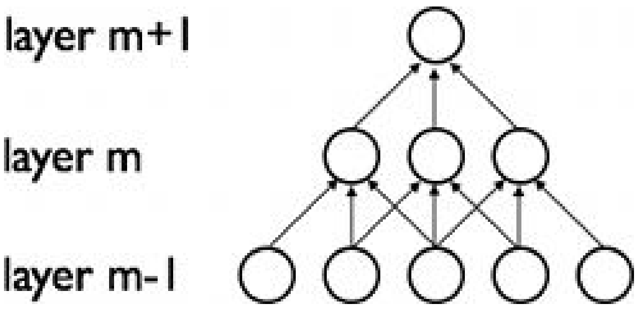
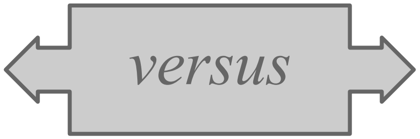
Unidad 4: Redes Convolucionales

Curso: Redes Neuronales Profundas

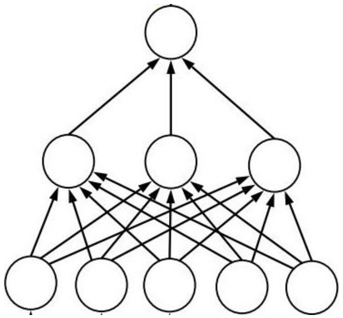
Convolutional Neural Networks

Sparse connectivity

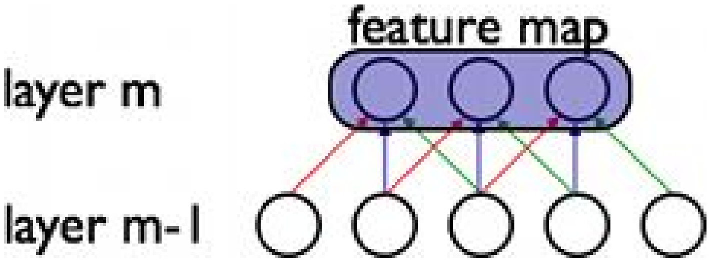




Dense connectivity



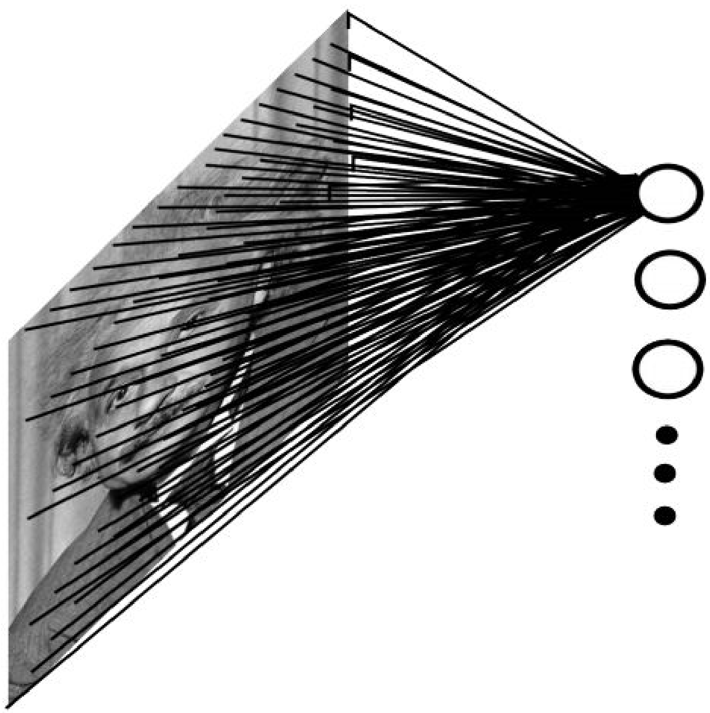
Shared weights



Convolutional Neural Networks

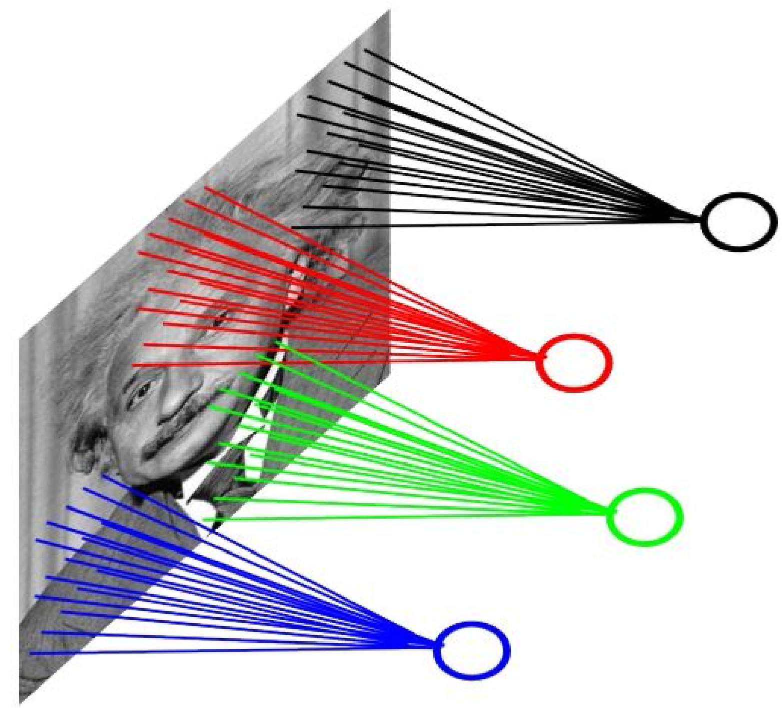
dot product + bias

$$h_k = \tanh(W_k^T x + b_k)$$



convolution + bias

$$h_{ij}^k = \tanh((W^k * x)_{ij} + b_k)$$

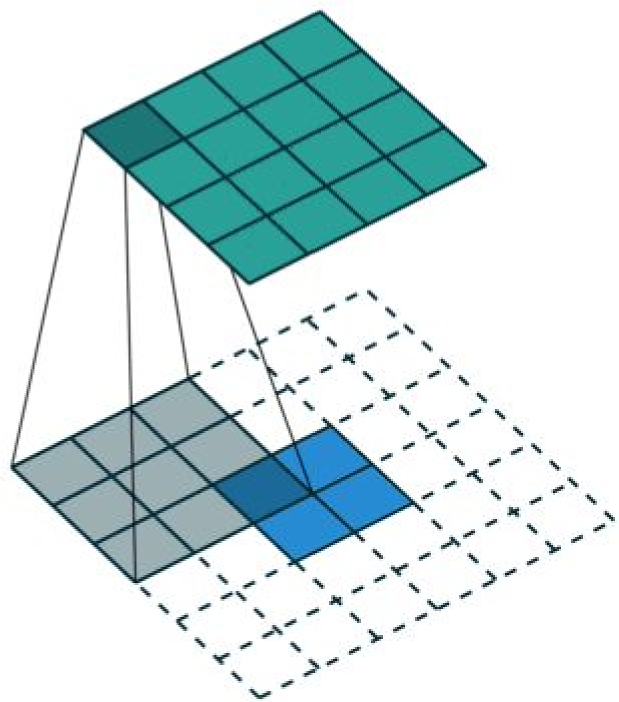


Convolución

$$s(t) = (x * w)(t) = \sum_{a=-\infty}^{\infty} x(a)w(t-a)$$

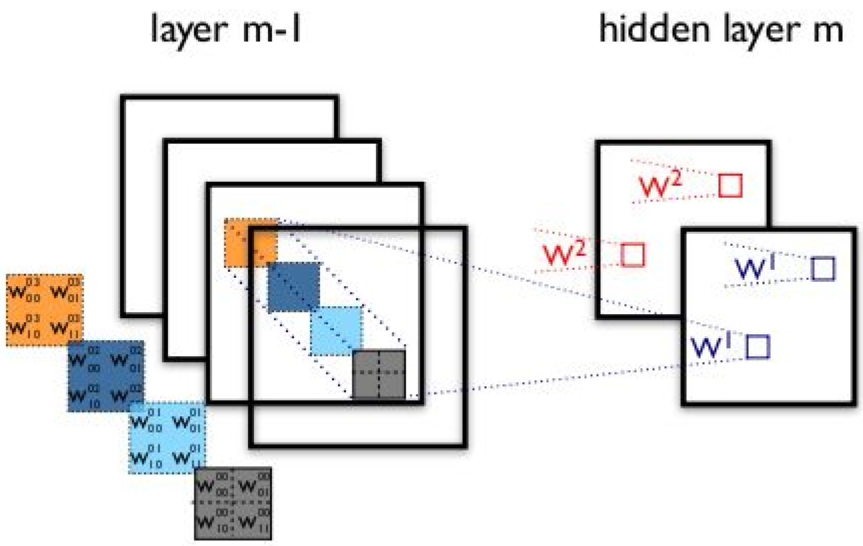
$$S(i,j) = (K * I)(i,j) = \sum_m \sum_n I(i-m, j-n)K(m,n)$$

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



Equivalentes en dos dimensiones

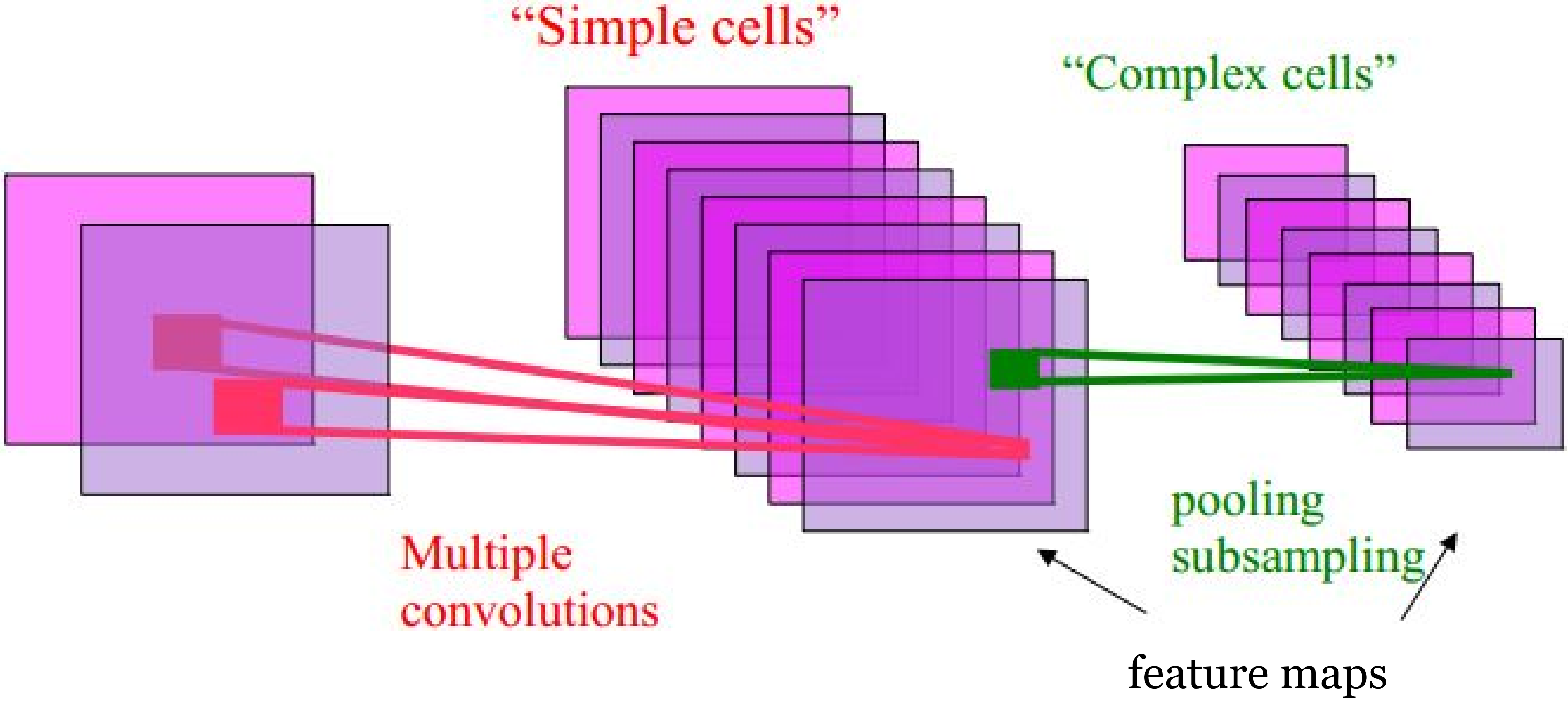
Convolutional Neural Networks



- Detects multiple motifs at each location
- The collection of units looking at the same patch is akin to a feature vector for that patch.
- The result is a 3D array, where each slice is a feature map.

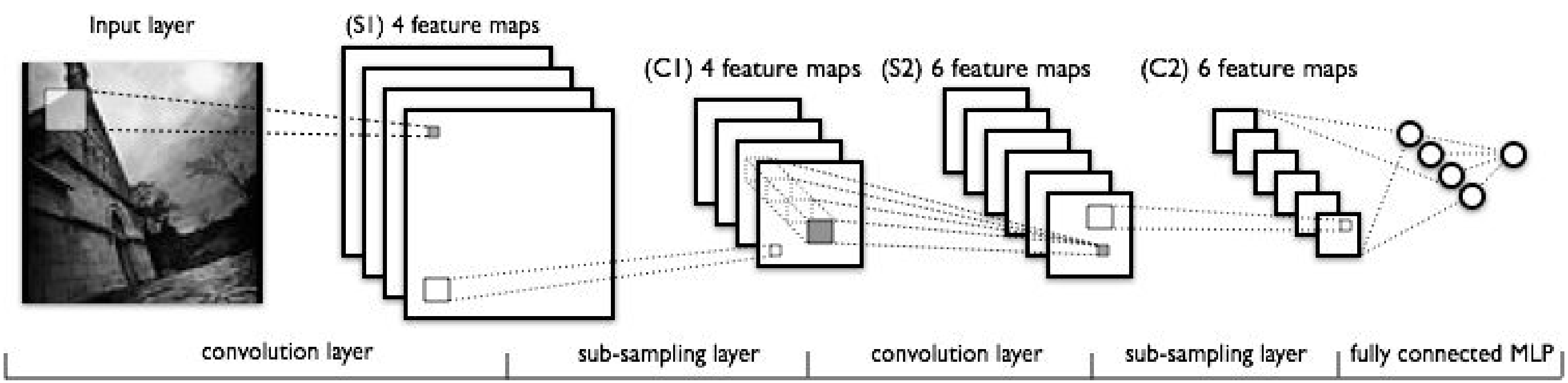
Convolutional Neural Networks

Pooling subsampling



Convolutional Neural Networks

- **Are deployed in many practical applications**
  - Image recognition, speech recognition, Google's and Baidu's photo taggers
- **Have won several competitions**
  - ImageNet, Kaggle Facial Expression, Kaggle Multimodal Learning, German Traffic Signs, Connectomics, Handwriting....
- **Are applicable to array data where nearby values are correlated**
  - Images, sound, time-frequency representations, video, volumetric images, RGB-Depth images,.....
- **One of the few deep models that can be trained purely supervised**

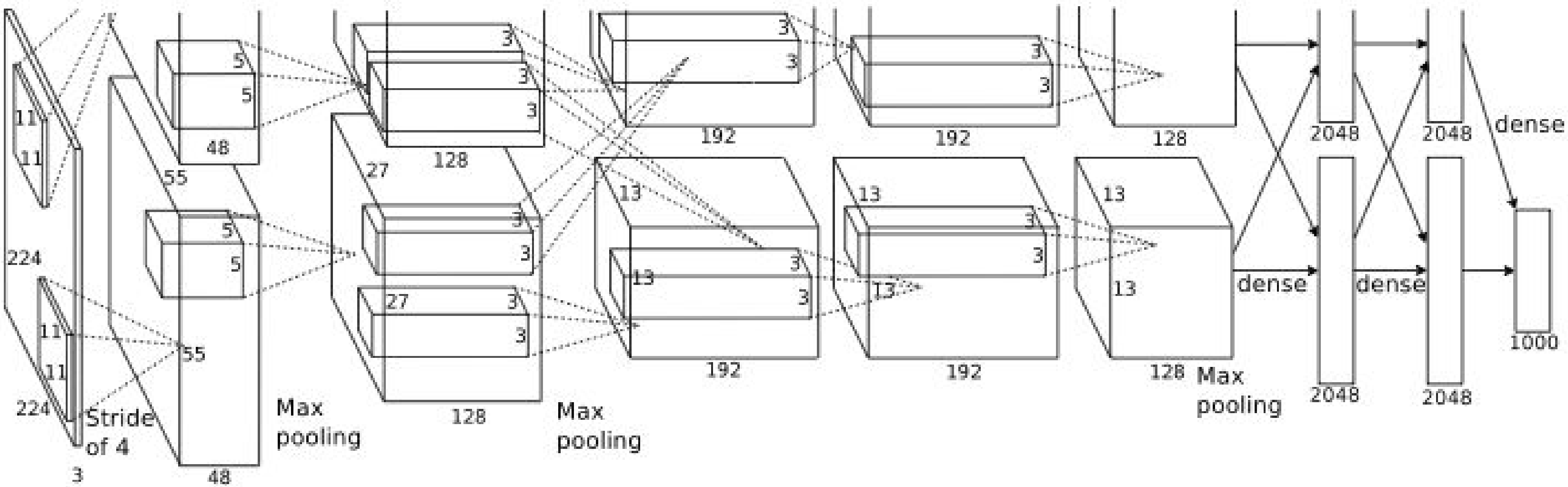


ImageNet Classification with Deep Convolutional Neural Networks

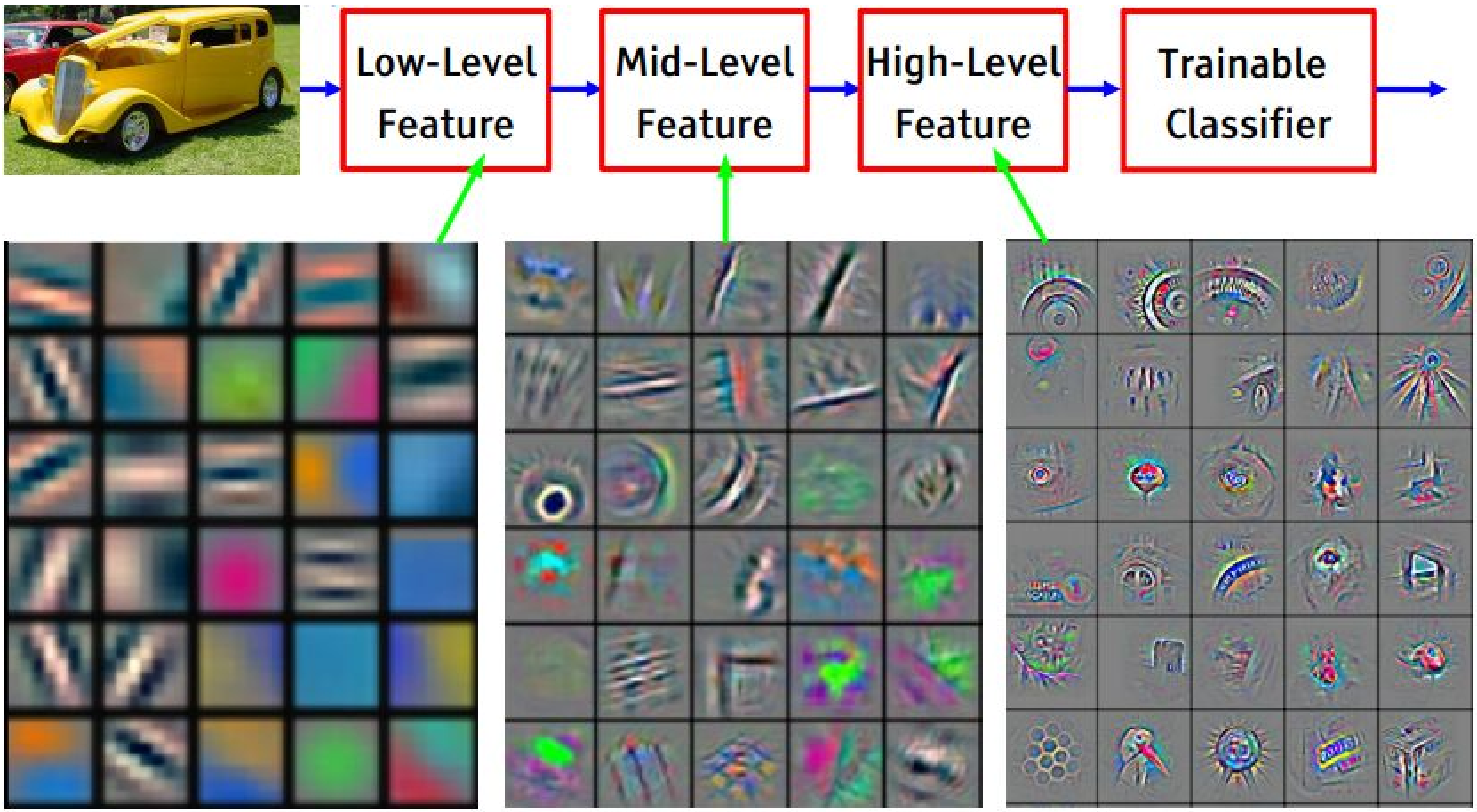
Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca

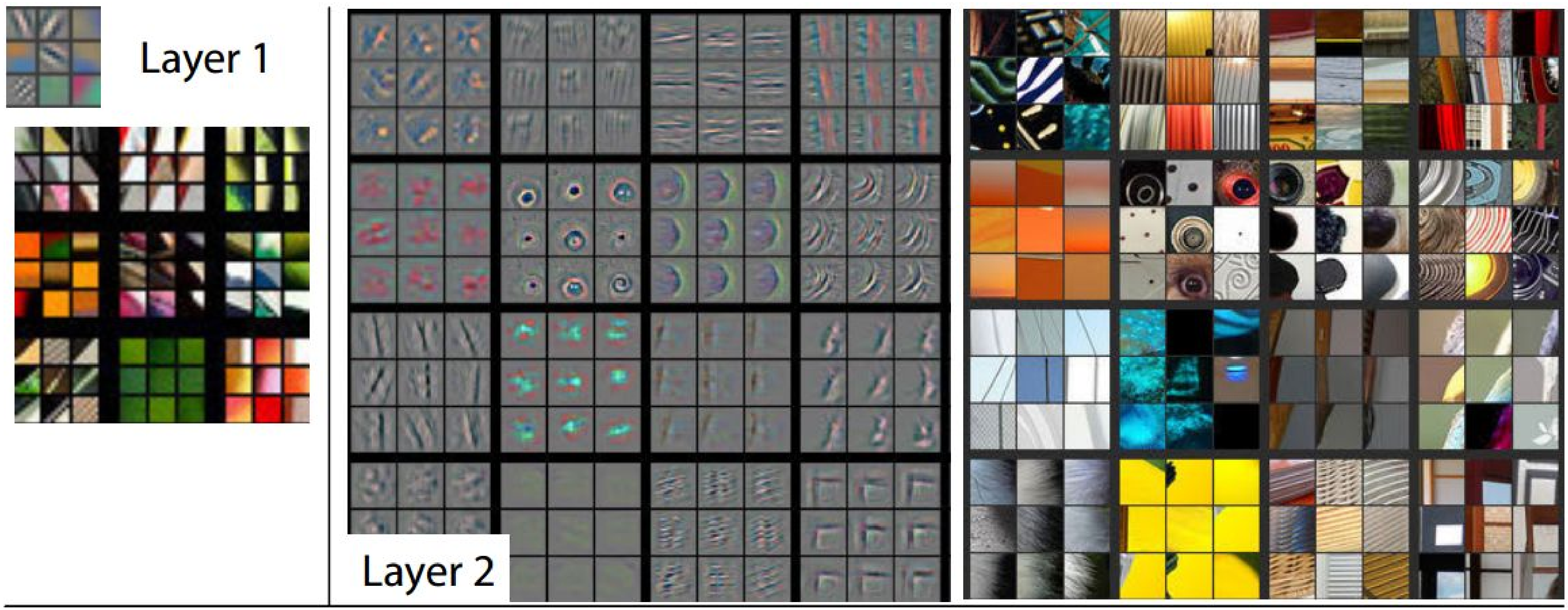


Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems* 2012: 1097-1105.



