our Model: EspiNet



$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}}\sum_i L_{cls}(p_i, p_i^*) + \lambda\frac{1}{N_{reg}}\sum_i p_i^* L_{reg}(t_i, t_i^*)$$

$$L_{reg}(t,t^*) = \sum_{i\in\{x,y,w,h\}} smoothL_1(t_i - t_i^*),$$

$$smoothL_1(x) = \begin{cases} 0{,}5x^2 & \text{if } |x| < 1 \\ |x| - 0{,}5 & \text{otherwise,} \end{cases}$$

$$t_x = (x - x_a)/w_a, \quad t_y = (y - y_a)/h_a$$
$$t_w = log(w/w_a), \quad t_h = log(h/h_a)$$
$$t_x^* = (x^* - x_a)/w_a, \quad t_y^* = (y^* - y_a)/h_a$$
$$t_w^* = log(w^*/w_a), \quad t_h^* = log(h^*/h_a)$$

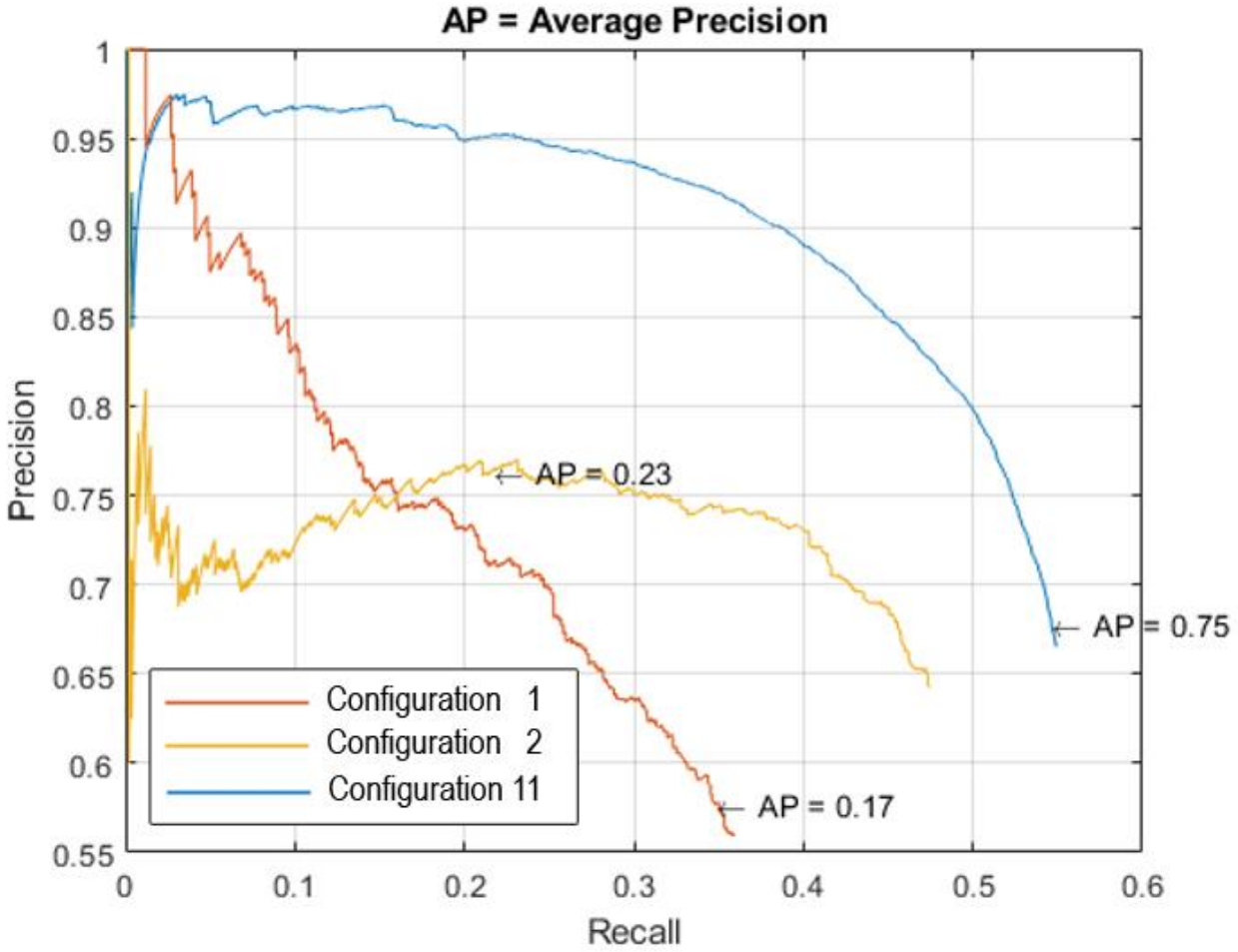
$$n_{out} = \left\lfloor \frac{n_{in} + 2p - k}{s} \right\rfloor + 1$$

- 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)

