Understanding Safety Based on Urban Perception

Felipe A. Moreno

About me

M.Sc. (c) Felipe A. Moreno
www.fmorenovr.com

Canadá

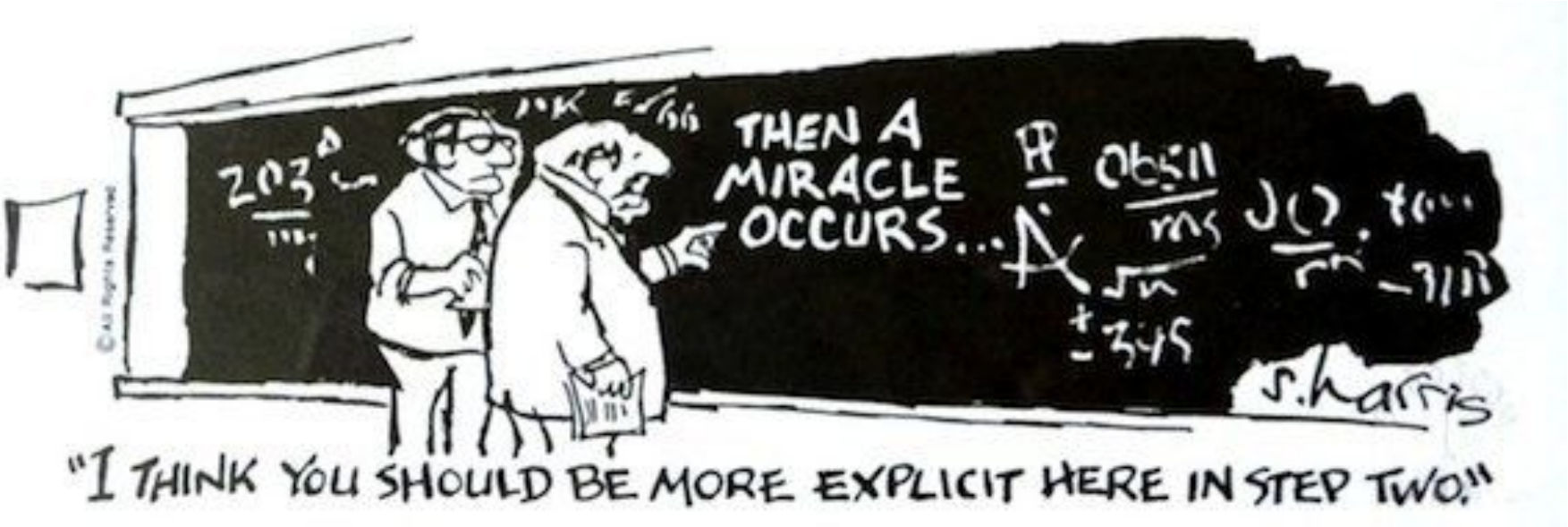
Perú

Universidad Católica San Pablo

Brasil

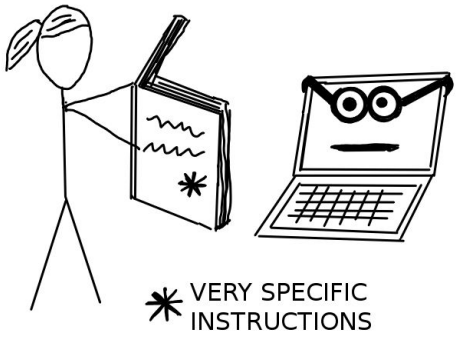
FGV FUNDAÇÃO GETULIO VARGAS

Interpretability Machine Learning

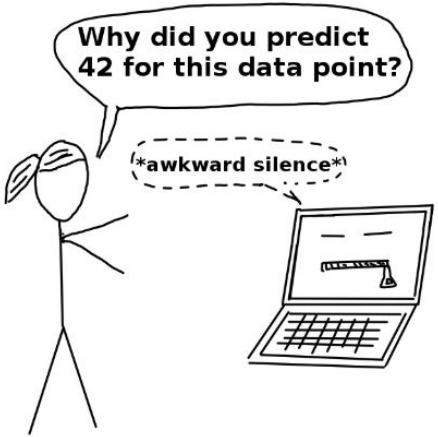


"I THINK YOU SHOULD BE MORE EXPLICIT HERE IN STEP TWO."

Without Machine Learning



With Machine Learning



Objetivo:

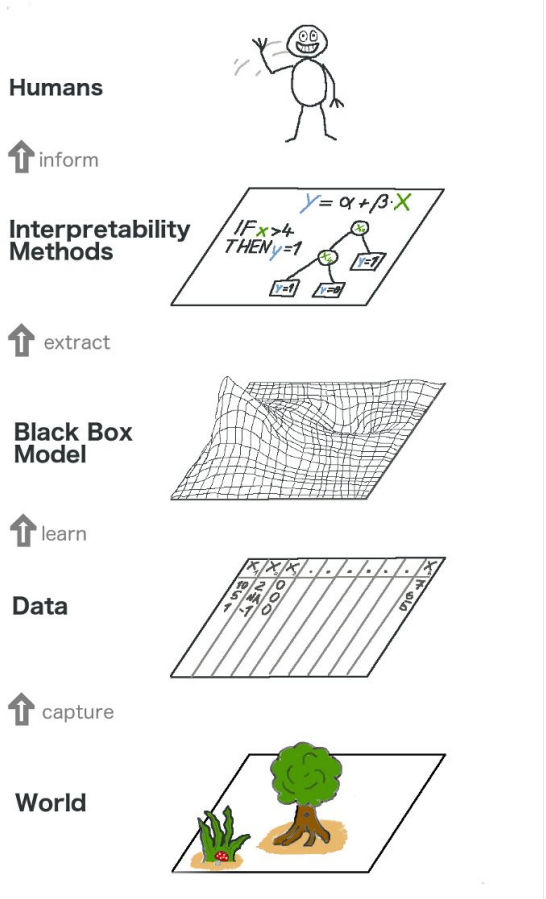
- ¿Por qué se hizo una predicción particular, a diferencia de otras?
- ¿Cuándo tiene éxito el modelo? ¿Cuándo confiar en un modelo?
- ¿Cuándo falla y por qué? ¿Cuándo ignorar la salida del modelo?

Un explainer system puede proporcionar explicaciones principalmente en dos formas:

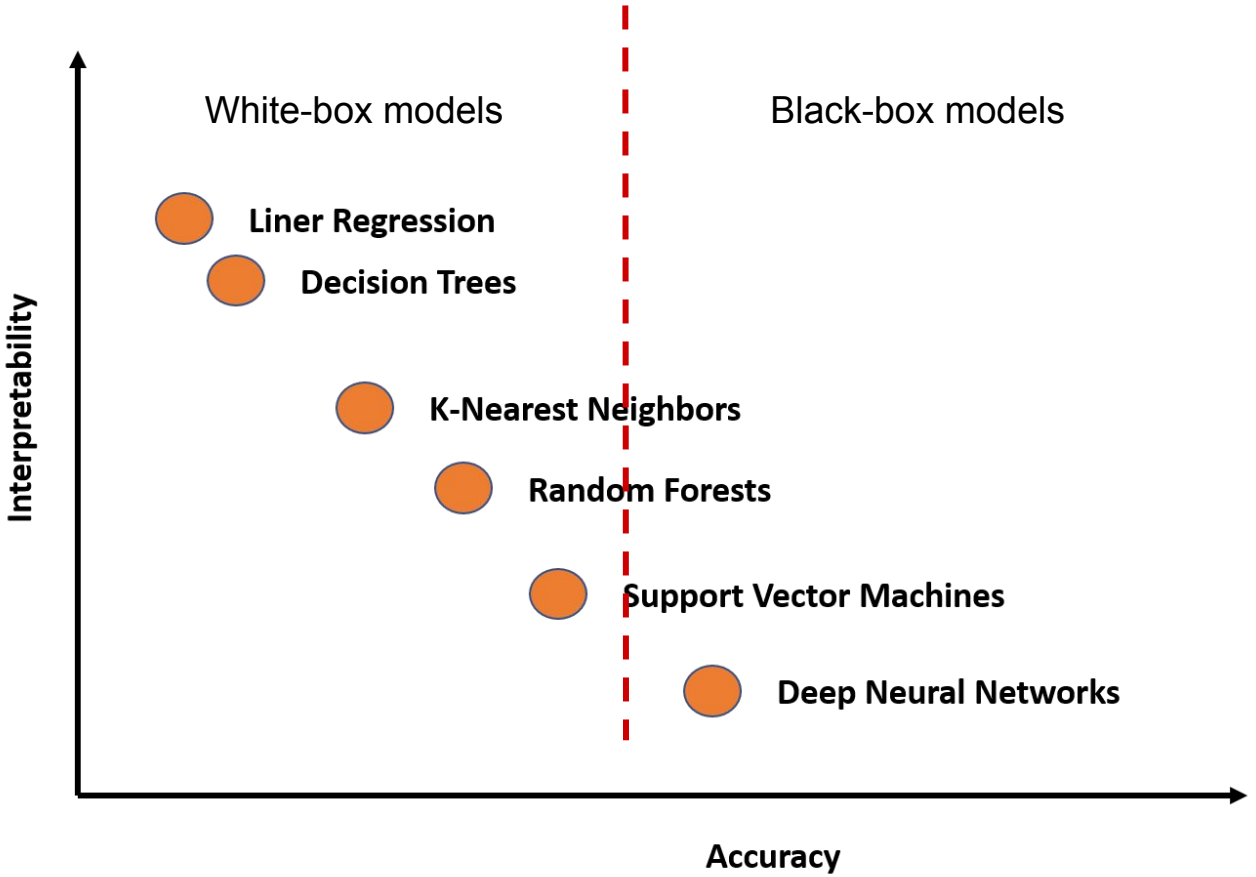
- Características relevantes que afectan las predicciones.
- Conjunto mínimo de instancias de entrenamiento relevantes críticas para las predicciones.