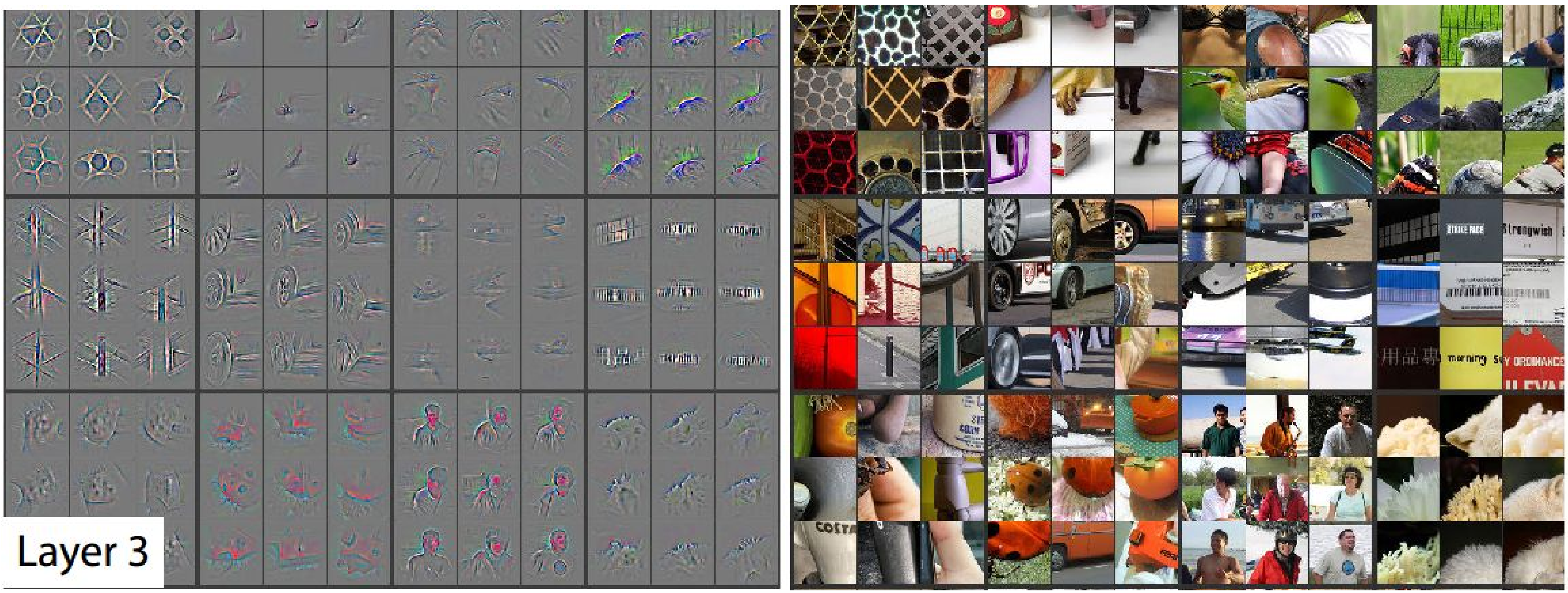
Zeiler, Matthew D, and Rob Fergus. "Visualizing and understanding convolutional neural networks." *arXiv preprint arXiv:1311.2901* (2013).

Visualizing and understanding convolutional neural networks



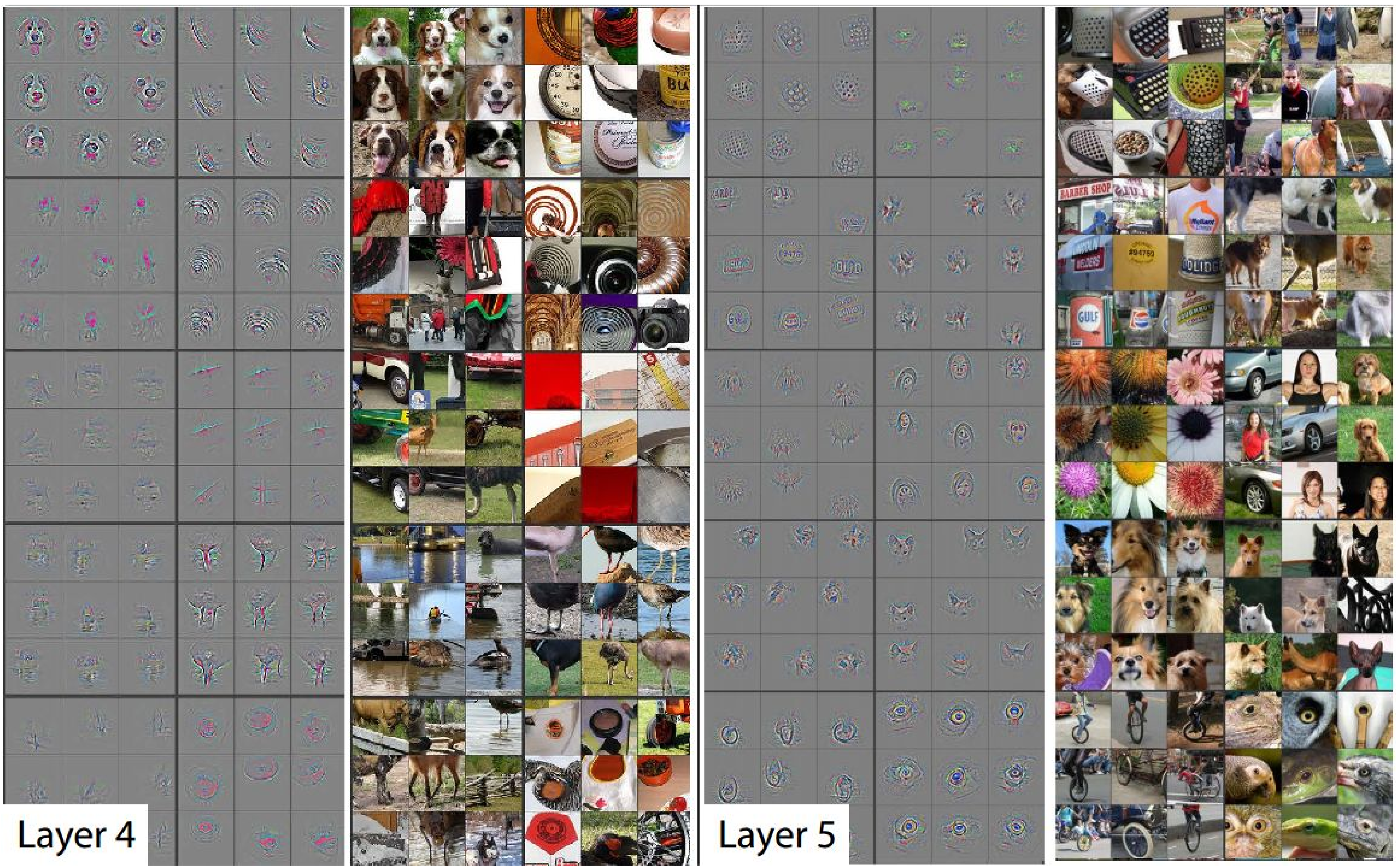
Zeiler, Matthew D, and Rob Fergus. "Visualizing and understanding convolutional neural networks." *arXiv preprint arXiv:1311.2901* (2013).

Visualizing and understanding convolutional neural networks

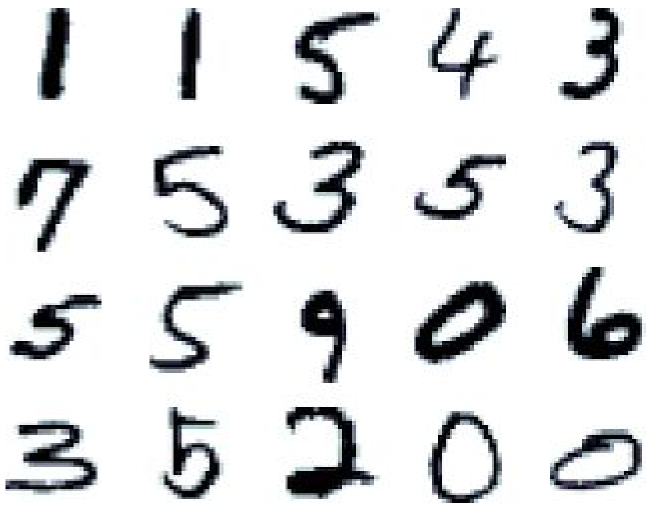


Zeiler, Matthew D, and Rob Fergus. "Visualizing and understanding convolutional neural networks." *arXiv preprint arXiv:1311.2901* (2013).

Visualizing and understanding convolutional neural networks

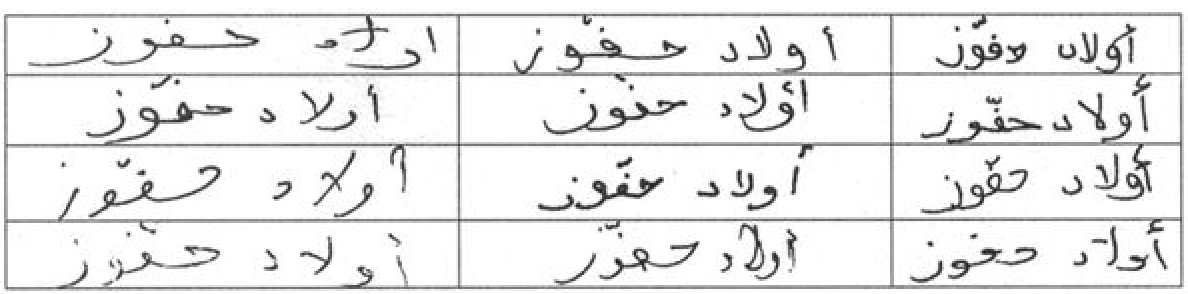


Tasks for Which Deep CNN are the Best



Handwriting recognition MNIST

Arabic Handwriting Recognition



Margner, Volker, and Haikal El Abed. "Arabic handwriting recognition competition." *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on* 23 Sep. 2007: 1274-1278.

Tasks for Which Deep CNN are the Best

StreetView House Numbers [2011]

94.3 % accuracy

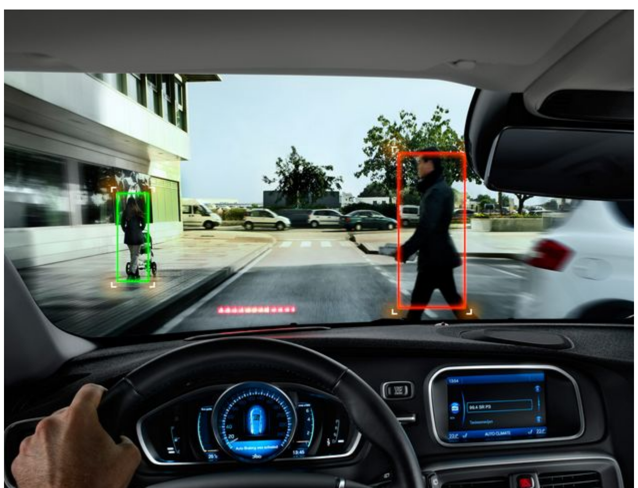
Netzer, Yuval et al. "Reading digits in natural images with unsupervised feature learning." *NIPS workshop on deep learning and unsupervised feature learning* 2011: 4.

Tasks for Which Deep CNN are the Best

Traffic Sign Contest, Silicon Valley, 2011(IDSIA)

0.56% ERROR

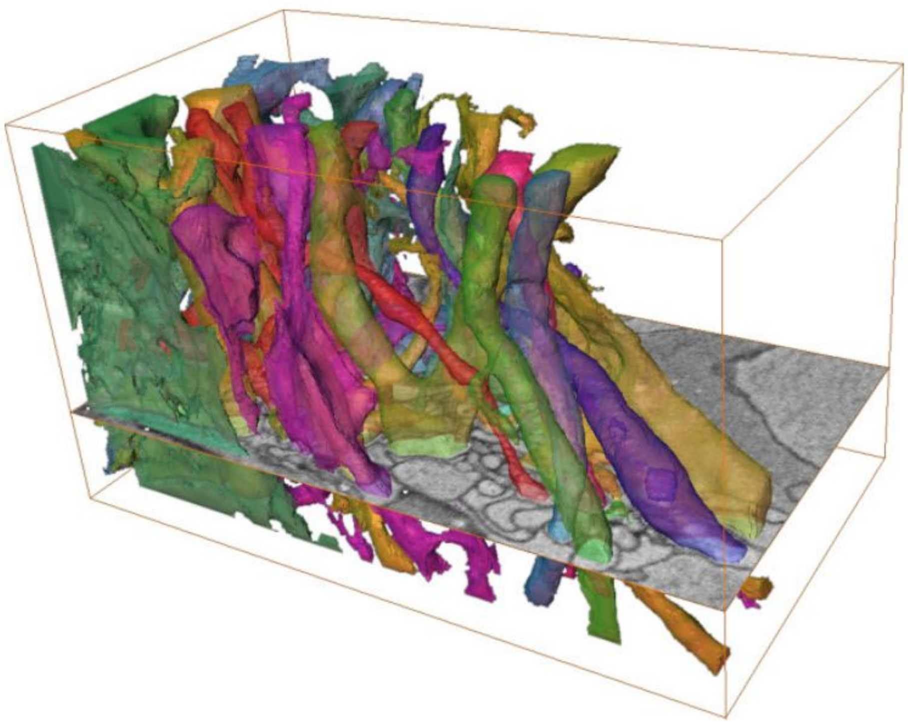
- first place
- twice better than humans
- three times better than the closest artificial competitor
- six times better than the best non-neural method



Pedestrian Detection [2013]: INRIA datasets and others (NYU)

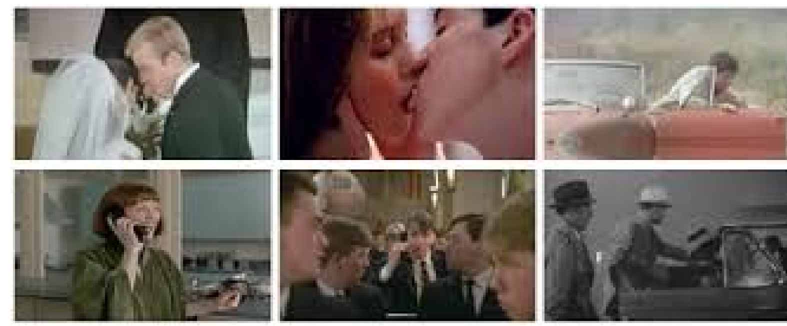
Tasks for Which Deep CNN are the Best

Volumetric brain image segmentation [2009] Connectomics (IDSIA, MIT)



Turaga, Srinivas C et al. "Convolutional networks can learn to generate affinity graphs for image segmentation." *Neural Computation* 22.2 (2010): 511-538.

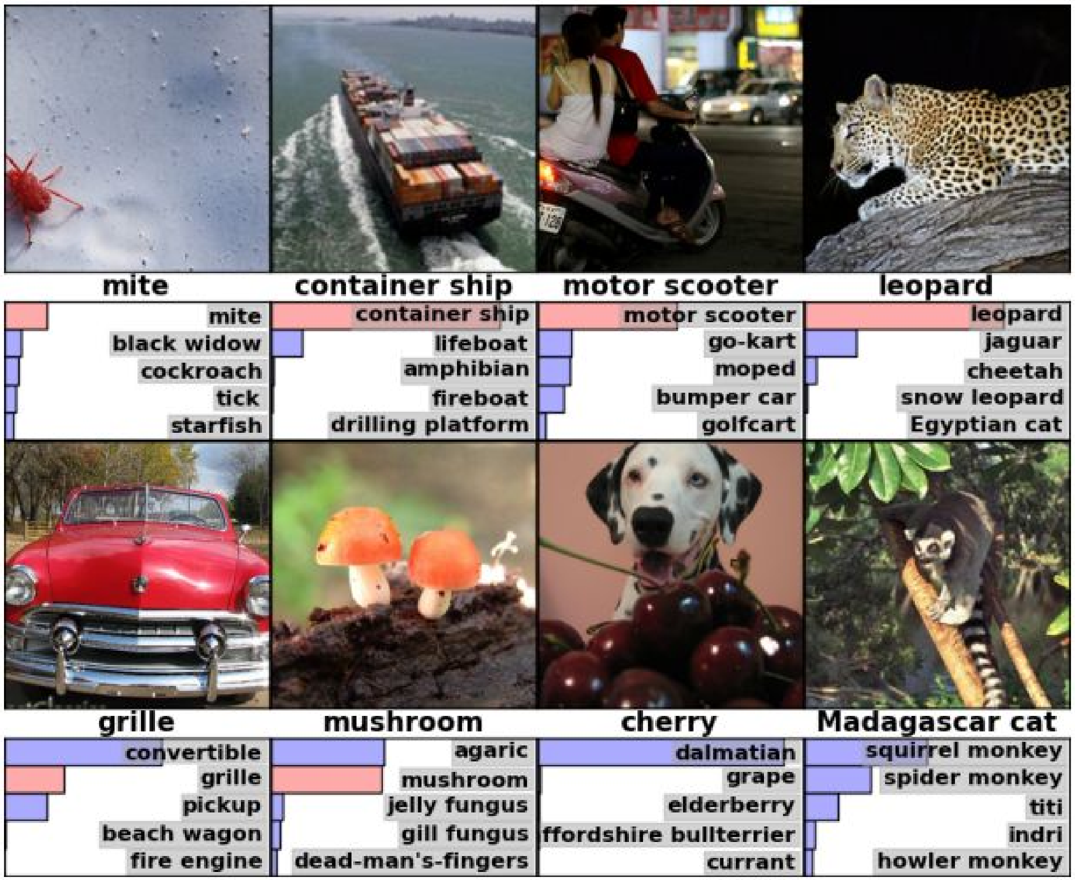
Human Action Recognition [2011] Hollywood II dataset (Stanford)



Le, Quoc V et al. "Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis." *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on* 20 Jun. 2011: 3361-3368.