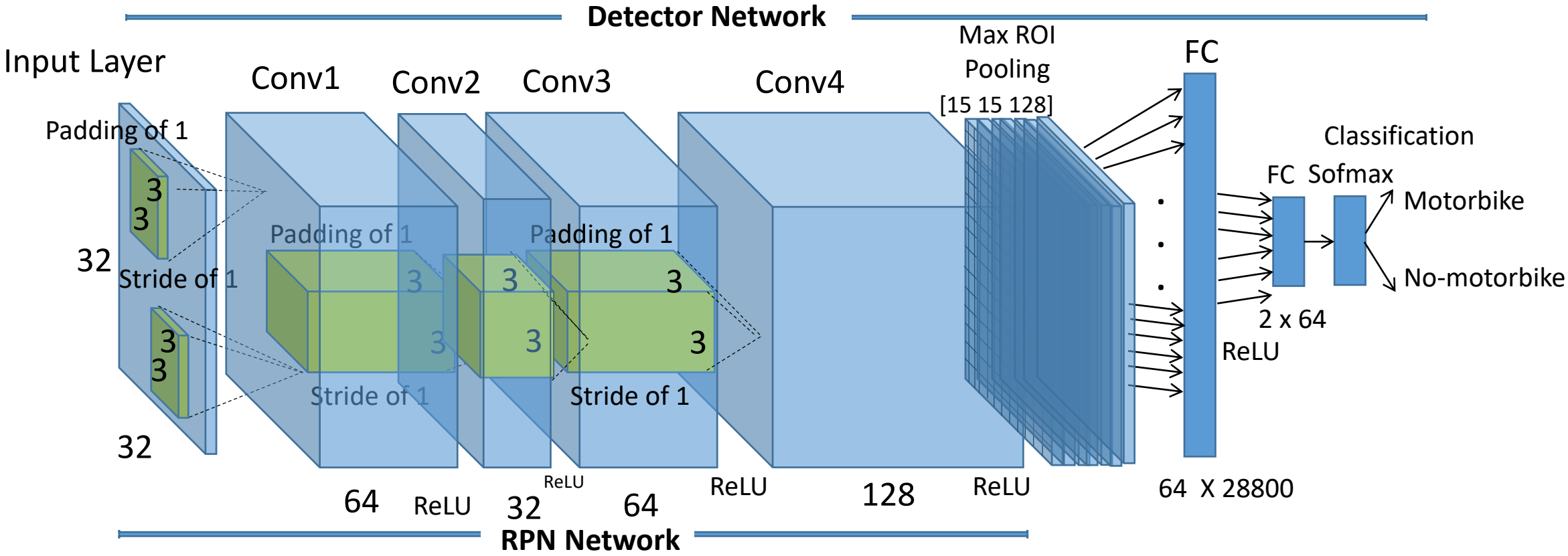
Experiments and Results



- 7,500 annotated images
- 220 motorcycles on urban traffic.
- 41,040 ROI annotated objects
- **60% Annotated object are occluded**

AP=75,23%

EspiNet V2



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

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

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

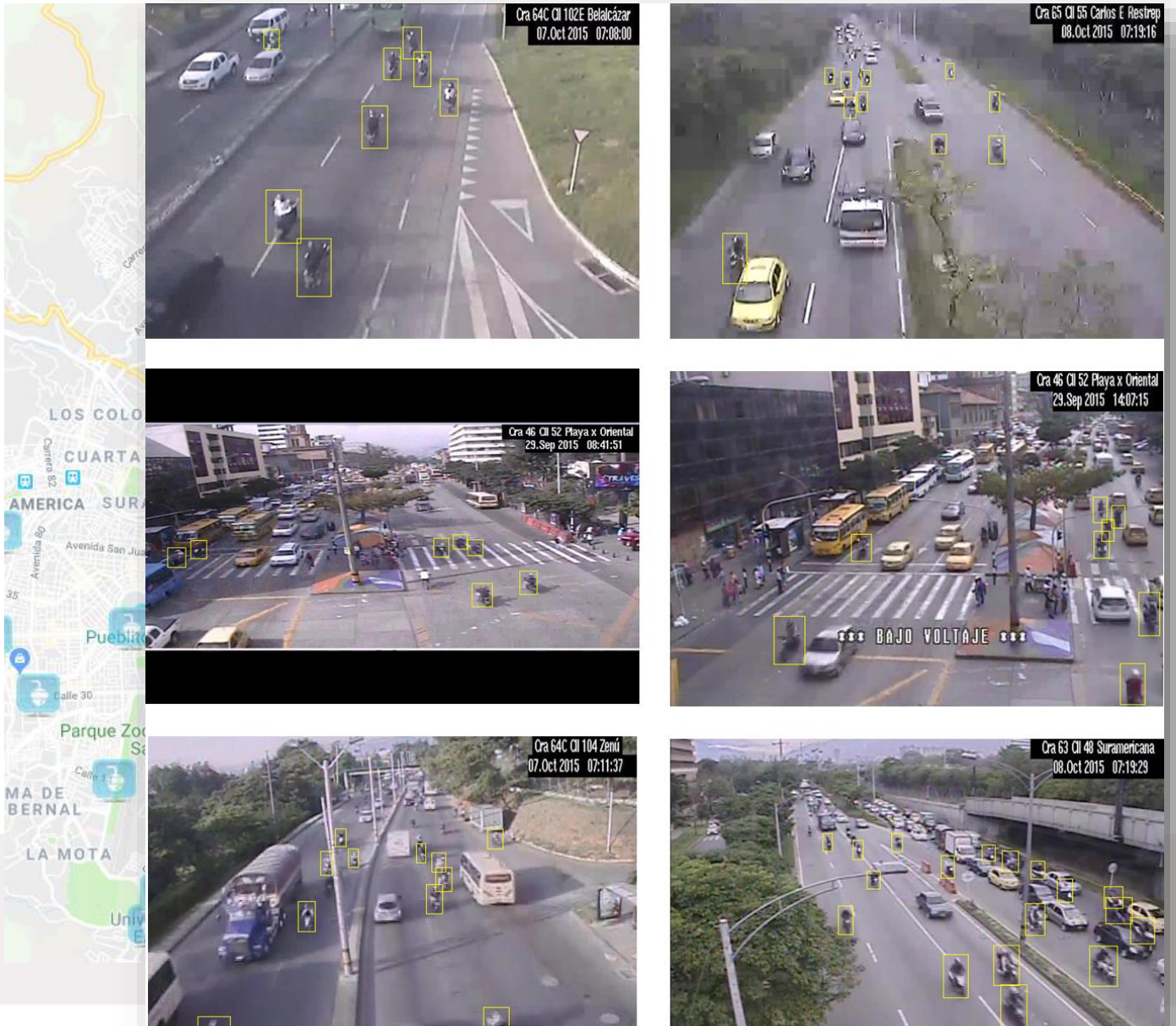
- Took 62 hours for training the dataset
  (90% Training – 5%Validating – 5%Testing)