Además, las explicaciones pueden ser en

- Global level: proporciona una vista de alto nivel del modelo.
- Local level: proporciona una justificación para una sola predicción



Interpretability vs Accuracy



Enfoques:

- **White-Box Explanations**: Explicaciones específicas del modelo, aprovechando la estructura interna y las ideas del modelo.

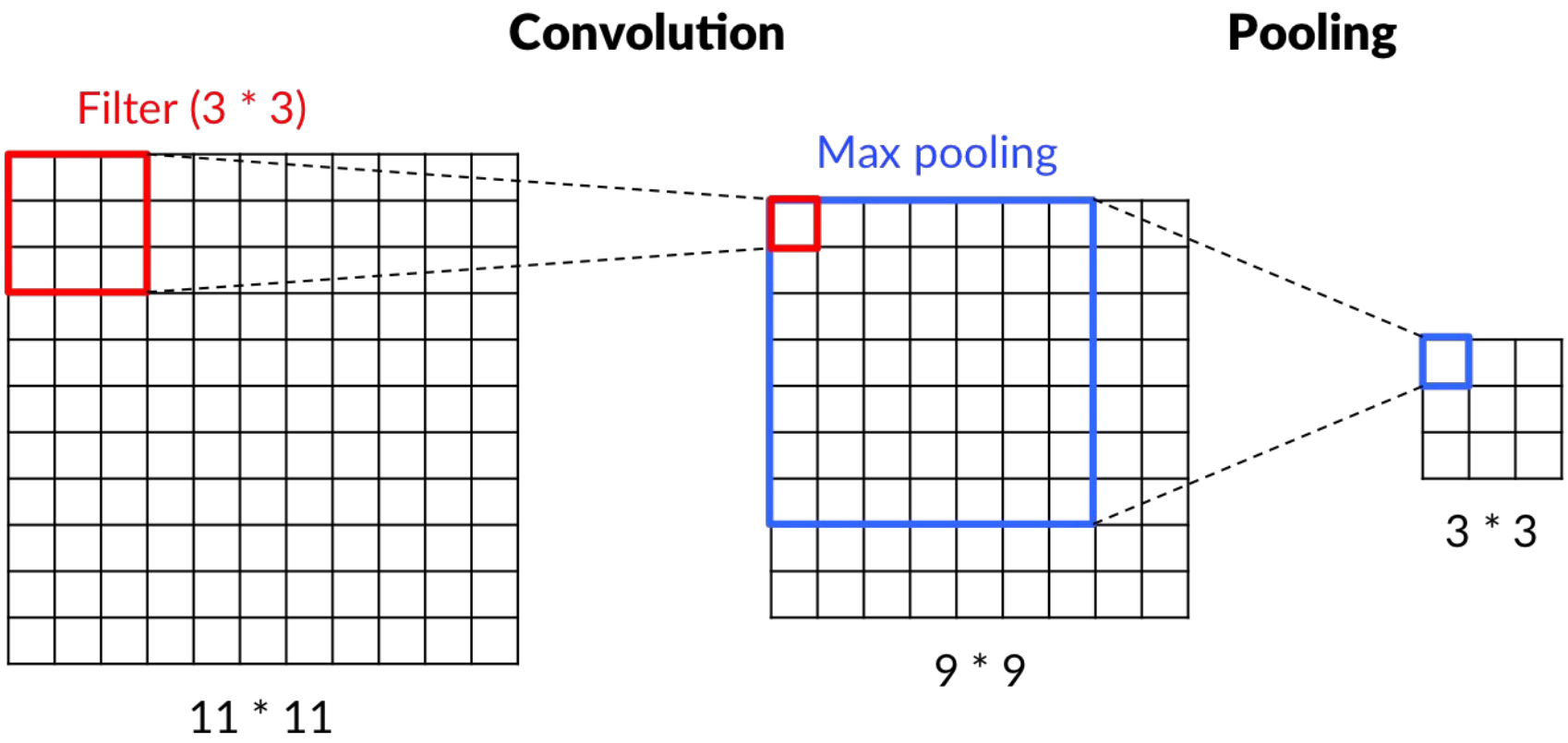
- **Black-Box Explanations**: Explicaciones agnósticas del modelo. Dichas explicaciones son relevantes para todos los modelos existentes, así como para los modelos futuros. - LIME, ANCHOR, SHARP, CAM

- **Out-of-box Interpretable models**: Desarrollar nuevos modelos que sean inherentemente interpretables.

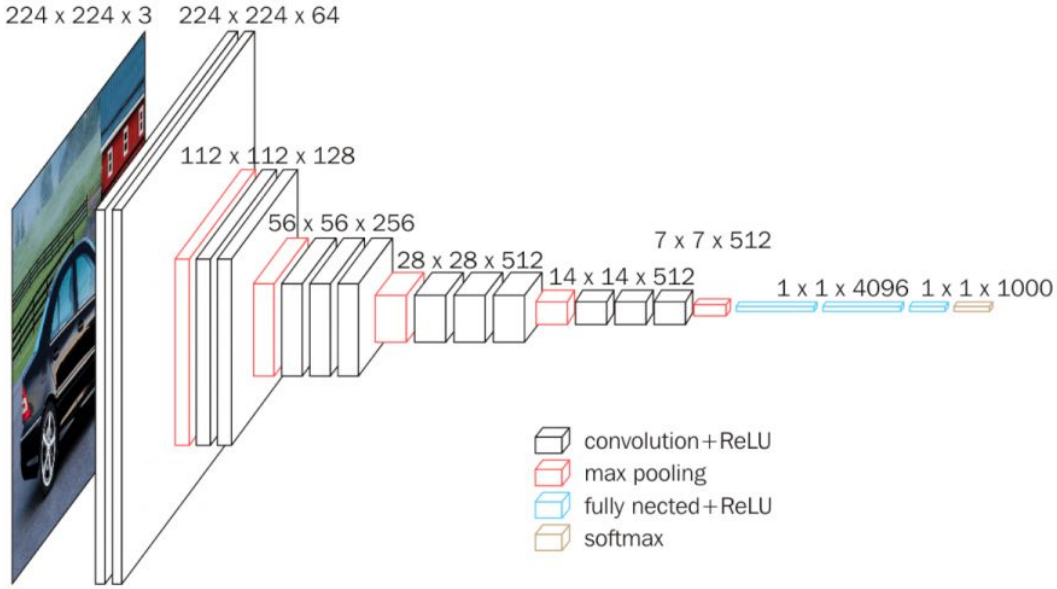
Deep Convolutional Networks

Partes de una CNN:
Una capa simple de una CNN incluye 3 tipos de operaciones:

- **Convolution**: Esta es la pata más importante de una CNN. La operación convolución usa solo sumas y multiplicaciones. Los filtros convolucionales escanean la imagen, realizando esta operación.
- **Nonlinearity**: Esta es una ecuación aplicada a la salida de un filtro convolucional. Nonlinearities permite a una CNN a aprender relaciones complicadas (curvas en vez de líneas) entre la entrada (imágenes de entrada y la clase resultante).
- **Pooling**: Este es el conocido “max pooling” el cual solo escoge el mayor número dentro de una determinada región (determinada por el tamaño del kernel). Pooling reduce el tamaño de la representación, de este modo reduciendo la cantidad de operaciones requeridas para la CNN.

