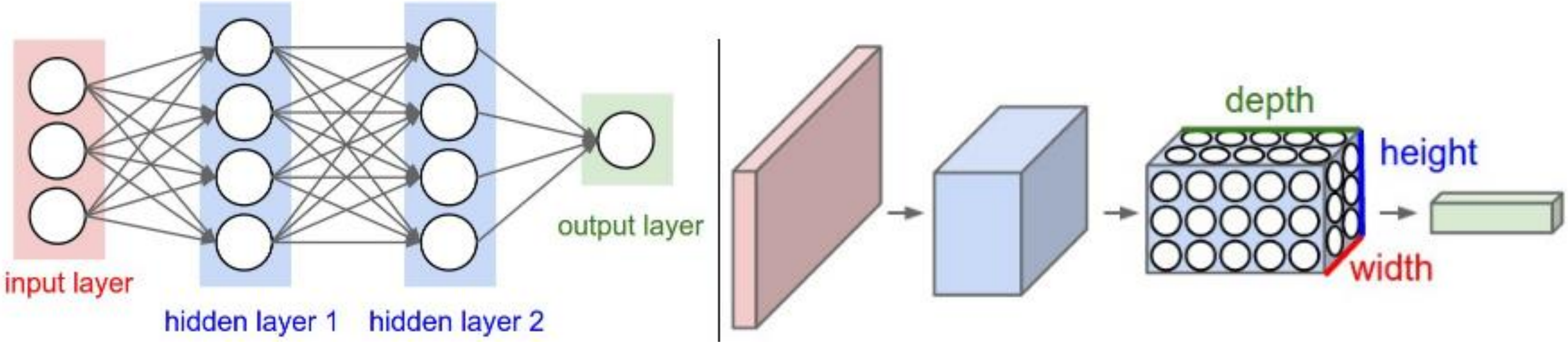
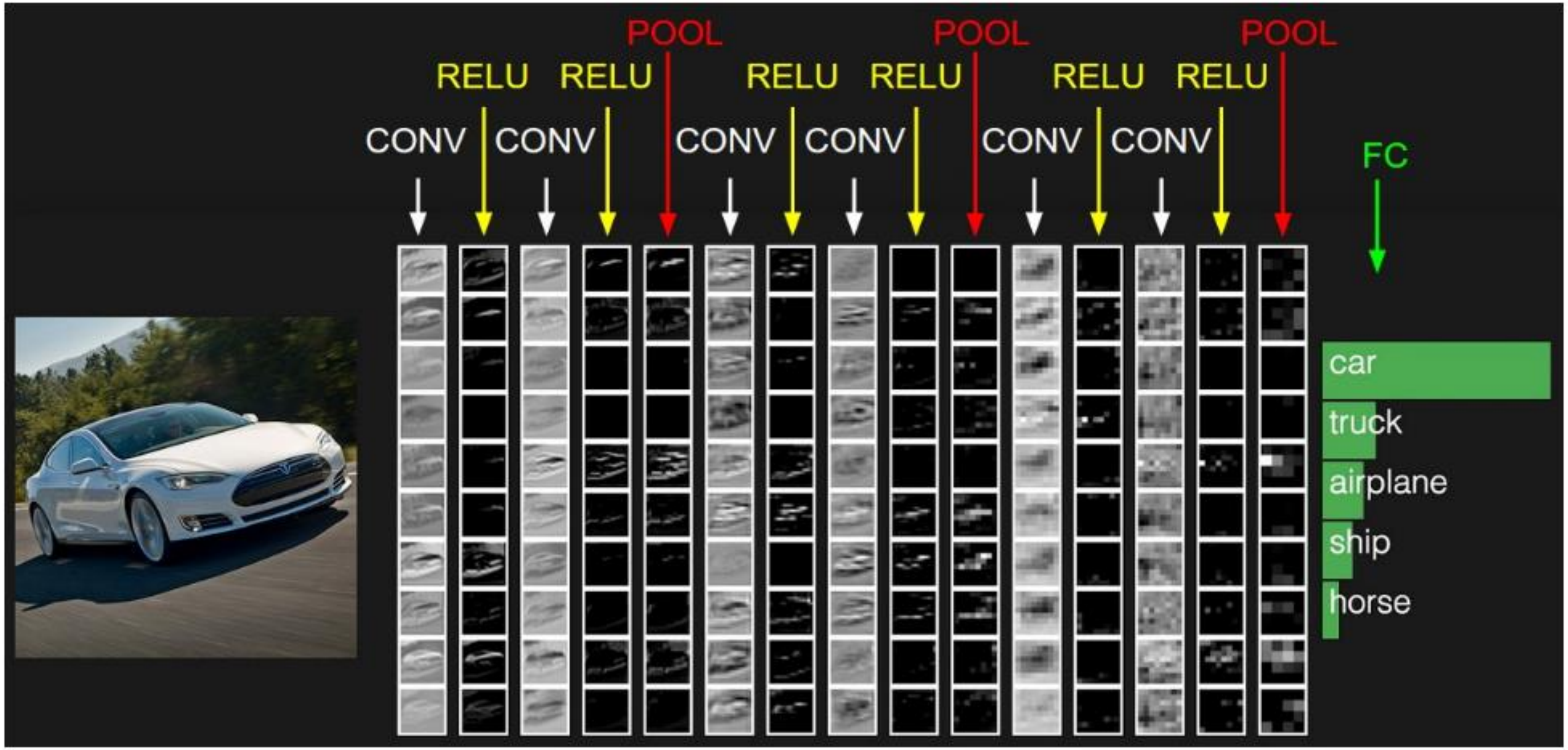