Tasks for Which Deep CNN are the Best

Object Recognition [2012] ImageNet competition

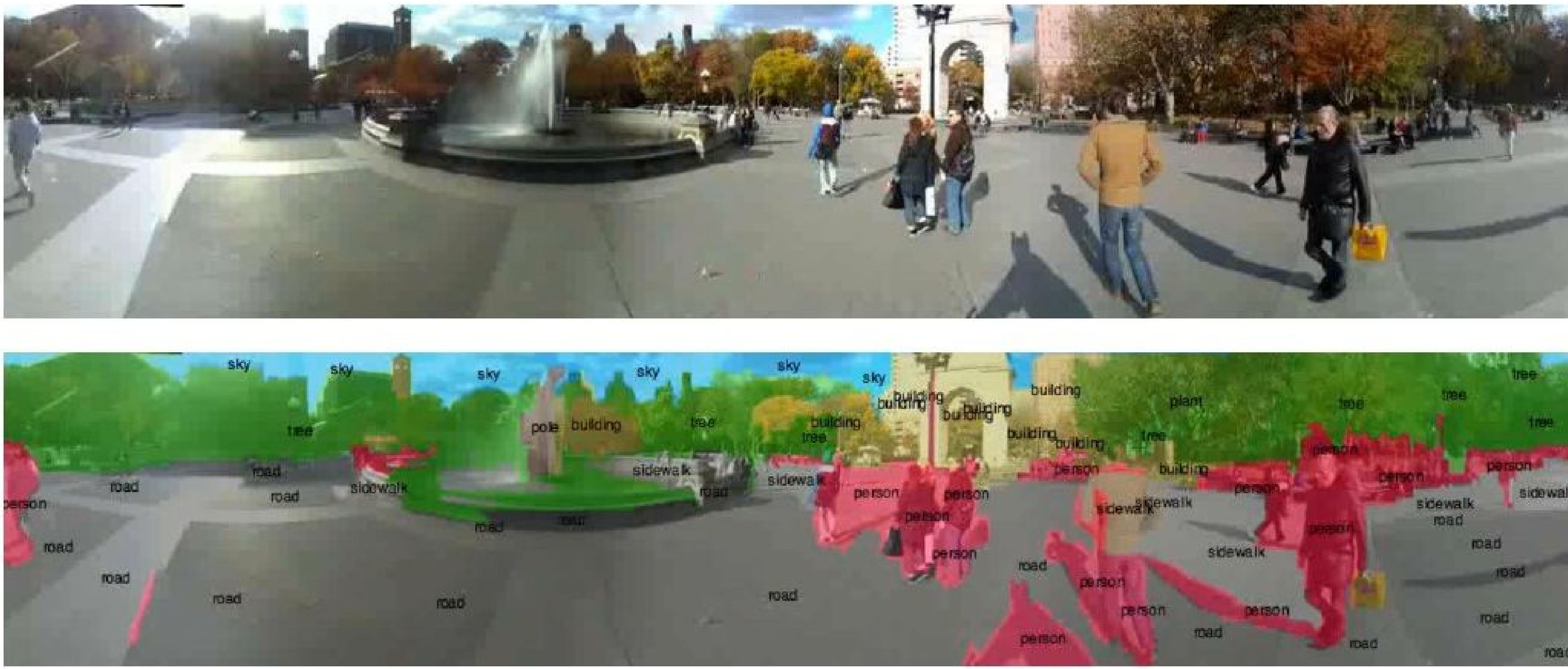


Error rate: 15% (whenever correct class isn't in top 5)
Previous state of the art: 25% error

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems* 2012: 1097-1105.

Tasks for Which Deep CNN are the Best

Scene Parsing [2012]



Farabet, Clément et al. "Scene parsing with multiscale feature learning, purity trees, and optimal covers." *arXiv preprint arXiv:1202.2160* (2012).

Google AI algorithm masters ancient game of Go

Deep-learning software defeats human professional for first time.

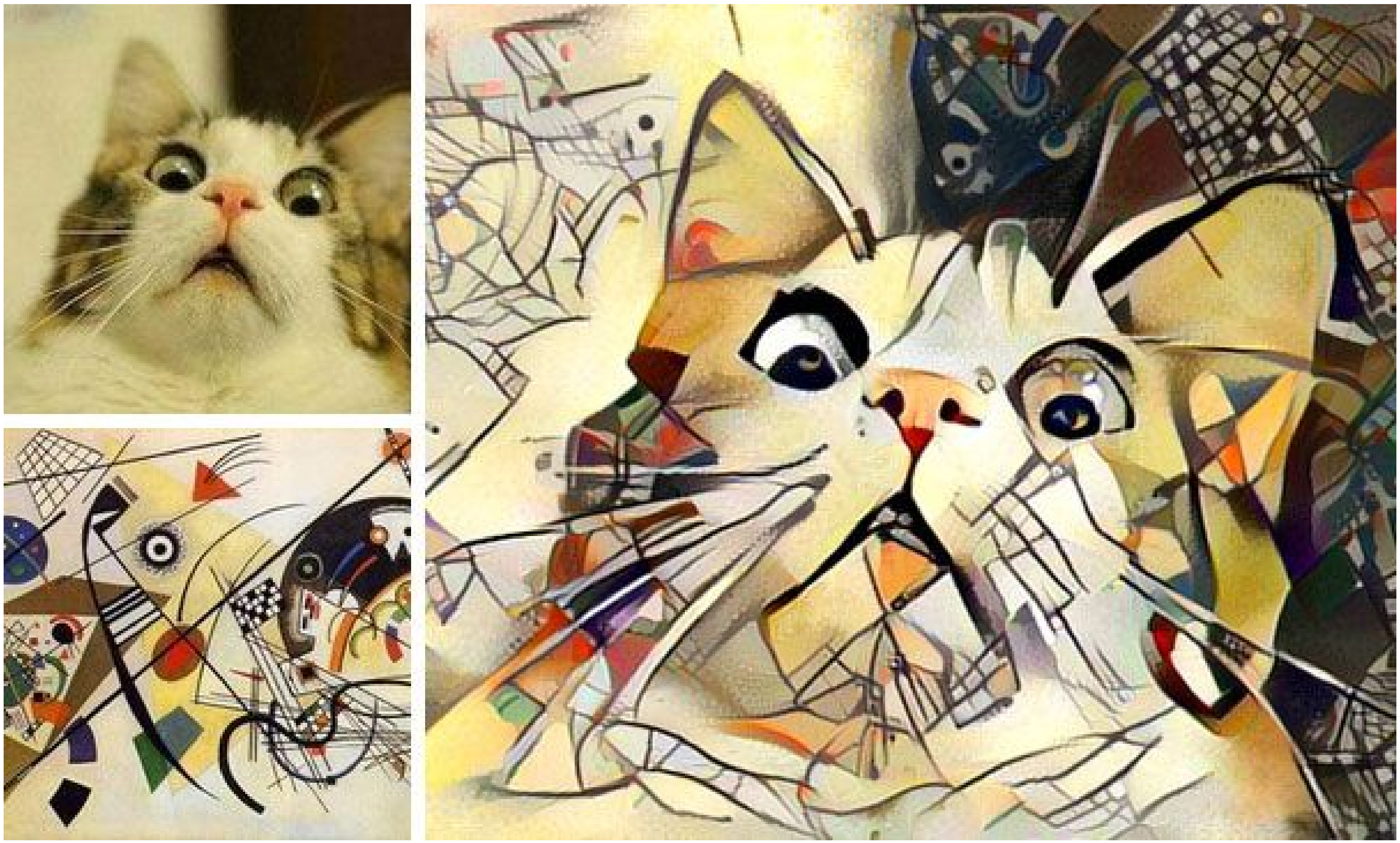
27 January 2016

A computer has beaten a human professional for the first time at Go — an ancient board game that has long been viewed as one of the greatest challenges for artificial intelligence (AI)

"We pass in the board position as a 19 × 19 image and use convolutional layers to construct a representation of the position."

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., ... & Dieleman, S. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, *529*(7587), 484-489.

Style transfer



Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge. "A neural algorithm of artistic style." *arXiv preprint arXiv:1508.06576* (2015).

Unidad 4: Redes Convolucionales

Curso: Redes Neuronales Profundas

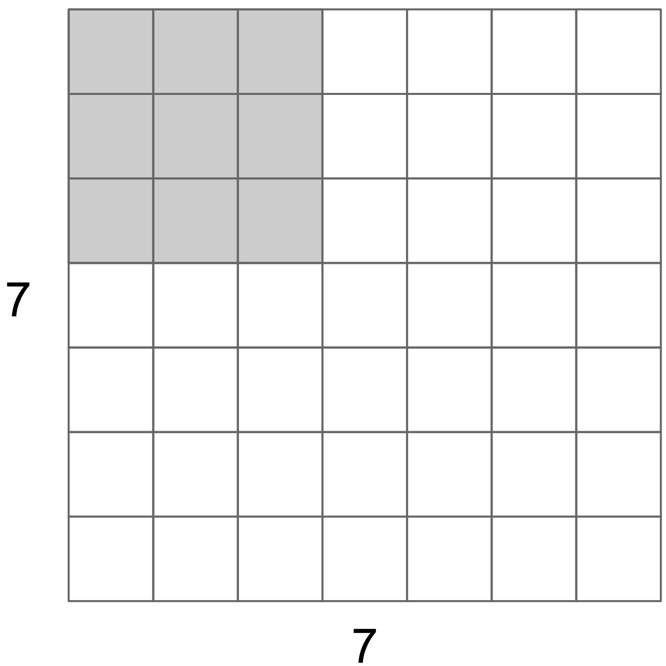
Se aproximan las JCC...

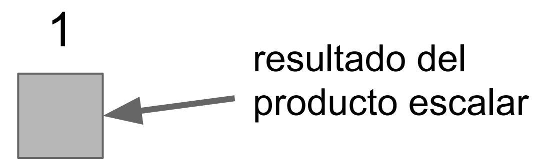
Charlas confirmadas

- **Andrés Rojas Paredes y Jheison López Restrepo (UNGS)** - Teoría de la complejidad y Heurísticas: Problema de optimización combinatoria QAP
- **Carlos Areces (FaMAF UNC)** - Optimizando Dominios de Planning
- **Daniel Fino (FUESMEN)** - Desafíos computacionales en radiodiagnóstico por MR y MR/PET
- **Ignacio Cassol (Universidad Austral)** - Refactorización de modelos estructurados de alto nivel a OO
- **Lucas Uzal (CIFASIS)** - Deep Learning en Machine Vision
- **Pablo Altamura (NeuralSoft)** - La Inteligencia Artificial aplicada a la gestión de las organizaciones
- **Carlos Luna (UdelaR)** - Análisis formal de modelos de seguridad para sistemas críticos: plataformas de virtualización y dispositivos móviles.
- **Uciel Pablo Chorostecki (IBR)** - La bioinformática como disciplina científica
- **Eugenia Simich (FCEIA - UNR)** - Construyendo tipos de datos con containers
- **Juan Edi (Manas)** - Introducción a Crystal: creando programas eficientes sin resignar simplicidad
- **Martín Ceresa (DCC - FCEIA - UNR)** - Charla de difusión: Elige tu propia LCC

Miércoles, jueves y viernes
http://fceia.unr.edu.ar/lcc/jcc/2016/

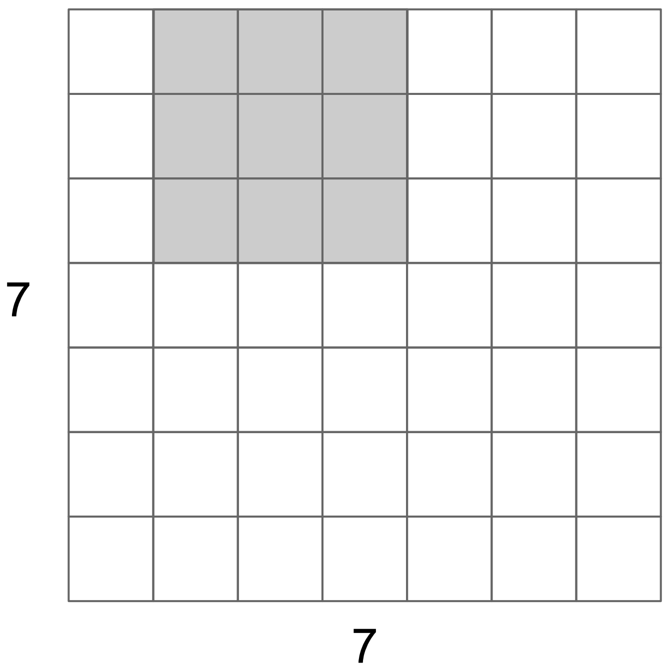
Tamaño de la salida

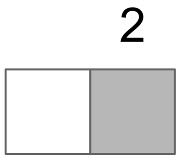





```
input_size = 7x7
filter_size = 3x3
stride = 1
output_size = ?
```

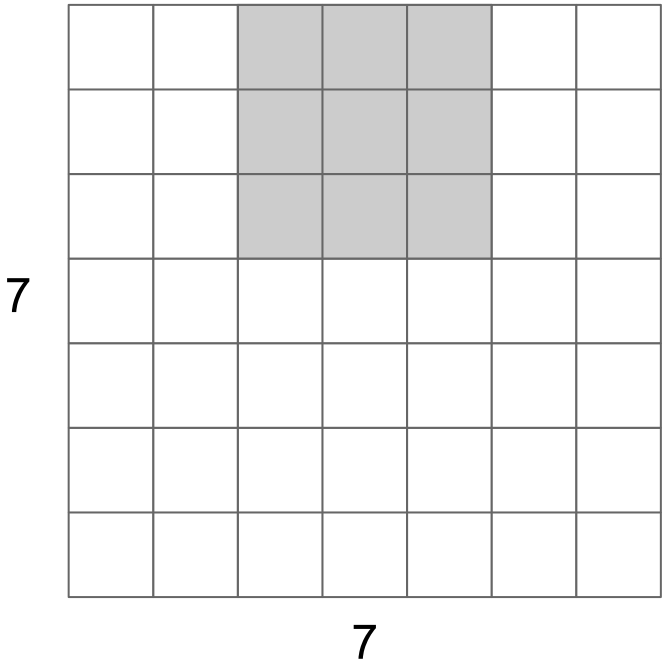
Tamaño de la salida

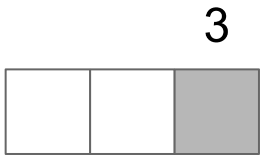





```
input_size = 7x7
filter_size = 3x3
stride = 1
output_size = ?
```

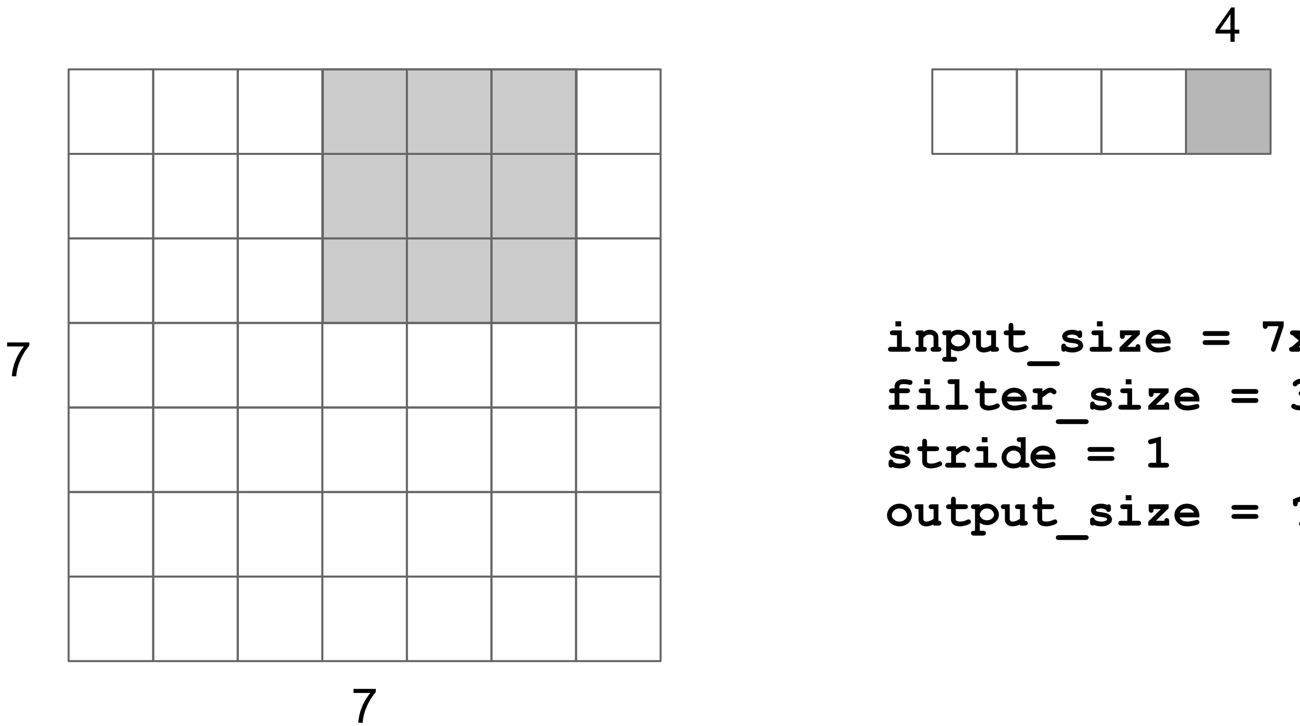
Tamaño de la salida





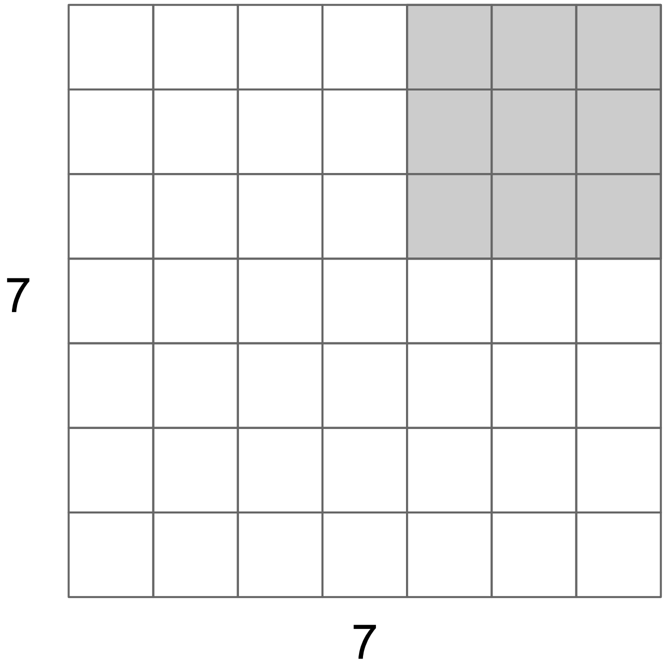

```
input_size = 7x7
filter_size = 3x3
stride = 1
output_size = ?
```

Tamaño de la salida



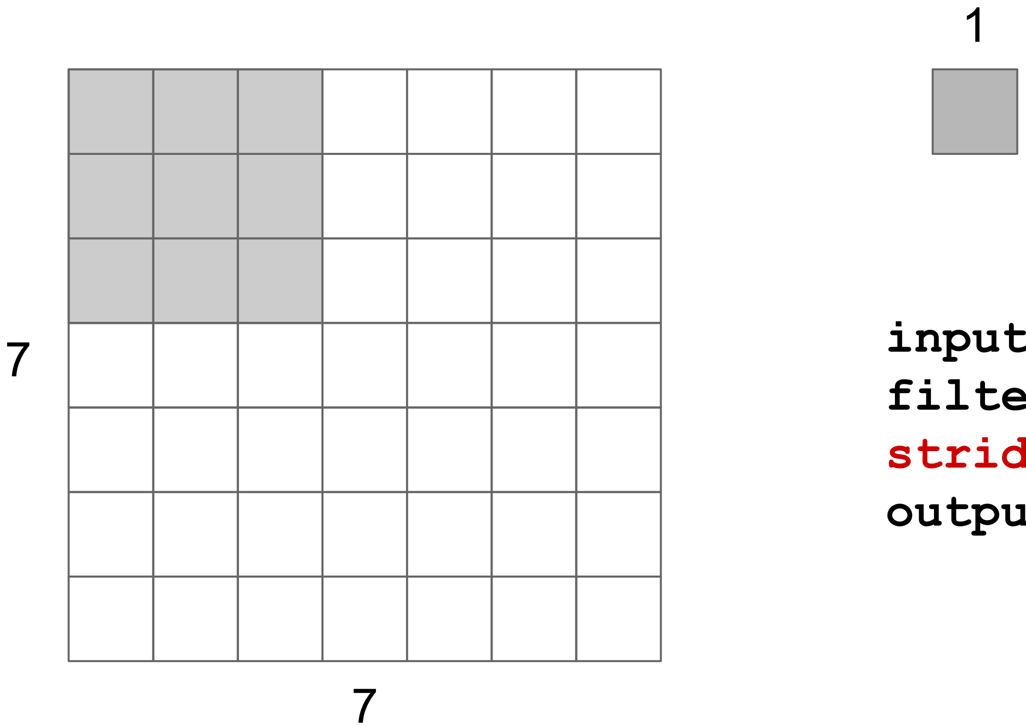

```
input_size = 7x7
filter_size = 3x3
stride = 1
output_size = ?
```

Tamaño de la salida



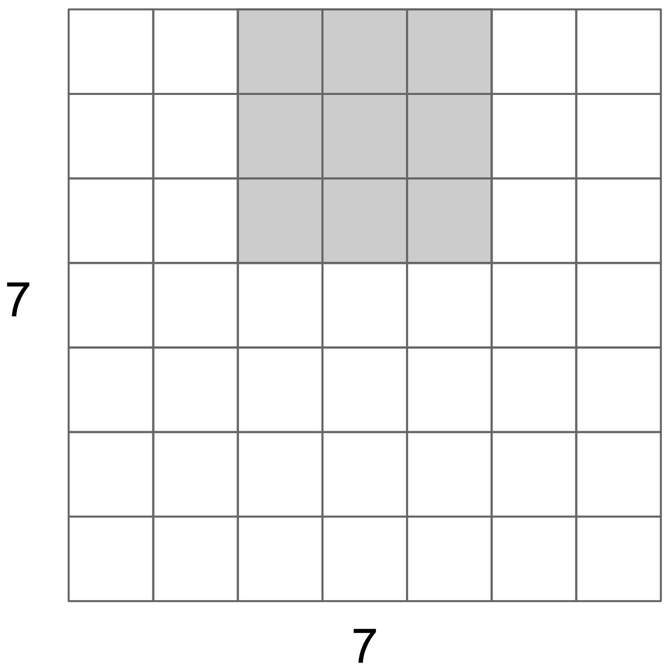

```
input_size = 7x7
filter_size = 3x3
stride = 1
output_size = 5x5
```

