From Human Vision to Computer Vision: Rolling in the deep with an image

ICGEB

सत्यमेव जयते
DEPARTMENT OF BIOTECHNOLOGY
GOVERNMENT OF INDIA

Submitted By :

Nimisha Tiwari

Project Associate

Translational Bioinformatics Group

ICGEB, New Delhi

Email: nimisha@icgeb.res.in

Content

- What is an Image, how does a computer perceives it?
- Intro Google Colab.
- Image Preprocessing.
- The algorithmic story of Convolution Neural Network.
- CNN architecture Models : Transfer Learning For Image Classification

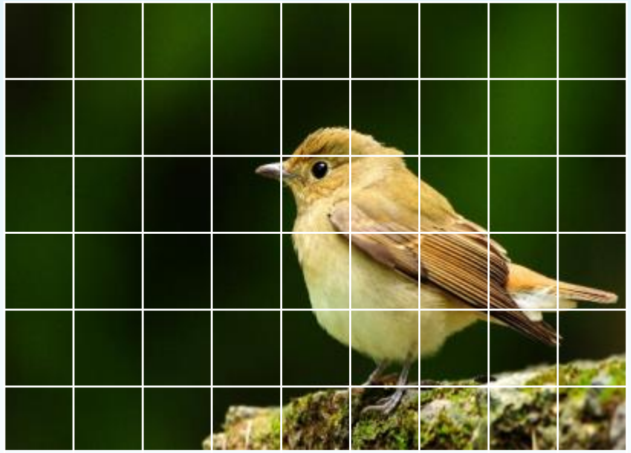
What is an Image ?

{What is difference in the vision of a human and a computer}

What a human sees

What a computer sees

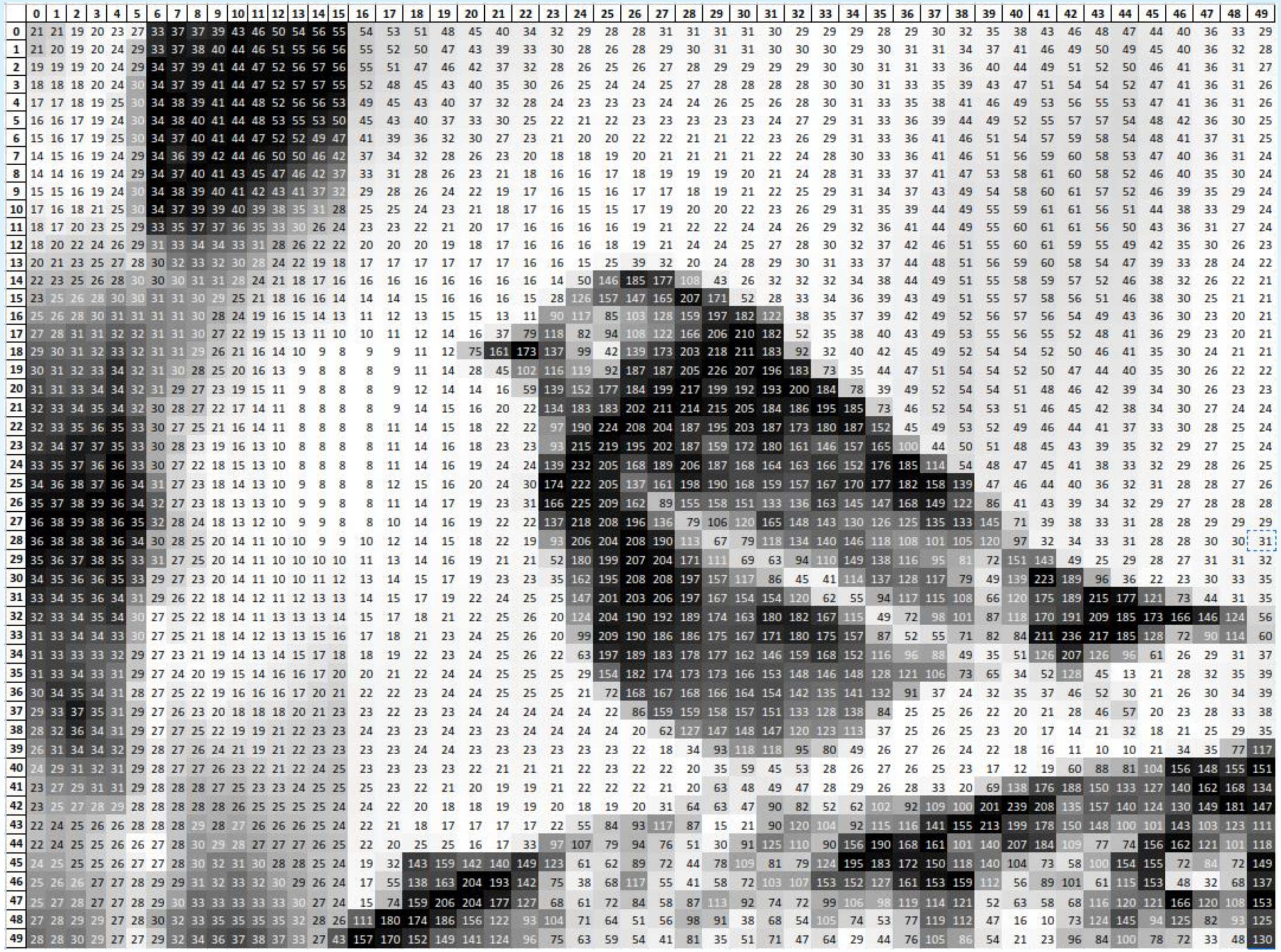
Color/RGB Image



6x9 pixels

28	34	32	30	29	30	31	33	30
31	33	32	31	34	31	35	34	31
32	31	28	29	90	88	79	33	32
32	30	27	31	99	75	64	48	33
31	32	29	30	45	68	54	50	36
30	31	30	52	48	54	55	56	58