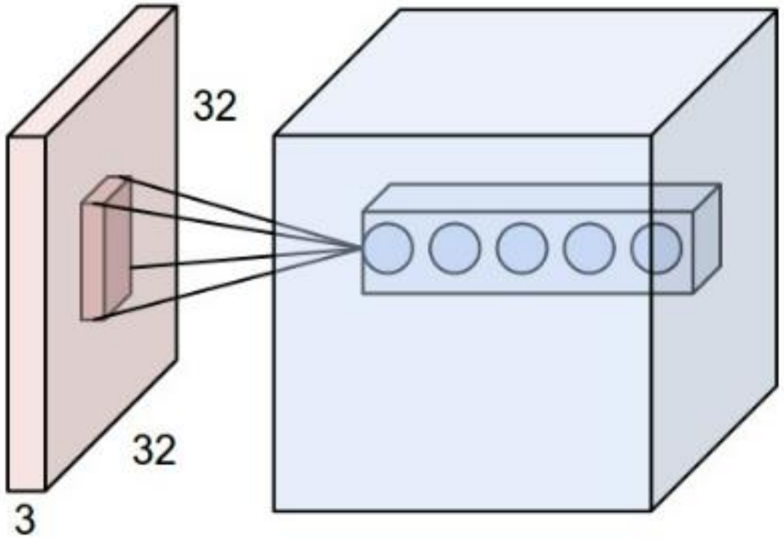
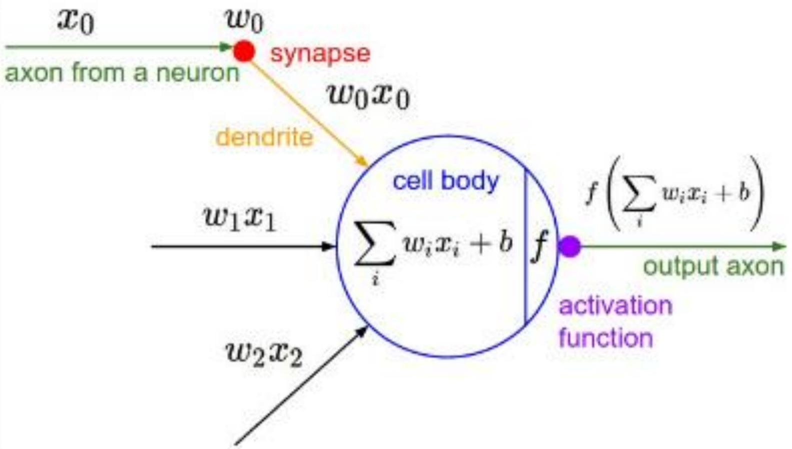