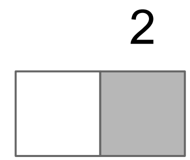
Tamaño de la salida




```
input_size = 7x7
filter_size = 3x3
stride = 2
output_size = ?
```

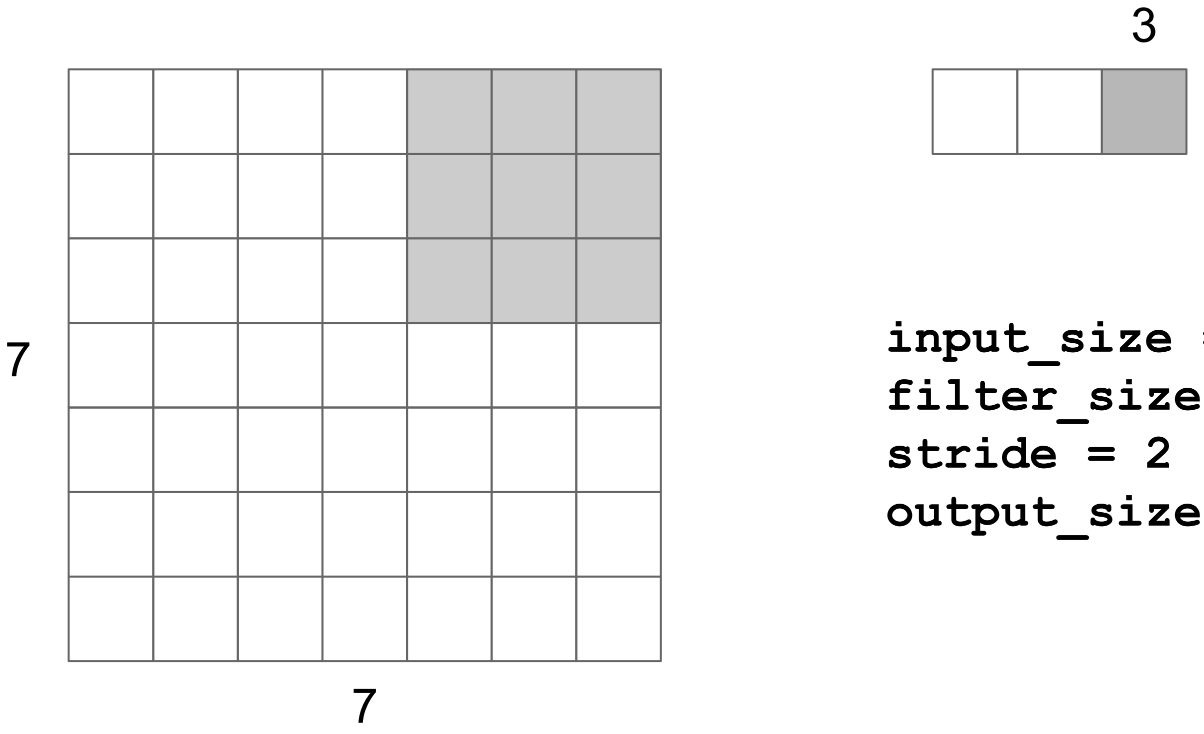
Tamaño de la salida





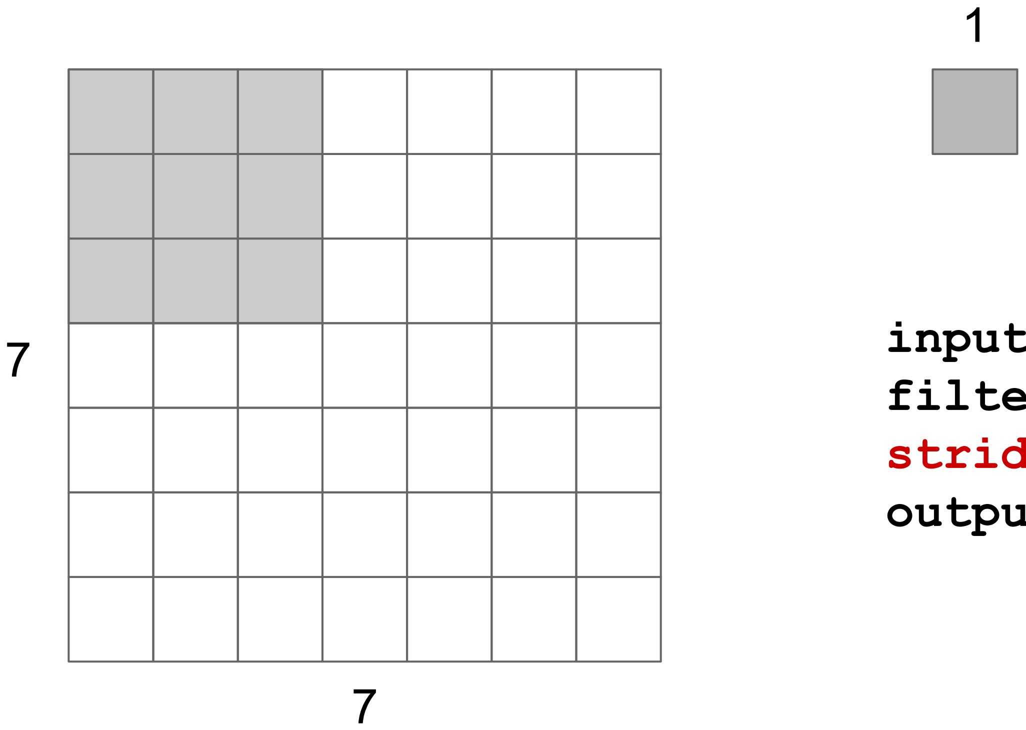

```
input_size = 7x7
filter_size = 3x3
stride = 2
output_size = ?
```

Tamaño de la salida



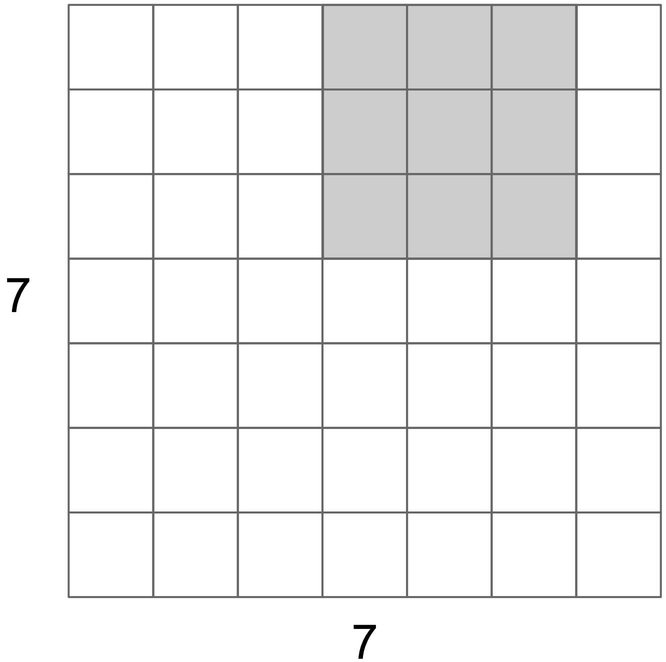

```
input_size = 7x7
filter_size = 3x3
stride = 2
output_size = 3x3
```

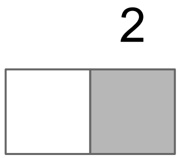
Tamaño de la salida




```
input_size = 7x7
filter_size = 3x3
stride = 3
output_size = ?
```

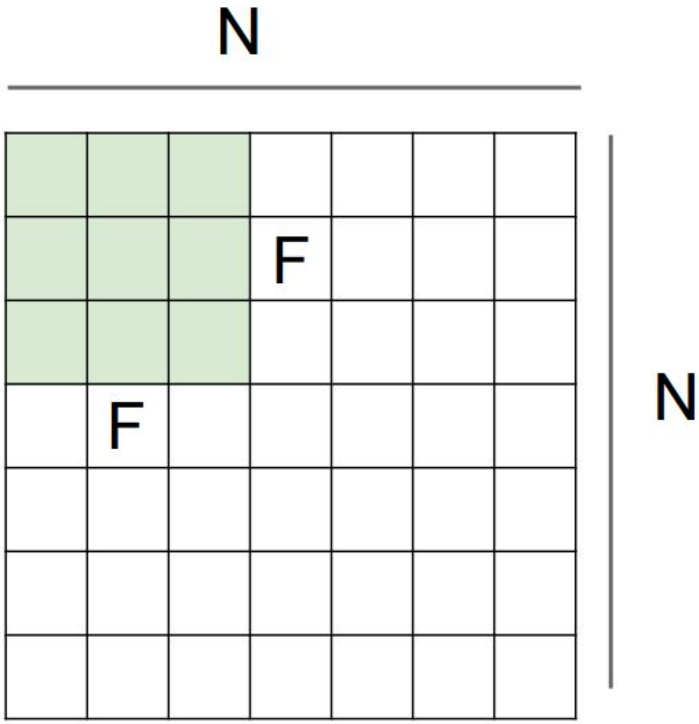
Tamaño de la salida






```
input_size = 7x7
filter_size = 3x3
stride = 3
output_size = 2x2 (!)
```

Tamaño de la salida

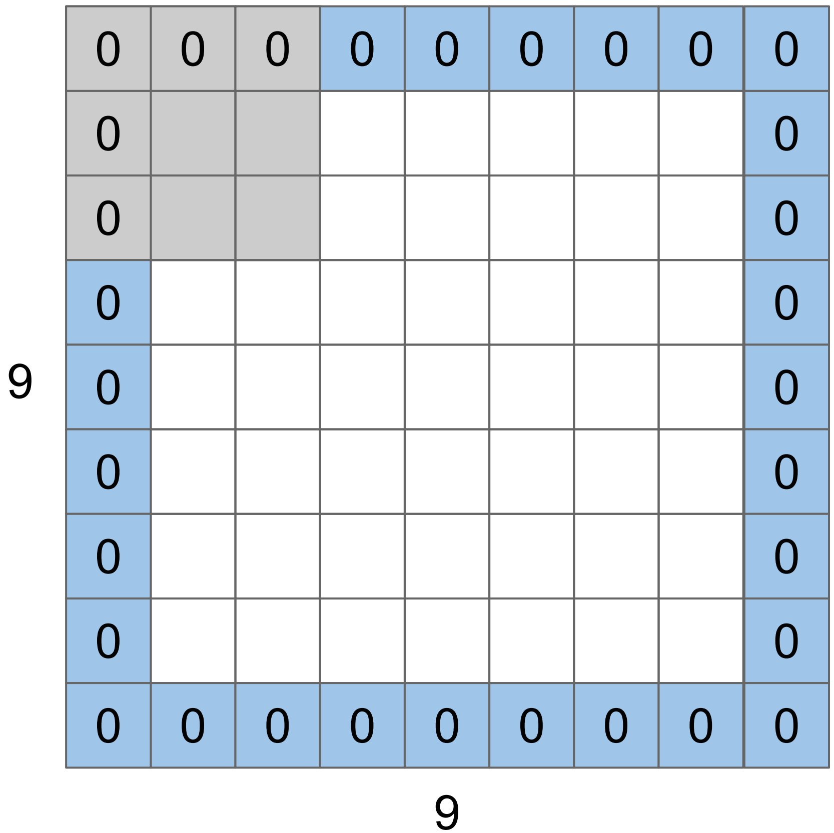


Output size:
(N - F) / stride + 1

e.g. N = 7, F = 3:
stride 1 => (7 - 3)/1 + 1 = 5
stride 2 => (7 - 3)/2 + 1 = 3
stride 3 => (7 - 3)/3 + 1 = 2.33

http://cs231n.stanford.edu/syllabus.html

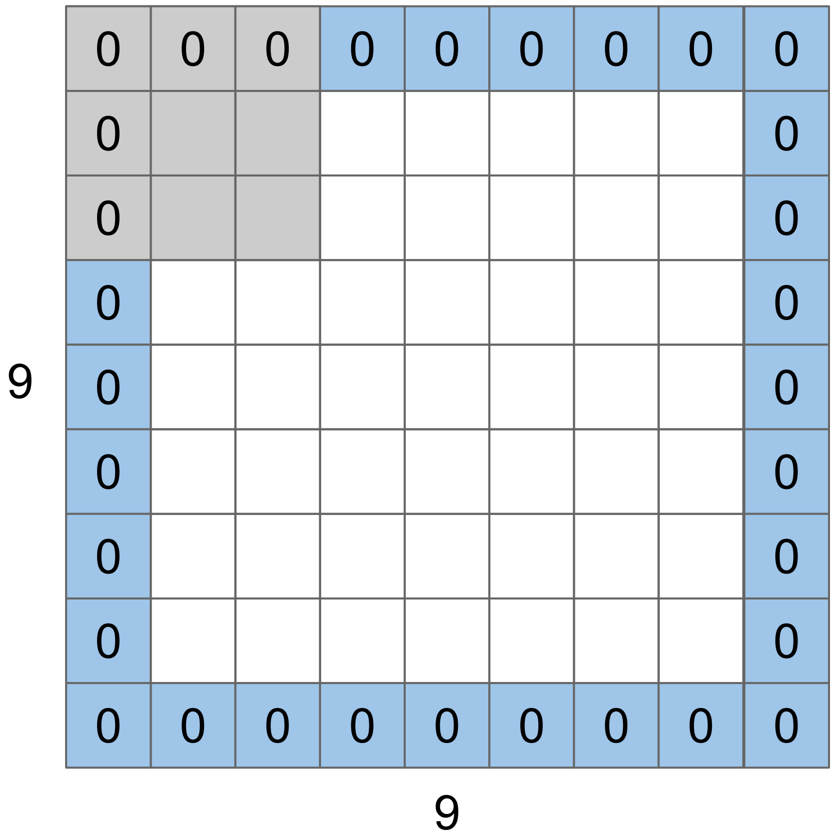
Zero padding




```
input_size = 7x7
filter_size = 3x3
stride = 1
padding = 1
output_size = 7x7
```

```
output_size = (input_size + 2*padding - filter_size) / stride + 1
```

Zero padding



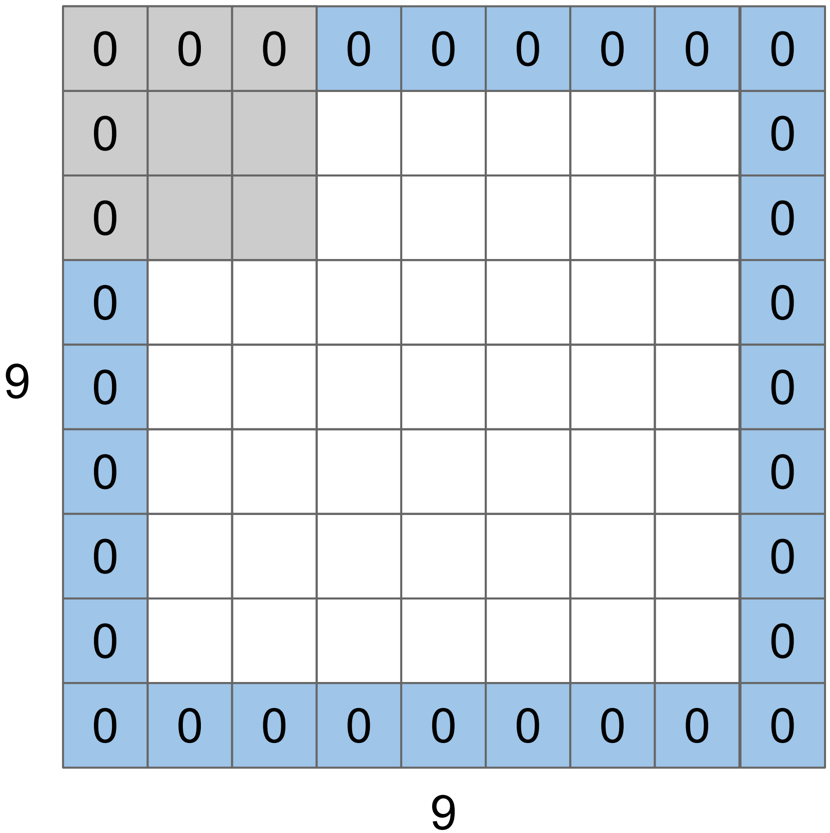
Algunos modos predefinidos

```
'valid':
    padding = 0
'same':
    padding = (filter_size - 1) / 2
'full':
    padding = filter_size - 1
```

```
input_size = 7x7
filter_size = 3x3
stride = 1
padding = 1
output_size = 7x7
```

output_size = (input_size + 2*padding - filter_size) / stride + 1

Razones para usar padding (='same')



- Controlar más fácilmente el tamaño de los mapas (independizarse de **`filter_size`**)
- Detectar patrones 'pegados a los bordes'

```
input_size = 7x7
filter_size = 3x3
stride = 1
padding = 1
output_size = 7x7
```

`output_size = (input_size + 2*padding - filter_size) / stride + 1`

`theano.tensor.nnet.conv2d`(*input*, *filters*, *input_shape=None*, *filter_shape=None*, *border_mode='valid'*,*subsample=(1, 1)*, *filter_flip=True*)

input (*symbolic 4D tensor*) – Mini-batch of feature map stacks, of shape (batch size, input channels, input rows, input columns).

filters (*symbolic 4D tensor*) – Set of filters used in CNN layer of shape (output channels, input channels, filter rows, filter columns).

input_shape (*None, tuple/list of len 4 of int or Constant variable*) – The shape of the input parameter. Optional, possibly used to choose an optimal implementation. You can give `None` for any element of the list to specify that this element is not known at compile time.

filter_shape (*None, tuple/list of len 4 of int or Constant variable*) – The shape of the filters parameter. Optional, possibly used to choose an optimal implementation. You can give `None` for any element of the list to specify that this element is not known at compile time.

subsample (*tuple of len 2*) – Factor by which to subsample the output. Also called strides elsewhere.

border_mode (*str, int or tuple of two int*) –
Either of the following:

- `'valid'`: **apply filter wherever it completely overlaps with the input**.
- `'full'`: **apply filter wherever it partly overlaps with the input.**
- `'half'`: **pad input with a symmetric border of** `filter rows / 2` rows and `filter columns / 2` columns.
- `int`: **pad input with a symmetric border of zeros of the given width**
- `(int1, int2)`: **pad input with a symmetric border of** `int1` **rows** and `int2` columns.

filter_flip (*bool*) – If `True`, will flip the filter rows and columns before sliding them over the input. This operation is normally referred to as a convolution, and this is the default. If `False`, the filters are not flipped and the operation is referred to as a cross-correlation.

```
keras.layers.convolutional.Convolution2D(nb_filter, nb_row, nb_col,
init='glorot_uniform', activation='linear', weights=None, border_mode='valid',
subsample=(1, 1), dim_ordering='default', W_regularizer=None, b_regularizer=None,
activity_regularizer=None, W_constraint=None, b_constraint=None, bias=True)
```

input_shape: When using this layer as the first layer in a model, provide this keyword argument, e.g. input_shape=(3, 128, 128) for 128x128 RGB pictures.

nb_filter: Number of convolution filters to use.

nb_row: Number of rows in the convolution kernel.

nb_col: Number of columns in the convolution kernel.

init: name of initialization function for the weights of the layer (see initializations), or alternatively, Theano function to use for weights initialization. This parameter is only relevant if you don't pass a `weights` argument.

activation: name of activation function to use (see activations), or alternatively, elementwise Theano function. If you don't specify anything, no activation is applied (ie. "linear" activation: a(x) = x).

weights: list of numpy arrays to set as initial weights.

border_mode: 'valid' or 'same'.

subsample: tuple of length 2. Factor by which to subsample output. Also called strides elsewhere.

W_regularizer: ...

b_regularizer: ...

activity_regularizer: ...

W_constraint: ...

b_constraint: ...

dim_ordering: 'th' or 'tf'. In 'th' mode, the channels dimension (the depth) is at index 1, in 'tf' mode is it at index 3. It defaults to the`image_dim_ordering` value found in your Keras config file at `~/.keras/keras.json`. If you never set it, then it will be "tf".

bias: whether to include a bias (i.e. make the layer affine rather than linear).