- 256 pixels
- 0 – 255
- Black - 0
- White - 255

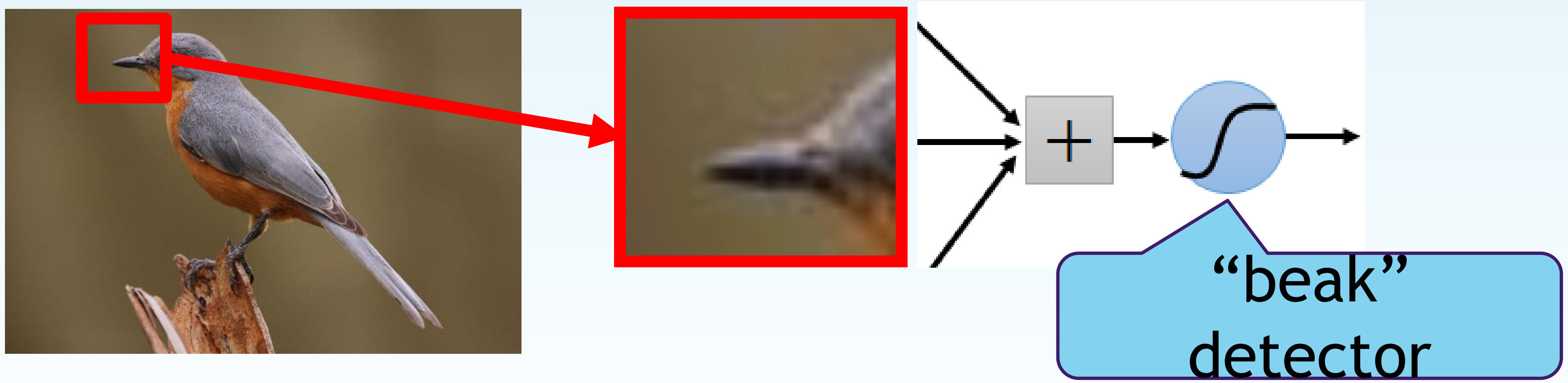


	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49
0	21	21	19	20	23	27	33	37	37	39	43	46	50	54	56	55	54	53	51	48	45	40	34	32	29	28	28	31	31	31	31	30	29	29	29	28	29	30	32	35	38	43	46	48	47	44	40	36	33	29
1	21	20	19	20	24	29	33	37	38	40	44	46	51	55	56	56	55	52	50	47	43	39	33	30	28	26	28	29	30	31	31	30	30	30	29	30	31	31	34	37	41	46	49	50	49	45	40	36	32	28
2	19	19	19	20	24	29	34	37	39	41	44	47	52	56	57	56	55	51	47	46	42	37	32	28	26	25	26	27	28	29	29	29	29	30	30	31	31	33	36	40	44	49	51	52	50	46	41	36	31	27
3	18	18	18	20	24	30	34	37	39	41	44	47	52	57	57	55	52	48	45	43	40	35	30	26	25	24	24	25	27	28	28	28	28	30	30	31	33	35	39	43	47	51	54	54	52	47	41	36	31	26
4	17	17	18	19	25	30	34	38	39	41	44	48	52	56	56	53	49	45	43	40	37	32	28	24	23	23	23	24	24	26	25	26	28	30	31	33	35	38	41	46	49	53	56	55	53	47	41	36	31	26
5	16	16	17	19	24	30	34	38	40	41	44	48	53	55	53	50	45	43	40	37	33	30	25	22	21	22	23	23	23	23	23	24	27	29	31	33	36	39	44	49	52	55	57	57	54	48	42	36	30	25
6	15	16	17	19	25	30	34	37	40	41	44	47	52	52	49	47	41	39	36	32	30	27	23	21	20	20	22	22	21	21	22	23	26	29	31	33	36	41	46	51	54	57	59	58	54	48	41	37	31	25
7	14	15	16	19	24	29	34	36	39	42	44	46	50	50	46	42	37	34	32	28	26	23	20	18	18	19	20	21	21	21	21	22	24	28	30	33	36	41	46	51	56	59	60	58	53	47	40	36	31	24
8	14	14	16	19	24	29	34	37	40	41	43	45	47	46	42	37	33	31	28	26	23	21	18	16	16	17	18	19	19	19	20	21	24	28	31	33	37	41	47	53	58	61	60	58	52	46	40	35	30	24
9	15	15	16	19	24	30	34	38	39	40	41	42	43	41	37	32	29	28	26	24	22	19	17	16	15	16	17	17	18	19	21	22	25	29	31	34	37	43	49	54	58	60	61	57	52	46	39	35	29	24
10	17	16	18	21	25	30	34	37	39	39	40	39	38	35	31	28	25	25	24	23	21	18	17	16	15	15	17	19	20	20	22	23	26	29	31	35	39	44	49	55	59	61	61	56	51	44	38	33	29	24
11	18	17	20	23	25	29	33	35	37	37	36	35	33	30	26	24	23	23	22	21	20	17	16	16	16	16	19	21	22	22	24	24	26	29	32	36	41	44	49	55	60	61	61	56	50	43	36	31	27	24
12	18	20	22	24	26	29	31	33	34	34	33	31	28	26	22	22	20	20	20	19	18	17	16	16	16	18	19	21	24	24	25	27	28	30	32	37	42	46	51	55	60	61	59	55	49	42	35	30	26	23
13	20	21	23	25	27	28	30	32	33	32	30	28	24	22	19	18	17	17	17	17	17	17	16	16	15	25	39	32	20	24	28	29	30	31	33	37	44	48	51	56	59	60	58	54	47	39	33	28	24	22
14	22	23	25	26	28	30	30	30	31	31	28	24	21	18	17	16	16	16	16	16	16	16	16	14	50	146	185	177	108	43	26	32	32	32	34	38	44	49	51	55	58	59	57	52	46	38	32	26	22	21
15	23	25	26	28	30	30	31	31	30	29	25	21	18	16	16	14	14	14	15	16	16	16	15	28	126	157	147	165	207	171	52	28	33	34	36	39	43	49	51	55	57	58	56	51	46	38	30	25	21	21
16	25	26	28	30	31	31	31	31	30	28	24	19	16	15	14	13	11	12	13	15	15	13	11	90	117	85	103	128	159	197	182	122	38	35	37	39	42	49	52	56	57	56	54	49	43	36	30	23	20	21
17	27	28	31	31	32	32	31	31	30	27	22	19	15	13	11	10	10	11	12	14	16	37	79	118	82	94	108	122	166	206	210	182	52	35	38	40	43	49	53	55	56	55	52	48	41	36	29	23	20	21
18	29	30	31	32	33	32	31	31	29	26	21	16	14	10	9	8	9	9	11	12	75	161	173	137	99	42	139	173	203	218	211	183	92	32	40	42	45	49	52	54	54	52	50	46	41	35	30	24	21	21
19	30	31	32	33	34	32	31	30	28	25	20	16	13	9	8	8	8	9	11	14	28	45	102	116	119	92	187	187	205	226	207	196	183	73	35	44	47	51	54	54	52	50	47	44	40	35	30	26	22	22
20	31	31	33	34	34	32	31	29	27	23	19	15	11	9	8	8	8	9	12	14	14	16	59	139	152	177	184	199	217	199	192	193	200	184	78	39	49	52	54	54	51	48	46	42	39	34	30	26	23	23
21	32	33	34	35	34	32	30	28	27	22	17	14	11	8	8	8	8	9	14	15	16	20	22	134	183	183	202	211	214	215	205	184	186	195	185	73	46	52	54	53	51	46	45	42	38	34	30	27	24	24
22	32	33	35	36	35	33	30	27	25	21	16	14	11	8	8	8	8	11	14	15	18	22	22	97	190	224	208	204	187	195	203	187	173	180	187	152	45	49	53	52	49	46	44	41	37	33	30	28	25	24
23	32	34	37	37	35	33	30	28	23	19	16	13	10	8	8	8	8	11	14	16	18	23	23	93	215	219	195	202	187	159	172	180	161	146	157	165	100	44	50	51	48	45	43	39	35	32	29	27	25	24
24	33	35	37	36	36	33	30	27	22	18	15	13	10	8	8	8	8	11	14	16	19	24	24	139	232	205	168	189	206	187	168	164	163	166	152	176	185	114	54	48	47	45	41	38	33	32	29	28	26	25
