Motorbikes Classification in Urban traffic using deep learning techniques

Prof. - Jorge Espinosa - PCJIC – UNAL

Prof. Sergio A. Velastin

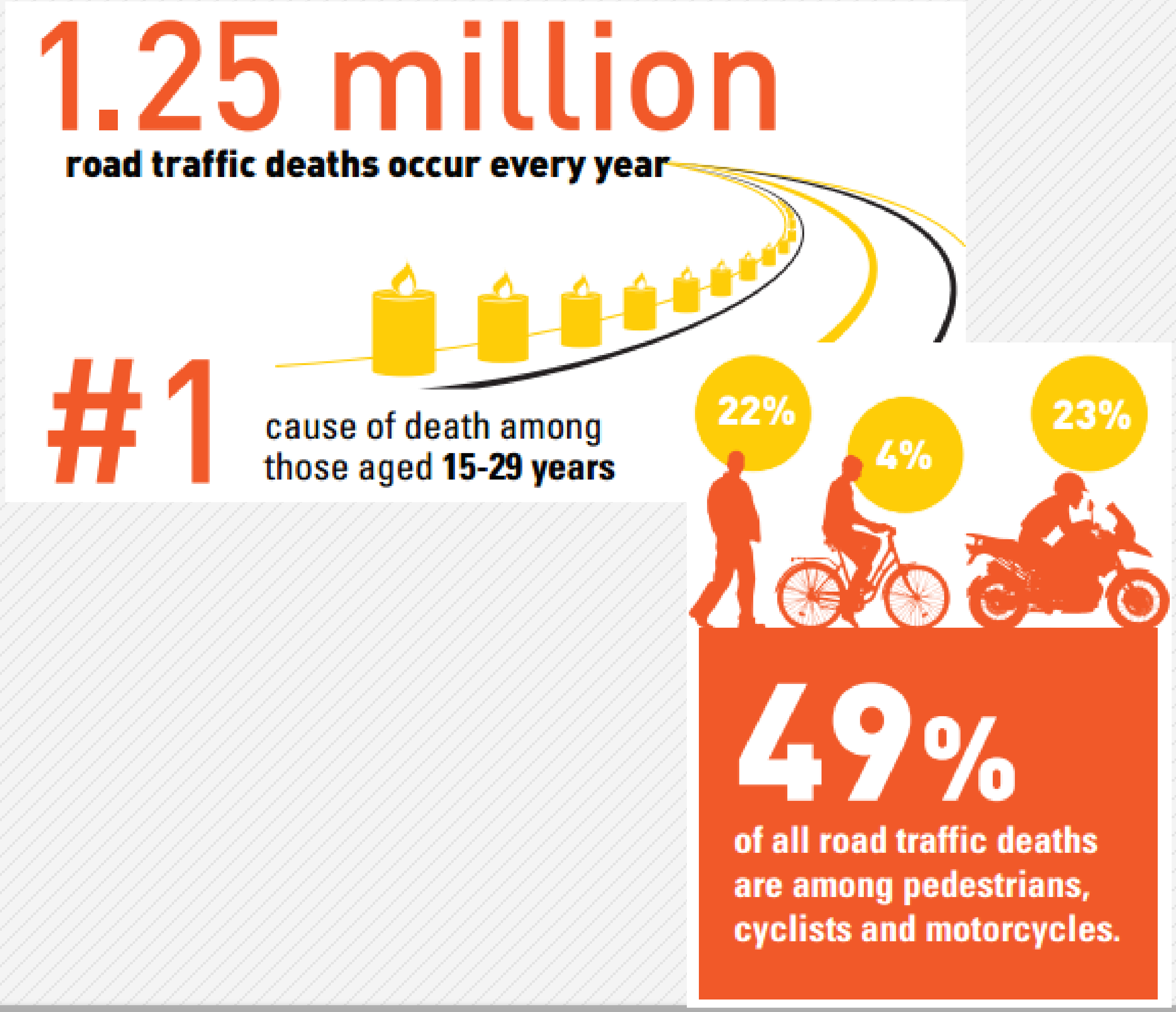
Prof. John William Branch, UNAL

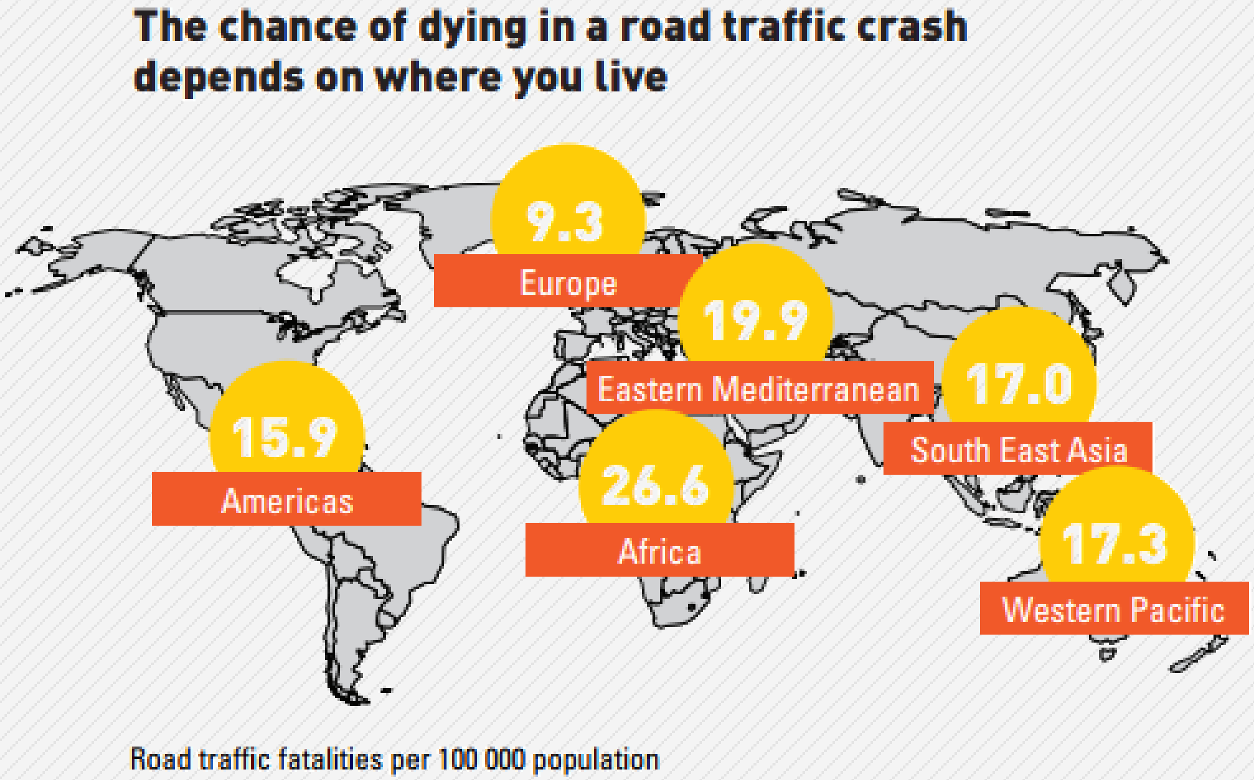
Content

1. Motivation
2. Deep Learning – Convolutional Neural Networks
3. Demo
4. Feature extraction by using Convolutional Neural Networks
5. Our approach
6. Conclusions and Future Work

Introduction

Why Motorcycles??