Configuraciones típicas

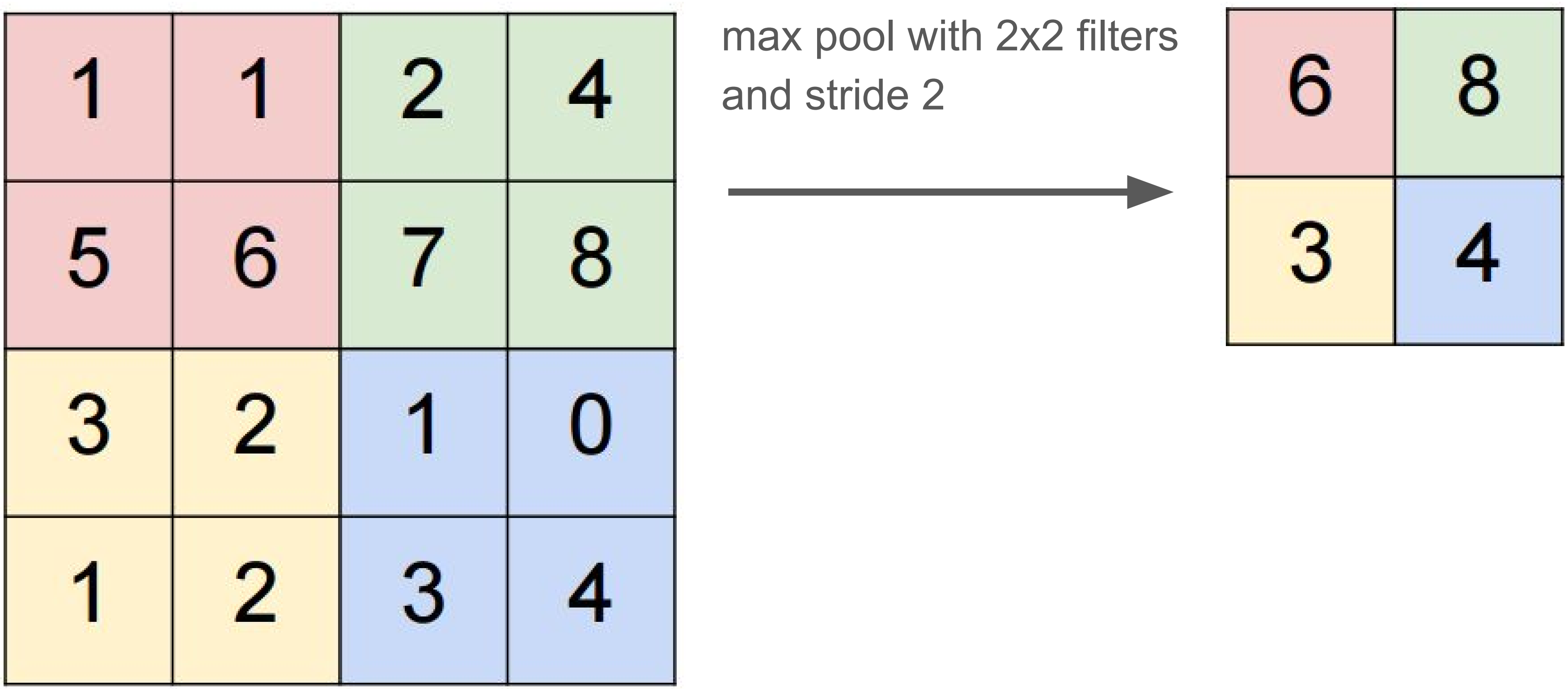
```
input_size = 2^N x 2^N
filter_size = 3x3
nb_filter = 2^M
stride = 1
padding = 1
   output_size = input_size
```

```
input_size = 2^N x 2^N
filter_size = 5x5
nb_filter = 2^M
stride = 1
padding = 2
   output_size = input_size
```

```
input_size = 2^N x 2^N
filter_size = 5x5
nb_filter = 2^M
stride = 2
padding = 'same'
   output_size = input_size/2
```

```
input_size = 2^N x 2^N
filter_size = 1x1
nb_filter = 2^M
stride = 1
padding = 0
   output_size = input_size
```

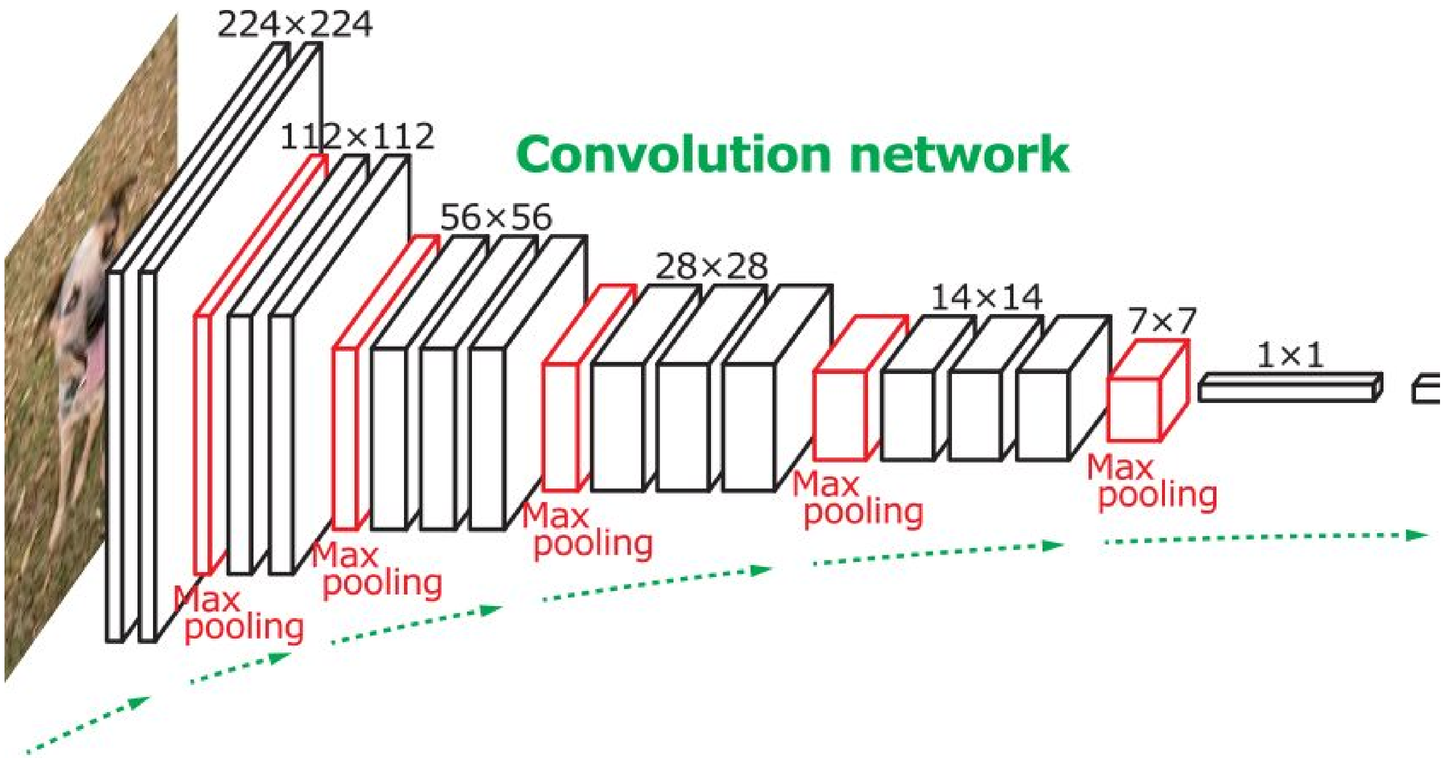
Pooling



`output_size = (input_size - pool_size) / stride + 1`

Si **`stride = pool_size:`** **`output_size = input_size / stride`**

Ejemplo: VGG net ("versión D")



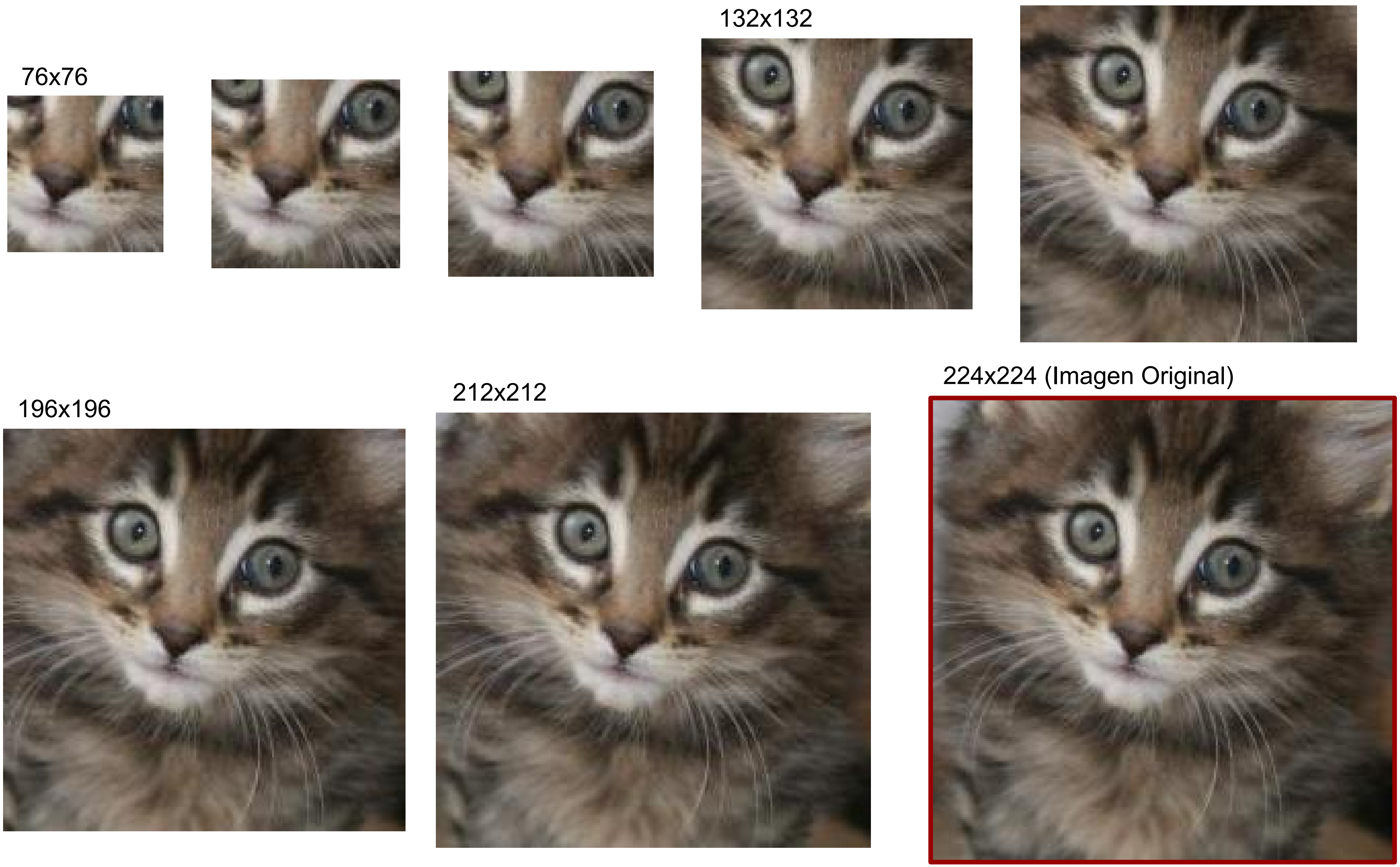
Input 224x224
conv3-64
conv3-64
maxpool
conv3-128
conv3-128
maxpool
conv3-256
conv3-256
conv3-256
maxpool
conv3-512
conv3-512
conv3-512
maxpool
conv3-512
conv3-512
conv3-512
maxpool
FC-4096
FC-4096
FC-1000
softmax

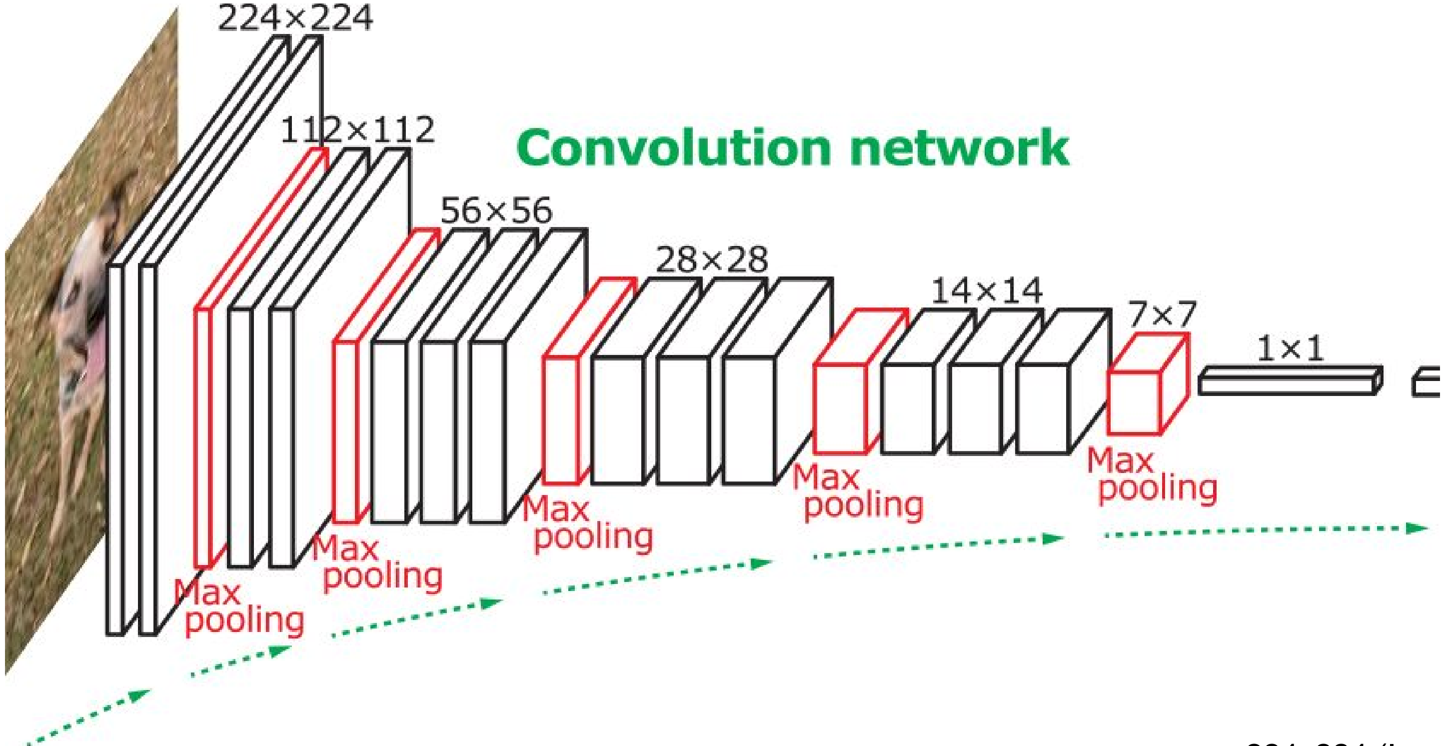
dataset: ImageNet (ILSVRC 2013)

Ejemplo: VGG net ("versión D")

```
INPUT:      [224x224x3]    memory: 224*224*3=150K    params: 0
CONV3-64:   [224x224x64]   memory: 224*224*64=3.2M   params: (3*3*3)*64 = 1,728
CONV3-64:   [224x224x64]   memory: 224*224*64=3.2M   params: (3*3*64)*64 = 36,864
POOL2:      [112x112x64]   memory: 112*112*64=800K   params: 0
CONV3-128:  [112x112x128]  memory: 112*112*128=1.6M  params: (3*3*64)*128 = 73,728
CONV3-128:  [112x112x128]  memory: 112*112*128=1.6M  params: (3*3*128)*128 = 147,456
POOL2:      [56x56x128]    memory: 56*56*128=400K    params: 0
CONV3-256:  [56x56x256]    memory: 56*56*256=800K    params: (3*3*128)*256 = 294,912
CONV3-256:  [56x56x256]    memory: 56*56*256=800K    params: (3*3*256)*256 = 589,824
CONV3-256:  [56x56x256]    memory: 56*56*256=800K    params: (3*3*256)*256 = 589,824
POOL2:      [28x28x256]    memory: 28*28*256=200K    params: 0
CONV3-512:  [28x28x512]    memory: 28*28*512=400K    params: (3*3*256)*512 = 1,179,648
CONV3-512:  [28x28x512]    memory: 28*28*512=400K    params: (3*3*512)*512 = 2,359,296
CONV3-512:  [28x28x512]    memory: 28*28*512=400K    params: (3*3*512)*512 = 2,359,296
POOL2:      [14x14x512]    memory: 14*14*512=100K    params: 0
CONV3-512:  [14x14x512]    memory: 14*14*512=100K    params: (3*3*512)*512 = 2,359,296
CONV3-512:  [14x14x512]    memory: 14*14*512=100K    params: (3*3*512)*512 = 2,359,296
CONV3-512:  [14x14x512]    memory: 14*14*512=100K    params: (3*3*512)*512 = 2,359,296
POOL2:      [7x7x512]      memory: 7*7*512=25K       params: 0
FC:         [1x1x4096]     memory: 4096              params: 7*7*512*4096 = 102,760,448
FC:         [1x1x4096]     memory: 4096              params: 4096*4096 = 16,777,216
FC:         [1x1x1000]     memory: 1000              params: 4096*1000 = 4,096,000
TOTAL memory: 24M * 4 bytes ~= 93MB / image (only forward! ~*2 for bwd)
TOTAL params: 138M parameters
```

Campo visual (VGG)





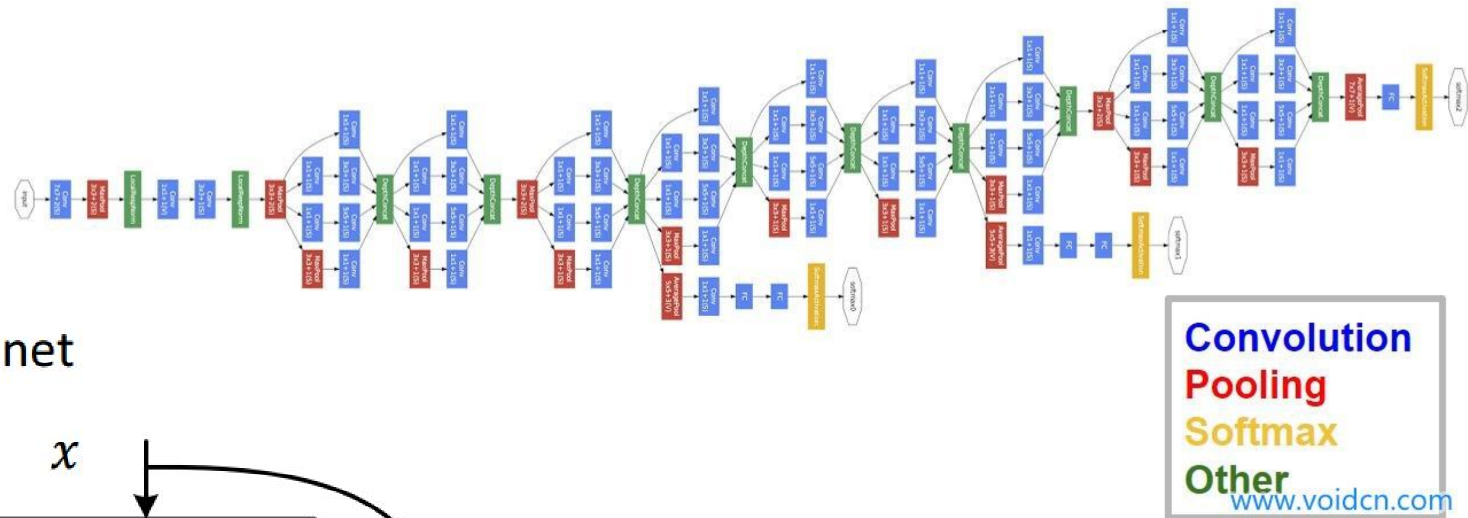
196x196

212x212

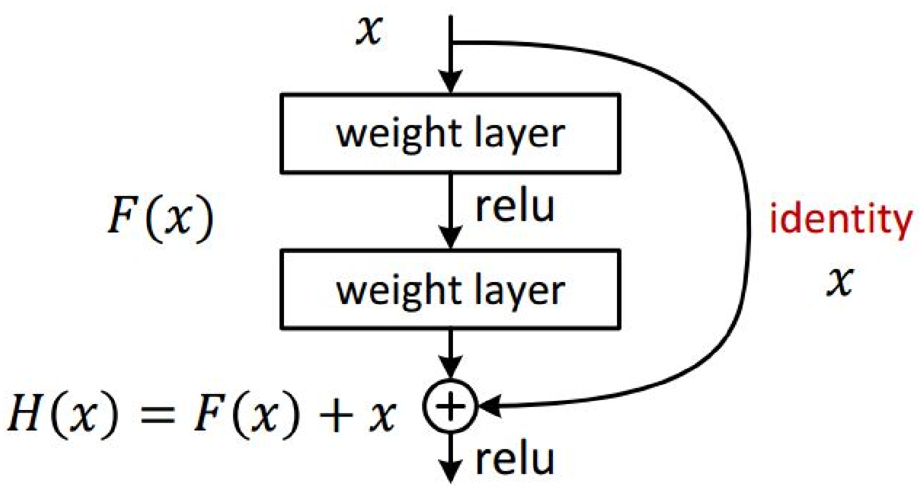
224x224 (Imagen Original)