25	34	36	38	37	36	34	31	27	23	18	14	13	10	9	8	8	8	12	15	16	20	24	30	174	222	205	137	161	198	190	168	159	157	167	170	177	182	158	139	47	46	44	40	36	32	31	28	28	27	26
26	35	37	38	39	36	34	32	27	23	18	13	13	10	9	9	8	8	11	14	17	19	23	31	166	225	209	162	89	155	158	151	133	136	163	145	147	168	149	122	86	41	43	39	34	32	29	27	28	28	28
27	36	38	39	38	36	35	32	28	24	18	13	12	10	9	9	8	8	10	14	16	19	22	22	137	218	208	196	136	79	106	120	165	148	143	130	126	125	135	133	145	71	39	38	33	31	28	28	29	29	29
28	36	38	38	38	36	34	30	28	25	20	14	11	10	10	9	9	10	12	14	15	18	22	19	93	206	204	208	190	113	67	79	118	134	140	146	118	108	101	105	120	97	32	34	33	31	28	28	30	30	31
29	35	36	37	38	35	33	31	27	25	20	14	11	10	10	10	10	11	13	14	16	19	21	21	52	180	199	207	204	171	111	69	63	94	110	149	138	116	95	81	72	151	143	49	25	29	28	27	31	31	32
30	34	35	36	36	35	33	29	27	23	20	14	11	10	10	11	12	13	14	15	17	19	23	23	35	162	195	208	208	197	157	117	86	45	41	114	137	128	117	79	49	139	223	189	96	36	22	23	30	33	35
31	33	34	35	36	34	31	29	26	22	18	14	12	11	12	13	13	14	15	17	19	22	24	25	25	147	201	203	206	197	167	154	154	120	62	55	94	117	115	108	66	120	175	189	215	177	121	73	44	31	35
32	32	33	34	35	34	30	27	25	22	18	14	11	13	13	13	14	15	17	18	21	22	25	26	20	124	204	190	192	189	174	163	180	182	167	115	49	72	98	101	87	118	170	191	209	185	173	166	146	124	56
33	31	33	34	34	33	30	27	25	21	18	14	12	13	13	15	16	17	18	21	23	24	25	26	20	99	209	190	186	186	175	167	171	180	175	157	87	52	55	71	82	84	211	236	217	185	128	72	90	114	60
34	31	33	33	33	32	29	27	23	21	19	14	13	14	15	17	18	18	19	22	23	24	25	26	22	63	197	189	183	178	177	162	146	159	168	152	116	96	88	49	35	51	126	207	126	96	61	26	29	31	37
35	31	33	34	33	31	29	27	24	20	19	15	14	16	16	17	20	20	21	22	24	24	25	25	25	29	154	182	174	173	173	166	153	148	146	148	128	121	106	73	65	34	52	128	45	13	21	28	32	35	39
36	30	34	35	34	31	28	27	25	22	19	16	16	16	17	20	21	22	22	23	24	24	25	25	25	21	72	168	167	168	166	164	154	142	135	141	132	91	37	24	32	35	37	46	52	30	21	26	30	34	39
37	29	33	37	35	31	29	27	26	23	20	18	18	18	20	21	23	23	22	23	23	24	24	24	24	24	22	86	159	159	158	157	151	133	128	138	84	25	25	26	22	20	21	28	46	57	20	23	28	33	38
38	28	32	36	34	31	29	27	27	25	22	19	19	21	22	23	23	24	23	23	24	23	23	24	24	24	24	20	62	127	147	148	147	120	123	113	37	25	26	25	23	20	17	14	21	32	18	21	25	29	35
39	26	31	34	34	32	29	28	27	26	24	21	19	21	22	23	23	23	23	24	24	23	23	23	23	23	23	22	18	34	93	118	118	95	80	49	26	27	26	24	22	18	16	11	10	10	21	34	35	77	117
40	24	29	31	32	31	29	28	27	27	26	23	22	21	22	24	25	23	23	23	23	22	21	21	21	22	23	22	22	20	35	59	45	53	28	26	27	26	25	23	17	12	19	60	88	81	104	156	148	155	151
41	23	27	29	31	31	29	28	28	28	27	25	23	23	24	25	25	25	23	22	21	20	19	19	21	22	22	22	21	20	63	48	49	47	28	29	26	28	33	20	69	138	176	188	150	133	127	140	162	168	134
42	23	25	27	28	29	28	28	28	28	28	26	25	25	25	25	24	24	22	20	18	18	19	19	20	18	19	20	31	64	63	47	90	82	52	62	102	92	109	100	201	239	208	135	157	140	124	130	149	181	147
43	22	24	25	26	26	28	28	28	29	28	27	26	26	26	25	24	22	21	18	17	17	17	17	22	55	84	93	117	87	15	21	90	120	104	92	115	116	141	155	213	199	178	150	148	100	101	143	103	123	111
44	22	24	25	25	26	26	27	28	30	29	28	27	27	27	26	25	22	20	25	25	16	17	33	97	107	79	94	76	51	30	91	125	110	90	156	190	168	161	101	140	207	184	109	77	74	156	162	121	101	118
45	24	25	25	25	26	27	27	28	30	32	31	30	28	28	25	24	19	32	143	159	142	140	149	123	61	62	89	72	44	78	109	81	79	124	195	183	172	150	118	140	104	73	58	100	154	155	72	84	72	149
46	25	26	26	27	27	28	29	29	31	32	33	32	30	29	26	24	17	55	138	163	204	193	142	75	38	68	117	55	41	58	72	103	107	153	152	127	161	153	159	112	56	89	101	61	115	153	48	32	68	137
47	25	27	28	27	27	28	29	30	33	33	33	33	33	30	27	24	15	74	159	206	204	177	127	68	61	72	84	58	87	113	92	74	72	99	106	98	119	114	121	52	63	58	68	116	120	121	166	120	108	153
48	27	28	29	29	27	28	30	32	33	35	35	35	35	32	28	26	111	180	174	186	156	122	93	104	71	64	51	56	98	91	38	68	54	105	74	53	77	119	112	47	16	10	73	124	145	94	125	82	93	125
49	28	28	30	29	27	27	29	32	34	36	37	38	37	33	27	43	157	170	152	149	141	124	96	75	63	59	54	41	81	35	51	71	47	64	29	44	76	105	86	54	21	23	96	84	100	78	72	33	48	130