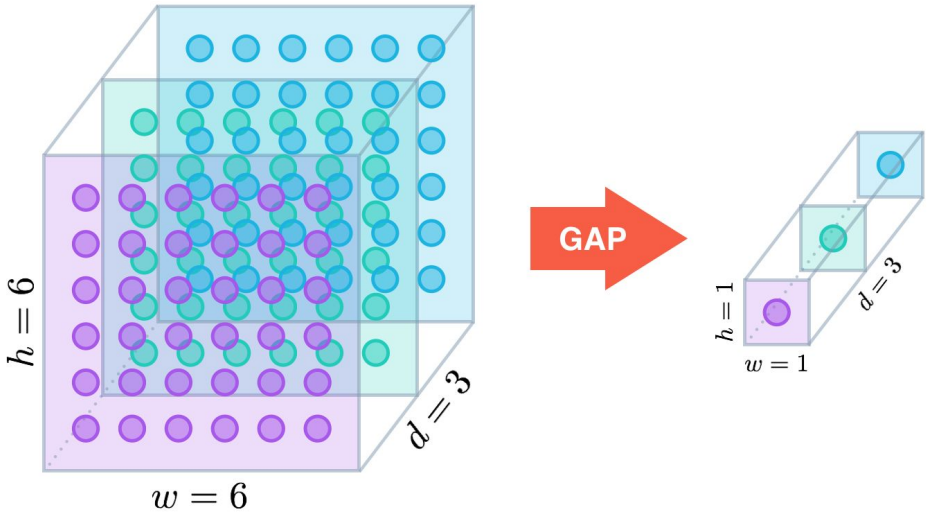

Convolution
Pooling
Filter (3 * 3)
Max pooling
11 * 11
9 * 9
3 * 3

VGG model



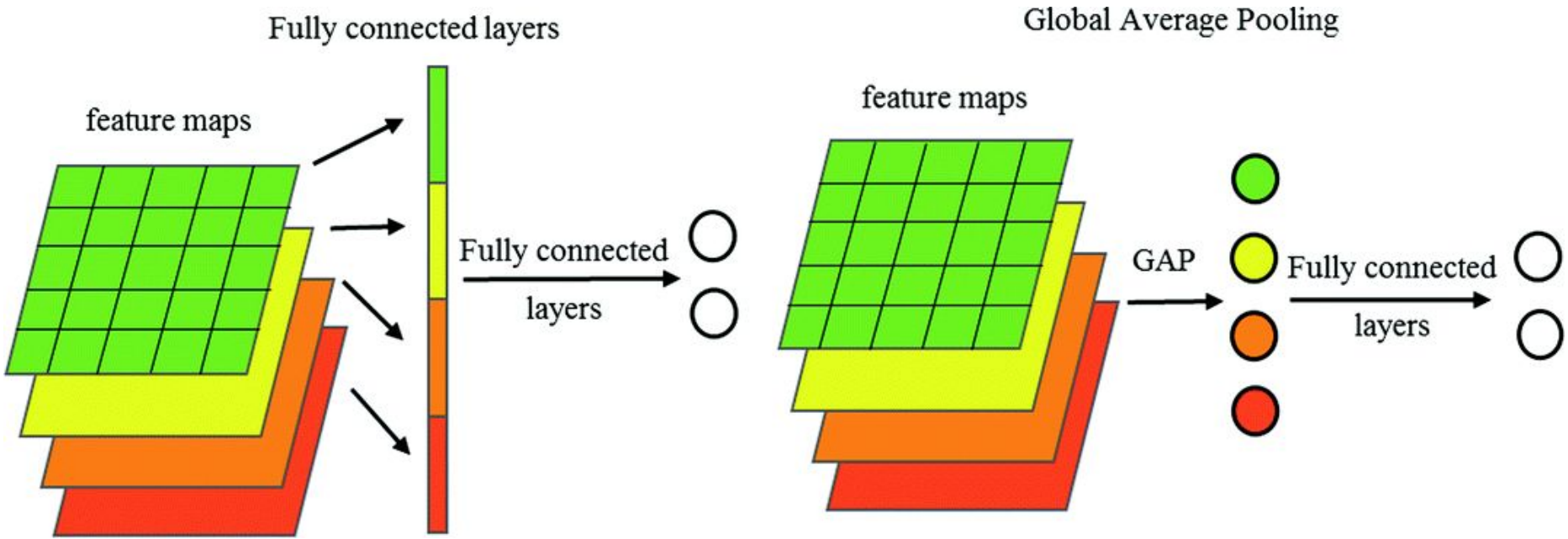
- 5 bloques de 2 convoluciones y un max pooling.
- 3 capas Fully-Connected. Notar que la mayoría de los parámetros están en estas 3 últimas capas.
- Debido a esto, es probable que tu modelo tenga un overfitting (pero para evitar eso, se usa dropout).

Global Average Pooling

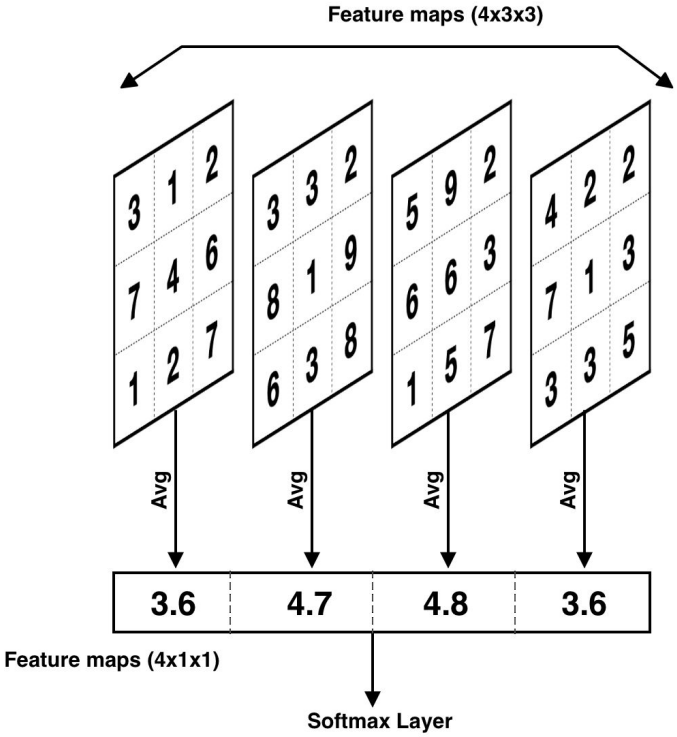


- Para evitar un overfitting en los modelos, se plantea un método para reducir la cantidad de parámetros.
- Similar a MaxPooling, GAP reduce la dimensionalidad de un tensor hxwhd a 1x1xd.
- En otras palabras, GAP promedia cada uno de los mapas de características 'n' de la última capa convolucional, produciendo un vector de tamaño n.