Combinaciones más complejas

GoogLeNet



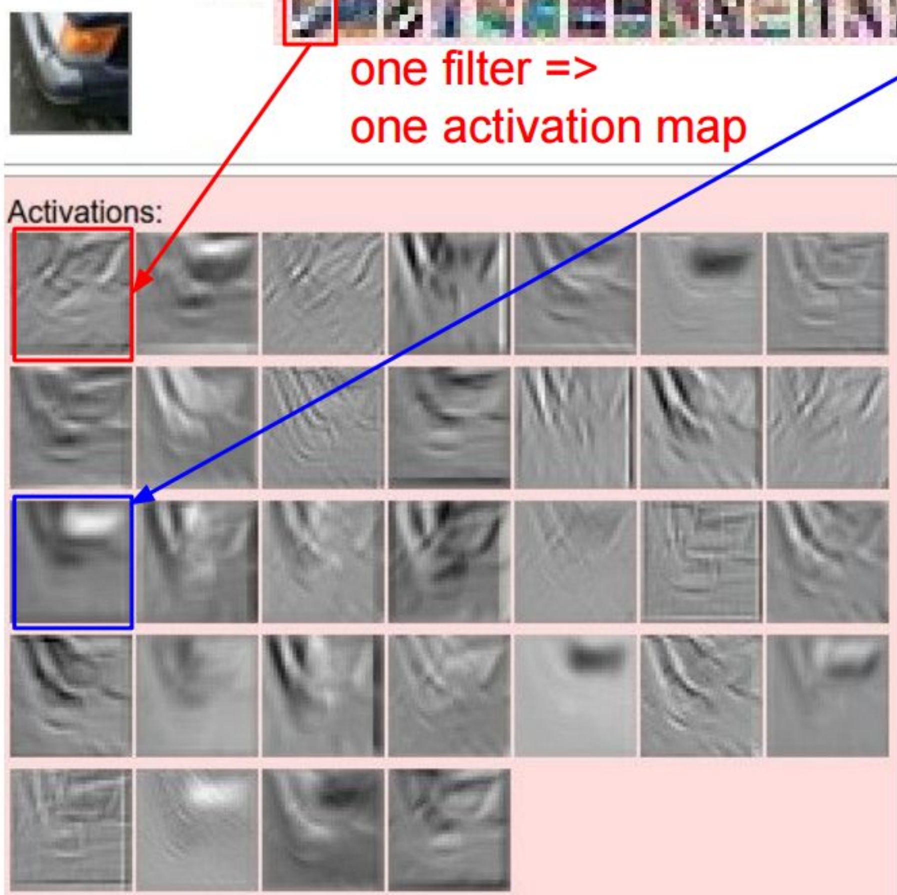
- Residual net



Entendiendo el modelo

- Ver cómo son las activaciones
- Visualizar los pesos (sólo interpretable la primera capa)
- Visualizar los parches que maximizan la activación de una dada neurona
- Visualizar el espacio de representación
- Oclusiones y optimización sobre imágenes

Mapas de activación



example 5x5 filters
(32 total)

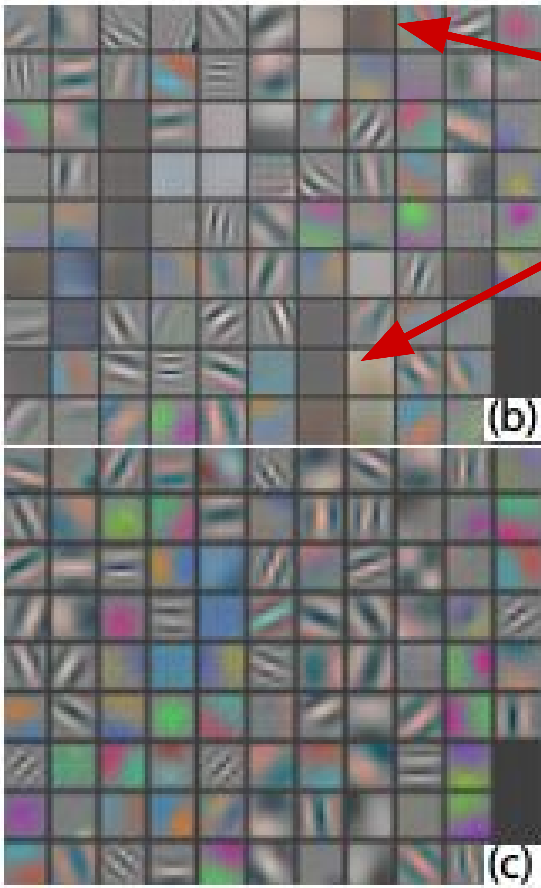
We call the layer convolutional because it is related to convolution of two signals:

$$f[x,y] * g[x,y] = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} f[n_1,n_2] \cdot g[x-n_1,y-n_2]$$

elementwise multiplication and sum of a filter and the signal (image)

http://cs231n.stanford.edu/syllabus.html

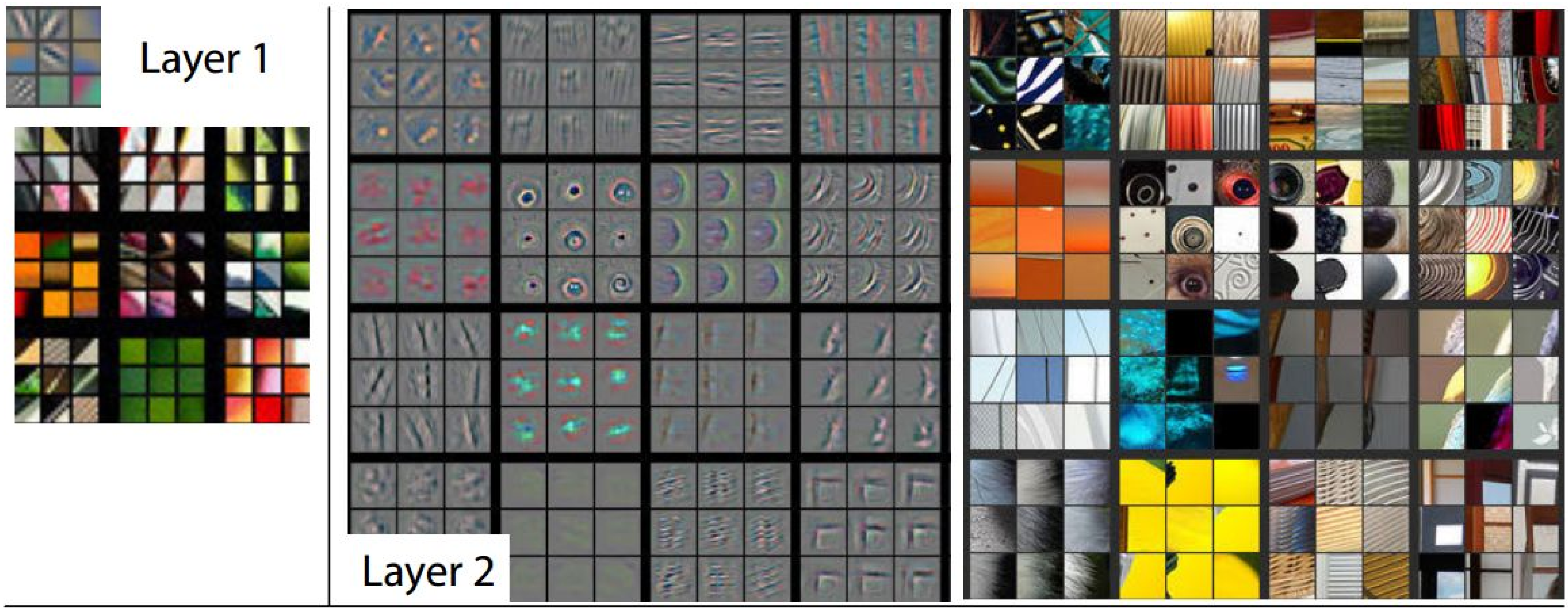
Visualizing and understanding convolutional neural networks



Viendo los filtros podemos detectar algunos problemas

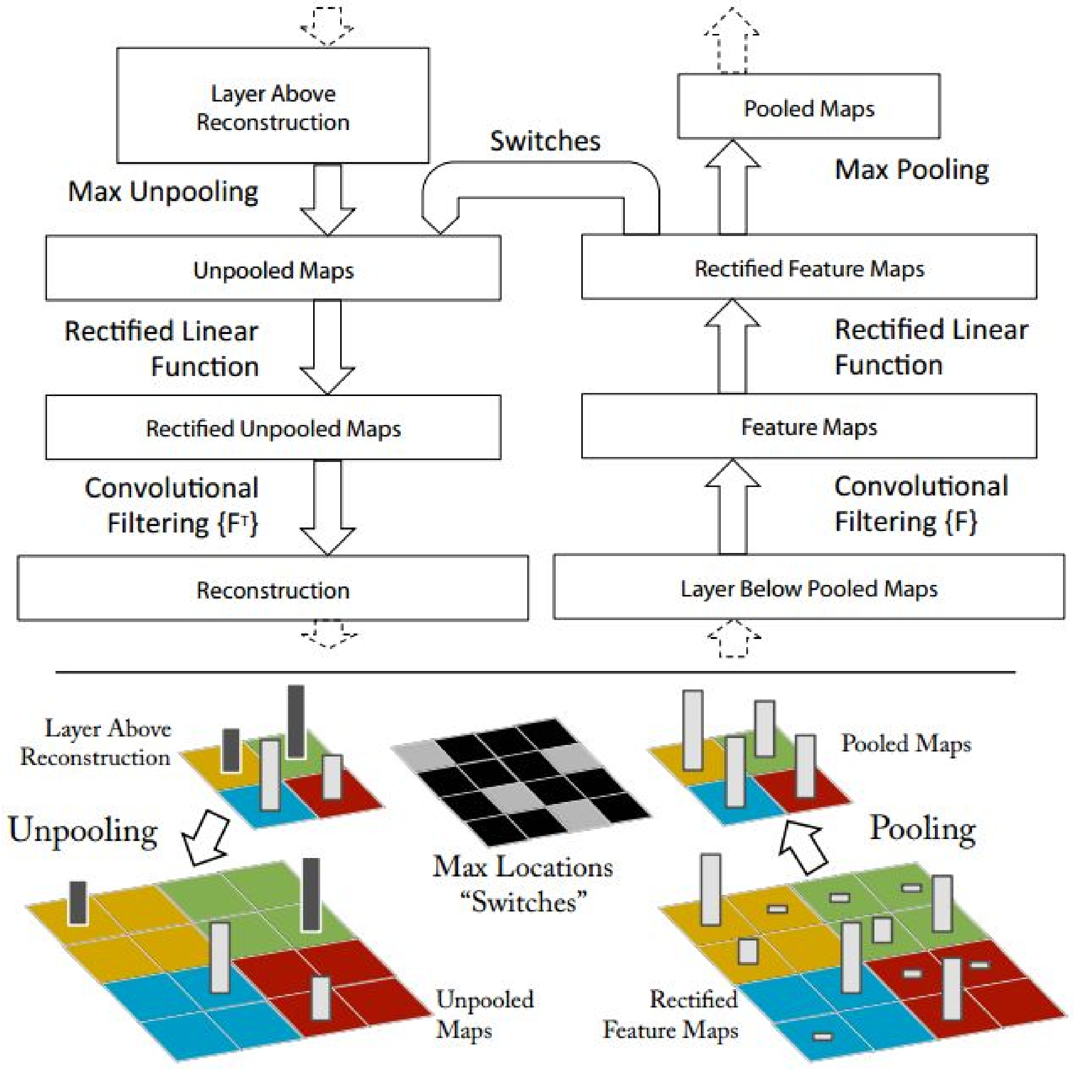
Zeiler, Matthew D, and Rob Fergus. "Visualizing and understanding convolutional neural networks." *arXiv preprint arXiv:1311.2901* (2013).

Visualizing and understanding convolutional neural networks



Zeiler, Matthew D, and Rob Fergus. "Visualizing and understanding convolutional neural networks." *arXiv preprint arXiv:1311.2901* (2013).

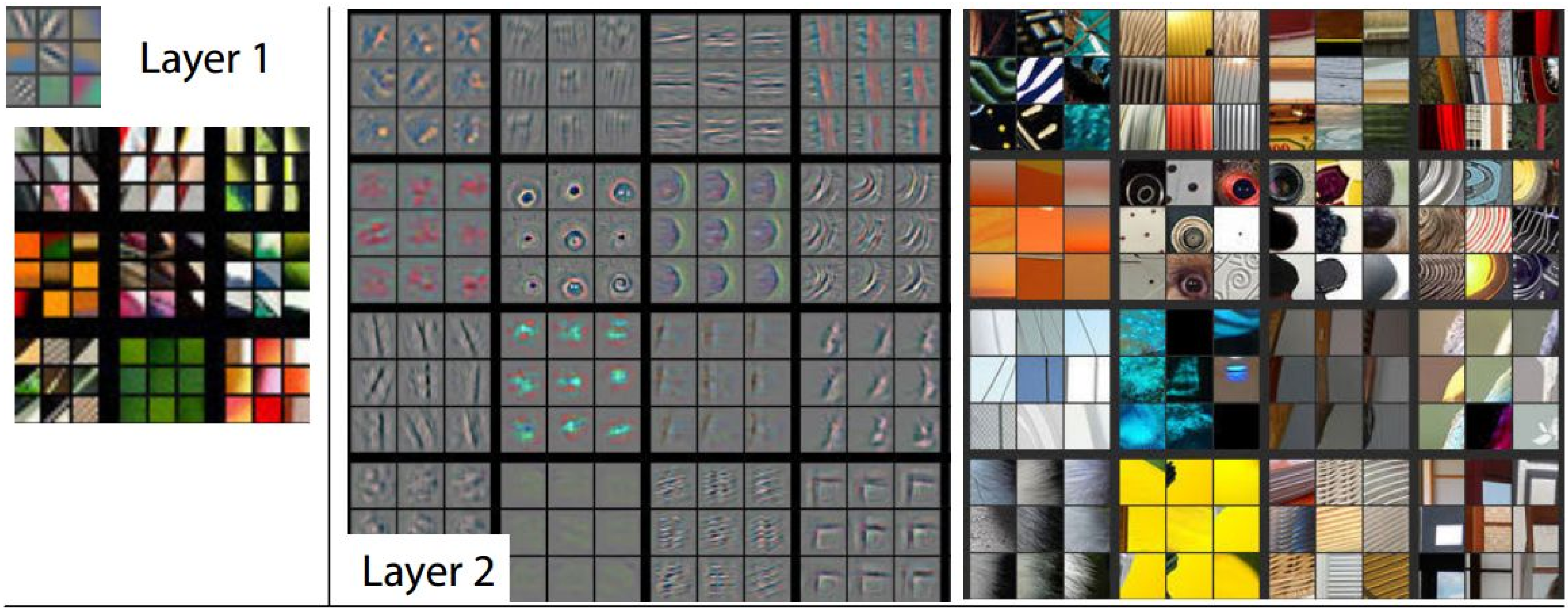
"Invertir" una red convolucional (Deconvolución)



"Invertir" una red convolucional (Deconvolución)

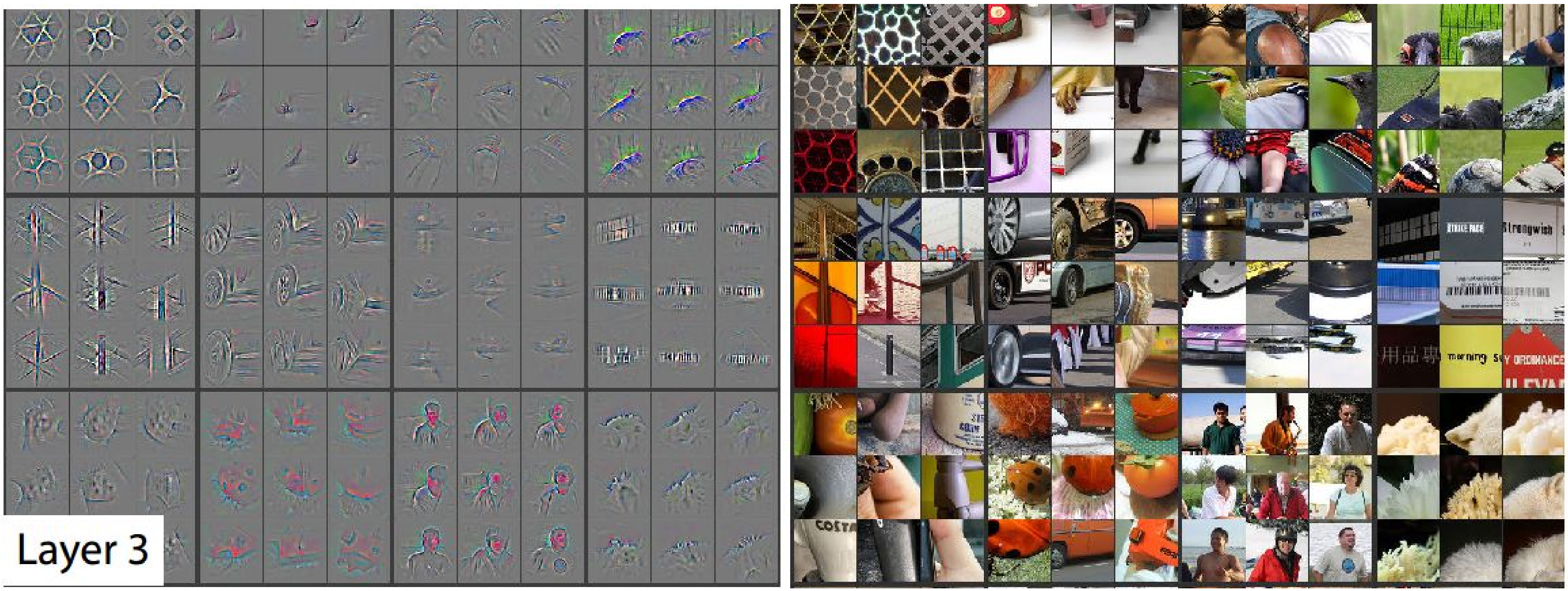
1. Presentar un ejemplo y computar las activaciones.
2. Para examinar una activación dada, poner en cero el resto de las activaciones de la capa.
3. Partiendo de ahí, aplicar sucesivamente unpool, rectify y deconv

Visualizing and understanding convolutional neural networks



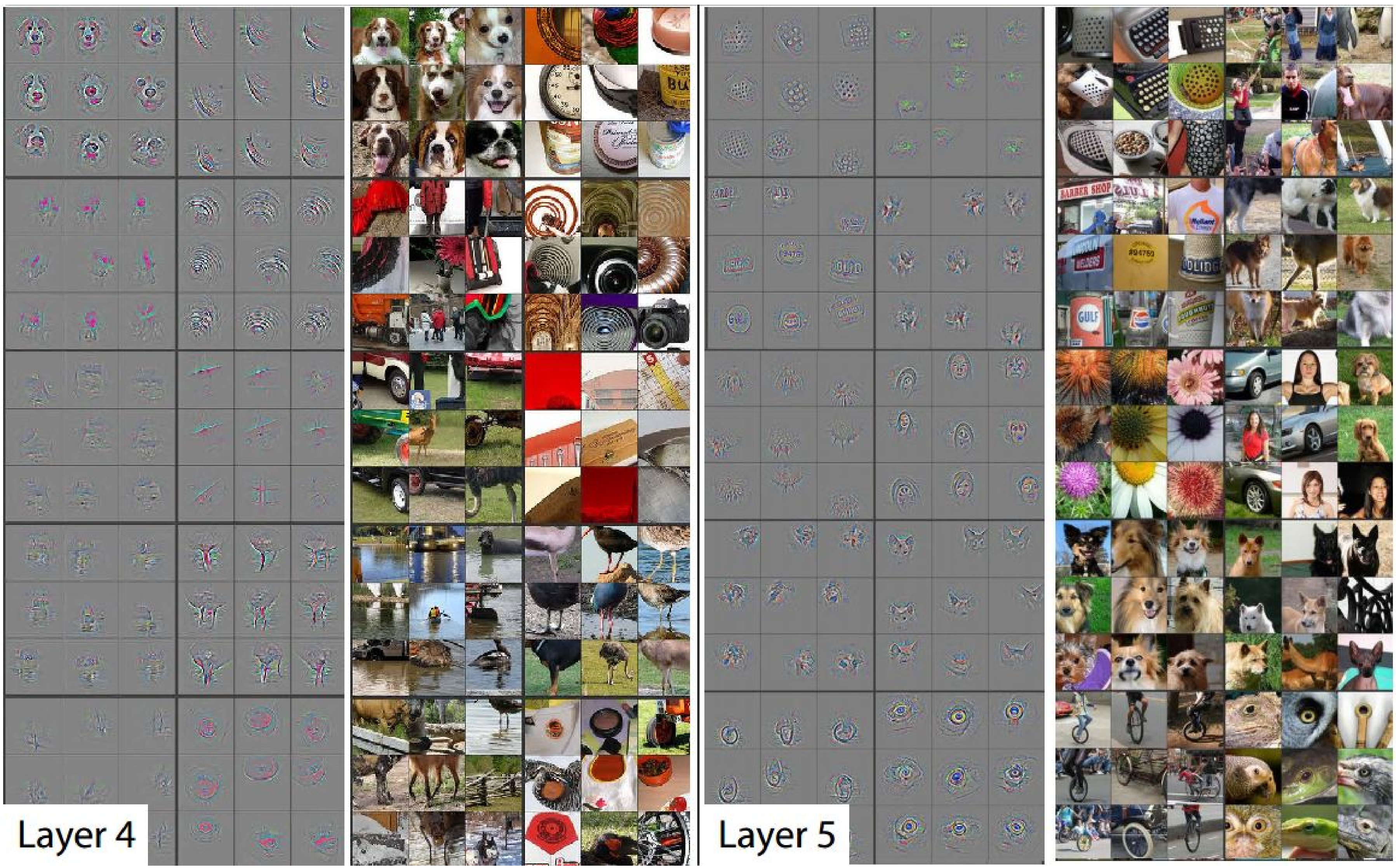
Zeiler, Matthew D, and Rob Fergus. "Visualizing and understanding convolutional neural networks." *arXiv preprint arXiv:1311.2901* (2013).