Consider learning an image:

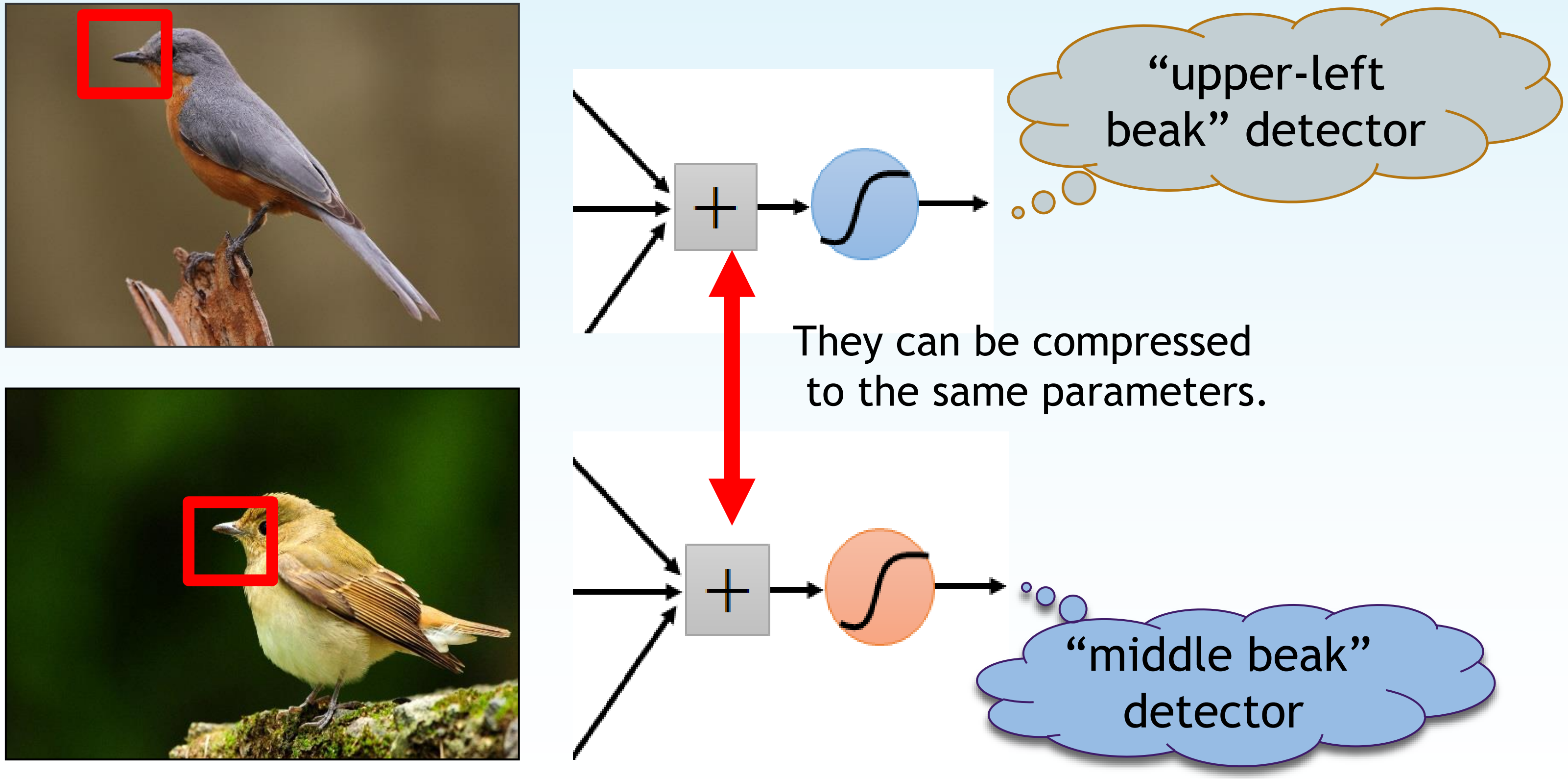
- Some patterns are much smaller than the whole image

Can represent a small region with fewer parameters



“beak” detector

Same pattern appears in different places:
They can be compressed!
What about training a lot of such “small” detectors
and each detector must “move around”.



“upper-left beak” detector

They can be compressed to the same parameters.

“middle beak” detector

Convolution Neural Network

- Deep learning explores the possibility of learning features directly from input data, avoiding hand-crafted features.

- A deep net is trained by feeding it input and letting it compute layer-by-layer to generate the final output for comparison with the correct answer.

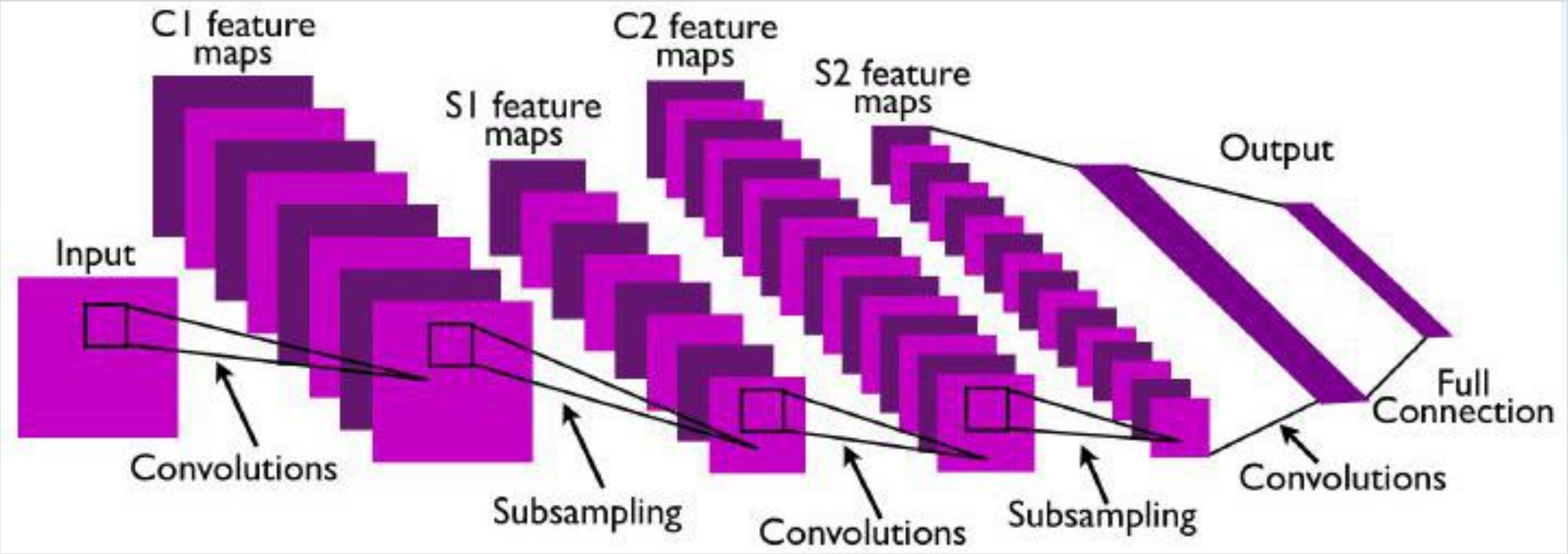
- After computing the error at the output, this error flows backward through the net by backpropagation.

- At each step backward the model parameters are tuned in a direction that tries to reduce the error.

- helps in model improvement, training is an iterative process that involves multiple passes of the input data until the model converges.

Convolution Neural Network: CNN

- There are three layers used to build CNN architectures:-
  - Convolutional layer,
  - Pooling layer, and
  - Fully connected layer.