Global status report on road safety 2015 - WHO

MOYCOT
Movilidad y Control del Tráfico

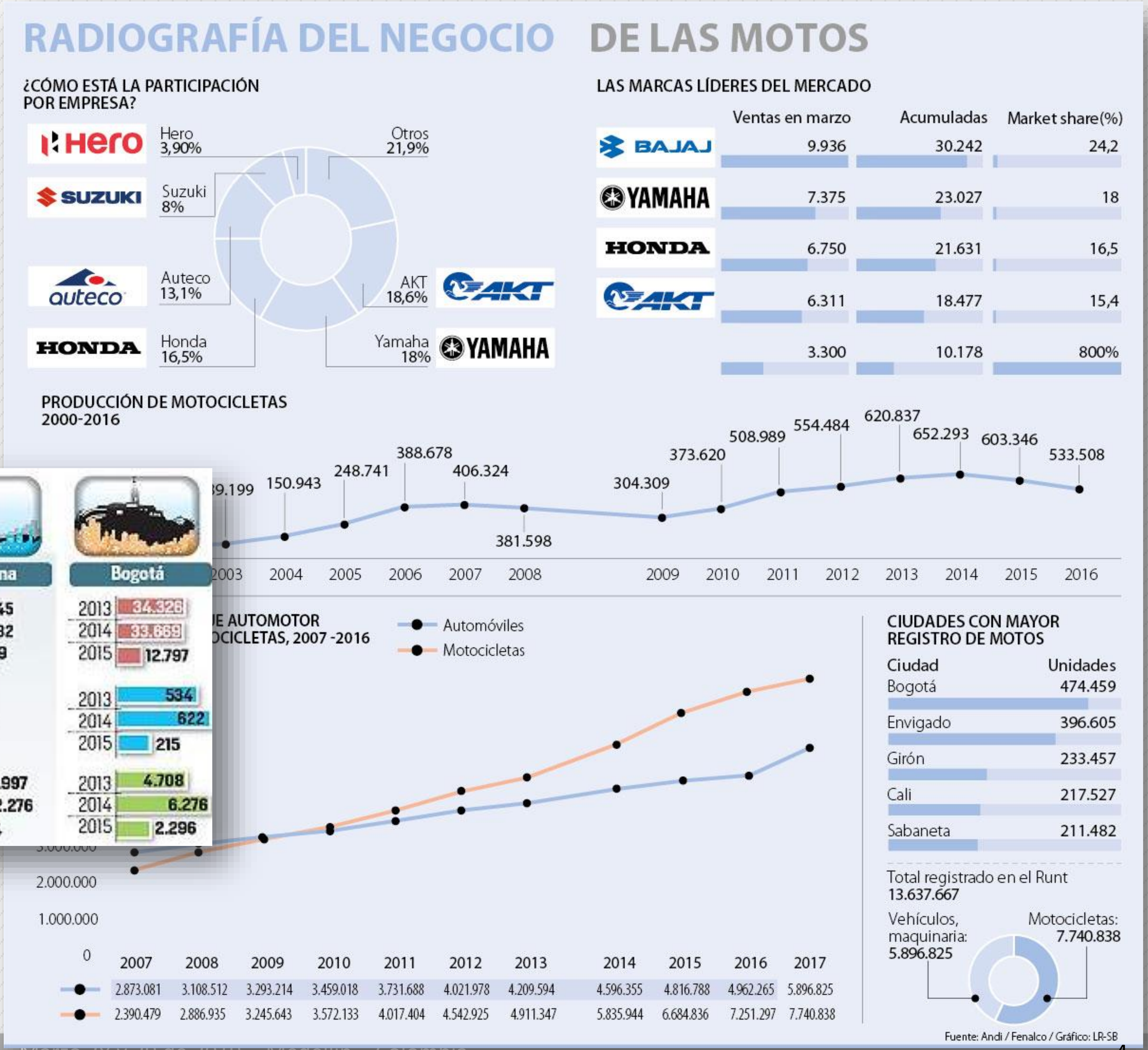
Introduction

Why Motorcycles??

High Accidentally Rate (2Q -2018)

Barranquilla		Cali		Medellín		Bucaramanga		Cartagena		Bogotá	
2013	5.653	2013	23.767	2013	48.750	2013	4.070	2013	5.345	2013	34.326
2014	5.664	2014	21.000	2014	45.409	2014	3.732	2014	5.682	2014	33.669
2015	2.427	2015	7.300	2015	18.358	2015	1.553	2015	2.079	2015	12.797
2013	69	2013	228	2013	306	2013	81	2013	62	2013	534
2014	108	2014	256	2014	289	2014	55	2014	85	2014	622
2015	43	2015	132	2015	95	2015	24	2015	19	2015	215
2013	1.204	2013	2.042	2013	2.832	2013	1.738	2013	1.997	2013	4.708
2014	1.186	2014	2.100	2014	2.301	2014	1.158	2014	2.276	2014	6.276
2015	421	2015	820	2015	376	2015	1.107	2015	844	2015	2.296

Sources: Secretarías de Movilidad y Medicina Legal



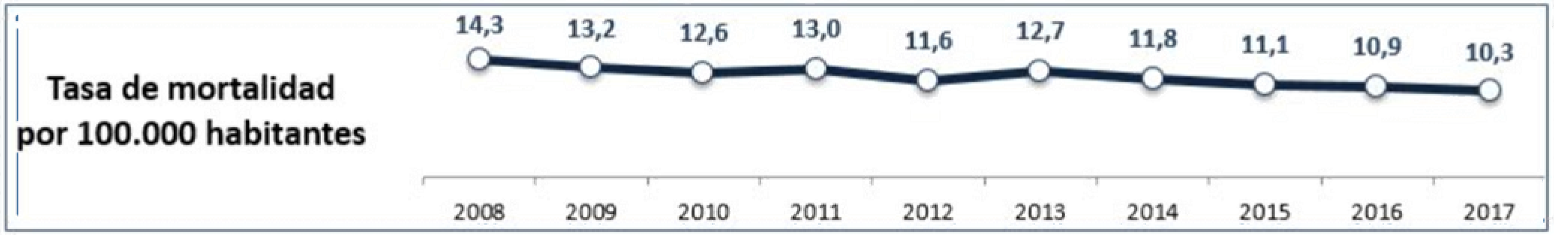
OBSERVATORIO DE MOVILIDAD DE MEDELLÍN

MOYCOT Movilidad y Control del Tráfico

Información general de incidentes, heridos y muertos

Alcaldía de Medellín Cuenta con vos

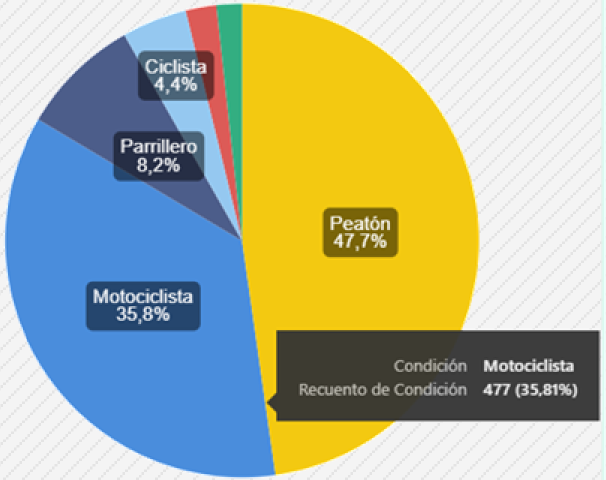
Tasa de mortalidad en incidentes de tránsito por cada 100.000 habitantes



