

Classical CV

Machine Learning

Denis Litvinov

December 8, 2022

Table of Contents

1 Image Features

2 Filters

3 Convolution

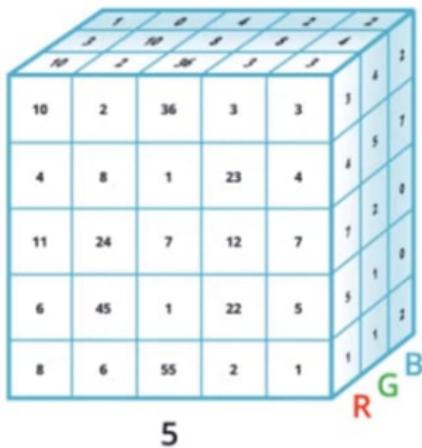
Features

RGB Image

$5 \times 5 \times 3$



3D Array



Interest Points



(a) Hessian detector



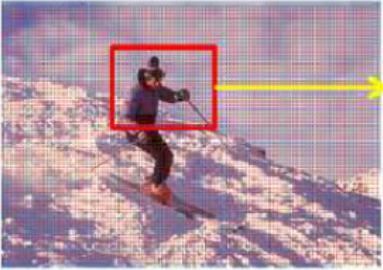
(b) DoG detector



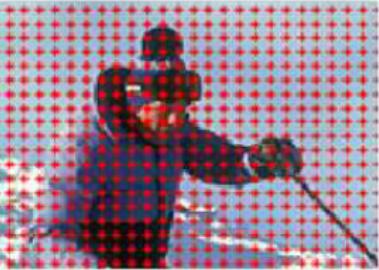
(c) Harris detector



(d) Random sampling

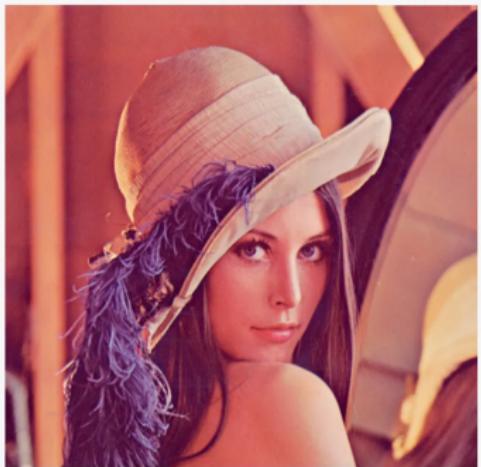


(e) Dense sampling

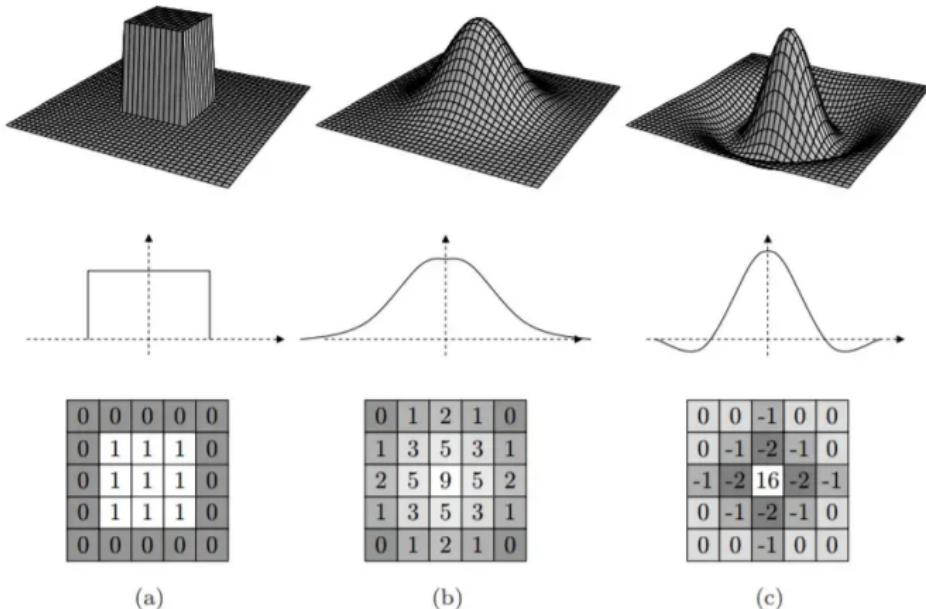


(f) Zooming

Lena



Filters



Average Smoothing

Input $f(x,y)$

$$\begin{matrix} * & \frac{1}{9} & \begin{array}{|c|c|c|} \hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline \end{array} & = & \text{Output } g(x,y) \end{matrix}$$

Convolution operator,
not multiplication!