Convolution

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

6 x 6 image

These are the network parameters to be learned

1	-1	-1
-1	1	-1
-1	-1	1

Filter 1

-1	1	-1
-1	1	-1
-1	1	-1

Filter 2

⋮ ⋮

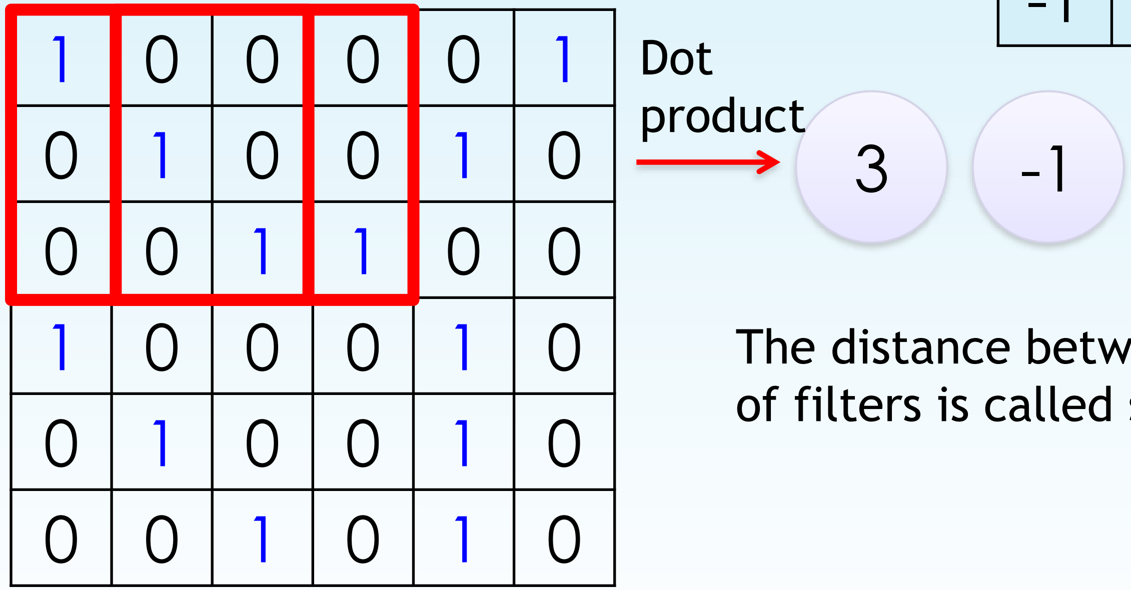
Each filter detects a small pattern (3 x 3).

Convolution

1	-1	-1
-1	1	-1
-1	-1	1

Filter 1

stride=1



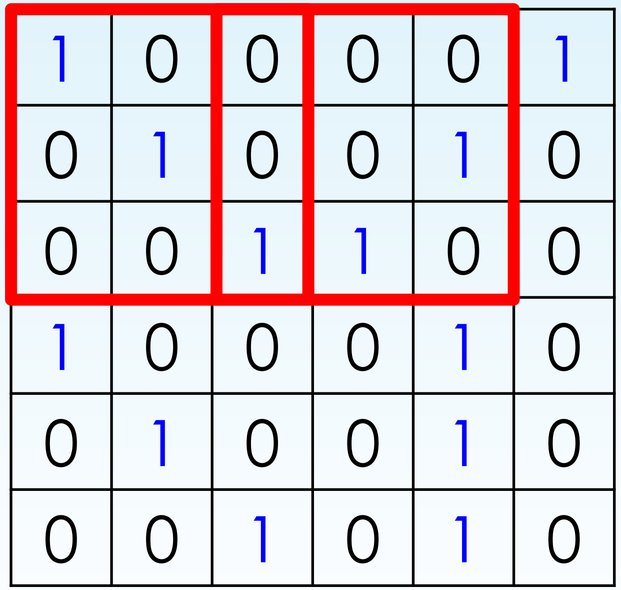
1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
1	0	0	0	1	0
0	1	0	0	1	0
0	0	1	0	1	0

6 x 6 image

The distance between the applications of filters is called stride.

Convolution

If stride=2



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

6 x 6 image

1	-1	-1
-1	1	-1
-1	-1	1

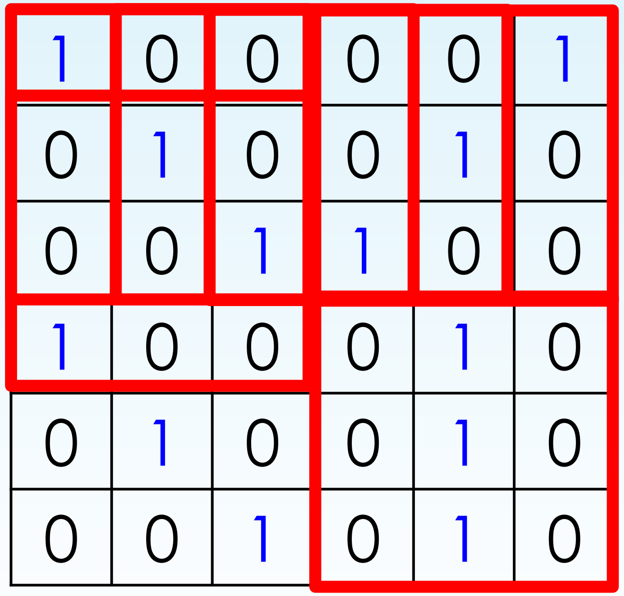
Filter 1

3 -3

Stride hyper parameter is smaller than the filter size the convolution is applied in overlapping windows