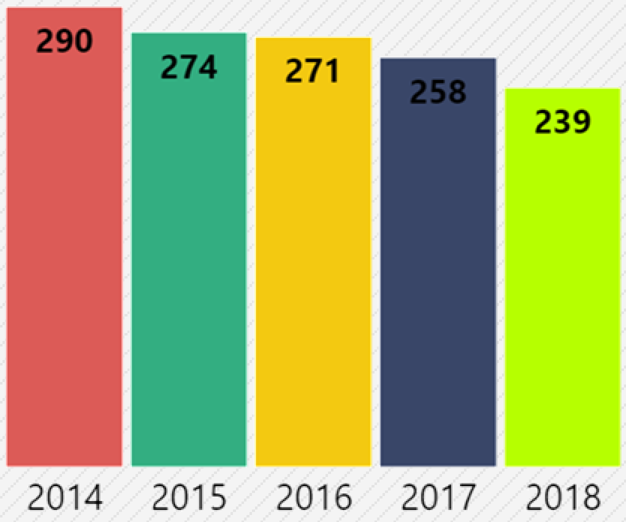
Medellín presenta en los últimos años una tendencia decreciente en el comportamiento de las tasas de accidentalidad y mortalidad por cada 100.000 habitantes.

En el año 2018 se registra una tasa de mortalidad de un solo dígito.

Heridos | Muertos



Víctimas por año en incidentes de tránsito



60% de los incidentes totales involucran un vehículo tipo motocicleta

Mes de más incidentalidad en el año **Agosto**

78% de las víctimas totales son usuarios de motocicleta (conductor y acompañante)