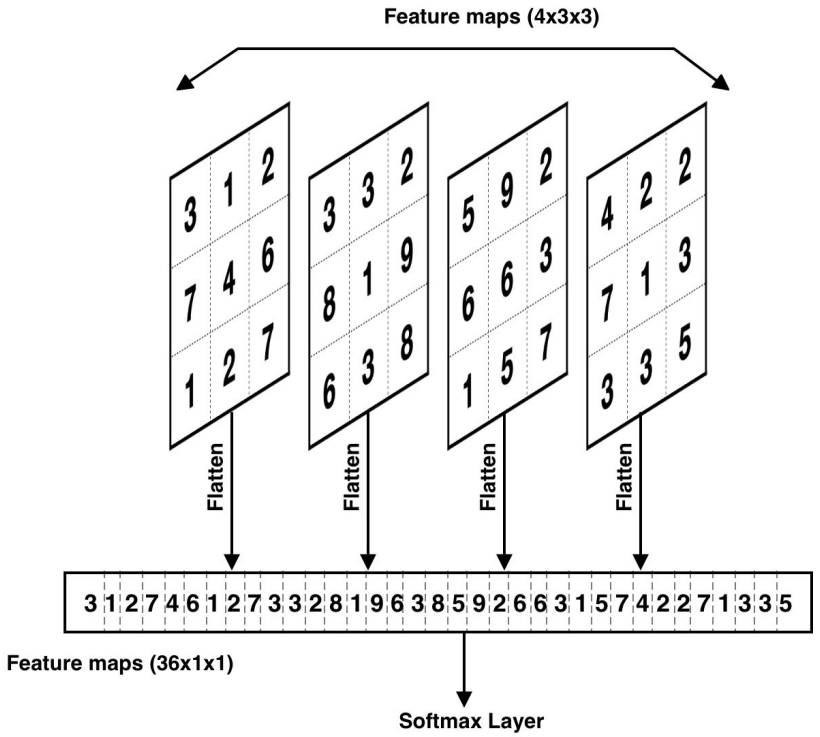
GAP vs FC


Fully connected layers
feature maps
Fully connected
layers
Global Average Pooling
feature maps
GAP
Fully connected
layers