Convolution

stride=1



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

6 x 6 image

1	-1	-1
-1	1	-1
-1	-1	1

Filter 1

3	-1	-3	-1
-3	1	0	-3
-3	-3	0	1
3	-2	-2	-1

Convolution

-1	1	-1
-1	1	-1
-1	1	-1

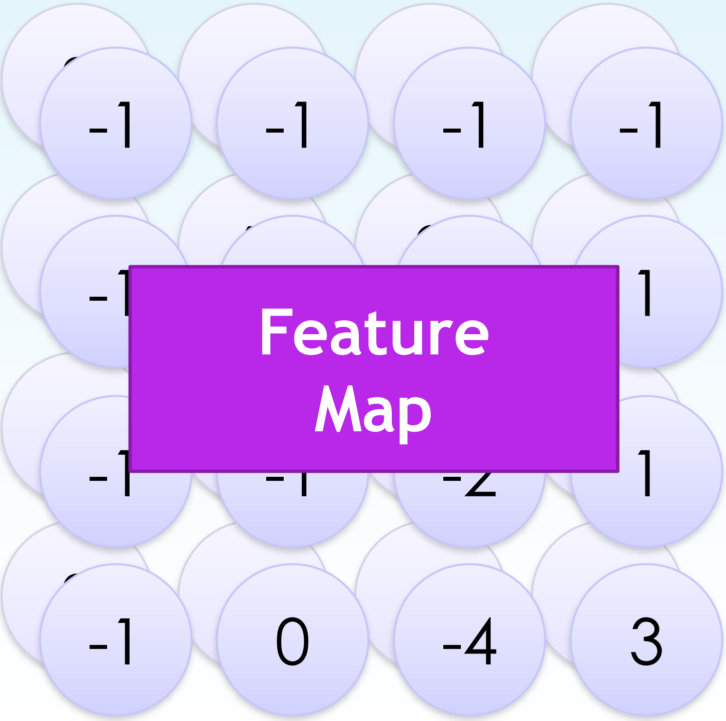
Filter 2

stride=1

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

6 x 6 image

Repeat this for each filter



Two 4 x 4 images
Forming 2 x 4 x 4 matrix

Color image: RGB 3 channels

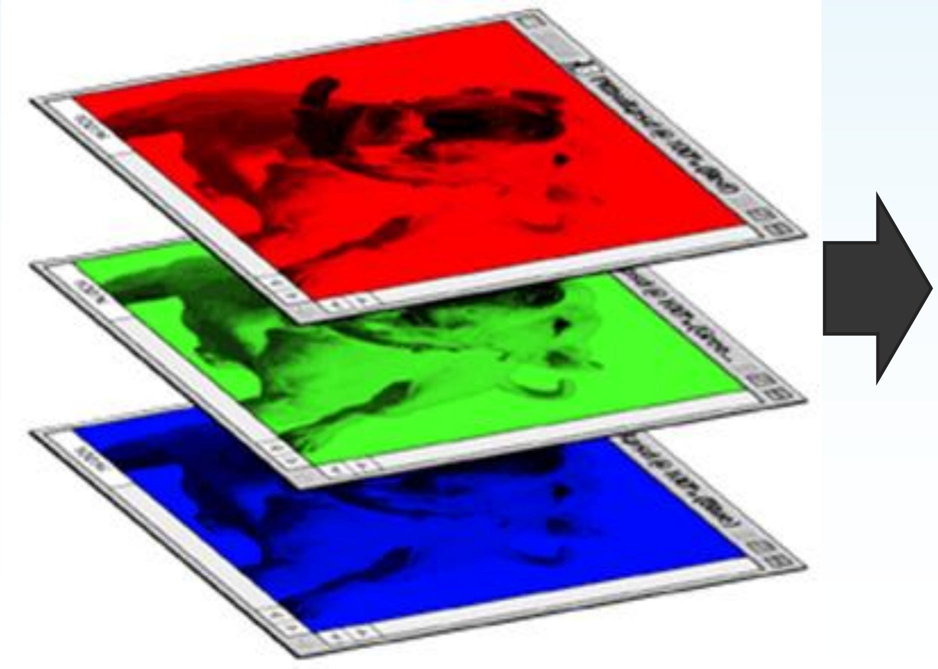
1	-1	-1
-1	1	-1
-1	-1	1

Filter 1

-1	1	-1
-1	1	-1
-1	1	-1

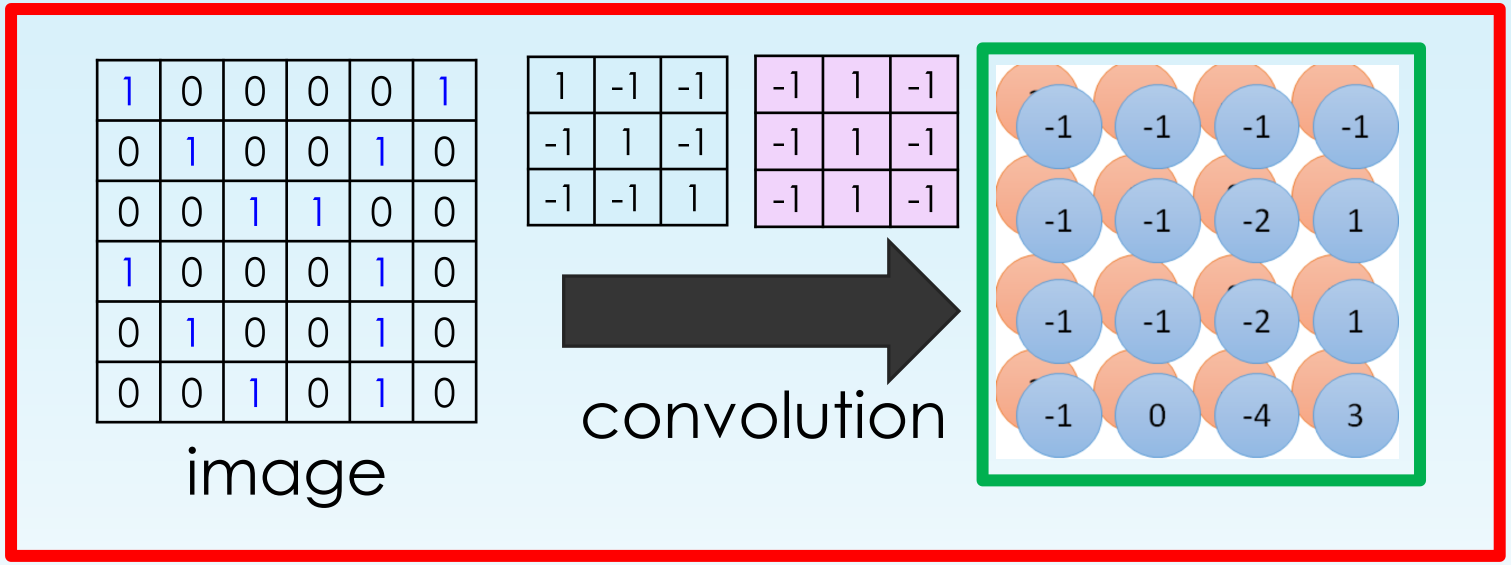
Filter 2