Visualizing and understanding convolutional neural networks

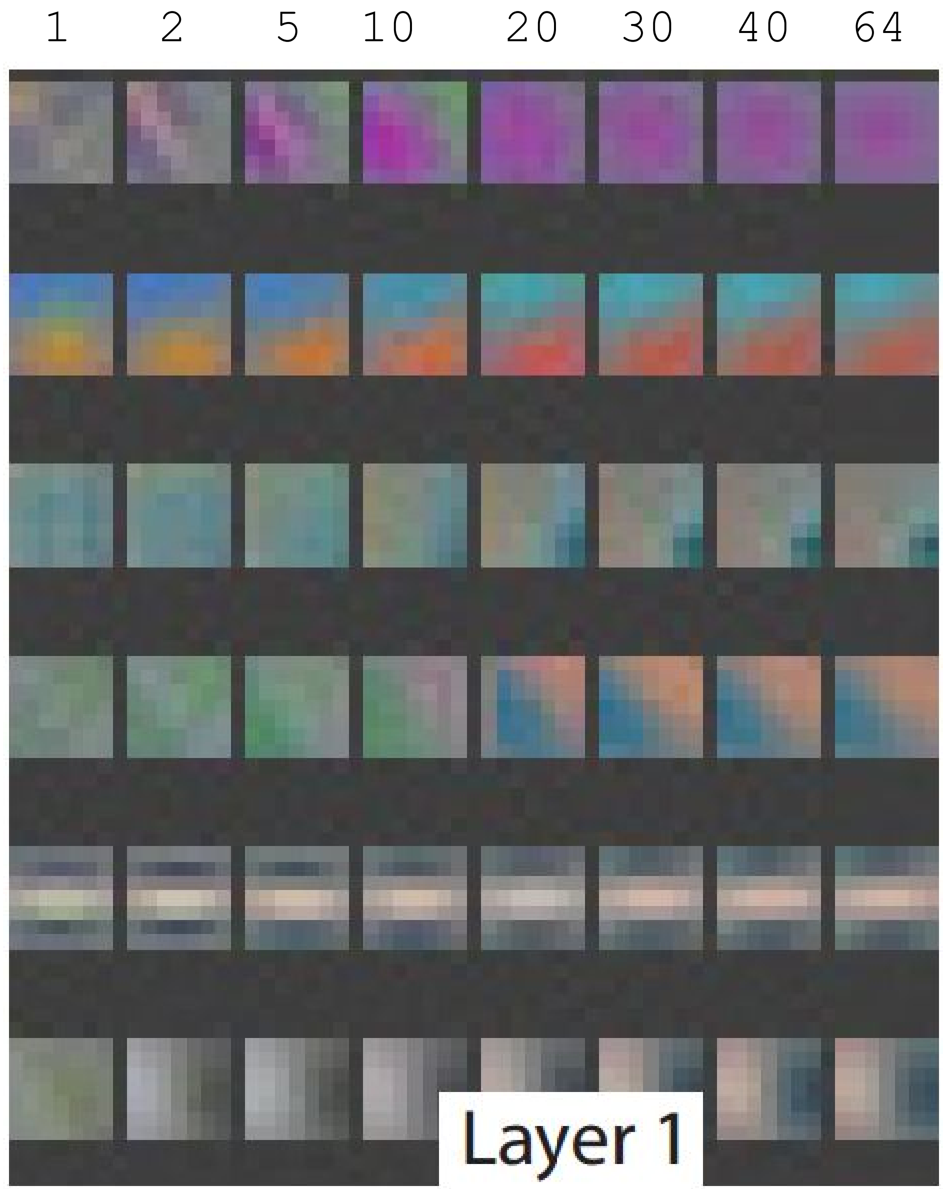


Zeiler, Matthew D, and Rob Fergus. "Visualizing and understanding convolutional neural networks." *arXiv preprint arXiv:1311.2901* (2013).

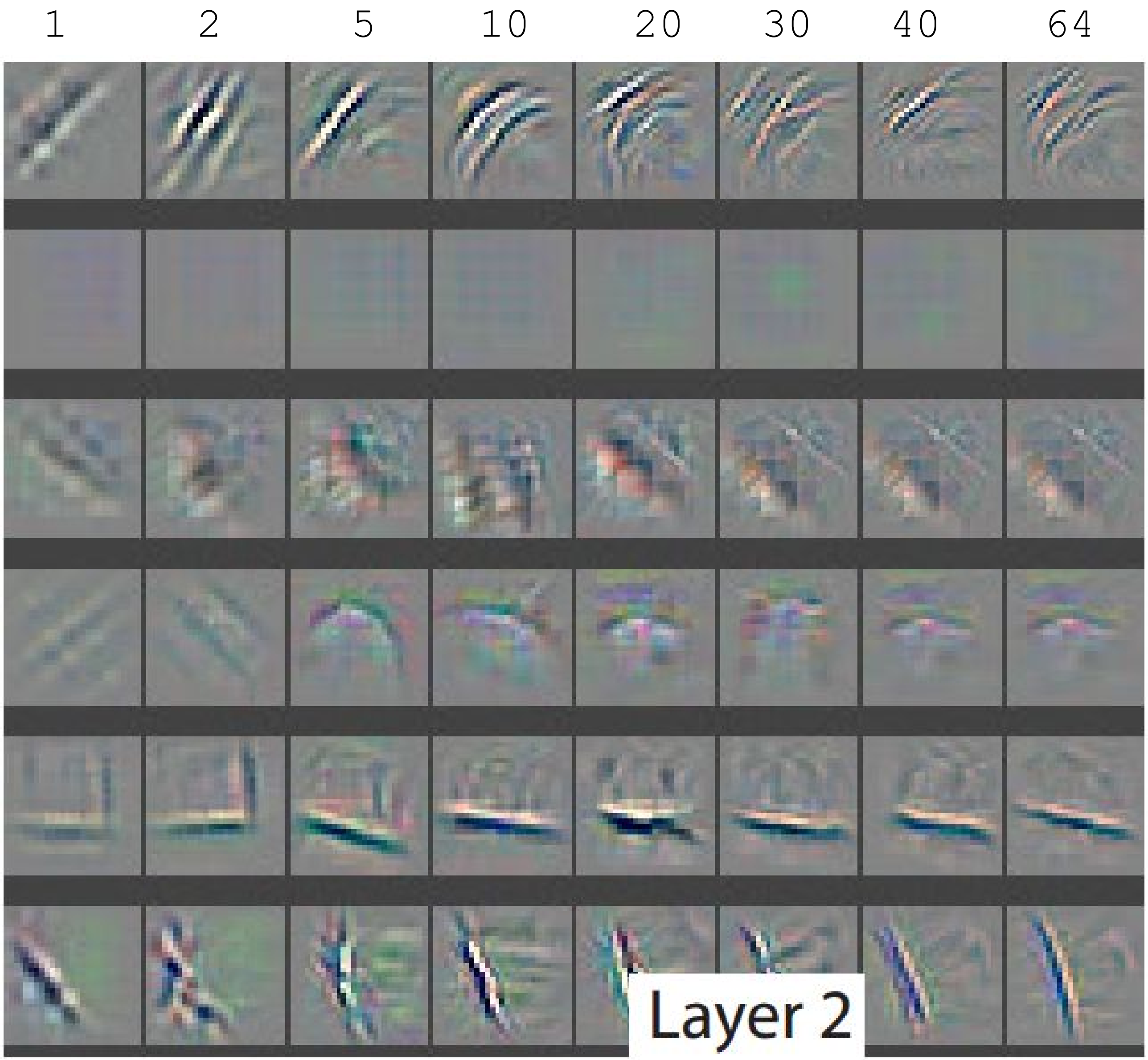
Visualizing and understanding convolutional neural networks