GAP vs FC

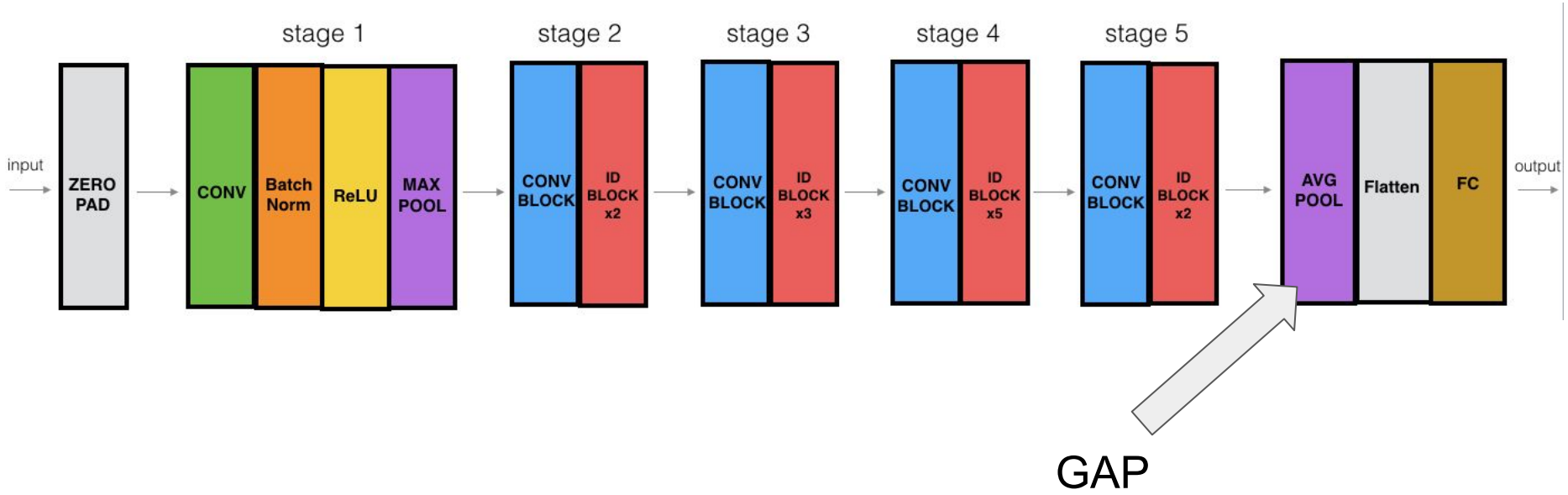


GLobal Avarage Pooling

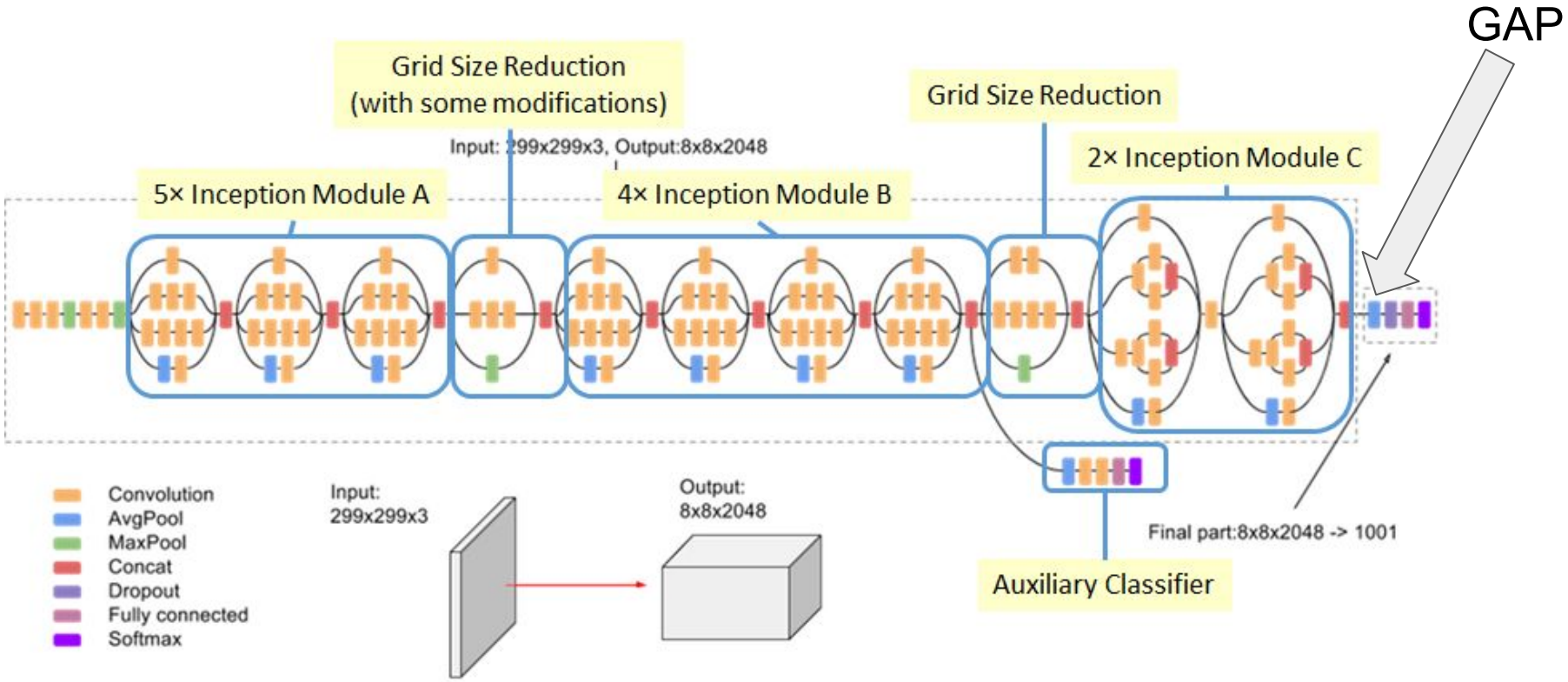


Fully Connected Layer

ResNet model

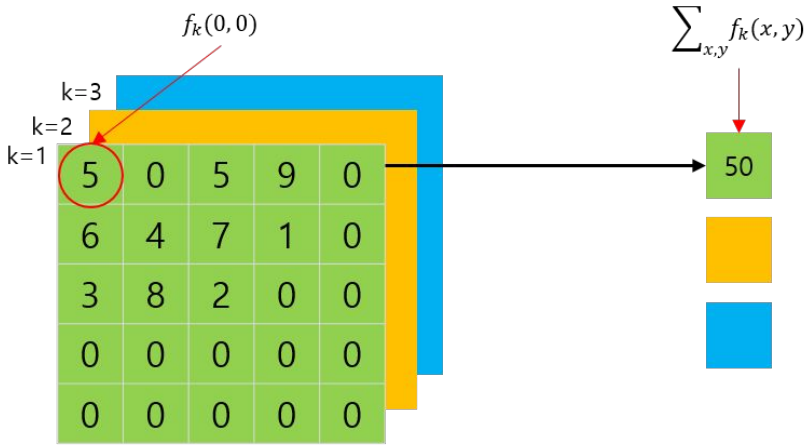


InceptionV3 (GoogleNet) model



AP vs MP

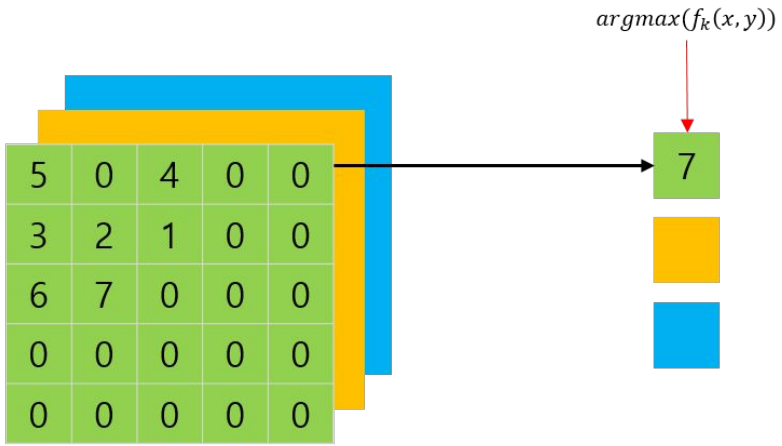
Global Average Pooling



Last convolution layer output

GAP output

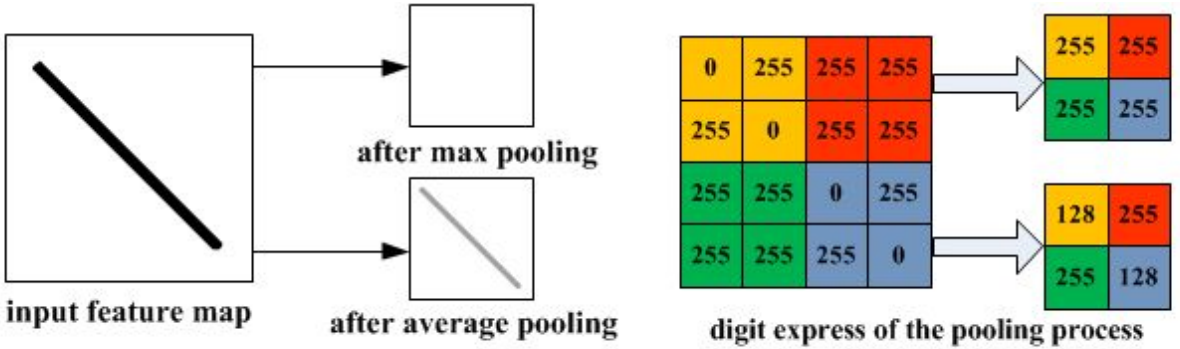
Global Max Pooling



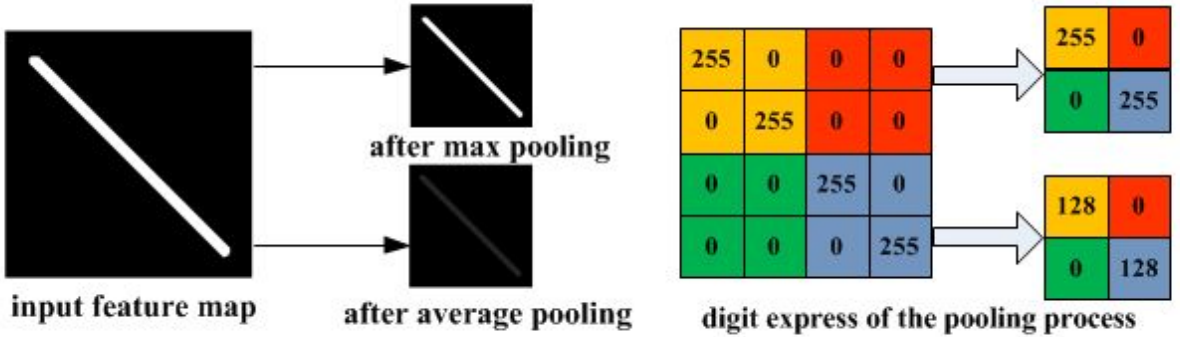
Last convolution layer output

GMP output

AP vs MP



(a) Illustration of max pooling drawback



(b) Illustration of average pooling drawback

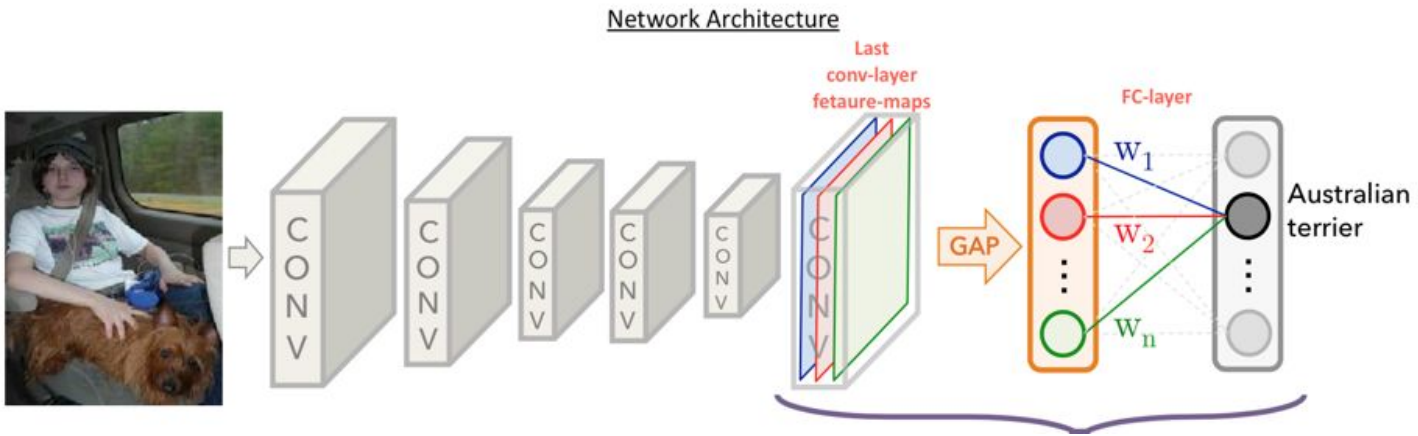
Class Activation Maps

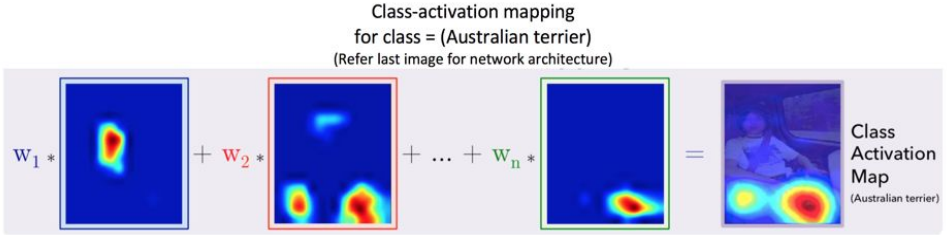
Class Activation Maps

Las capas Fully-Connected son similares como black-box models entre las capas convolucionales y el clasificador, lo que lleva a la pérdida de la información espacial de la imagen.

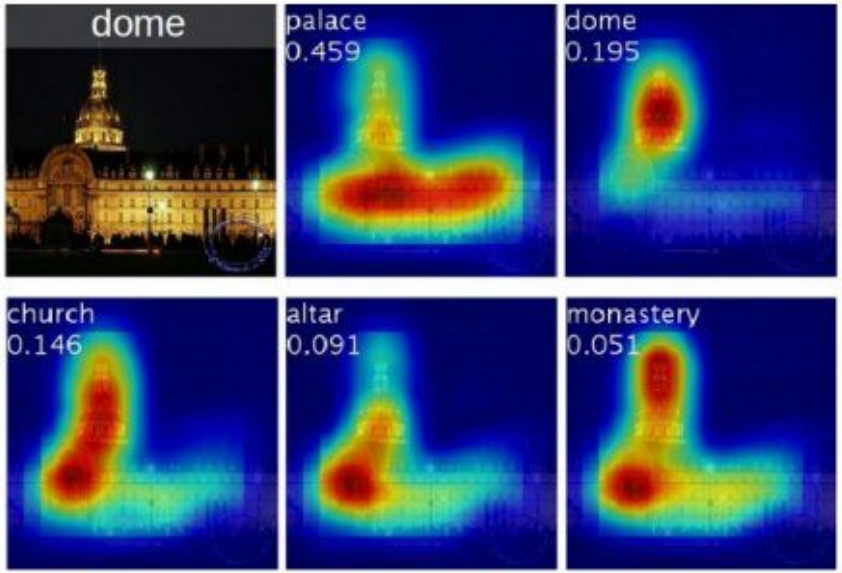
En este enfoque se reemplaza las capas FC por un "Global Average Pooling (GAP)". Este vector de salida de capa GAP está conectado además a una capa completamente conectada para producir la salida deseada (puntajes en caso de clasificación) como se muestra a continuación:

Class Activation Maps (CAM)





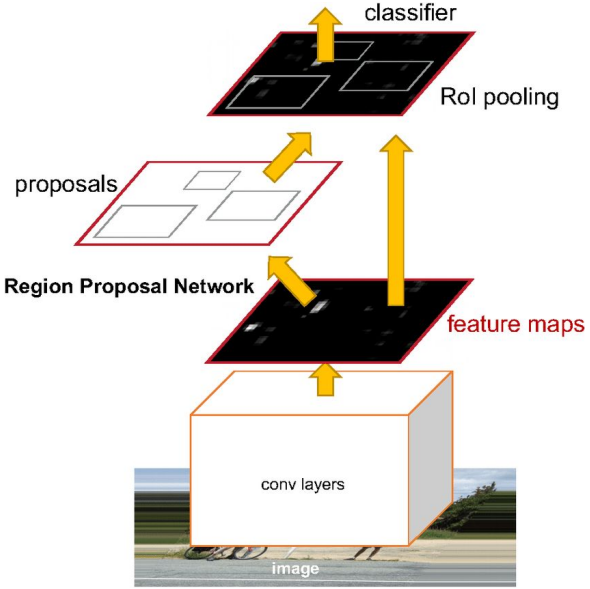
CAM - classification



Class activation maps of top 5 predictions

Class activation maps for one object class

CAM - object detection



Faster R-CNN

Faster R-CNN + CAM

Urban Perception

Which looks more safety?

Which place looks **livelier** ?

For this question: **362,708** clicks collected

Goal: **500,000** clicks

SEE REAL-TIME RANKINGS

RANK	CITY	CLICKS	TREND	RANK	CITY	CLICKS	TREND
1	Washington DC	6296		54	Cape Town	16228	
2	London	17982		55	Belo Horizonte	12728	
3	New York	22424		56	Gaborone	4717	

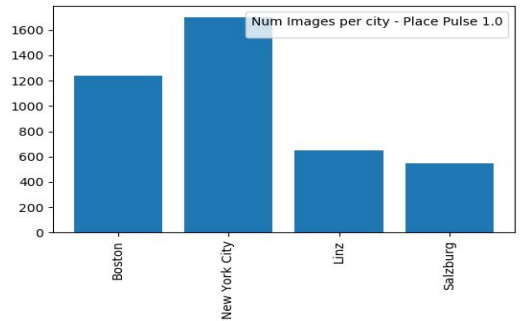
Dataset

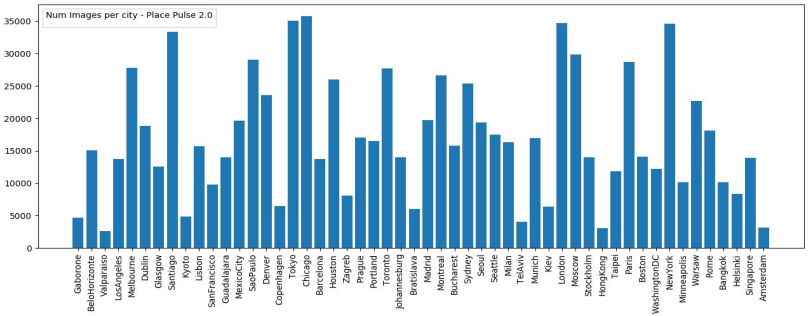
Place Pulse 1.0

- Contiene un total de 73,806 comparaciones de 4,109 imágenes de 4 ciudades (New York City (incl. Manhattan y partes de Queens, Brooklyn & The Bronx), Boston (incl. partes de Cambridge), Linz y Salzburg) de dos países (US y Austria)
- Tres tipos de comparaciones: Safe, Wealth, Unique.

Place Pulse 2.0

- Contiene un total de 1.17 millones de comparaciones de 110,988 imagenes de 56 ciudades de 28 países entre los 5 continentes.
- Seis tipos de comparaciones: Safe, Wealth, Depress, Beautiful, Boring, Lively.





Dataset

Place Pulse 1.0				
Ciudades	# de imágenes	*safe mean*	*wealth mean*	*unique mean*
Linz	650	4.85	5.01	4.83
Boston	1237	4.93	4.97	4.76
New York	1705	4.47	4.31	4.46
Salzburg	544	4.75	4.89	5.04
Total	4136			

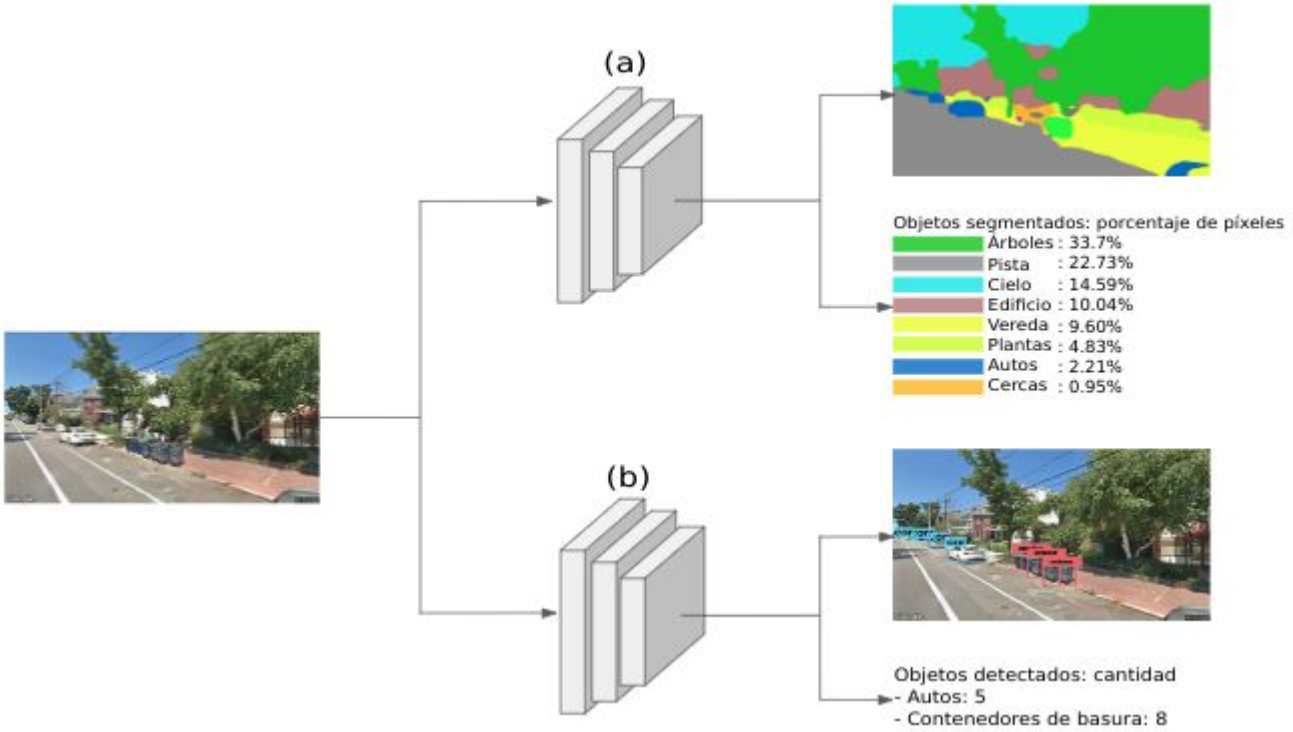
Place Pulse 2.0		
Continente	# de ciudades	# imágenes
America	22	50,028
Europa	22	38,747
Asia	7	11,417
Oceania	2	6,097
Africa	3	5,101
Total	56	111,390

(a)

