MODERN COMPUTER VISION

BY RAJEEV RATAN

Convolutions

How We Detect Features in Images

The Convolution Operation

- Mathematical term to describe the process of combining **two** functions to produce a **third**
- In our situation, the output is called a Feature Map
- We use a matrix, called a Filter or Kernel that is applied to our Image
- So the first 'Function' is the image that is combined with the Kernel or Filter which produces a Feature Map

$$\text{Image} \times \text{Kernel} = \text{Feature Map}$$

Example of a Convolution Operation

$$
\begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix}
*
\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}
=
\begin{bmatrix} 2 & 1 & -1 \\ -1 & 1 & 3 \\ 2 & 1 & 1 \end{bmatrix}
$$

Input Image | Filter or Kernel | Output or Feature Map

Recall grayscale images are just intensity values ranging from black to white

255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	255	220	146	237	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	255	91		170	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	253	68	[illegible]	220	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	238	30	62	241	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	218		64	241	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	145		64	241	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	81		72	244	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	72		128	255	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	73		146	255	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	64	[illegible]	184	255	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	63	[illegible]	188	255	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	71	[illegible]	187	255	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	68	[illegible]	190	255	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	83		160	255	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	181	48	184	255	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255

Example of a Convolution Operation

$$(1\times 0)+(0\times 1)+(1\times 0)+(1\times 1)+(0\times 0)+(0\times -1)+(0\times 0)+(1\times 1)+(1\times 0)=2$$

1x0	0x1	1x0	0	1
1x1	0x0	0x-1	1	1
0x0	1x1	1x0	0	0
1	0	0	1	0
0	0	1	1	0

Input Image

*

0	1	0
1	0	-1
0	1	0

Filter or Kernel

=

2		

Output or Feature Map

Example of a Convolution Operation

$$(0 \times 0) + (1 \times 1) + (0 \times 0) + (0 \times 1) + (0 \times 0) + (1 \times -1) + (1 \times 0) + (1 \times 1) + (0 \times 0) = 1$$

1	0x0	1x1	0x0	1
1	0x1	0x0	1x-1	1
0	1x0	1x1	0x0	0
1	0	0	1	0
0	0	1	1	0

Input Image

*

0	1	0
1	0	-1
0	1	0

Filter or Kernel

=

2	1	

Output or Feature Map

Example of a Convolution Operation

1	0	1	0	1
1	0	0	1	1
0	1	1	0	0
1	0	0	1	0
0	0	1	1	0

Input Image

*

0	1	0
1	0	-1
0	1	0

Filter or Kernel

=

2	1	-1

Output or Feature Map

Example of a Convolution Operation

1	0	1	0	1
1	0	0	1	1
0	1	1	0	0
1	0	0	1	0
0	0	1	1	0

Input Image

*

0	1	0
1	0	-1
0	1	0

Filter or Kernel

=

2	1	-1
-1		

Output or Feature Map

Example of a Convolution Operation

1	0	1	0	1
1	0	0	1	1
0	1	1	0	0
1	0	0	1	0
0	0	1	1	0

Input Image

*

0	1	0
1	0	-1
0	1	0

Filter or Kernel

=

2	1	-1
-1	1	

Output or Feature Map

Example of a Convolution Operation

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} * \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 & -1 \\ -1 & 1 & 3 \\ & & \end{bmatrix}$$

Input Image

Filter or Kernel

Output or Feature Map

Example of a Convolution Operation

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} * \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 & -1 \\ -1 & 1 & 3 \\ 2 & & \end{bmatrix}$$

Input Image — Filter or Kernel — Output or Feature Map

Example of a Convolution Operation

1	0	1	0	1
1	0	0	1	1
0	1	1	0	0
1	0	0	1	0
0	0	1	1	0

Input Image

*

0	1	0
1	0	-1
0	1	0

Filter or Kernel

=

2	1	-1
-1	1	3
2	1	

Output or Feature Map

Example of a Convolution Operation

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} * \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 & -1 \\ -1 & 1 & 3 \\ 2 & 1 & 1 \end{bmatrix}$$

Input Image

Filter or Kernel

Output or Feature Map

Image Features

- Our feature maps are actually **Feature Detectors**
- **Why did we do this?**
- Because Convolution Filters or Kernels **detect features** in images