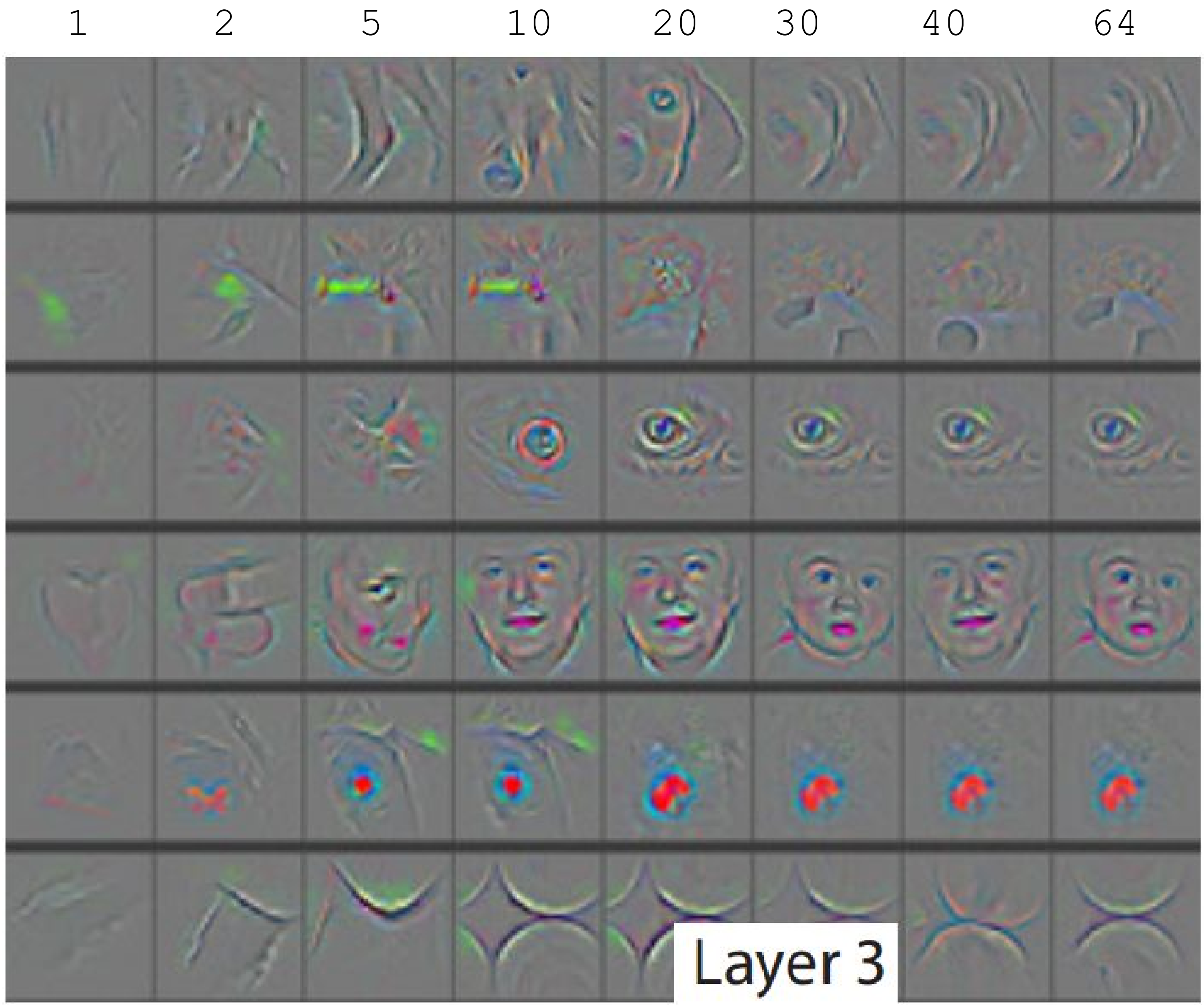
Evolución de la red con las épocas



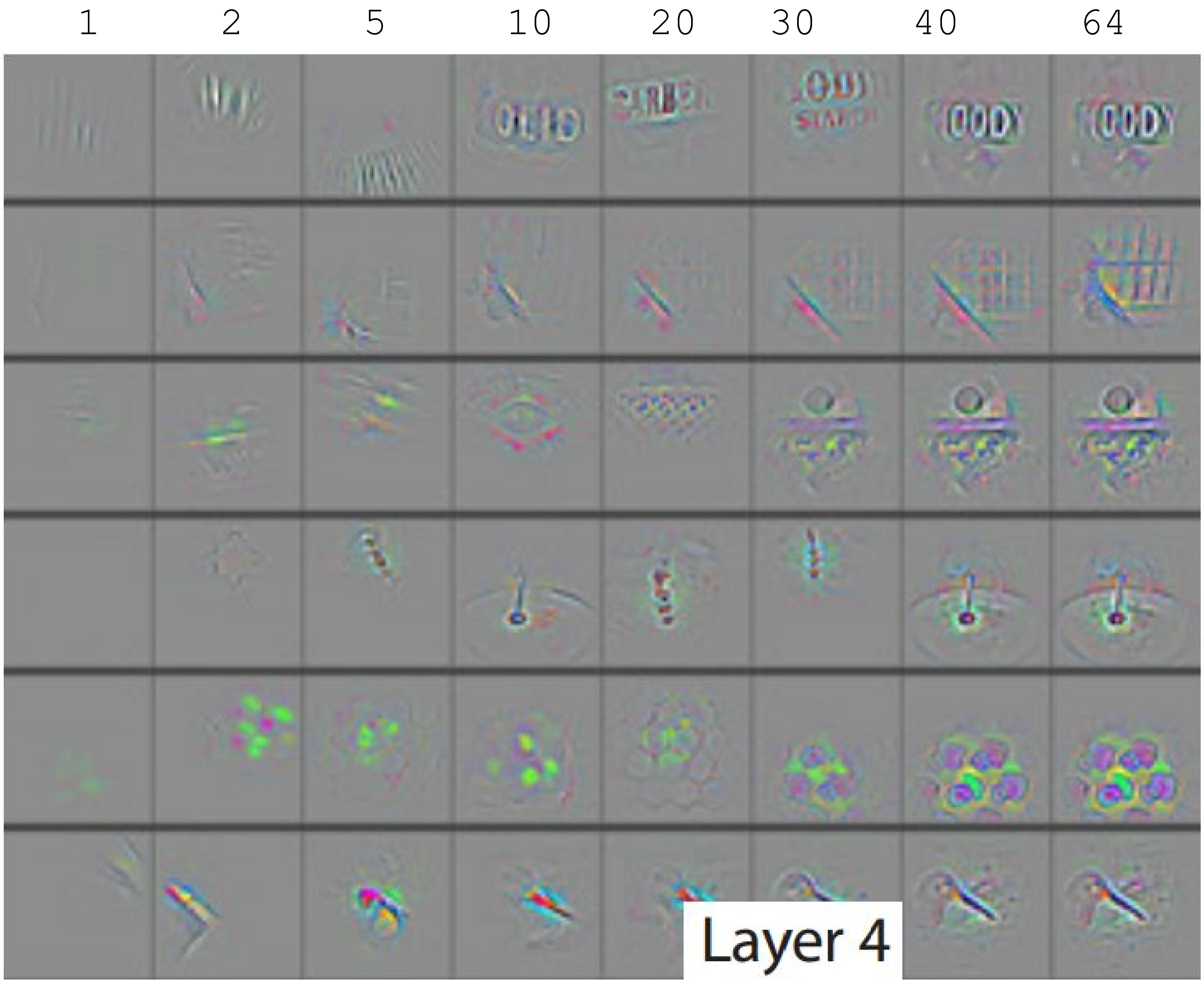
Evolución de la red con las épocas



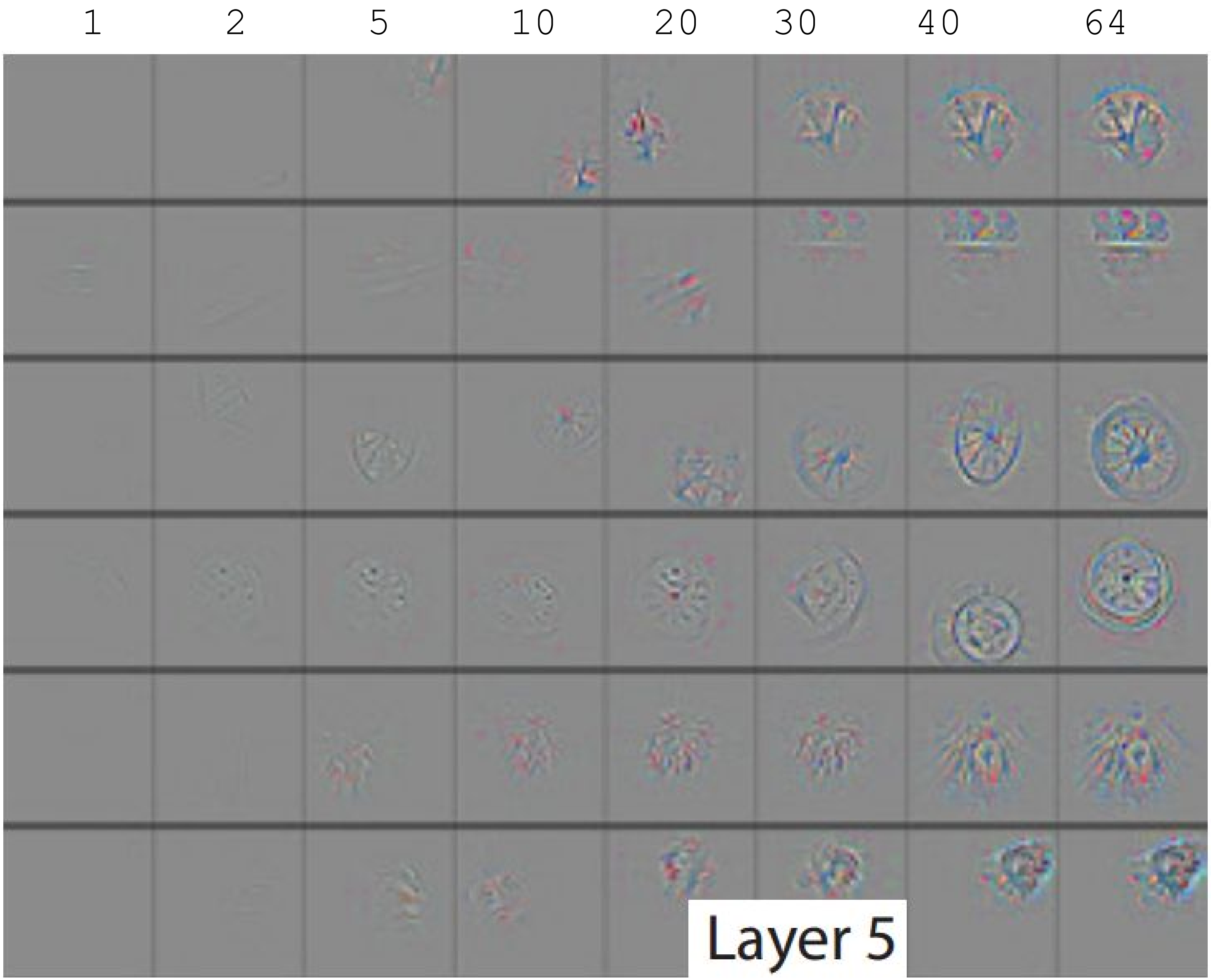
Evolución de la red con las épocas



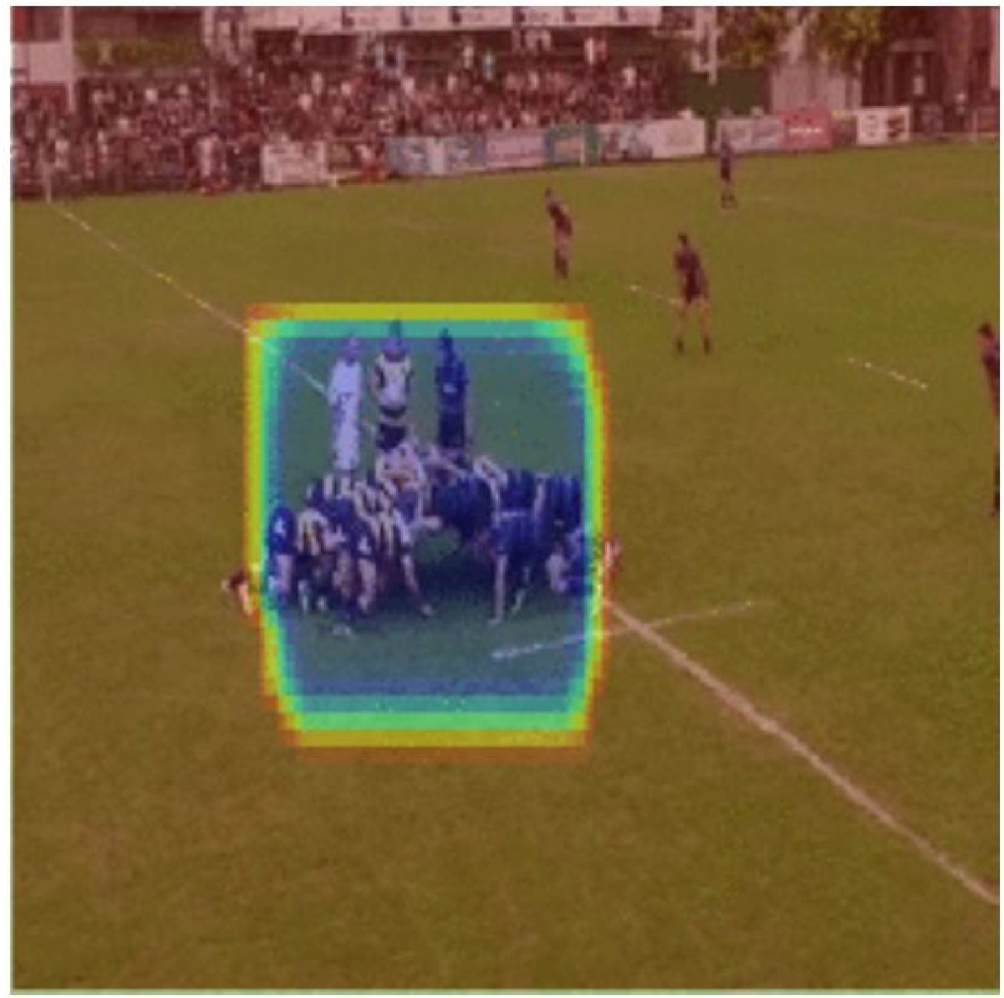
Evolución de la red con las épocas



Evolución de la red con las épocas

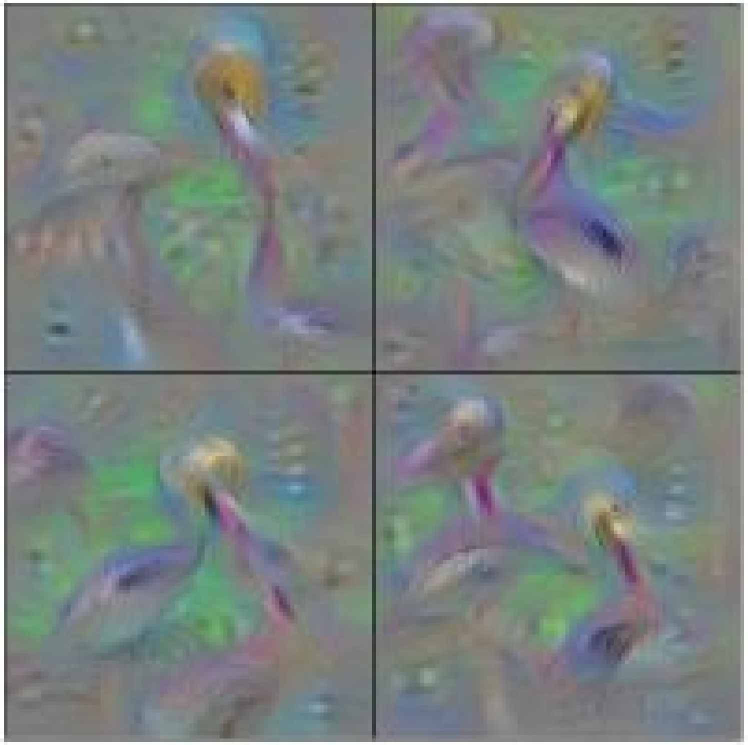


Visualización por oclusión



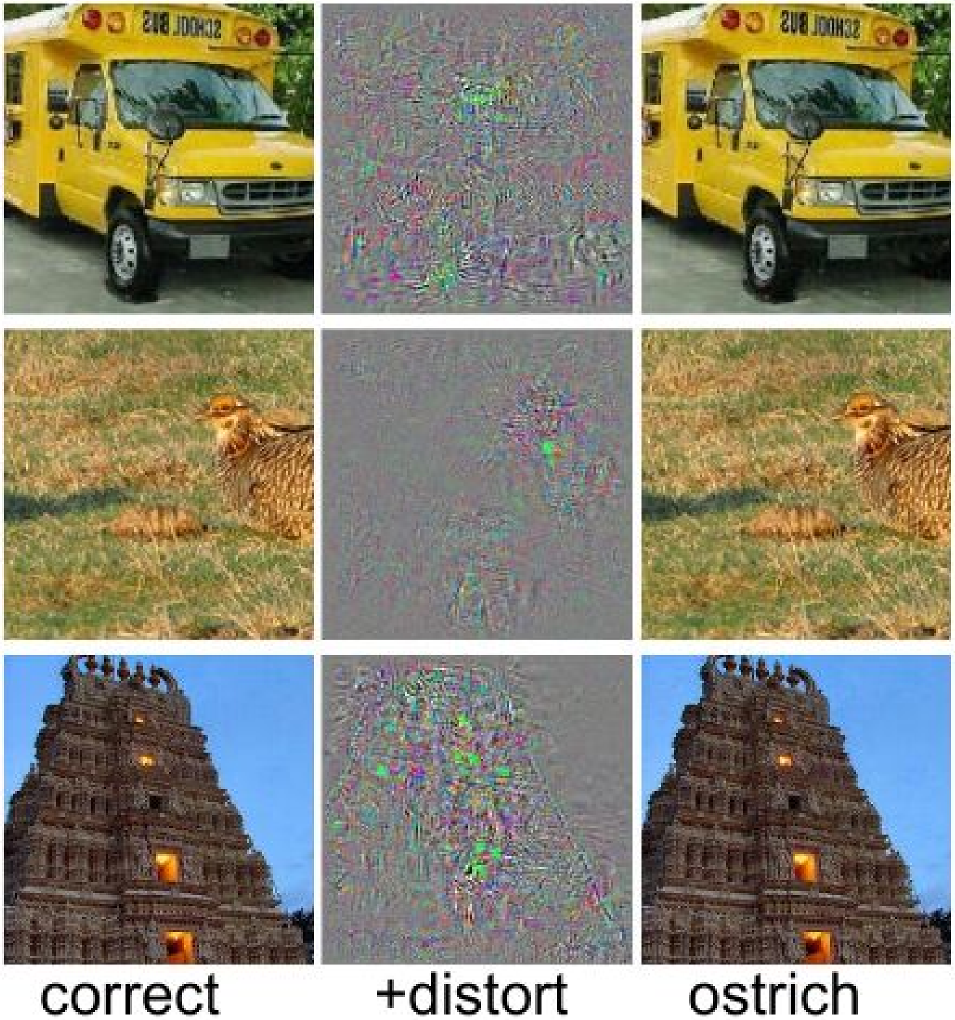
Optimización sobre imágenes

Partiendo de una imagen formada por ceros, maximizar la probabilidad de la clase “Pelicano”



Optimización sobre imágenes

Partiendo de una imagen bien clasificada, hacer que se clasifique como “Avestruz”



Trabajo Práctico 3

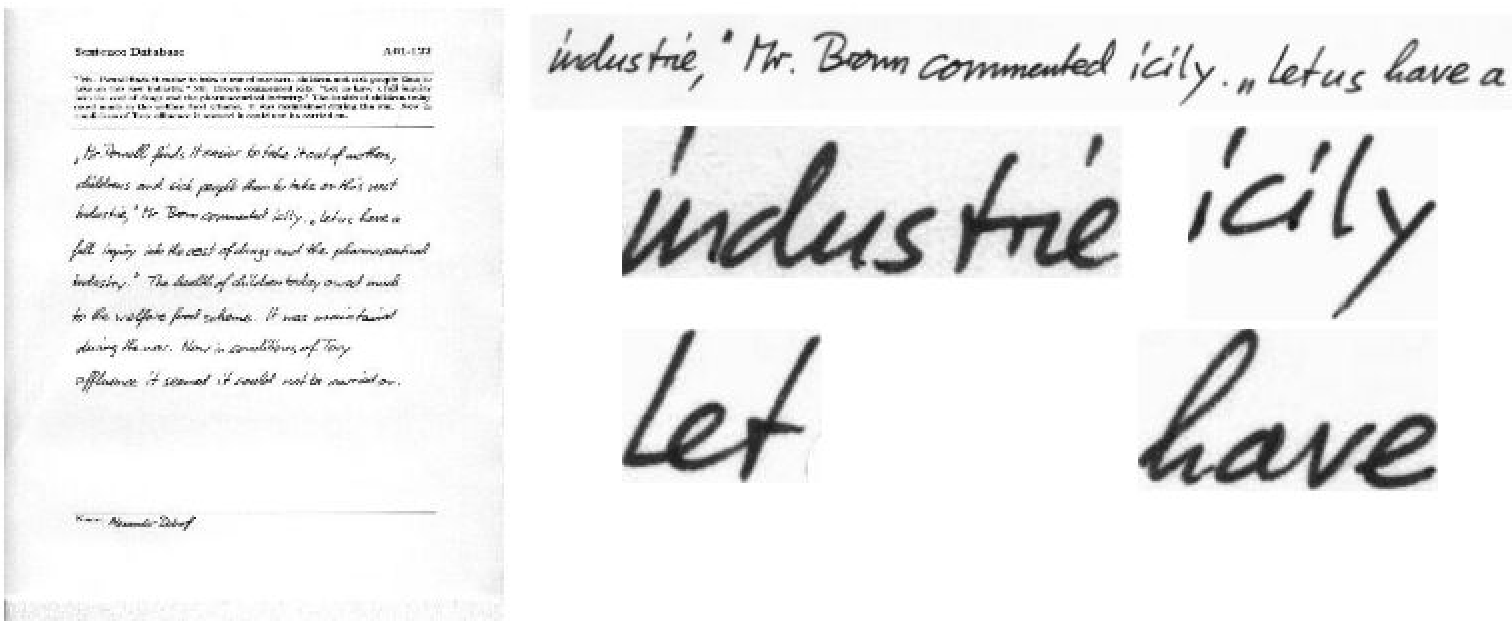
Observaciones sobre el dataset

IAM Handwriting Database

http://www.fki.inf.unibe.ch/databases/iam-handwriting-database

- forms of handwritten English text
- year 2002
- were scanned at a resolution of 300dpi
- PNG images with 256 gray levels.

The figure below provides samples of a complete form, a text line and some extracted words.



IAM Handwriting Database

- 657 writers contributed samples of their handwriting
- 1'539 pages of scanned text
- 5'685 isolated and labeled sentences
- 13'353 isolated and labeled text lines
- 115'320 isolated and labeled words

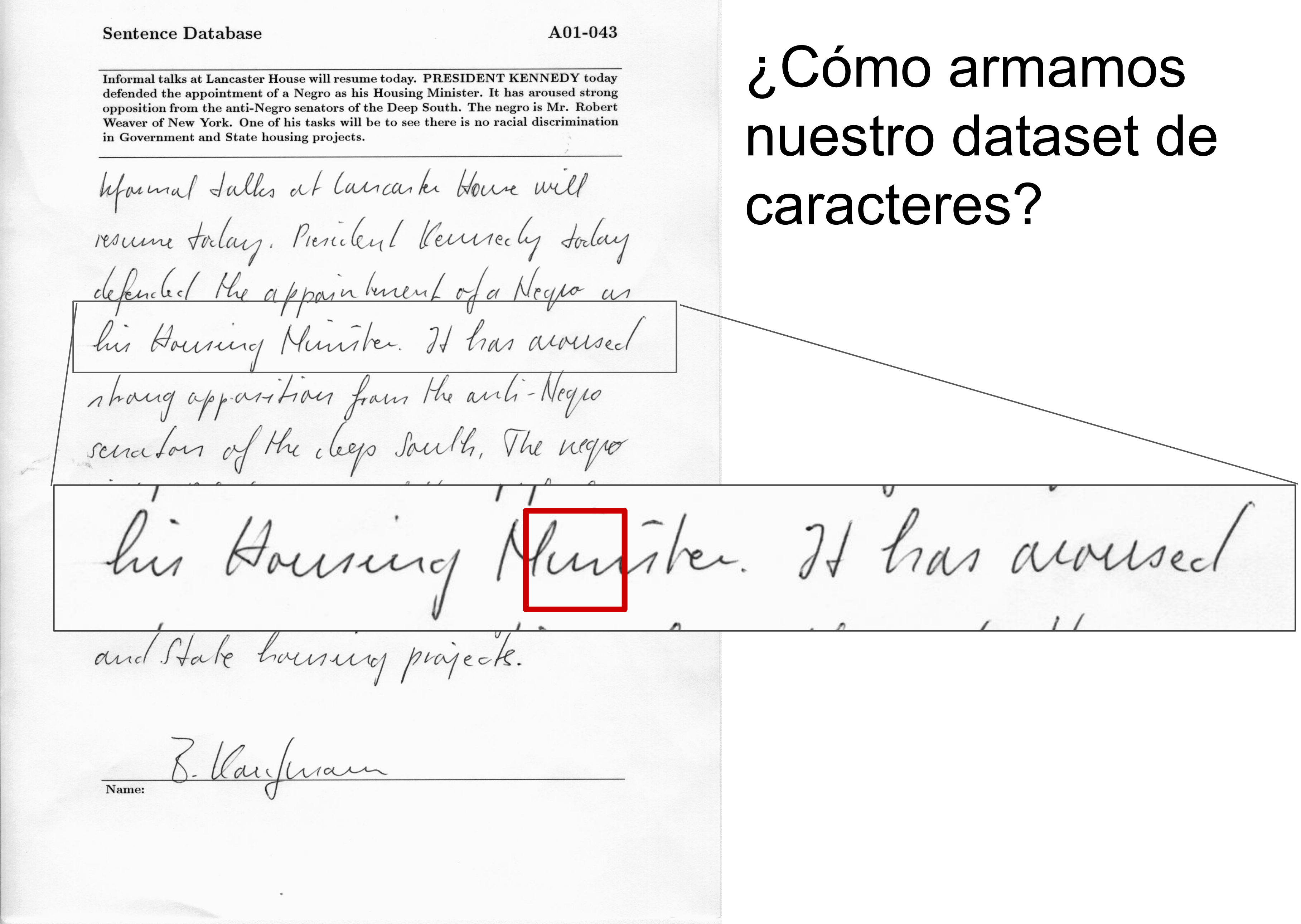
No presenta segmentación al nivel de caracteres

Large Writer Independent Text Line Recognition Task

This task consists of a total number of 9'862 text lines. It provides one training, one testing, and two validation sets. The text lines of all data sets are mutually exclusive, thus each writer has contributed to one set only.

Train	6161	283
Validation 1	900	46
Validation 2	940	43
Test	1861	128

¿Cómo armamos nuestro dataset de caracteres?


Sentence Database A01-043

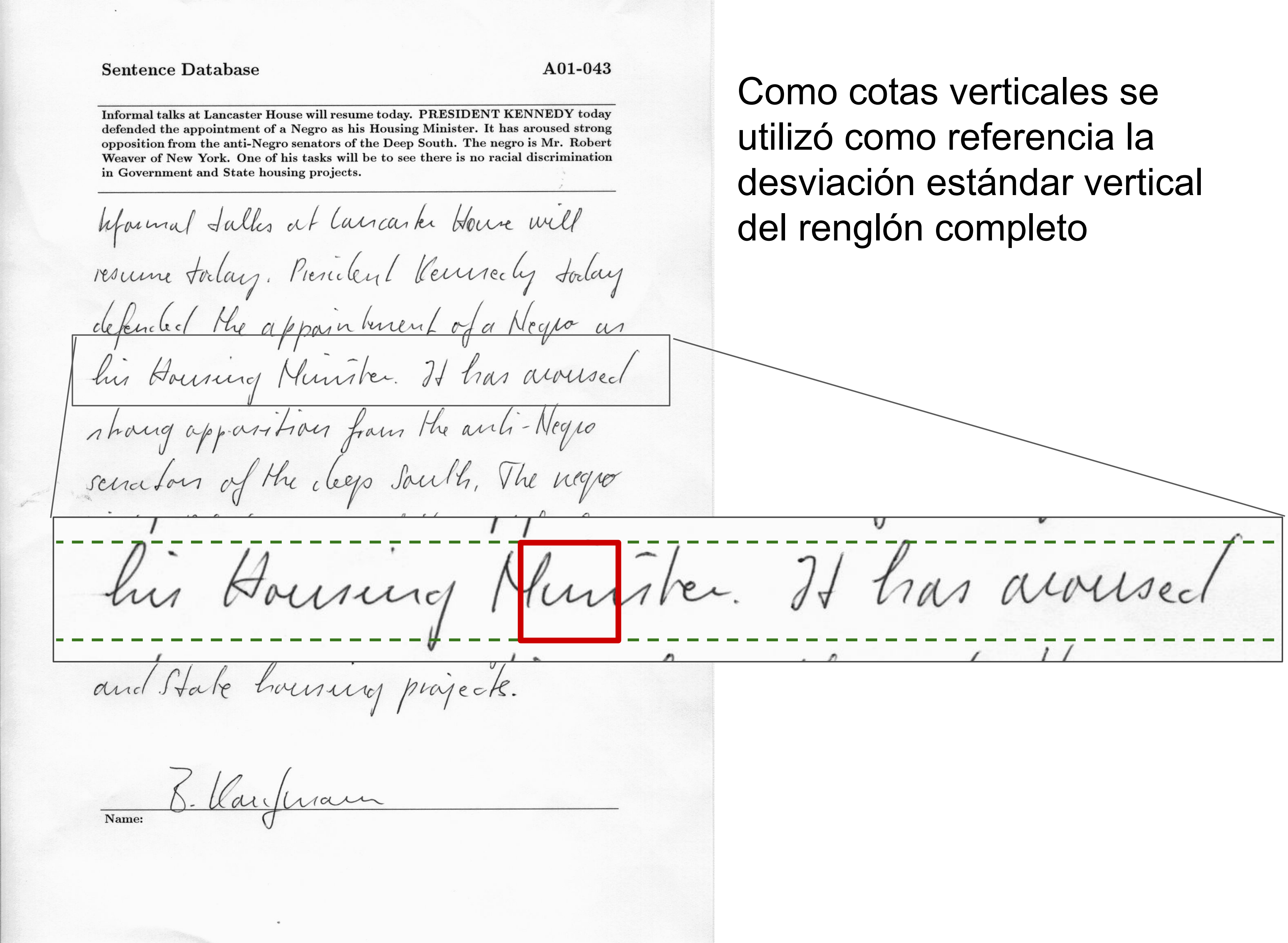
Informal talks at Lancaster House will resume today. PRESIDENT KENNEDY today defended the appointment of a Negro as his Housing Minister. It has aroused strong opposition from the anti-Negro senators of the Deep South. The negro is Mr. Robert Weaver of New York. One of his tasks will be to see there is no racial discrimination in Government and State housing projects.

Informal talks at Lancaster House will
resume today. President Kennedy today
defended the appointment of a Negro as
his Housing Minister. It has aroused
strong opposition from the anti-Negro
senators of the deep South. The negro

his Housing Minister. It has aroused

and State housing projects.

Name: B. Kaufmann


Sentence Database A01-043

Informal talks at Lancaster House will resume today. PRESIDENT KENNEDY today defended the appointment of a Negro as his Housing Minister. It has aroused strong opposition from the anti-Negro senators of the Deep South. The negro is Mr. Robert Weaver of New York. One of his tasks will be to see there is no racial discrimination in Government and State housing projects.

Informal talks at Lancaster House will
resume today. President Kennedy today
defended the appointment of a Negro as
his Housing Minister. It has aroused
strong opposition from the anti-Negro
senators of the deep South. The negro

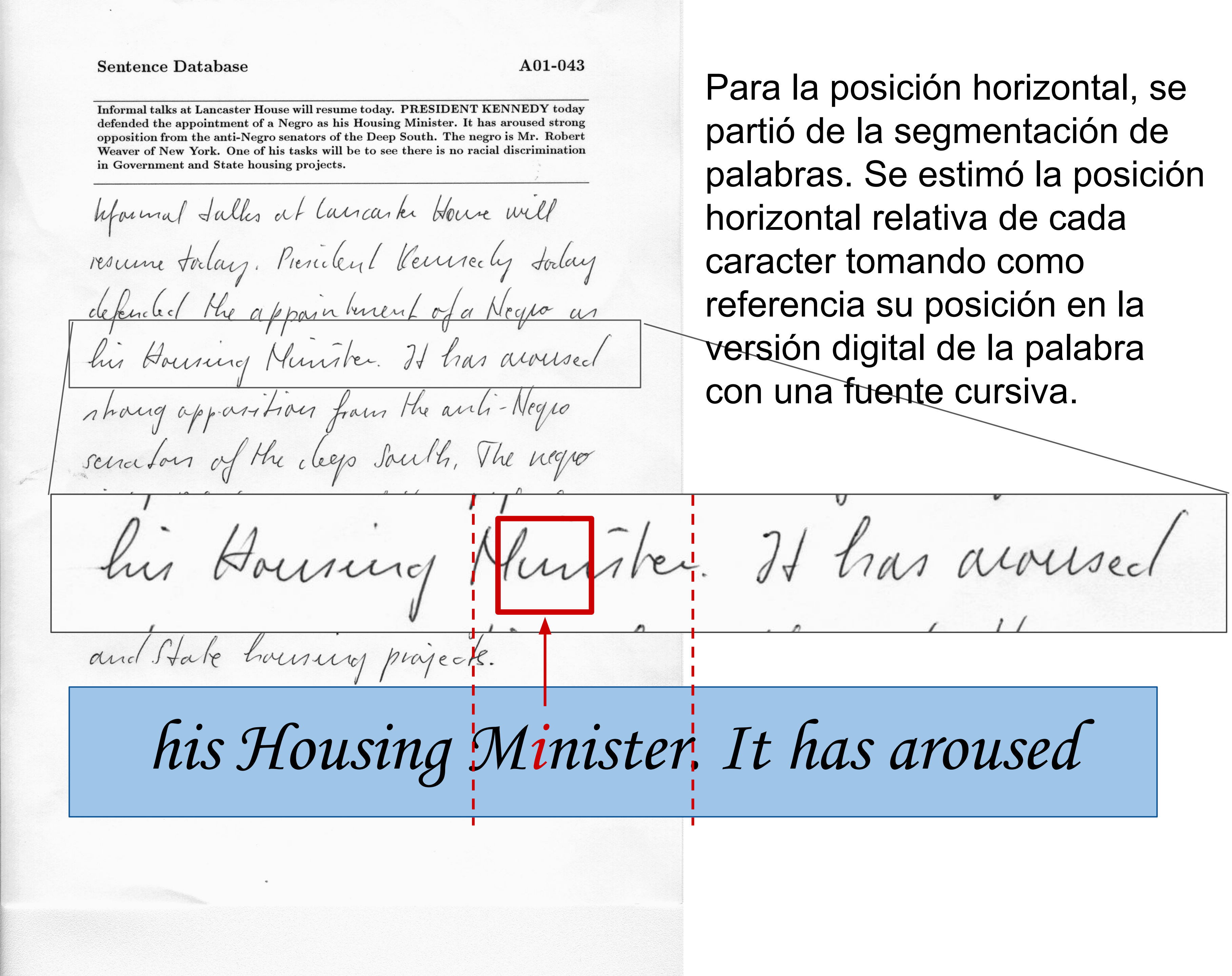
and State housing projects.

Name:

his Housing Minister. It has aroused

Como cotas verticales se utilizó como referencia la desviación estándar vertical del renglón completo

Para la posición horizontal, se partió de la segmentación de palabras. Se estimó la posición horizontal relativa de cada caracter tomando como referencia su posición en la versión digital de la palabra con una fuente cursiva.



Muestras del dataset de caracteres obtenido

x (equis minúscula)

8 (ocho)

" (comillas)

l (l minúscula)