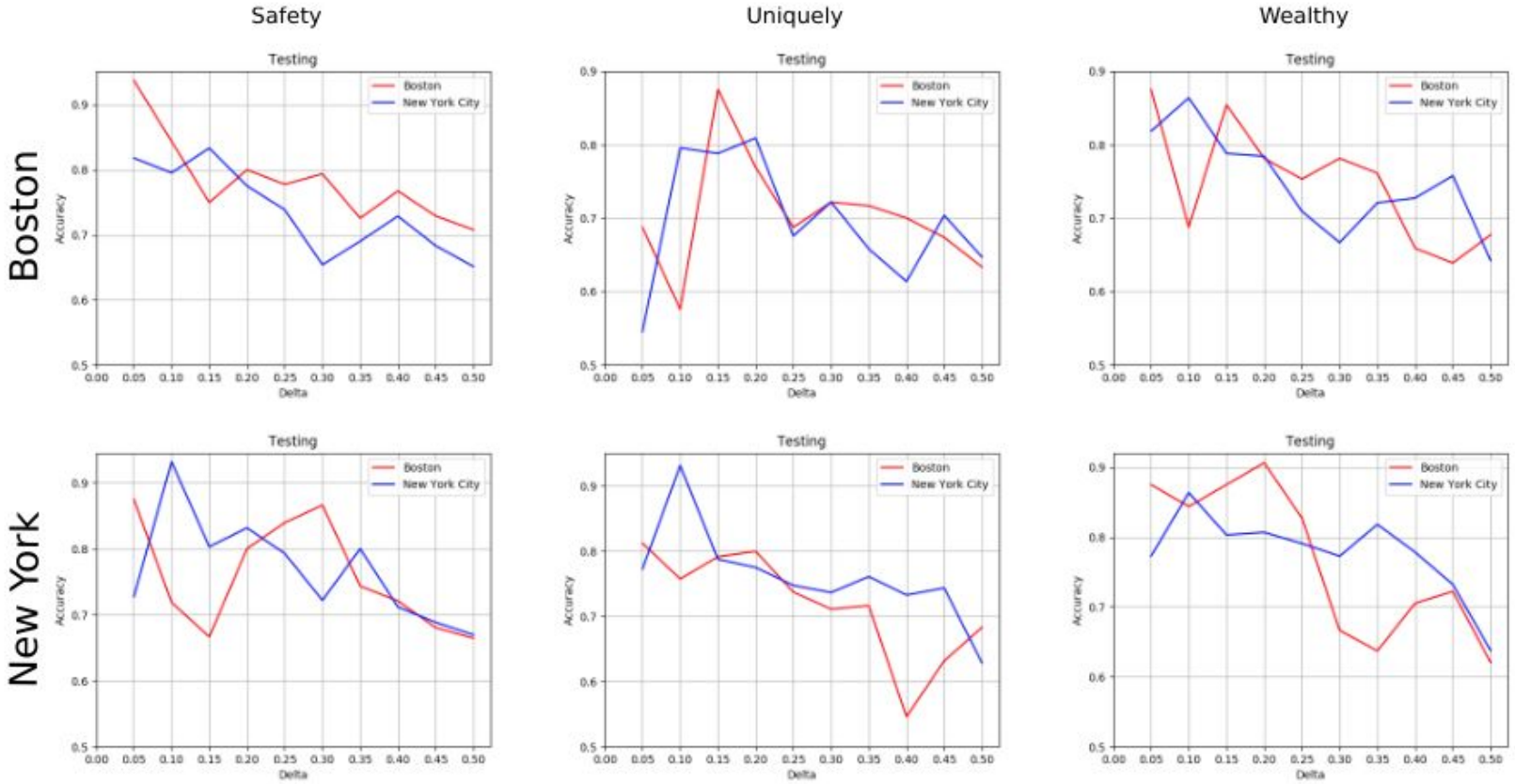
Place Pulse 2.0		
Categoría	# de imágenes	*mean*
Safety	368,926	5.188
Lively	267,292	5.085
Beautiful	175,361	4.920
Wealthy	152,241	4.890
Depressing	132,467	4.816
Boring	127,362	4.810
Total	1,223,649	

(b)