Output $g(x,y)$

Average Smoothing

IMAGE WITH SALT AND PEPPER NOISE

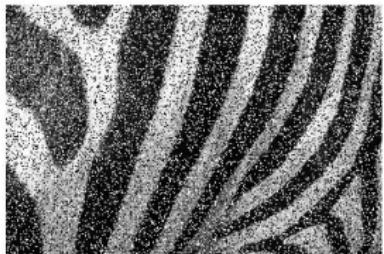


IMAGE AFTER MEDIAN FILTERING, WINDOW SIZE 3x3



IMAGE AFTER MEDIAN FILTERING, WINDOW SIZE 5x5



IMAGE AFTER MEDIAN FILTERING, WINDOW SIZE 7x7



Sobel Filters

X – Direction Kernel

-1	0	1
-2	0	2
-1	0	1

Y – Direction Kernel

-1	-2	-1
0	0	0
1	2	1

Sobel Filters



Laplacian Filter

0	1	0
1	-4	1
0	1	0

1	1	1
1	-8	1
1	1	1

-1	2	-1
2	-4	2
-1	2	-1

Laplacian Filter

Y



Laplacian



Gaussian Blur

1	2	1
2	4	2
1	2	1

1/16

1	4	7	4	1
4	16	26	16	4
7	26	41	26	7
4	16	26	16	4
1	4	7	4	1

1/273

0	0	1	2	1	0	0
0	3	13	22	13	3	0
1	13	59	97	59	13	1
2	22	97	159	97	22	2
1	13	59	97	59	13	1
0	3	13	22	13	3	0
0	0	1	2	1	0	0