input layer
hidden layer 1
hidden layer 2
output layer
depth
height
width

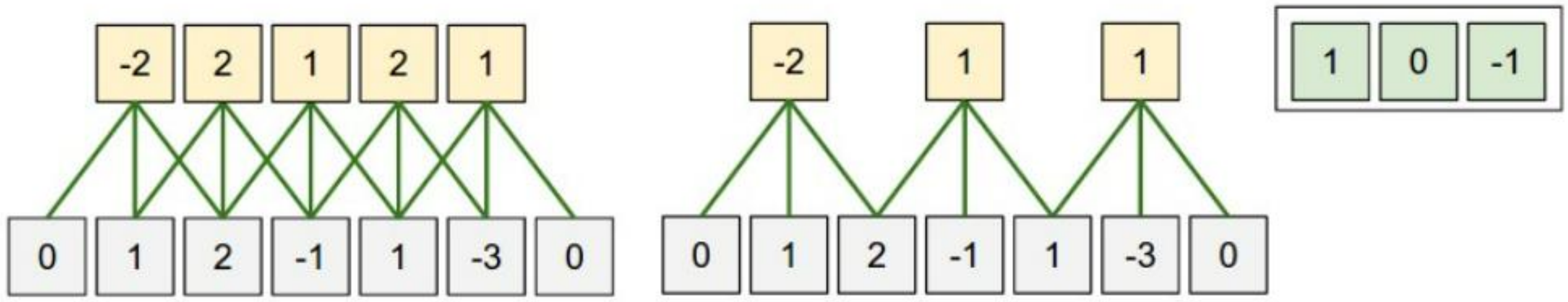

POOL
POOL
POOL
RELU
RELU
RELU
RELU
RELU
RELU
CONV
CONV
CONV
CONV
CONV
CONV
FC
car
truck
airplane
ship
horse

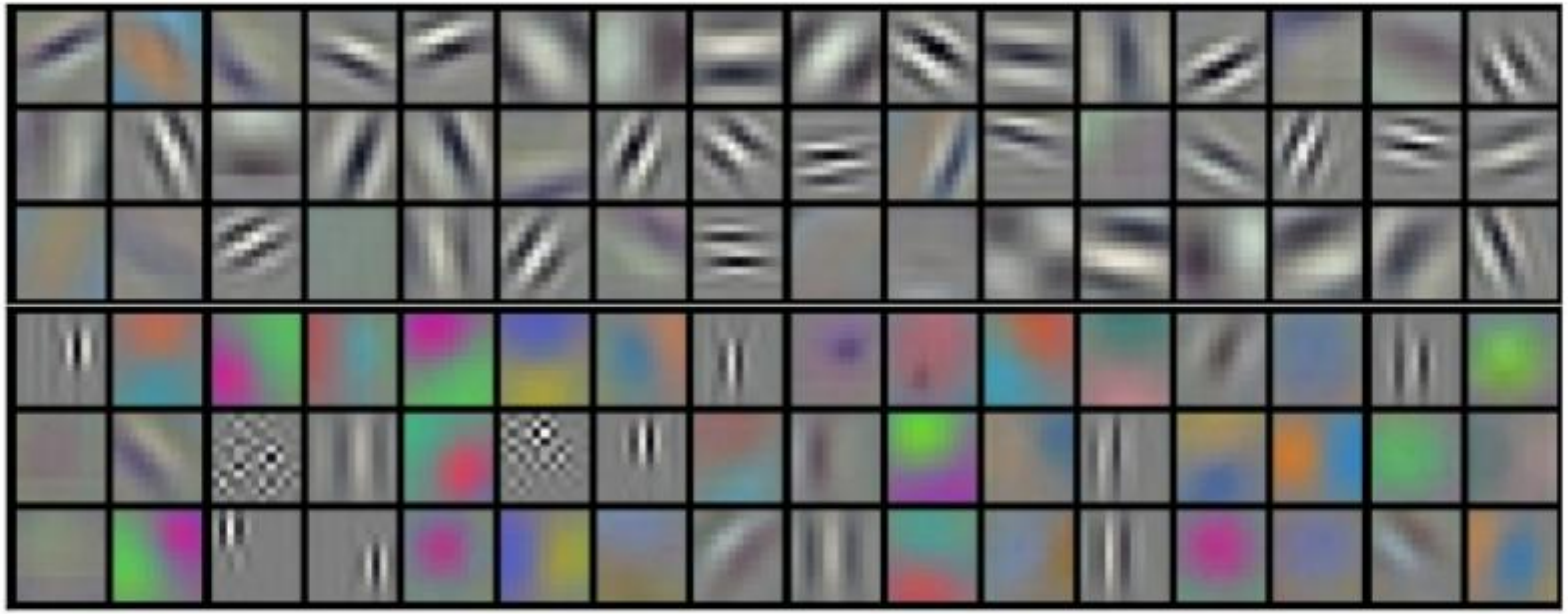

32
32
3


x_0
w_0
synapse
axon from a neuron
$w_0 x_0$
dendrite
cell body
$f\left(\sum_i w_i x_i + b\right)$
$w_1 x_1$
$\sum_i w_i x_i + b$
f
output axon
activation
function
$w_2 x_2$