Color image



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

Convolution v.s. Fully Connected



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

image

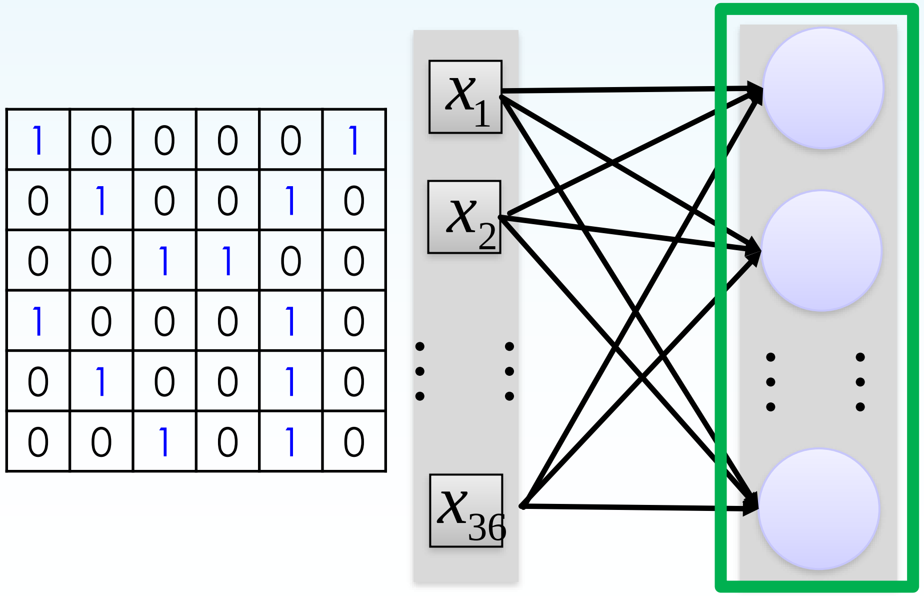
1	-1	-1
-1	1	-1
-1	-1	1

-1	1	-1
-1	1	-1
-1	1	-1

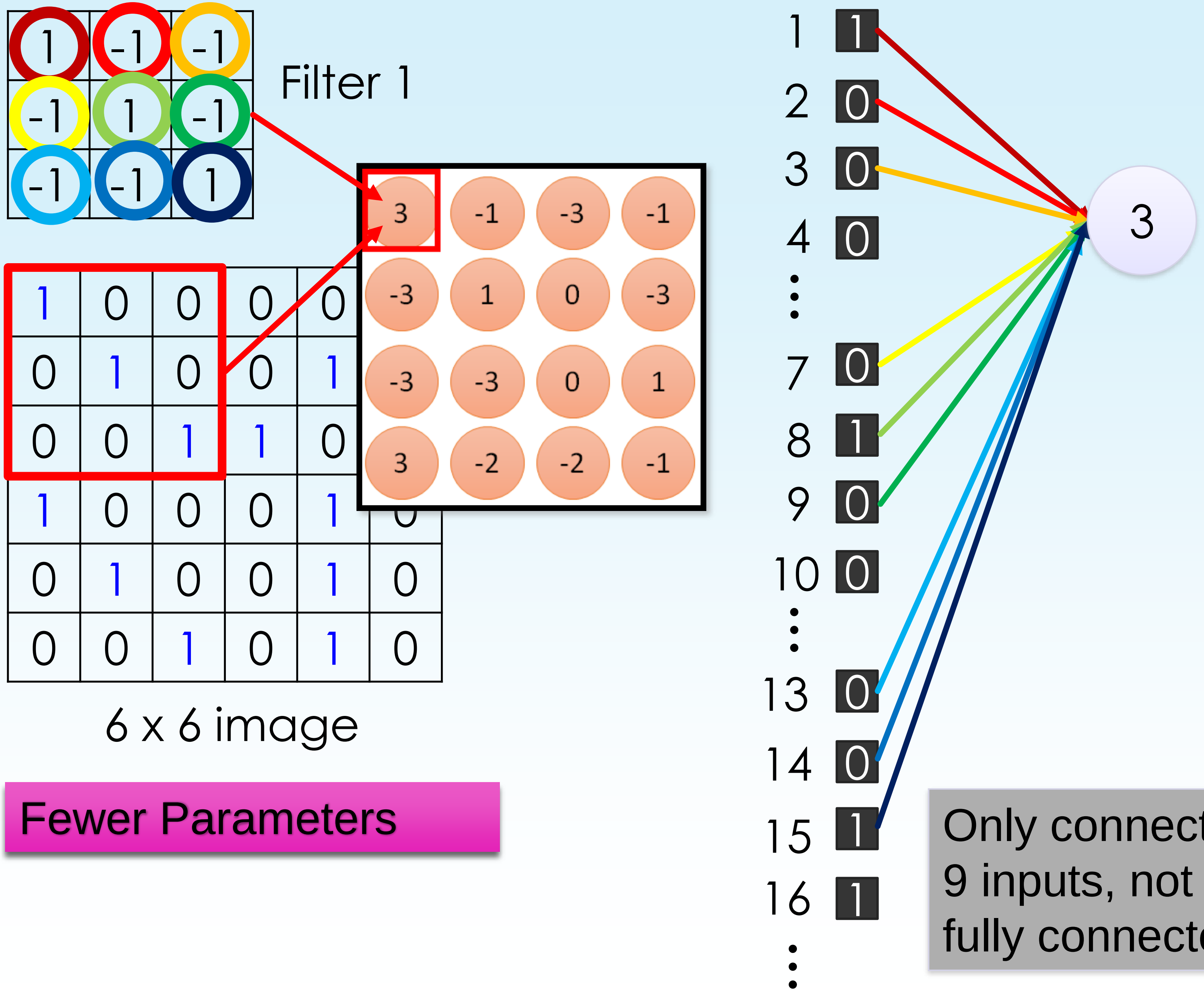
convolution

-1	-1	-1	-1
-1	-1	-2	1
-1	-1	-2	1
-1	0	-4	3

Fully-connected



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



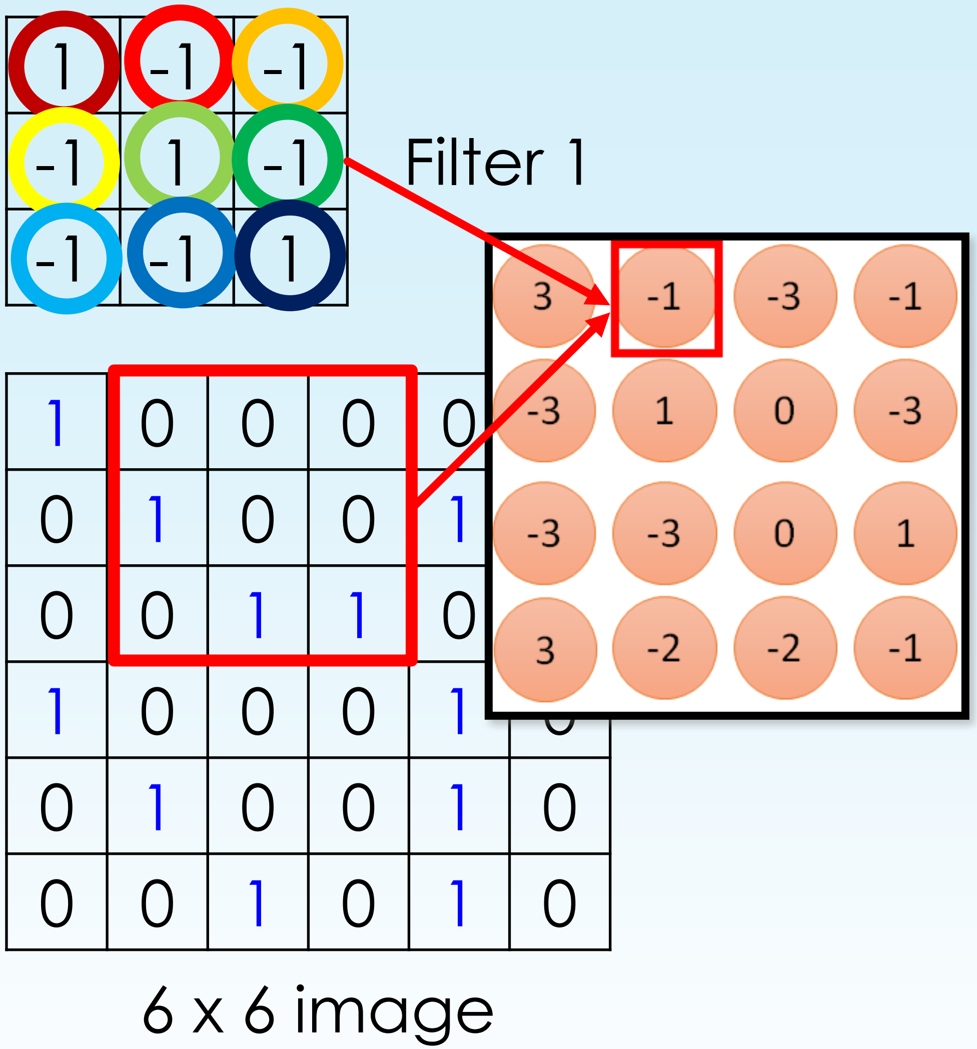
1	-1	-1
-1	1	-1
-1	-1	1

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

3	-1	-3	-1
-3	1	0	-3
-3	-3	0	1
3	-2	-2	-1