YOLO vs EspiNet (UMD 10K)

YOLO V3 AP = 80%

EspiNet AP = 90%

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
  Minimal H size 25 px (c.f. KITTI)
- **40 % Annotated object are occluded**

Available Soon at: http://videodatasets.org

YOLO vs EspiNet (Sec5k)

YOLO V3 AP = 77%

EspiNet = 80%

Tracking by Detection

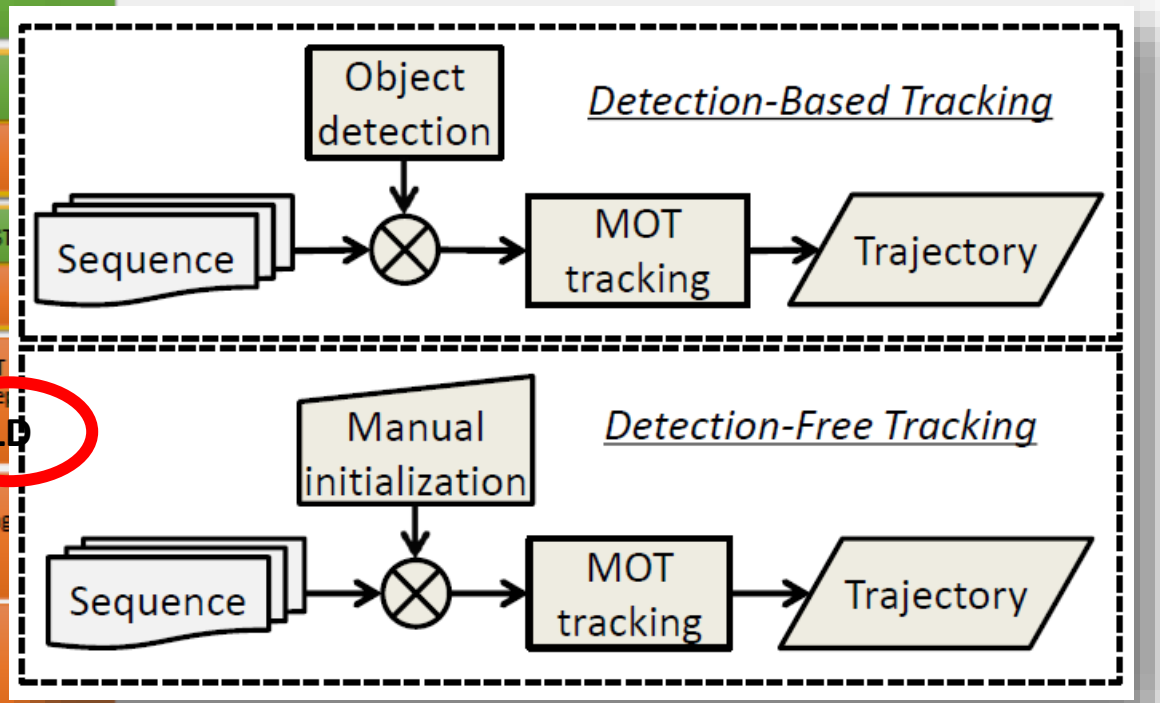
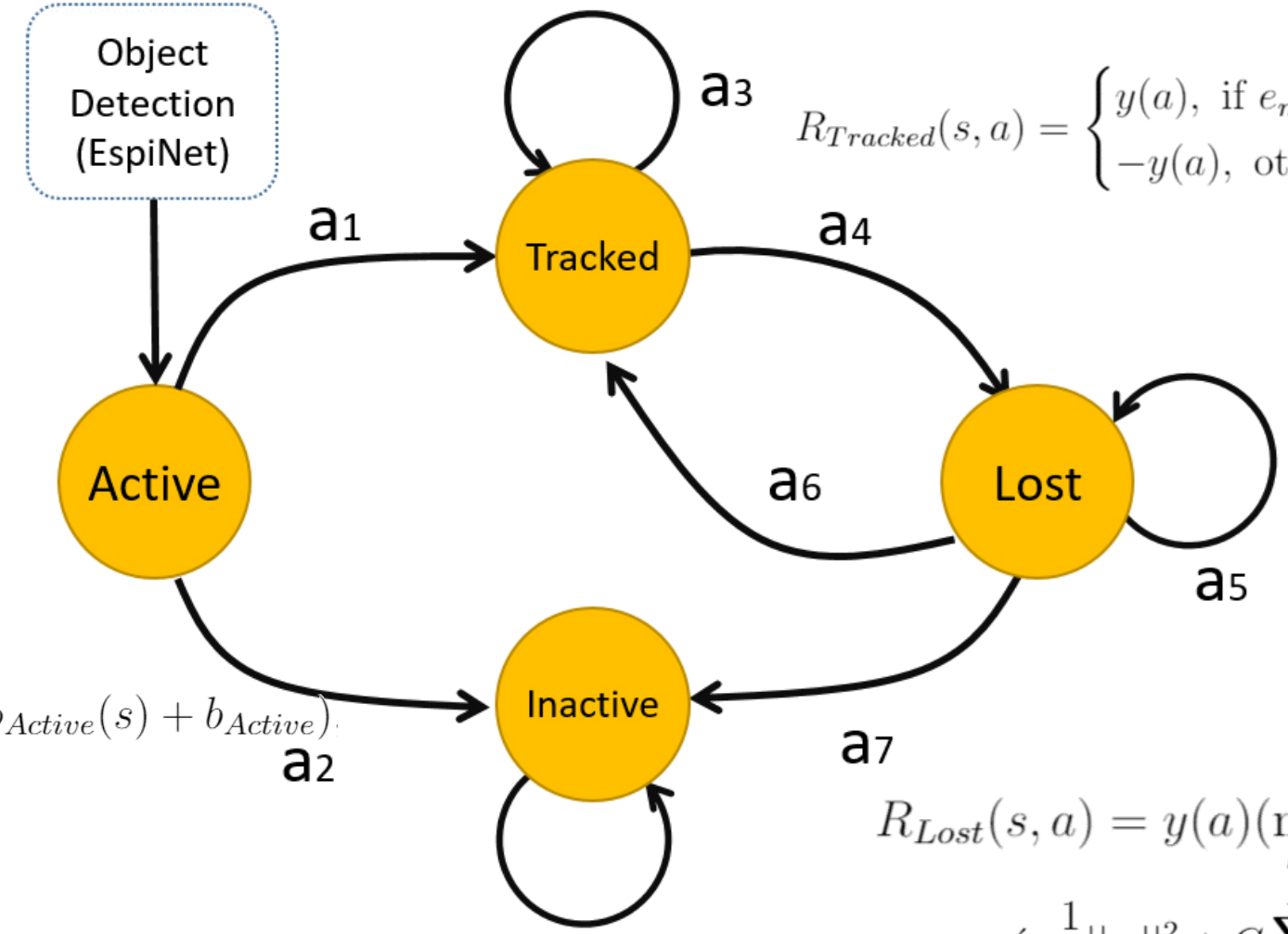