Our Convolution Filter as an Edge Detector

1	1	0	0	0
1	1	0	0	0
1	1	0	0	0
1	1	0	0	0
1	1	0	0	0

Input Image

*

1	0	-1
1	0	-1
1	0	-1

Filter or Kernel

=

3	3	0
3	3	0
3	3	0

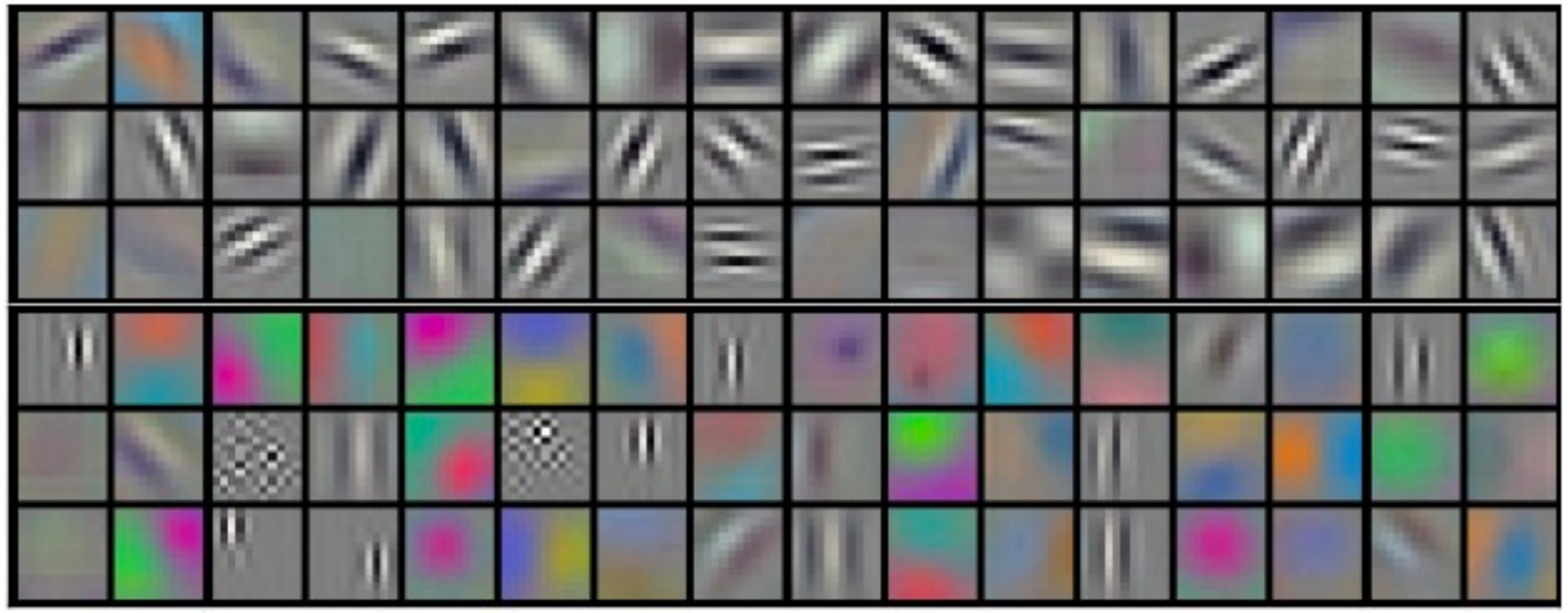
Output or Feature Map

Convolution Filters Detect Many Features

Operation	Filter	Convolved Image
Identity	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	
Edge detection	$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$	
	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	
	$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$	
Sharpen	$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$	
Box blur (normalized)	$\frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$	
Gaussian blur (approximation)	$\frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$	

https://en.wikipedia.org/wiki/Kernel_(image_processing)

Convolution Filters Detect Many Features



Example filters learned by Krizhevsky et al.

Convolution Filters Detect Many Features

Source – Deep Learning Methods for Vision
https://cs.nyu.edu/~fergus/tutorials/deep_learning_cvpr12/

Calculating Feature Map Size

Feature Map Size = $n - f + 1 = m$

Feature Map Size = $5 - 3 + 1 = 3$

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} * \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 & -1 \\ -1 & 1 & 3 \\ 2 & 1 & 1 \end{bmatrix}$$

5 x 5 — $n \times n$

3 x 3 — $f \times f$

3 x 3 — $m \times m$

Next…

Feature Detectors