Día de mayor incidentalidad en el año **Viernes**

65% del total de peatones víctimas fatales tenía más de 60 años en 2018

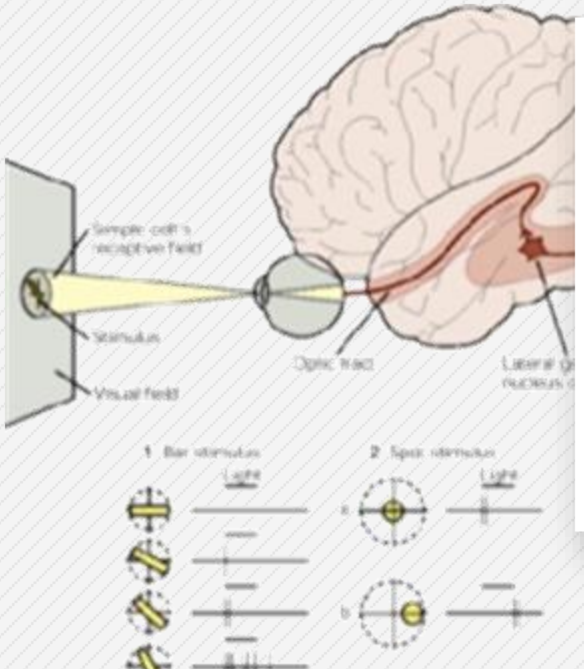
DEMO

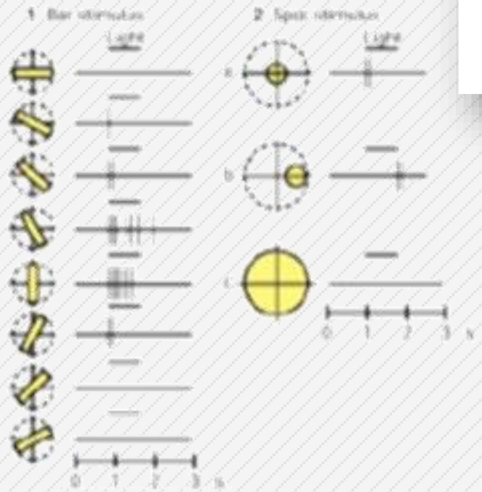
YOLO

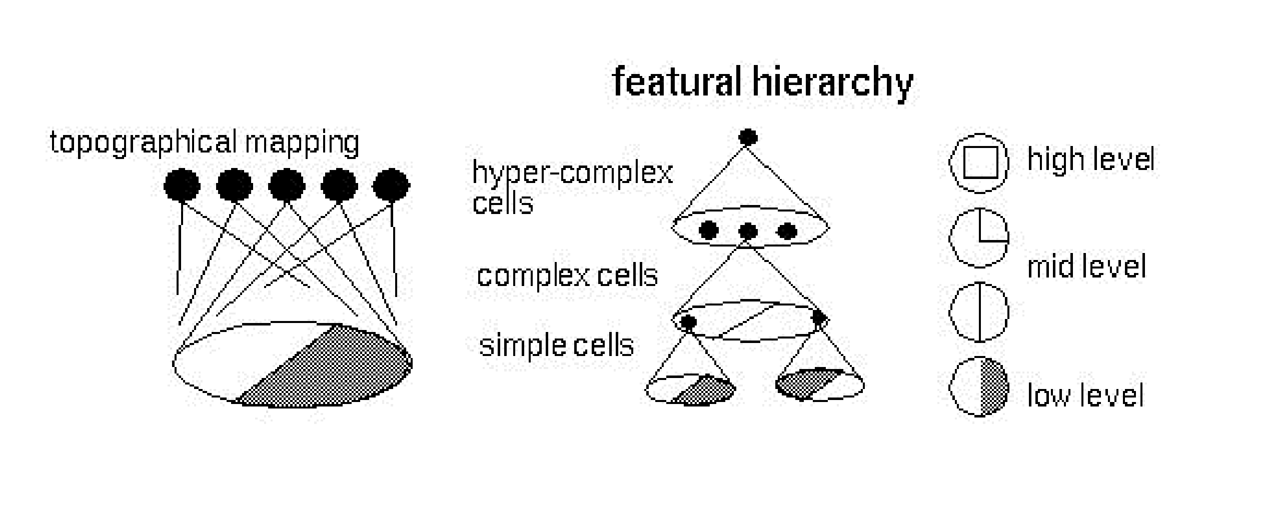
You Look Only Once

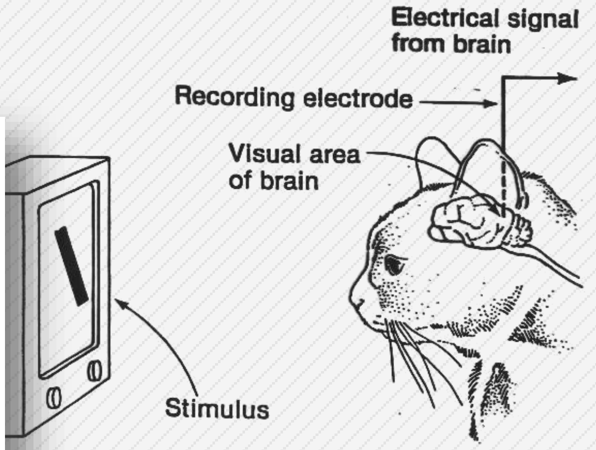
Convolutional Neural Networks

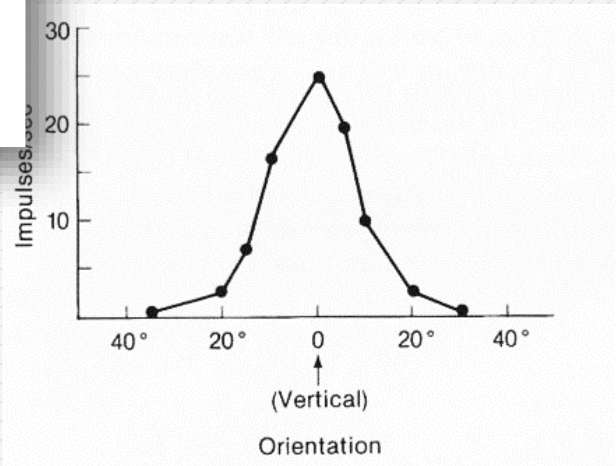
A bit of History







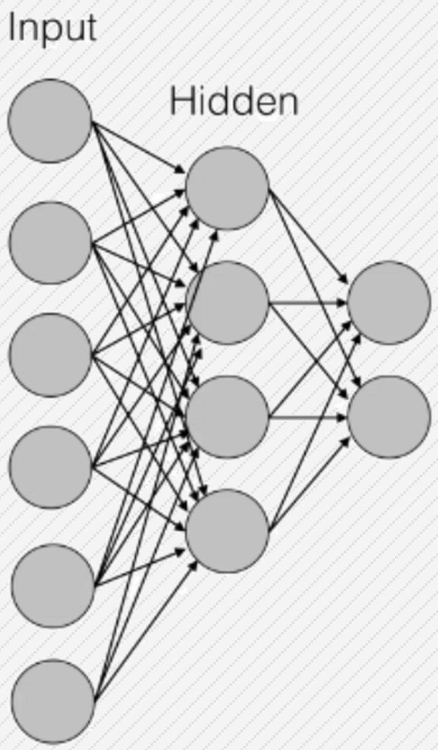




Hubel & Wiesel, 1959, 1962, 1965,1968

Convolutional Neural Networks

Computational Implications

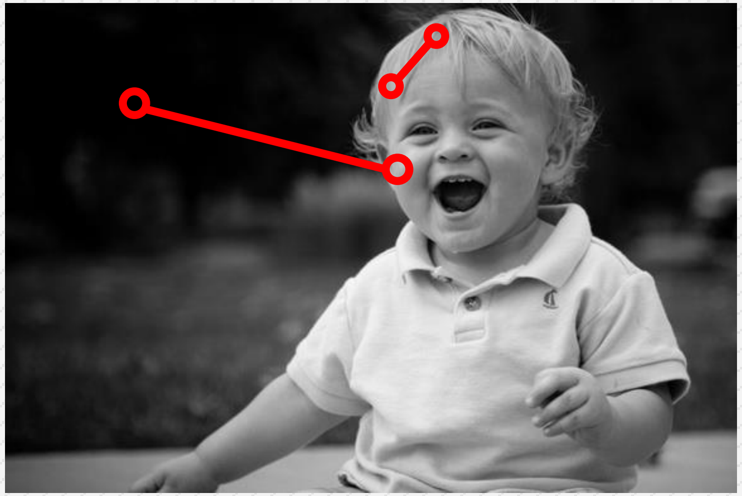


NN MLP Tradtional Architecture

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

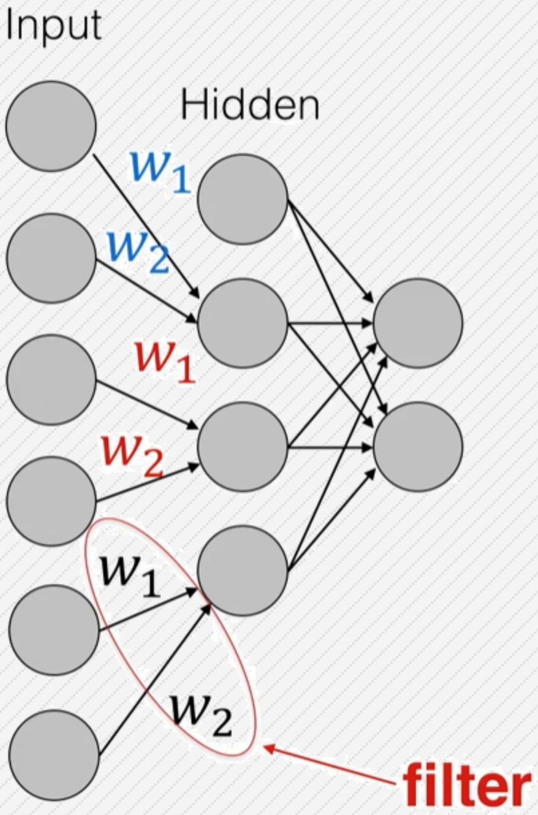
Typical Image: 256 x 256 (56,000 pixels)
First layer weights: 560k parameters !

Too many parameters!!



Space Matters!!

Convolutional Neural Networks
Why Convolution?