Fig. 1. Taxonomy of tracking algorithms

[9] Mustansar Fiaz, Arif Mahmood, and Soon Ki Jung. Tracking Noisy Targets: A Reviewof Recent Object Tracking Approaches.arXiv preprint arXiv:1802.03098, 2018.

Tracking by Detection



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

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

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

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

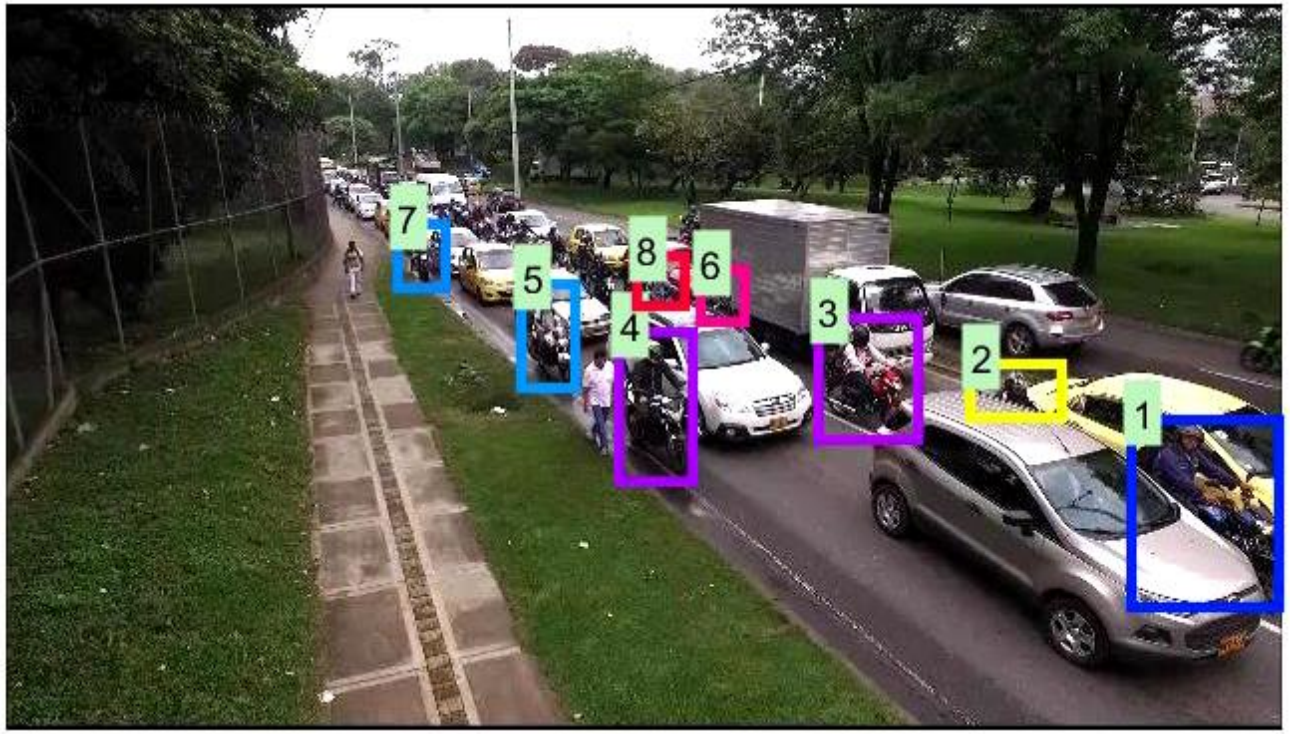
$$s.t. \quad y_k(\mathrm{w}^T \phi(t_k,d_k) + b) \geq 1 - \xi k, \xi k \geq 0, \forall k$$