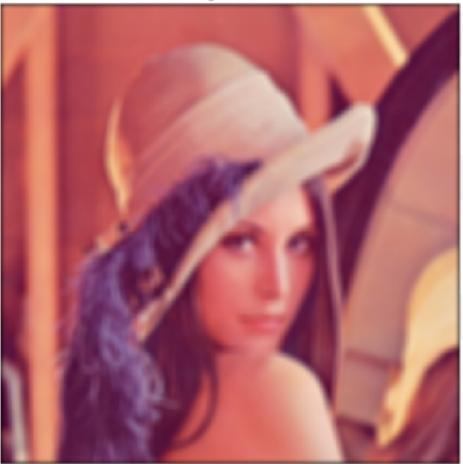
1/1003

Gaussian Blur

image source



image blurred



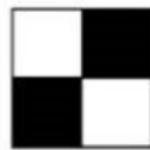
Viola-Jones Face Detection



Edge Features



Line Features



Four-rectangle Features

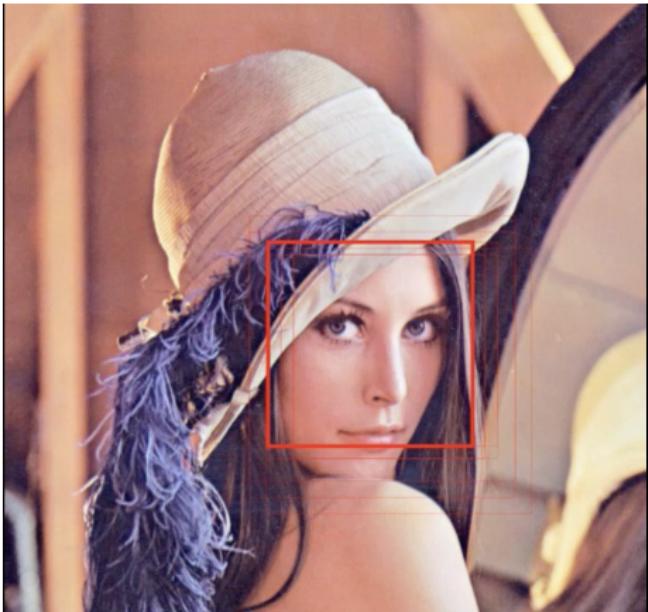
Important Features for Face Detection



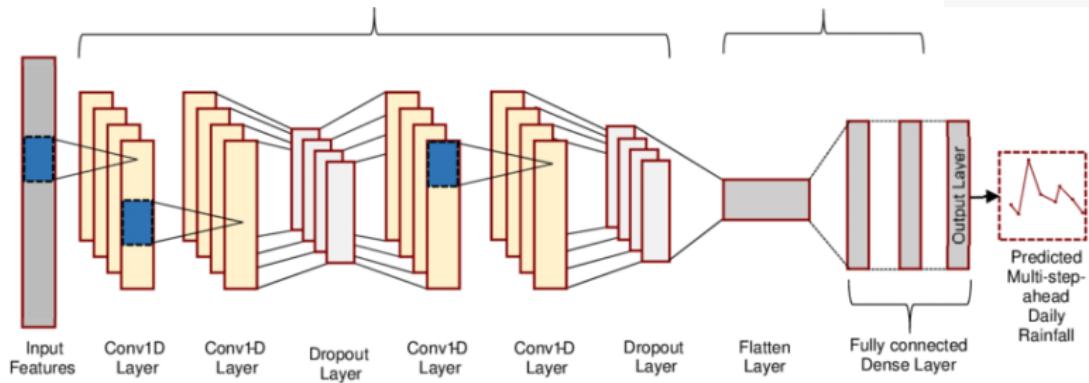
Viola-Jones Face Detection



Viola-Jones Face Detection



CNN



Convolution

Formal definition of convolution of functions f and g

$$(fg)(t) = \sum_{j=1}^m g(j)f(i-j+m/2)$$

$$f \quad \boxed{1 \quad 4 \quad 2 \quad 5}$$

$$g \quad \boxed{3 \quad 4 \quad 1}$$

$$c = f * g$$

$$\begin{array}{c} \boxed{1 \quad 4 \quad 2 \quad 5} \\ \boxed{1 \quad 4 \quad 3} \end{array}$$

$$C[0] = 1*3 = 3$$

$$\begin{array}{c} \boxed{1 \quad 4 \quad 2 \quad 5} \\ \boxed{1 \quad 4 \quad 3} \end{array}$$

$$C[1] = 1*4 + 4*3 = 16$$

$$\begin{array}{c} \boxed{1 \quad 4 \quad 2 \quad 5} \\ \boxed{1 \quad 4 \quad 3} \end{array}$$

$$C[2] = 1*1 + 4*4 + 2*3 = 23$$

$$\begin{array}{c} \boxed{1 \quad 4 \quad 2 \quad 5} \\ \boxed{1 \quad 4 \quad 3} \end{array}$$

$$C[3] = 4*1 + 2*4 + 5*3 = 27$$

$$\begin{array}{c} \boxed{1 \quad 4 \quad 2 \quad 5} \\ \boxed{1 \quad 4 \quad 3} \end{array}$$

$$C[4] = 2*1 + 5*4 = 22$$

$$\begin{array}{c} \boxed{1 \quad 4 \quad 2 \quad 5} \\ \boxed{1 \quad 4 \quad 3} \end{array}$$

$$C[5] = 5*1 = 5$$

Convolution

$$(fg)(m, n) = \sum_{i=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} g(m, n)f(m - i, n - j)$$

I(0,0)	I(1,0)	I(2,0)	I(3,0)	I(4,0)	I(5,0)	I(6,0)
I(0,1)	I(1,1)	I(2,1)	I(3,1)	I(4,1)	I(5,1)	I(6,1)
I(0,2)	I(1,2)	I(2,2)	I(3,2)	I(4,2)	I(5,2)	I(6,2)
I(0,3)	I(1,3)	I(2,3)	I(3,3)	I(4,3)	I(5,3)	I(6,3)
I(0,4)	I(1,4)	I(2,4)	I(3,4)	I(4,4)	I(5,4)	I(6,4)
I(0,5)	I(1,5)	I(2,5)	I(3,5)	I(4,5)	I(5,5)	I(6,5)
I(0,6)	I(1,6)	I(2,6)	I(3,6)	I(4,6)	I(5,6)	I(6,6)

Input image

×

H(0,0)	H(1,0)	H(2,0)
H(0,1)	H(1,1)	H(2,1)
H(0,2)	H(1,2)	H(2,2)

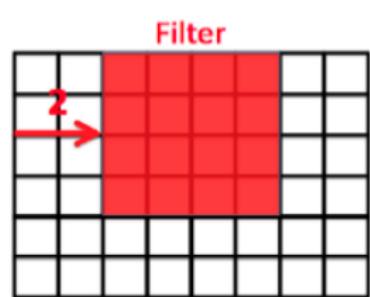
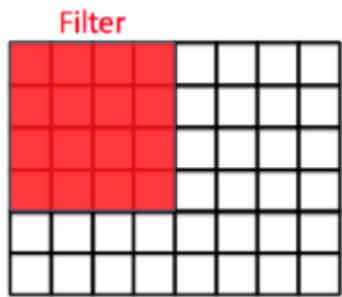
Filter