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

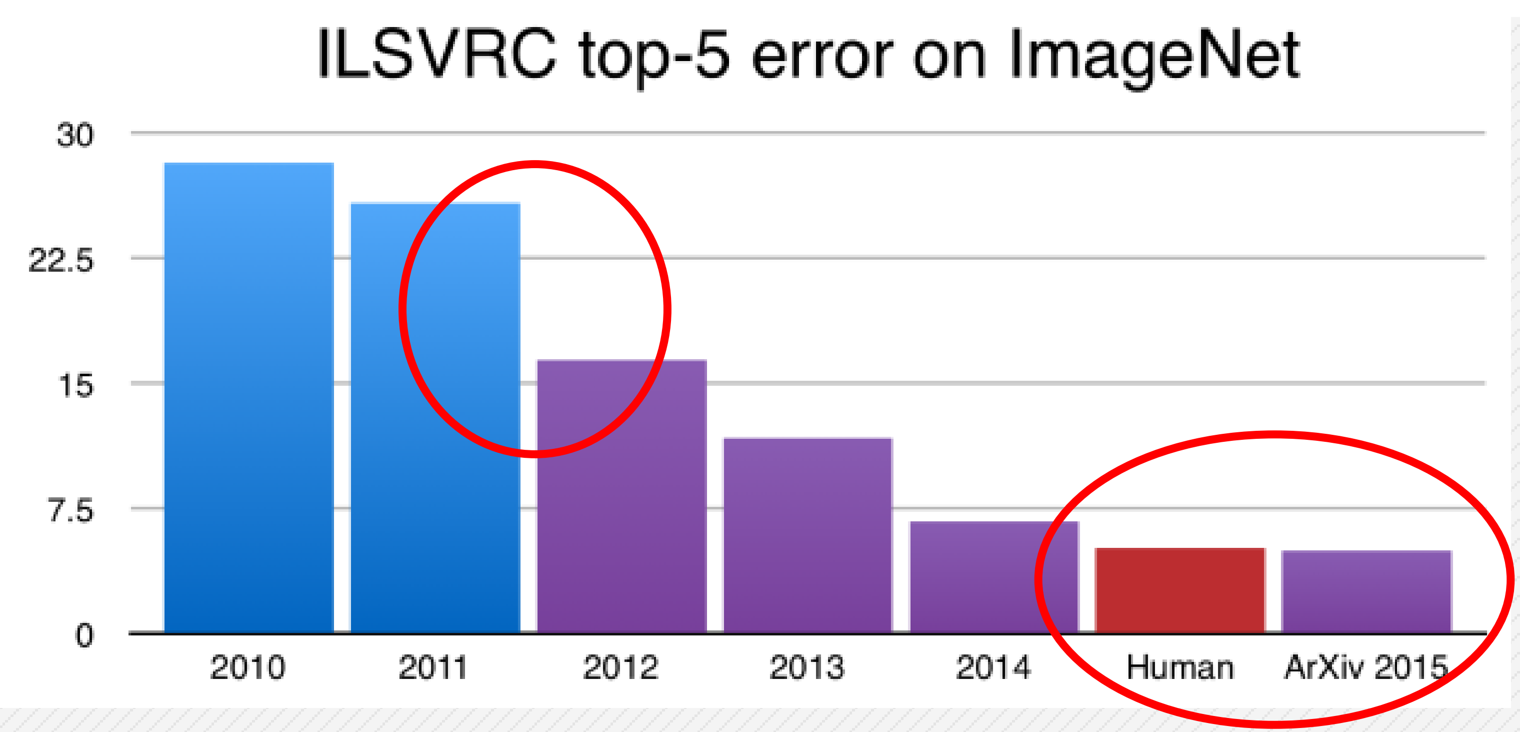
$$y = w_1 x_1 + w_2 x_2$$

If $(\mathrm{w}_1, \mathrm{w}_2) = (1, -1)$: $y = x_1 - x_2$

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

Deep Learning for Vehicle Classification

IMAGENET Image Large Scale Visual Recognition Challenge



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

Deep Learning for Motorbike Classification

Pretraining CNN

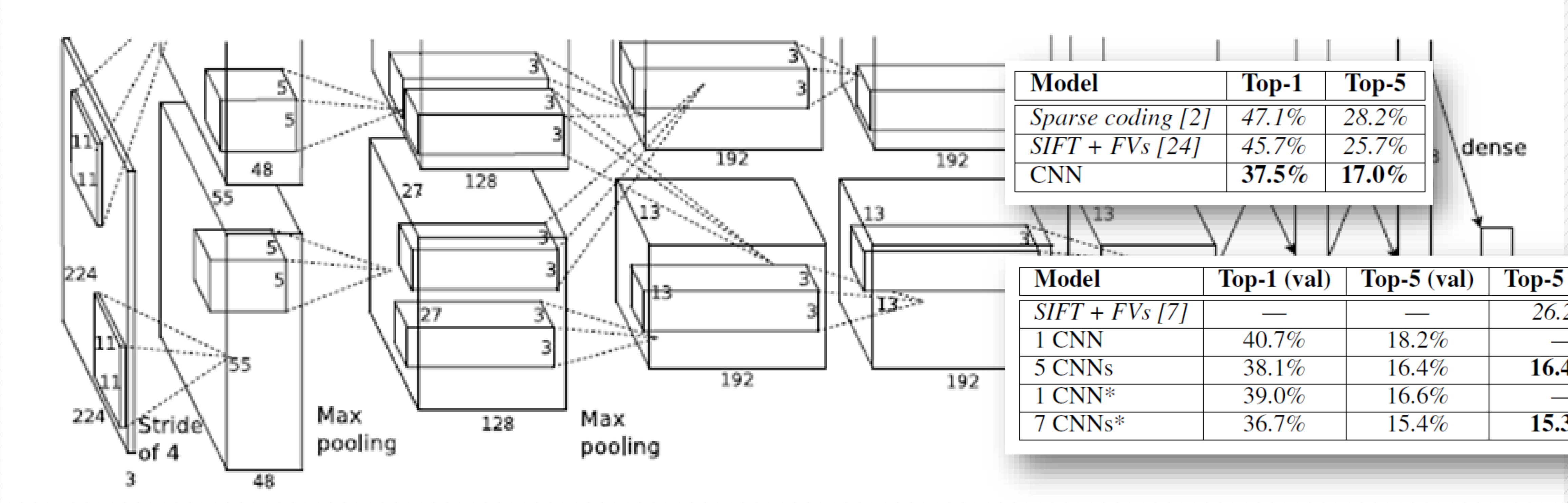
- AlexNet (ImageNet [2]) took almost a **entire week** for traininig , running in two GPUs GTX 580 3GB, **ImagineNet** dataset contains more than 15 millions of high res images, distributed in around 22 thousand categories and labelled and classified on 1000 categories.

- 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.

Deep Learning for Motorbike Classification

AlexNet



Model	Top-1	Top-5
Sparse coding [2]	*47.1%*	*28.2%*
SIFT + FVs [24]	*45.7%*	*25.7%*
CNN	**37.5%**	**17.0%**

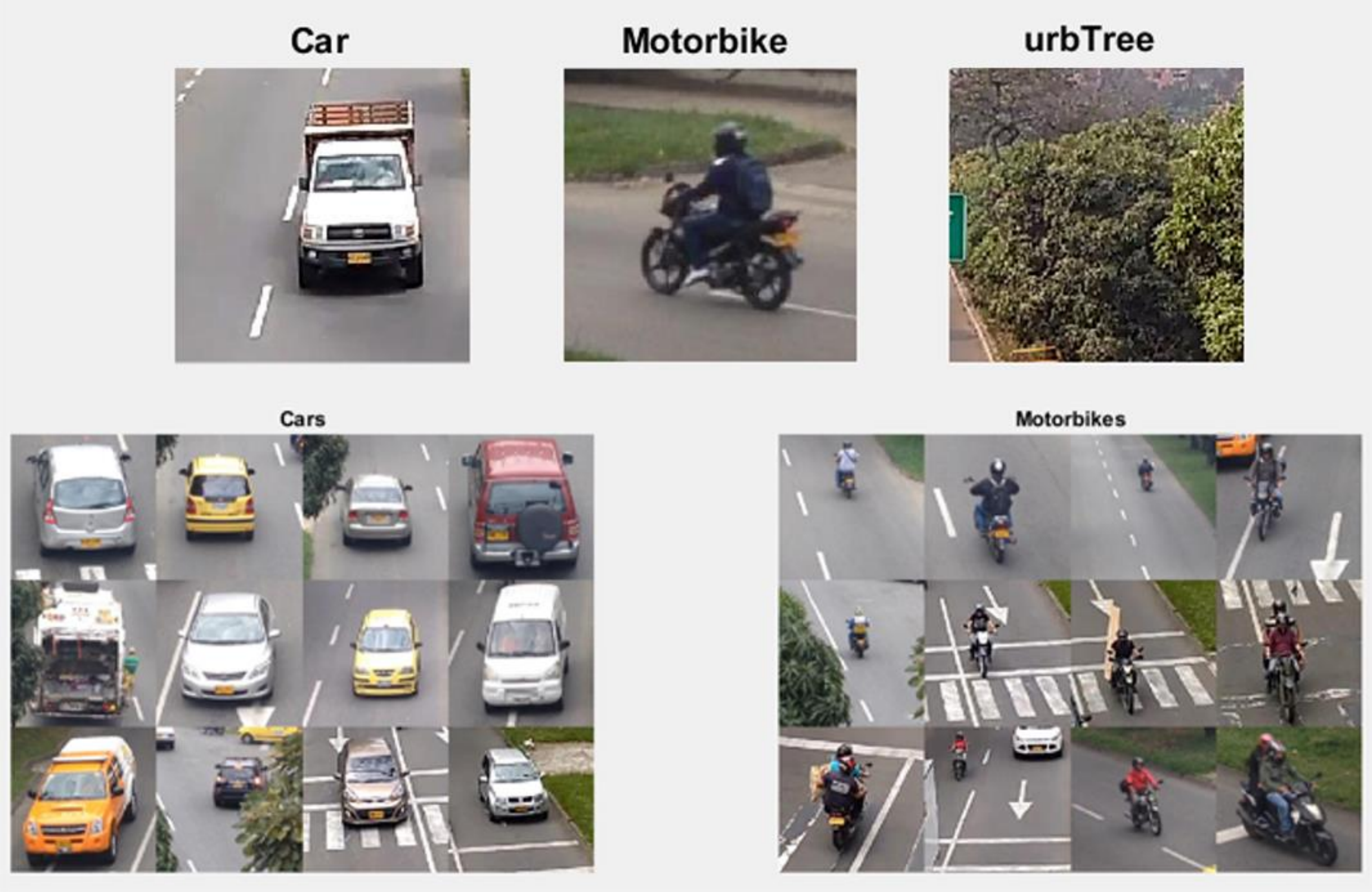
Model	Top-1 (val)	Top-5 (val)	Top-5 (test)
SIFT + FVs [7]	—	—	*26.2%*
1 CNN	40.7%	18.2%	—
5 CNNs	38.1%	16.4%	**16.4%**
1 CNN*	39.0%	16.6%	—
7 CNNs*	36.7%	15.4%	**15.3%**

8 Layers network, first 5 of Convolution last three Fully Connected. The output of the last fully-connected layer is fed to a 1000-way softmax which produces a distribution over the 1000 class labels.