The MDP of the algorithm

[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
UMD Dataset

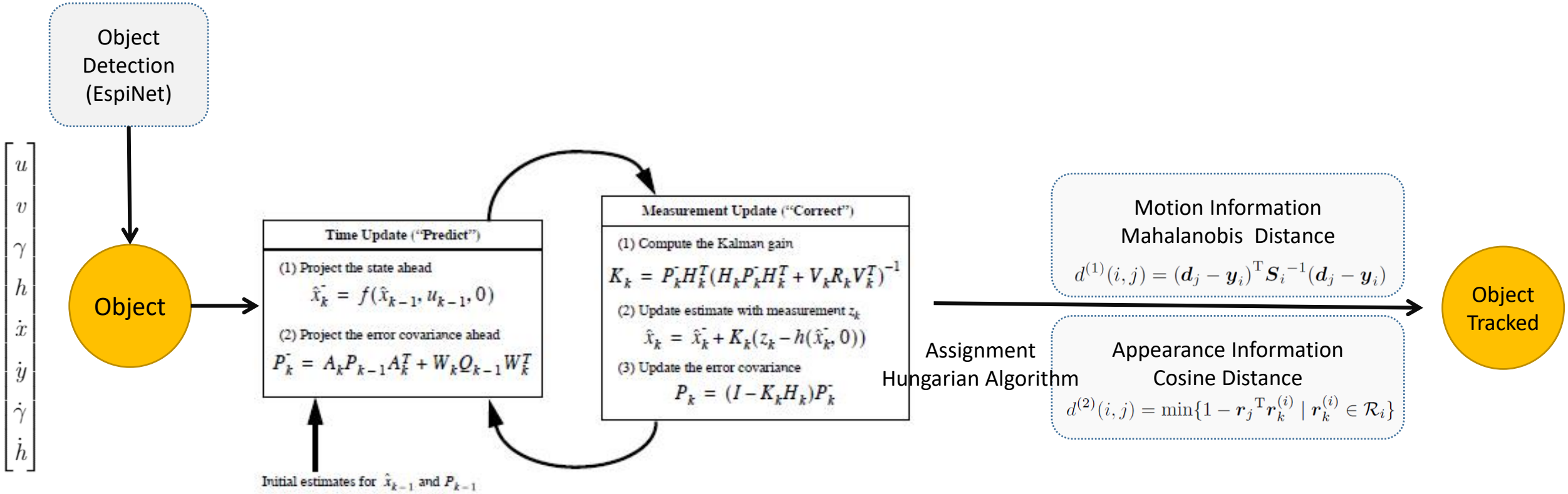
Metrics	EspiNet	Faster R-CNN	YOLO
Recall	**92.5**	73.4	69.6
Precision	93.6	68.7	**95.7**
F1-Score	**93.0**	70.9	80.6
False Alarm Rate	0.33	1.76	**0.17**
GT Tracks	318	318	318
Mostly Tracked	**285**	107	69
Mostly Lost	**1**	7	7
False Positives	3,318	17,578	**1,661**
False Negatives	**3,922**	13,951	15,996
ID Swiches	**75**	662	80
Fragmentations	415	1,630	**299**
MOTA	**86.1**	38.7	66.2
MOTP	**77.7**	72.6	76.7



Rcll	Prcn	FAR	GT	MT	ML	IDs	MOTA	MOTP
92,5	93,6	0.33	318	285	1	75	**86,1**	**77,7**

Comparative results for **MDP tracking** on **UMD dataset** using detectors **EspiNet**, **Faster-RCNN** (based on VGG16) and **YOLO V3**

Tracking by Detection



The DeepSort Tracker algorithm

[82] Nicolai Wojke, Alex Bewley, and Dietrich Paulus. Simple Online and Realtime Trackingwith a Deep Association Metric.arXiv:1703.07402 [cs], March 2017. arXiv: 1703.07402