Spatial arrangement

- $(W-F+2P)/S+1$, W - the input volume size, F - the receptive field size of the Conv Layer neurons, S - the stride with which they are applied, P – the amount of zero padding used on the border



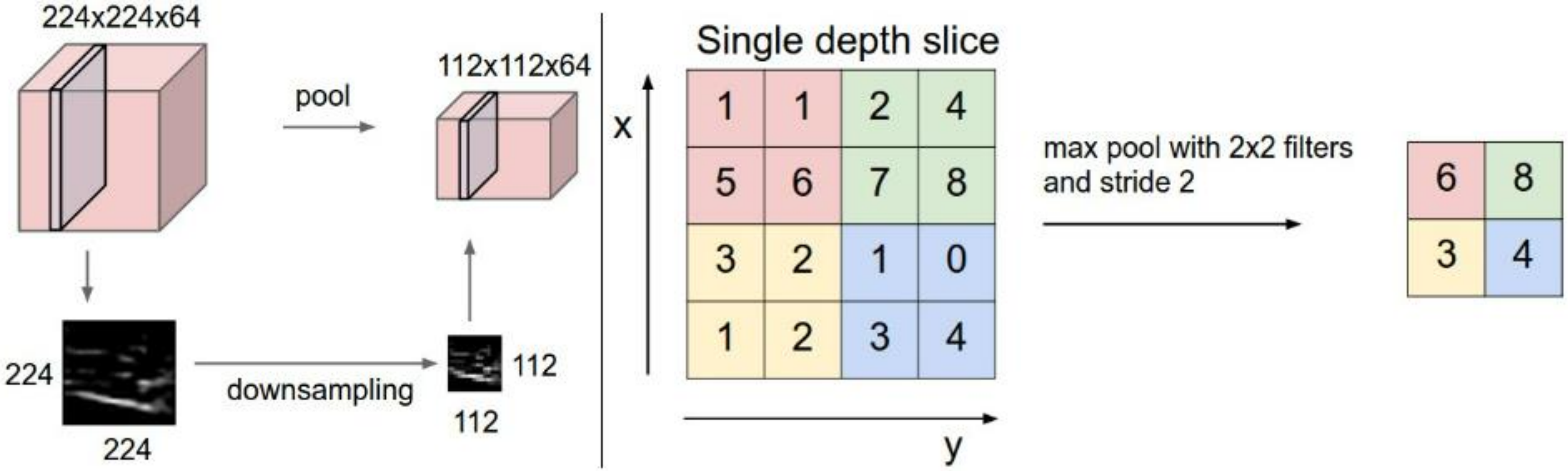


Conv Layer

- Accepts a volume of size $W1 \times H1 \times D1$
- Requires four hyperparameters:
  - Number of filters K,
  - their spatial extent F,
  - the stride S,
  - the amount of zero padding P.
- Produces a volume of size $W2 \times H2 \times D2$ where:
  - $W2=(W1-F+2P)/S+1$
  - $H2=(H1-F+2P)/S+1$ (i.e. width and height are computed equally by symmetry)
  - $D2=K$
- With parameter sharing, it introduces $F \cdot F \cdot D1$ weights per filter, for a total of $(F \cdot F \cdot D1) \cdot K$ weights and K biases.
- In the output volume, the d-th depth slice (of size $W2 \times H2$) is the result of performing a valid convolution of the d-th filter over the input volume with a stride of S, and then offset by d-th bias.

Pooling Layer

- Accepts a volume of size $W1 \times H1 \times D1$
- Requires two hyperparameters:
  - their spatial extent F,
  - the stride S,
- Produces a volume of size $W2 \times H2 \times D2$ where:
  - $W2=(W1-F)/S+1$
  - $H2=(H1-F)/S+1$
  - $D2=D1$
- Introduces zero parameters since it computes a fixed function of the input
- For Pooling layers, it is not common to pad the input using zero-padding


224x224x64
pool
112x112x64
224
downsampling
112
112
224
Single depth slice
x
1 1 2 4
5 6 7 8
3 2 1 0
1 2 3 4
y
max pool with 2x2 filters
and stride 2
6 8
3 4