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

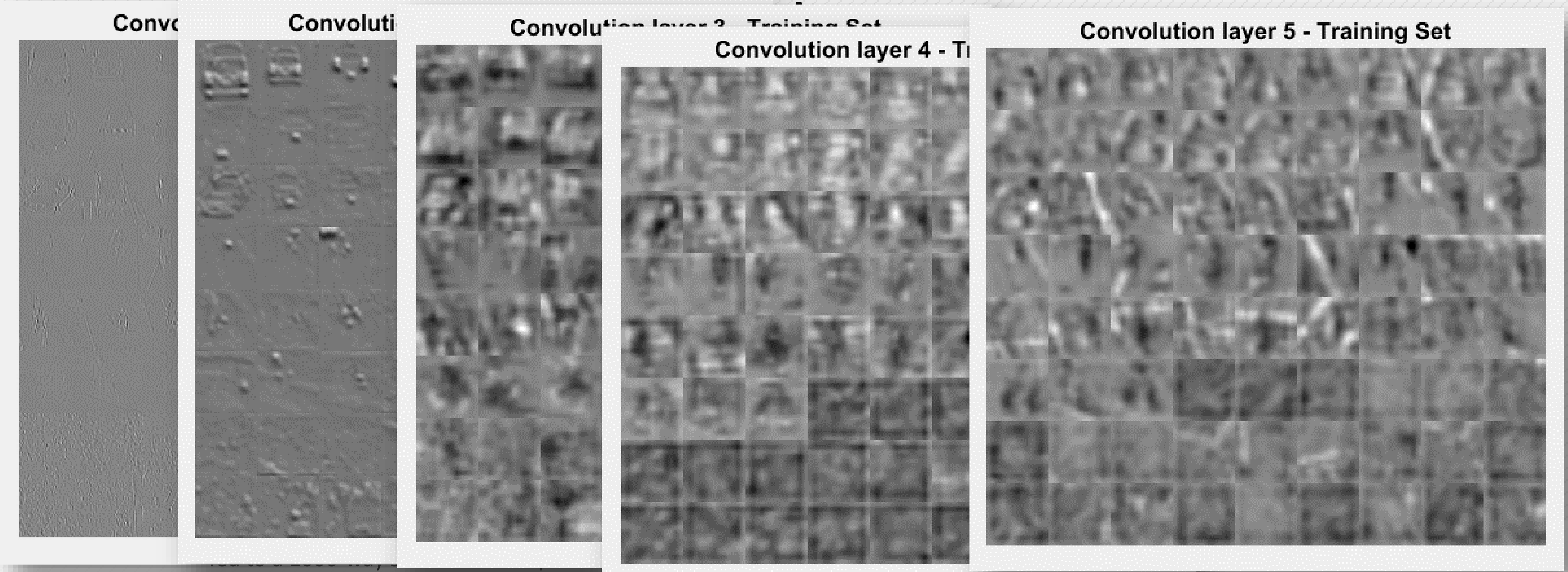
AlexNet - A. Krizhevsky, I. Sutskever, y G. E. Hinton, [2]

Deep Learning for Motorbike Classification



The three categories for Dataset created for classification
Only 80 Examples per Category = 240 Total

Deep Learning for Motorbike Classification

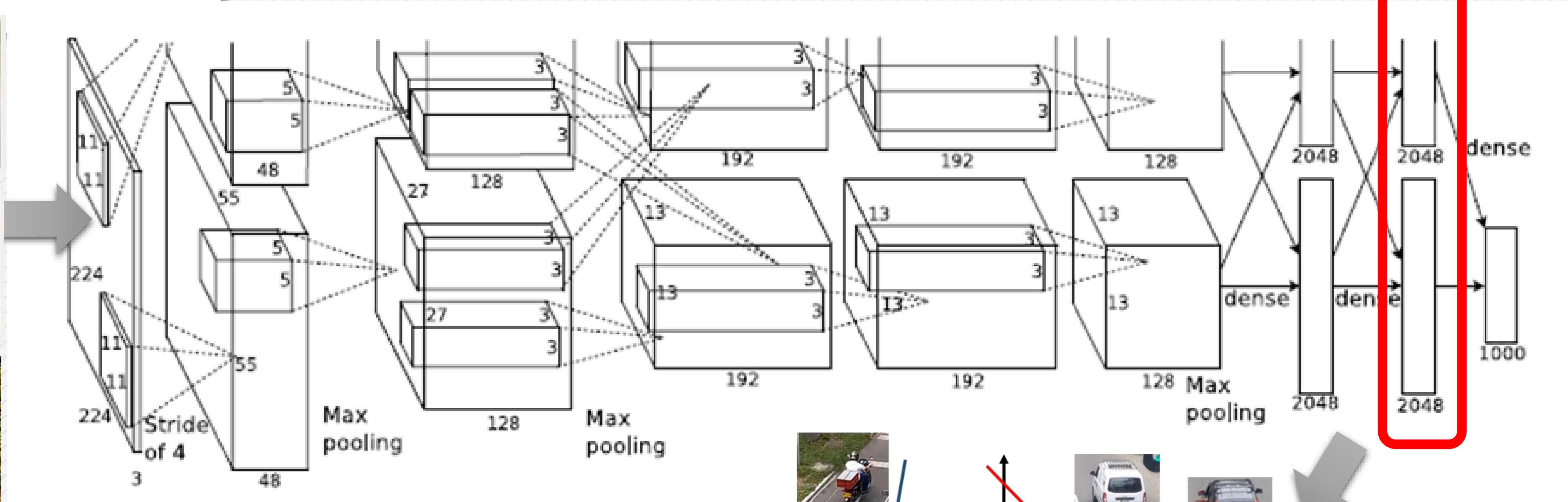


The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers with a final 1000-way softmax