Tracking by Detection
UMD Dataset

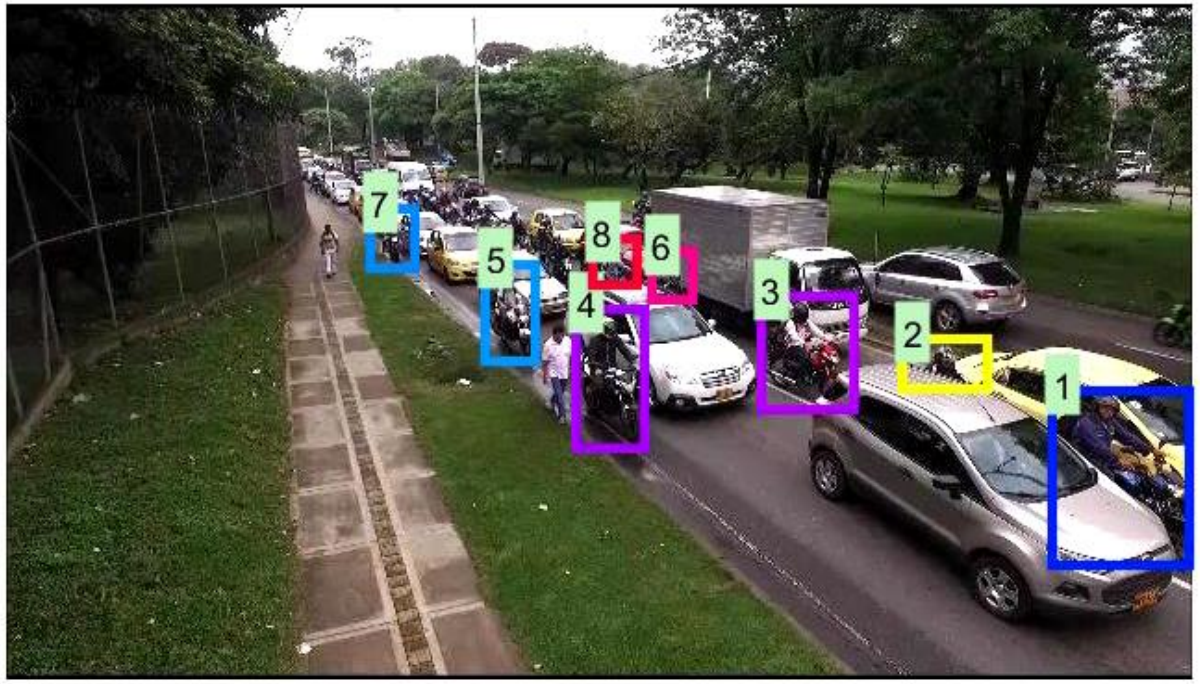
Metrics	EspiNet	Faster R-CNN	YOLO
Recall	**91.1**	77.3	78.9
Precision	**96.6**	58.1	93.3
F1-Score	**93.7**	66.3	85.5
False Alarm Rate	**0.17**	2.93	0.30
GT Tracks	318	318	318
Mostly Tracked	**273**	140	159
Mostly Lost	5	6	**1**
False Positives	**1,704**	29,255	2,959
False Negatives	**4,698**	11,904	11,083
ID Swiches	**112**	1,958	286
Fragmentations	**555**	2,084	635
MOTA	**87.6**	17.9	72.7
MOTP	**77.2**	71.6	76.1

Rcll	Prcn	FAR	GT	MT	ML	IDs	MOTA	MOTP
91,1	96,6	0.17	318	273	5	112	**87,6**	**77,2**

Comparative results for **Deep Sort tracking** on **UMD dataset** using detectors **EspiNet**, **Faster-RCNN** (based on VGG16) and **YOLO V3**

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



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

Tracking by Detection

Secretaría de Medellín Dataset

Conclusiones

- Las redes Convolucionales resultan ser ideales para la selección y extracción de caracteríticas obteniendo así un excelente desempeño en la detección de objetos.
- La Inteligencia Artificial, y más concretamente: las técnicas de reconocimiento de patrones aplicadas a la vision por computador, se encuentran en un estadio muy avanzado de madurez tecnológica.
- El “cuello de botella” para implementar algoritmos de reconocimiento supervisado, sigue siendo la creación y anotación de enormes conjuntos de datos. Es importante explorar alternativas no supervisadas.
- Muchas actividades que requieren constante monitoreo y supervisión pueden ser asumidas con estas técnicas con un muy alto nivel de precisión.

Conclusiones

- Existen implicaciones de seguridad y privacidad que deben ser discutidas ampliamente antes de implementar muchos de los escenarios de monitoreo descritos (Big Brother is Watching You!!!).
- El auge de tecnologías de este tipo aceleran notoriamente las oportunidades tecnologicas, que obligan a un riguroso ejercicio de reflexión respecto al impacto que pueden generar en el desarrollo humano.

PhD Jorge E. Espinosa
jeespinosa@elpoli.edu.co

References

[1] ‘Chapter 3, Traffic Detector Handbook: Third Edition—Volume I - FHWA-HRT-06-108’. [Online]. Available: https://www.fhwa.dot.gov/publications/research/operations/its/06108/03.cfm. [Accessed: 20-Sep-2017].

[2] ‘Vehicle Detection, Tracking and Counting on Behance’. [Online]. Available: https://www.behance.net/gallery/4057777/Vehicle-Detection-Tracking-and-Counting. [Accessed: 20-Sep-2017].

[3] L. W. Tsai, J. W. Hsieh, and K. C. Fan, ‘Vehicle Detection Using Normalized Color and Edge Map’, IEEE Trans. Image Process., vol. 16, no. 3, pp. 850–864, Mar. 2007.

[4] X. Ma and W. E. L. Grimson, ‘Edge-based rich representation for vehicle classification’, in Tenth IEEE International Conference on Computer Vision (ICCV’05) Volume 1, 2005, vol. 2, p. 1185–1192 Vol. 2.

[5] N. Buch, J. Orwell, and S. A. Velastin, ‘3d extended histogram of oriented gradients (3dhog) for classification of road users in urban scenes’, 2009.

[6] R. S. Feris et al., ‘Large-Scale Vehicle Detection, Indexing, and Search in Urban Surveillance Videos’, IEEE Trans. Multimed., vol. 14, no. 1, pp. 28–42, Feb. 2012.

[7] Z. Chen and T. Ellis, ‘Multi-shape Descriptor Vehicle Classification for Urban Traffic’, in 2011 International Conference on Digital Image Computing Techniques and Applications (DICTA), 2011, pp. 456–461.

[8] Z. Chen, T. Ellis, and S. A. Velastin, ‘Vehicle detection, tracking and classification in urban traffic’, in 2012 15th International IEEE Conference on Intelligent Transportation Systems, 2012, pp. 951–956.