=

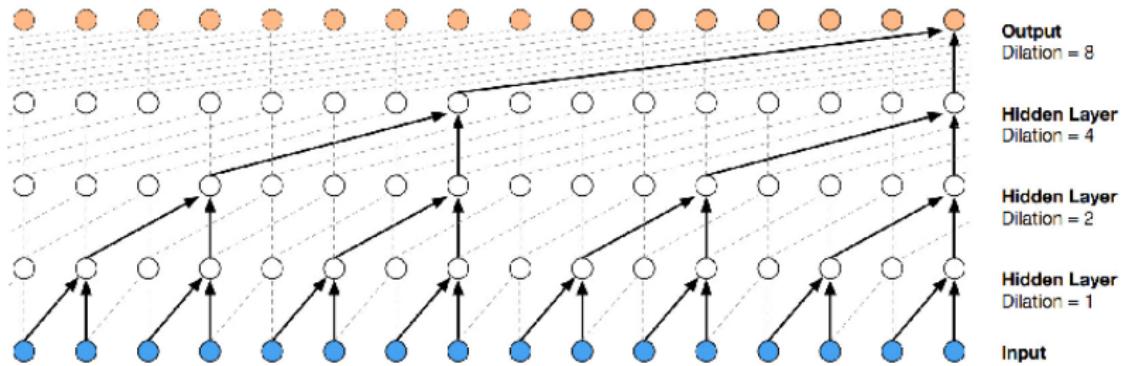
O(0,0)				

Output image

Stride

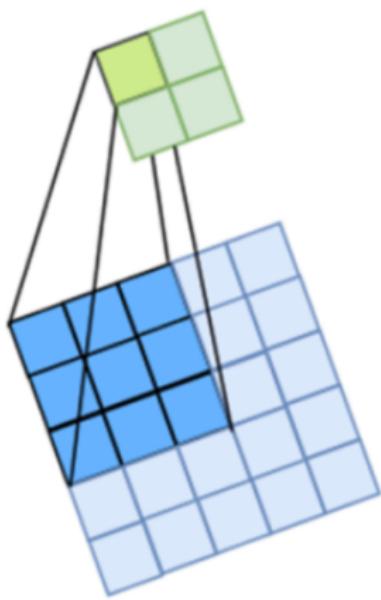


Dilation



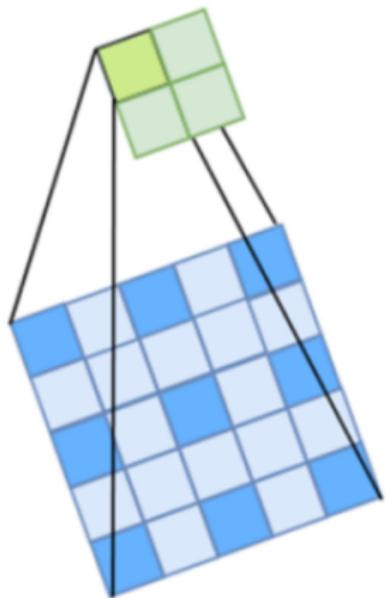
Dilation

Output:



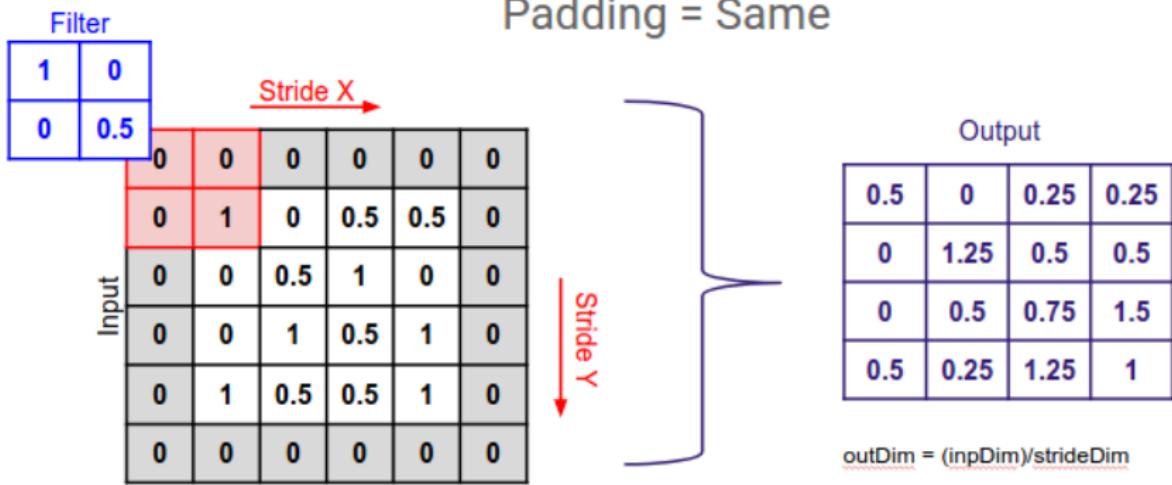
Input:

Convolution

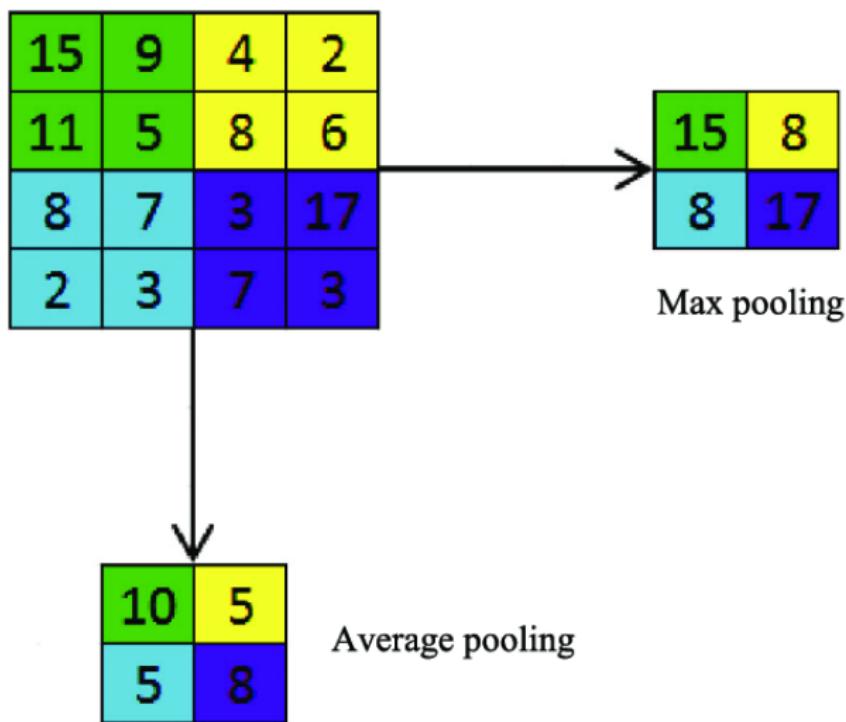


Dilated Convolution

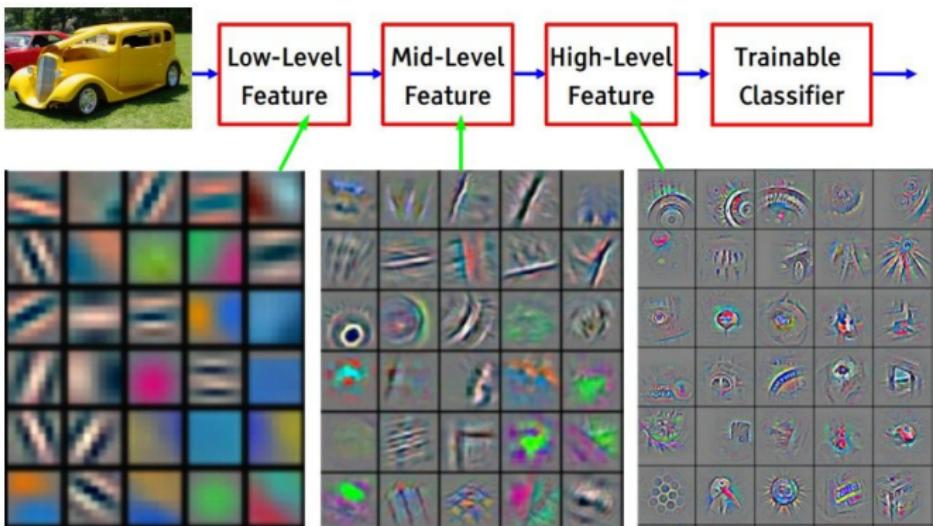
Padding



Pooling



Learned Kernels



Localization



True Label: Pomeranian



True Label: Car Wheel



True Label: Afghan Hound