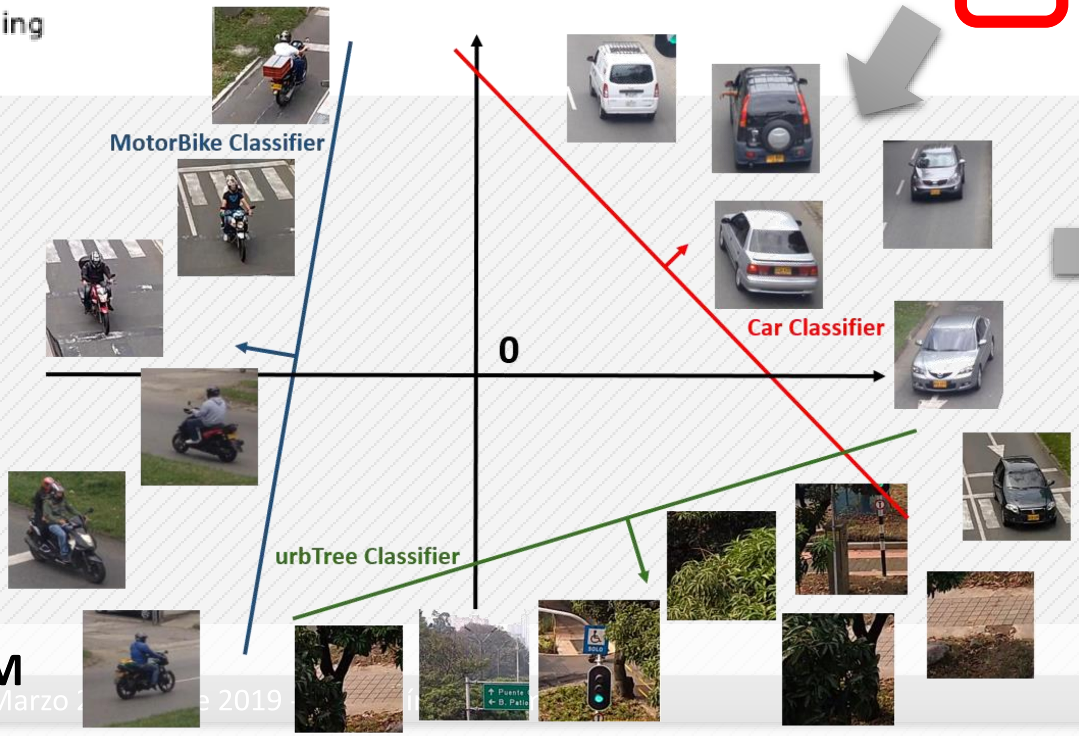
AlexNet - A. Krizhevsky, I. Sutskever, y G. E. Hinton, [2]

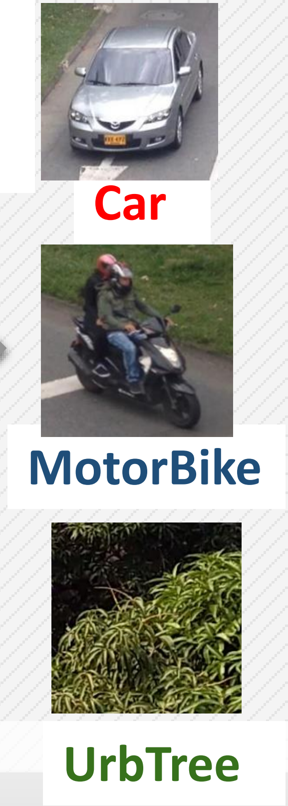
MOYCOT
Movilidad y Control del Tráfico

Our Approach



AlexNet





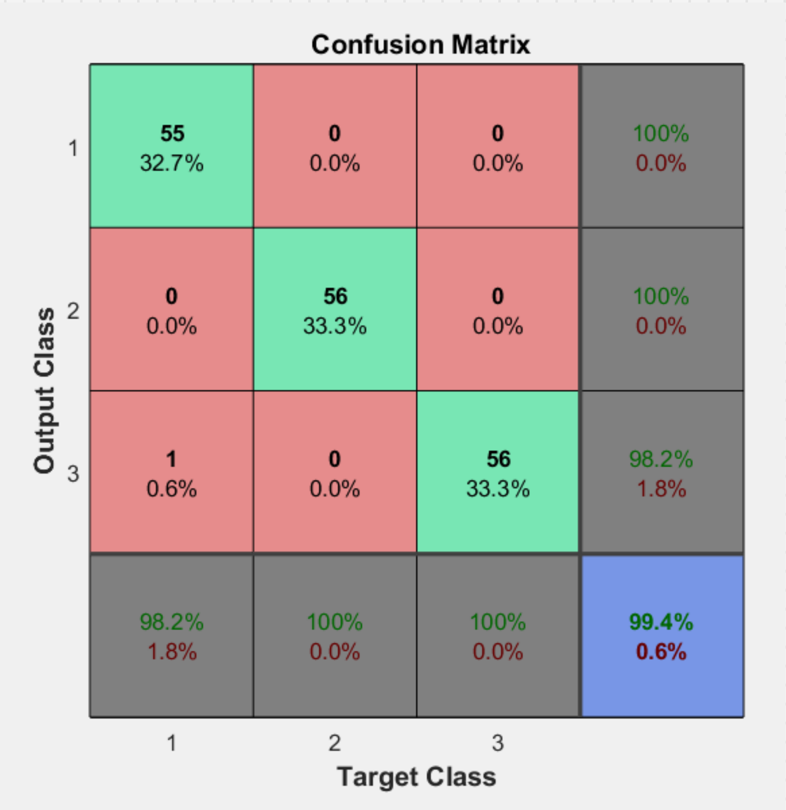
$$\min_{w,b,\xi} \frac{1}{2}||w||^2 + C\sum_{k=1}^{M} \xi k$$

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

Linear SVM

Deep Learning for Motorbike Classification

Results



1. Cars 2. Motorbikes 3. UrbTree

- Mean Accuracy: 99,40%
  (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 %)

Deep Learning for Motorbike Classification

5 Classes





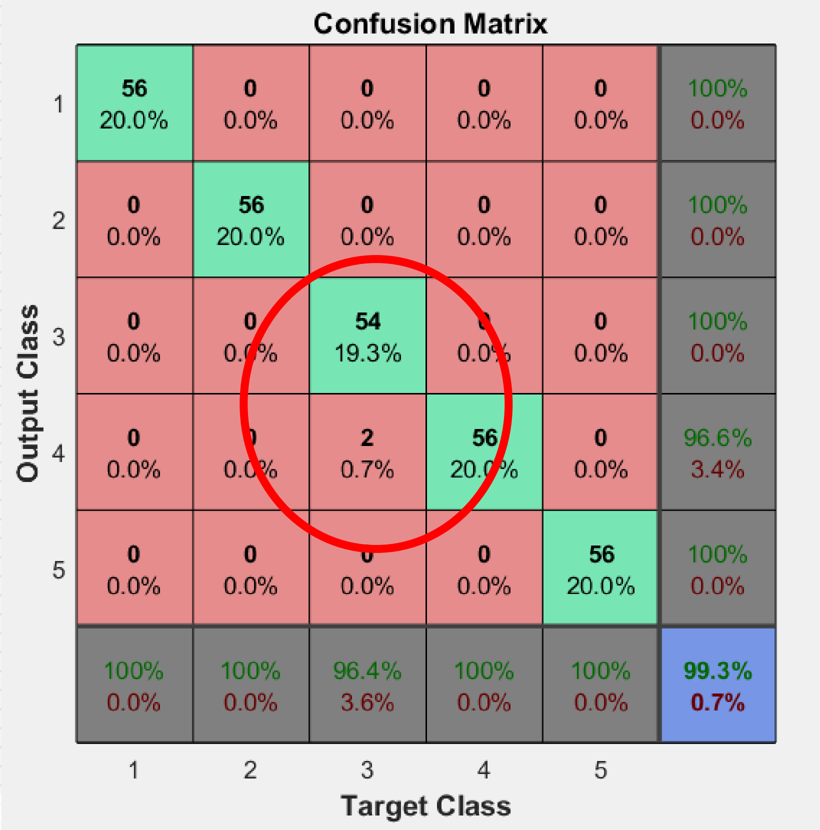
Experiment Set extension. (Class 1=Cars, 2 = Motorbikes, 3 = urbTree, 4=Car side, 5=Motorbikes side (Las two from Caltech-101) [12]

- Try to evaluate if the Features obtained for Motorcycles forces the classifier to treat all motorbikes as single class, and if it I also happen with cars.