[9] S. Gupte, O. Masoud, R. F. Martin, and N. P. Papanikolopoulos, ‘Detection and classification of vehicles’, IEEE Trans. Intell. Transp. Syst., vol. 3, no. 1, pp. 37–47, 2002.

[10] R. Cucchiara, M. Piccardi, and P. Mello, ‘Image analysis and rule-based reasoning for a traffic monitoring system’, IEEE Trans. Intell. Transp. Syst., vol. 1, no. 2, pp. 119–130, Jun. 2000.

[11] 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.

References

[12] C.-L. Huang and W.-C. Liao, ‘A vision-based vehicle identification system’, in Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on, 2004, vol. 4, pp. 364–367.

[13] A. Ottlik and H.-H. Nagel, ‘Initialization of model-based vehicle tracking in video sequences of inner-city intersections’, Int. J. Comput. Vis., vol. 80, no. 2, pp. 211–225, 2008.

[14] B. Tian et al., ‘Hierarchical and Networked Vehicle Surveillance in ITS: A Survey’, IEEE Trans. Intell. Transp. Syst., vol. 16, no. 2, pp. 557–580, Apr. 2015.

[15] ‘ImageNet Large Scale Visual Recognition Competition (ILSVRC)’. [Online]. Available: http://www.image-net.org/challenges/LSVRC/. [Accessed: 24-Oct-2016].

[16] H. Wang, Y. Cai, and L. Chen, ‘A vehicle detection algorithm based on deep belief network’, Sci. World J., vol. 2014, 2014.

[17] Z. Dong, M. Pei, Y. He, T. Liu, Y. Dong, and Y. Jia, ‘Vehicle Type Classification Using Unsupervised Convolutional Neural Network’, in 2014 22nd International Conference on Pattern Recognition (ICPR), 2014, pp. 172–177.

[18] X. Chen, S. Xiang, C. L. Liu, and C. H. Pan, ‘Vehicle Detection in Satellite Images by Hybrid Deep Convolutional Neural Networks’, IEEE Geosci. Remote Sens. Lett., vol. 11, no. 10, pp. 1797–1801, Oct. 2014.

[19] C. Hu, X. Bai, L. Qi, P. Chen, G. Xue, and L. Mei, ‘Vehicle Color Recognition With Spatial Pyramid Deep Learning’, IEEE Trans. Intell. Transp. Syst., vol. 16, no. 5, pp. 2925–2934, Oct. 2015.

[20] B. Su, J. Shao, J. Zhou, X. Zhang, and L. Mei, ‘„Vehicle Color Recognition in The Surveillance with Deep Convolutional Neural Networks “’, 2015.

[21] F. Zhang, X. Xu, and Y. Qiao, ‘Deep classification of vehicle makers and models: The effectiveness of pre-training and data enhancement’, in 2015 IEEE International Conference on Robotics and Biomimetics (ROBIO), 2015, pp. 231–236.

[22] C. M. Bautista, C. A. Dy, M. I. Mañalac, R. A. Orbe, and M. Cordel, ‘Convolutional neural network for vehicle detection in low resolution traffic videos’, in 2016 IEEE Region 10 Symposium (TENSYMP), 2016, pp. 277–281.

[23] R. Girshick, ‘Fast r-cnn’, in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1440–1448.

References

[24] S. Wang, F. Liu, Z. Gan, and Z. Cui, 'Vehicle type classification via adaptive feature clustering for traffic surveillance video', in 2016 8th International Conference on Wireless Communications Signal Processing (WCSP), 2016, pp. 1–5.

[25] F. Chabot, M. Chaouch, J. Rabarisoa, C. Teulière, and T. Chateau, 'Deep MANTA: A Coarse-to-fine Many-Task Network for joint 2D and 3D vehicle analysis from monocular image', ArXiv Prepr. ArXiv170307570, 2017.

[26] Q. Fan, L. Brown, and J. Smith, 'A closer look at Faster R-CNN for vehicle detection', in 2016 IEEE Intelligent Vehicles Symposium (IV), 2016, pp. 124–129.

[27] S. Ren, K. He, R. Girshick, and J. Sun, 'Faster r-cnn: Towards real-time object detection with region proposal networks', in Advances in neural information processing systems, 2015, pp. 91–99.

[28] K. Simonyan and A. Zisserman, 'Very deep convolutional networks for large-scale image recognition', ArXiv Prepr. ArXiv14091556, 2014.

[29] D. Liu and Y. Wang, 'Monza: Image Classification of Vehicle Make and Model Using Convolutional Neural Networks and Transfer Learning'.

[30] Y. Gao and H. J. Lee, 'Local Tiled Deep Networks for Recognition of Vehicle Make and Model', Sensors, vol. 16, no. 2, p. 226, Feb. 2016.

[31] X. Liu, W. Liu, T. Mei, and H. Ma, 'A Deep Learning-Based Approach to Progressive Vehicle Re-identification for Urban Surveillance', in European Conference on Computer Vision, 2016, pp. 869–884.

[32] J. Bromley et al., 'Signature Verification Using A "Siamese" Time Delay Neural Network', IJPRAI, vol. 7, no. 4, pp. 669–688, 1993.

[33] B. Su, J. Shao, J. Zhou, X. Zhang, L. Mei, and C. Hu, 'The Precise Vehicle Retrieval in Traffic Surveillance with Deep Convolutional Neural Networks', Int. J. Inf. Electron. Eng., vol. 6, no. 3, p. 192, 2016.