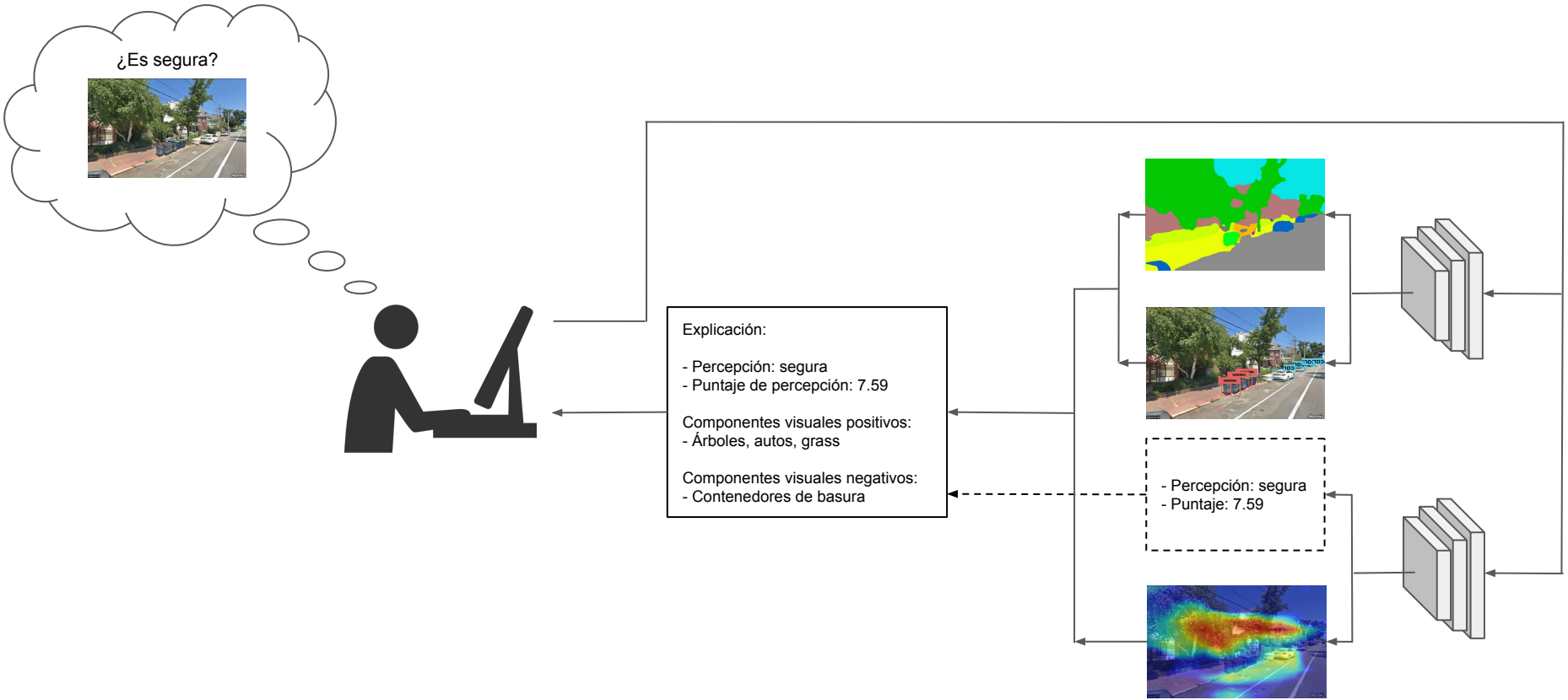
Identifying Visual Components



Training

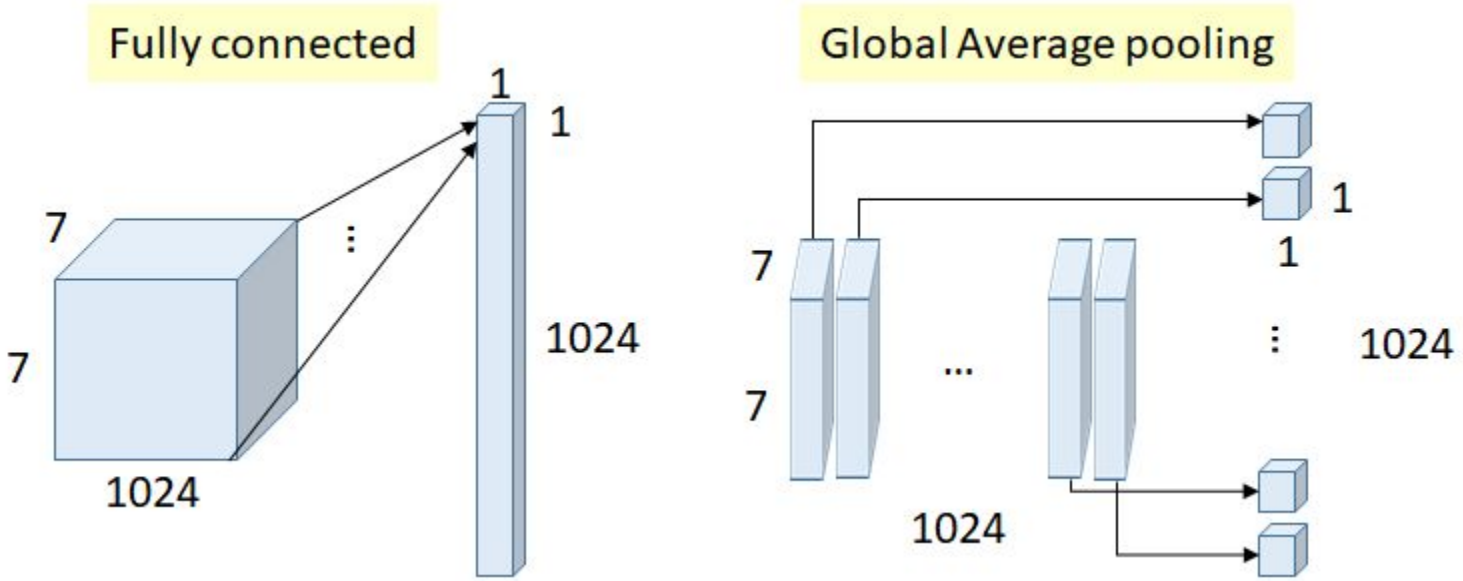

Safety
Uniquely
Wealthy
Boston
New York
Testing
Boston
New York City
Accuracy
Delta

Understanding Results



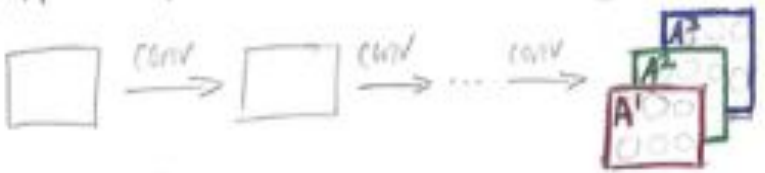
Thanks

Global Average Pooling



Class Activation Maps

$CAM^{cat} = w_1 A^1 + w_2 A^2 + w_3 A^3$, i.e. $\sum_k w_k^{cat} A^k$



The score for class cat, y^{cat}, is thus calculated as

$$y^{cat} = \sum_{k=1}^{K} w_k^{cat} \frac{1}{Z} \sum_{i=1}^{u} \sum_{j=1}^{v} A_{ij}^k$$

$\frac{1}{Z} \sum_{i=1}^{u} \sum_{j=1}^{v} A_{ij}^k$: the result of GAP for A^k

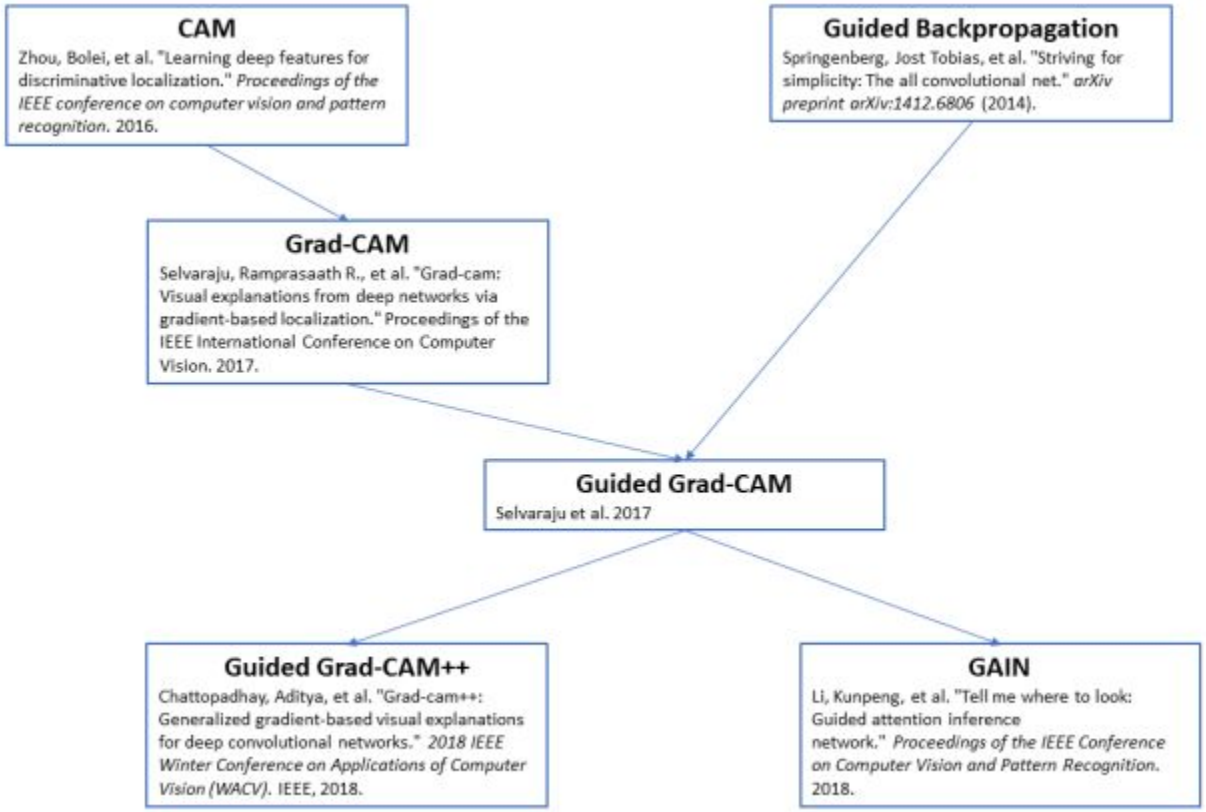
penultimate conv layer produces $k=3$ feature maps $A^k \in \mathbb{R}^{u \times v}$, for example:

GAP of A^k means taking the average of all the elements of A^k. That is,

$$\frac{1}{Z} \sum_{i=1}^{u} \sum_{j=1}^{v} A_{ij}^k$$

and $Z = uv$ (i.e. the total number of elements in the feature map)

if $A^1 = \begin{bmatrix} -2 & 6 \\ 3 & -1 \end{bmatrix}$ then GAP yields $\frac{-2+6+3-1}{4} = 1.5$


CAM
Zhou, Bolei, et al. "Learning deep features for discriminative localization." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
Guided Backpropagation
Springenberg, Jost Tobias, et al. "Striving for simplicity: The all convolutional net." arXiv preprint arXiv:1412.6806 (2014).
Grad-CAM
Selvaraju, Ramprasaath R., et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization." Proceedings of the IEEE International Conference on Computer Vision. 2017.
Guided Grad-CAM
Selvaraju et al. 2017
Guided Grad-CAM++
Chattopadhay, Aditya, et al. "Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks." 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2018.
GAIN
Li, Kunpeng, et al. "Tell me where to look: Guided attention inference network." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

Year	CNN	Developed by	Place	Top-5 error rate	No. of parameters
1998	LeNet(8)	Yann LeCun et al			60 thousand
2012	AlexNet(7)	Alex Krizhevsky, Geoffrey Hinton, Ilya Sutskever	1st	15.3%	60 million
2013	ZFNet()	Matthew Zeiler and Rob Fergus	1st	14.8%	
2014	GoogLeNet(19)	Google	1st	6.67%	4 million
2014	VGG Net(16)	Simonyan, Zisserman	2nd	7.3%	138 million
2015	ResNet(152)	Kaiming He	1st	3.6%	

Problem type	Last-layer activation	Loss function
Binary classification	`sigmoid`	`binary_crossentropy`
Multiclass, single-label classification	`softmax`	`categorical_crossentropy`
Multiclass, multilabel classification	`sigmoid`	`binary_crossentropy`
Regression to arbitrary values	None	`mse`
Regression to values between 0 and 1	`sigmoid`	`mse` or `binary_crossentropy`