MOYCOT
Movilidad y Control del Tráfico

Deep Learning for Motorbike Classification

Results



1. Cars 2. Motorbikes 3. UrbTree 4. Car_side 5. Motos_Side

- Mean Accuracy: 99,85 %

Deep Learning for Vehicle Classification

Results

Source Internet

Cars (Top_Carro)	82 %
Motorbike (Top_Moto)	40 %
UrbTree	58 %

MOYCOT
Movilidad y Control del Tráfico

Deep Learning for Vehicle Classification

Results

Source Internet

Carros (Cars_top)	73, 5%
Motos (Motos_top)	62,3%
UrbTree	22 %

MOYCOT
Movilidad y Control del Tráfico

Deep Learning for Vehicle Classification

Results

Source Google Street View at Medellin

Cars (Top_Carro)	61 %
Motorbike (Top_Moto)	85 %
UrbTree	31 %

MOYCOT
Movilidad y Control del Tráfico

Deep Learning for Vehicle Classification

Results

Source Internet

Cars (Top_Carro)	34 %
Motorbike (Top_Moto)	21 %
UrbTree	32%

Conclusions

- In this research it is proposed the implementation of a motorbike classification scheme in urban scenarios using CNNs for feature extraction.

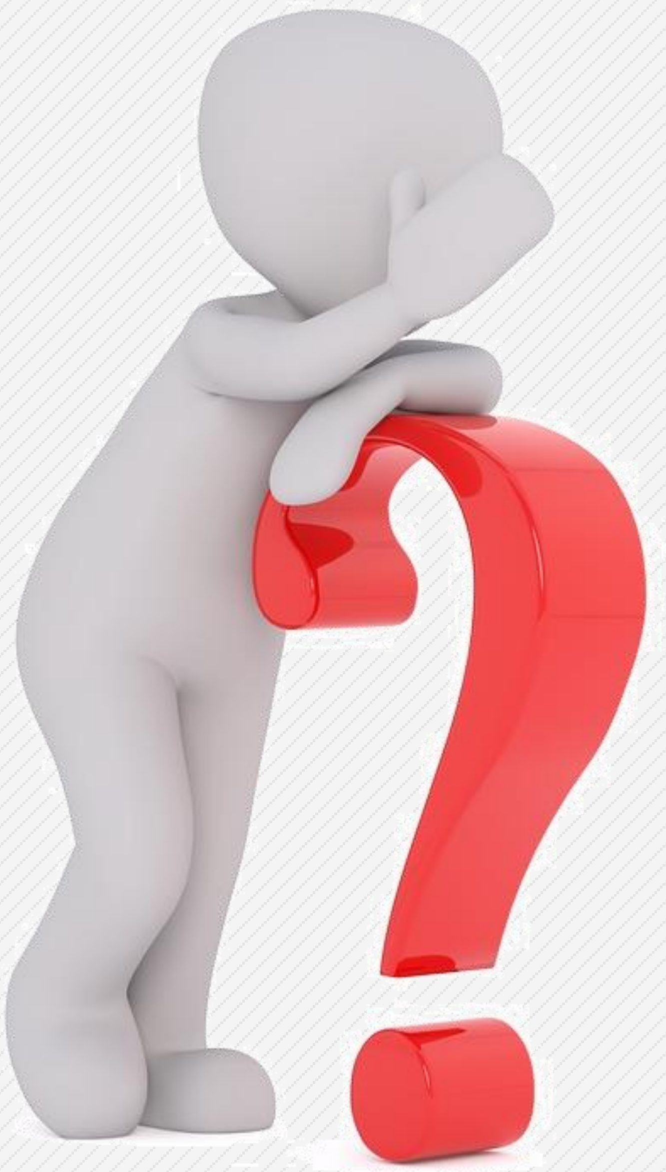
- CNNs already trained with millions of examples and able to classify 1000 categories can be used for feature extraction to train a linear SVM and then classifying for instance three different classes.

- GPUs use on CNN are critical: For instance Benchmark reports [15]: The Pascal Titan X with cuDNN is 49x to 74x faster than dual Xeon E5-2630 v3 CPUs.

- Region of Interest (ROI) can accelerate the speed of CNN analysis; this can be evaluated overall in video detection and classification analysis.

Future Work (this afternoon.. ;)

- This research will move toward the analysis of urban traffic videos, where detection and classification will be applied enriched with a wider set of urban road user classes (e.g. trucks, vans, cyclists, pedestrians.
- How to individualize objects (detection): Evaluation of different background substraction algorithms (Gaussian Mixture Models **GMM – MOG-MOG2** - Zivkovic-Heijden Gaussian mixture model (**ZHGMM**) -Self-Adaptive Gaussian Mixture Model - **SAGMM**
- Evaluate other Deep Learning Strategies for vehicle detection and classification: Move from Classification localization to Object Detection
- Implement tracking by using Deep Learning Strategies.



© Pixabay Question, Question Mark, Help, Response

Referencias

[1] “Caffe | Deep Learning Framework”. [En línea]. Disponible en: http://caffe.berkeleyvision.org/. [Accedido: 05-sep-2016].

[2] A. Krizhevsky, I. Sutskever, y G. E. Hinton, “Imagenet classification with deep convolutional neural networks”, en Advances in neural information processing systems, 2012, pp. 1097–1105.

[3] I. Arel, D. C. Rose, y T. P. Karnowski, “Deep Machine Learning - A New Frontier in Artificial Intelligence Research [Research Frontier]”, IEEE Comput. Intell. Mag., vol. 5, no 4, pp. 13–18, nov. 2010.

[4] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, y L. Fei-Fei, “ImageNet: A large-scale hierarchical image database”, en IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009, 2009, pp. 248–255.

[5] M. D. Zeiler y R. Fergus, “Visualizing and understanding convolutional networks”, en European Conference on Computer Vision, 2014, pp. 818–833.

[6] K. J. Lang, A. H. Waibel, y G. E. Hinton, “A time-delay neural network architecture for isolated word recognition”, Neural Netw., vol. 3, no 1, pp. 23–43, 1990.

[7] S. Hochreiter, “The vanishing gradient problem during learning recurrent neural nets and problem solutions”, Int. J. Uncertain. Fuzziness Knowl.-Based Syst., vol. 6, no 2, pp. 107–116, 1998.

Referencias

[8] H. Lee, R. Grosse, R. Ranganath, y A. Y. Ng, “Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations”, en Proceedings of the 26th annual international conference on machine learning, 2009, pp. 609–616.

[9] “ImageNet Large Scale Visual Recognition Competition (ILSVRC)”. [En línea]. Disponible en: http://www.image-net.org/challenges/LSVRC/. [Accedido: 24-oct-2016].

[10] A. S. Razavian, H. Azizpour, J. Sullivan, y S. Carlsson, “CNN Features Off-the-Shelf: An Astounding Baseline for Recognition”, en 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2014, pp. 512–519.

[11] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, y Y. LeCun, “Overfeat: Integrated recognition, localization and detection using convolutional networks”, ArXiv Prepr. ArXiv13126229, 2013.

[12] L. Fei-Fei, R. Fergus, y P. Perona, “Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories”, Comput. Vis. Image Underst., vol. 106, no 1, pp. 59–70, 2007.

[13] A. Vedaldi y K. Lenc, “Matconvnet: Convolutional neural networks for matlab”, en Proceedings of the 23rd ACM international conference on Multimedia, 2015, pp. 689–692.

[14] J. Donahue et al., “DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition.”, en ICML, 2014, pp. 647–655.

[15] J. Johnson, *cnn-benchmarks: Benchmarks for popular CNN models*. 2017.