[34] Y. Cai, X. Sun, H. Wang, L. Chen, and H. Jiang, 'Night-Time Vehicle Detection Algorithm Based on Visual Saliency and Deep Learning', J. Sens., vol. 2016, 2016.

References

[35] Y. Y. Wu and C. M. Tsai, 'Pedestrian, bike, motorcycle, and vehicle classification via deep learning: Deep belief network and small training set', in 2016 International Conference on Applied System Innovation (ICASI), 2016, pp. 1–4.

[36] B.-J. Huang, J.-W. Hsieh, and C.-M. Tsai, 'Vehicle Detection in Hsuehshan Tunnel Using Background Subtraction and Deep Belief Network', in Asian Conference on Intelligent Information and Database Systems, 2017, pp. 217–226.

[37] Y. Zhou, L. Liu, L. Shao, and M. Mellor, 'DAVE: A Unified Framework for Fast Vehicle Detection and Annotation', in European Conference on Computer Vision, 2016, pp. 278–293.

[38] R. You and J.-W. Kwon, 'VoNet: vehicle orientation classification using convolutional neural network', in Proceedings of the 2nd International Conference on Communication and Information Processing, 2016, pp. 195–199.

[39] 'Caffe | Deep Learning Framework'. [Online]. Available: http://caffe.berkeleyvision.org/. [Accessed: 05-Sep-2016].

[40] X. Luo, R. Shen, J. Hu, J. Deng, L. Hu, and Q. Guan, 'A Deep Convolution Neural Network Model for Vehicle Recognition and Face Recognition', Procedia Comput. Sci., vol. 107, pp. 715–720, 2017.

[41] C. Szegedy et al., 'Going deeper with convolutions', in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.

[42] A. Krizhevsky, I. Sutskever, and G. E. Hinton, 'Imagenet classification with deep convolutional neural networks', in Advances in neural information processing systems, 2012, pp. 1097–1105.

[43] M. Lin, Q. Chen, and S. Yan, 'Network in network', ArXiv Prepr. ArXiv13124400, 2013.

[44] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, 'Gradient-based learning applied to document recognition', Proc. IEEE, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[45] 'ImageNet Large Scale Visual Recognition Competition 2012 (ILSVRC2012)'. [Online]. Available: http://www.image-net.org/challenges/LSVRC/2012/. [Accessed: 30-Aug-2017].

[46] L. M. Brown, Q. Fan, and Y. Zhai, 'Self-calibration from vehicle information', in 2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2015, pp. 1–6.

References

[47] C. H. Lampert, M. B. Blaschko, and T. Hofmann, 'Efficient Subwindow Search: A Branch and Bound Framework for Object Localization', IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 12, pp. 2129–2142, Dec. 2009.

[48] J. R. Uijlings, K. E. Van De Sande, T. Gevers, and A. W. Smeulders, 'Selective search for object recognition', Int. J. Comput. Vis., vol. 104, no. 2, pp. 154–171, 2013.

[49] K. He, X. Zhang, S. Ren, and J. Sun, 'Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition', IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

[50] C. L. Zitnick and P. Dollár, 'Edge boxes: Locating object proposals from edges', in European Conference on Computer Vision, 2014, pp. 391–405.

[51] J. Hosang, R. Benenson, P. Dollár, and B. Schiele, 'What makes for effective detection proposals?', IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 4, pp. 814–830, Apr. 2016.

[52] R. Girshick, J. Donahue, T. Darrell, and J. Malik, 'Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation', in 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 580–587.

[53] 'ILSVRC2016'. [Online]. Available: http://image-net.org/challenges/LSVRC/2016/results. [Accessed: 30-Aug-2017].

[54] Z. Zivkovic, 'Improved adaptive Gaussian mixture model for background subtraction', in Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on, 2004, vol. 2, pp. 28–31.

[55] Z. Zivkovic and F. Van Der Heijden, 'Efficient adaptive density estimation per image pixel for the task of background subtraction', Pattern Recognit. Lett., vol. 27, no. 7, pp. 773–780, 2006.

[56] 'Image Category Classification Using Deep Learning - MATLAB & Simulink Example'. [Online]. Available: https://www.mathworks.com/help/vision/examples/image-category-classification-using-deep-learning.html. [Accessed: 28-Feb-2017].

[57] '8th International Conference on Pattern Recognition Systems |Universidad Carlos III de Madrid | Madrid, Spain'. [Online]. Available: http://velastin.dynu.com/icprs17/programme.php. [Accessed: 30-Aug-2017].

References

[58] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, 'CNN Features Off-the-Shelf: An Astounding Baseline for Recognition', in 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2014, pp. 512–519.
[59] M. D. Zeiler and R. Fergus, 'Visualizing and understanding convolutional networks', in European Conference on Computer Vision, 2014, pp. 818–833.
[60] 'The PASCAL Visual Object Classes Challenge 2007 (VOC2007)'. [Online]. Available: http://host.robots.ox.ac.uk/pascal/VOC/voc2007/index.html. [Accessed: 31-Aug-2017].
[61] 'ViPER: The Video Performance Evaluation Resource'. [Online]. Available: http://viper-toolkit.sourceforge.net/. [Accessed: 31-Aug-2017].
[62] F. Yin, D. Makris, and S. A. Velastin, 'Performance evaluation of object tracking algorithms', in IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, Rio De Janeiro, Brazil, 2007, p. 25.
[63] '▶ Deep Learning:Theoretical Motivations - VideoLectures.NET'. [Online]. Available: http://videolectures.net/deeplearning2015_bengio_theoretical_motivations/?q=deep%20learning. [Accessed: 20-Sep-2017].