Fewer Parameters

Only connect to 9 inputs, not fully connected



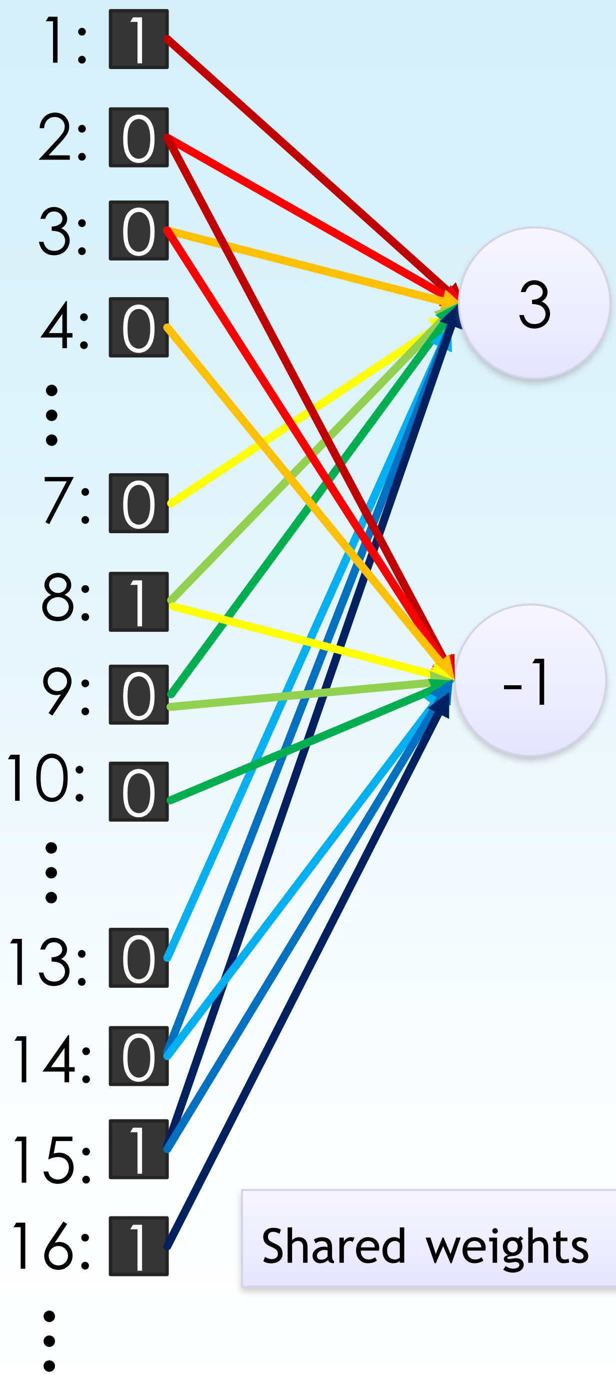
Filter 1

1	-1	-1
-1	1	-1
-1	-1	1

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

3	-1	-3	-1
-3	1	0	-3
-3	-3	0	1
3	-2	-2	-1

6 x 6 image



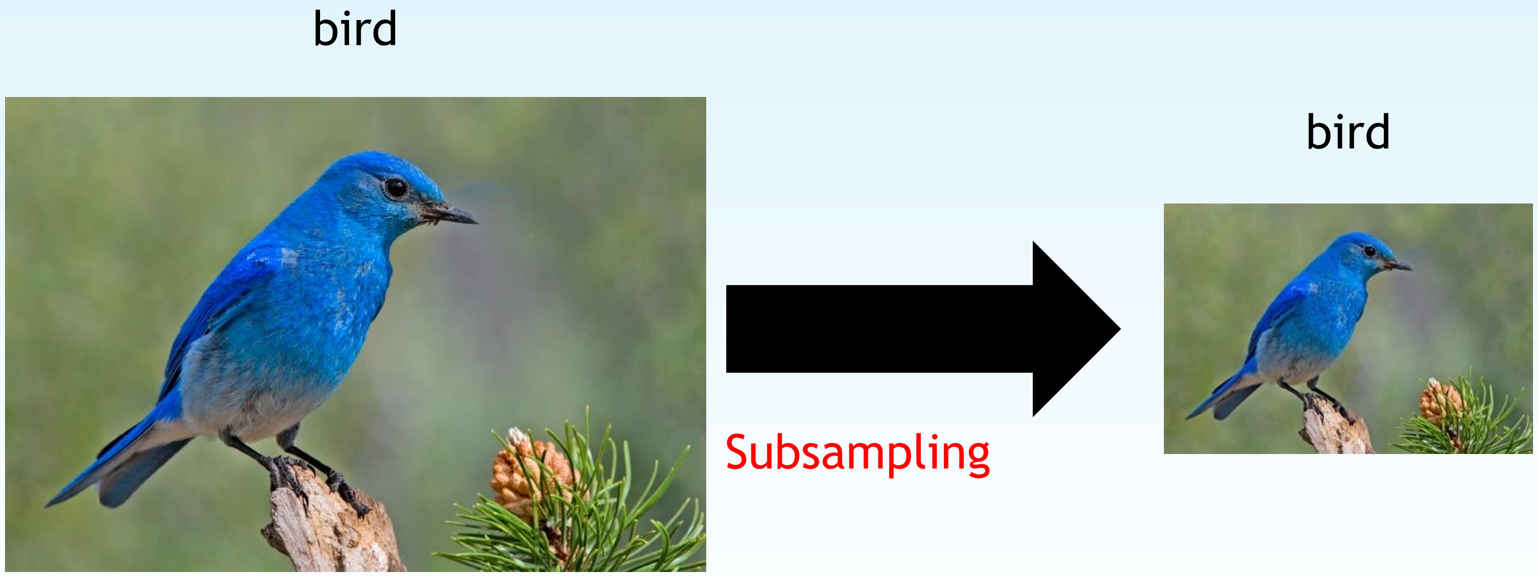
Fewer parameters

Even fewer parameters

Shared weights

Why Pooling

- Subsampling pixels will not change the object



We can subsample the pixels to make image smaller

→ fewer parameters to characterize the image

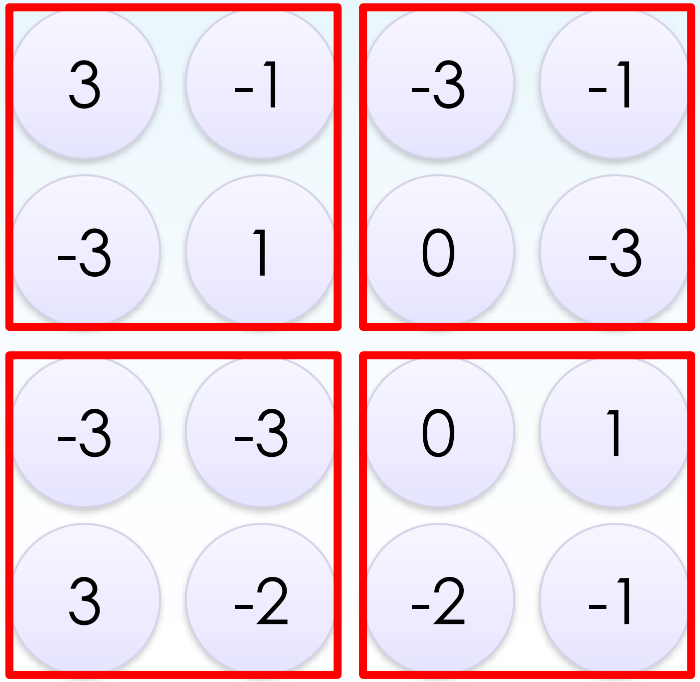
Max Pooling

1	-1	-1
-1	1	-1
-1	-1	1

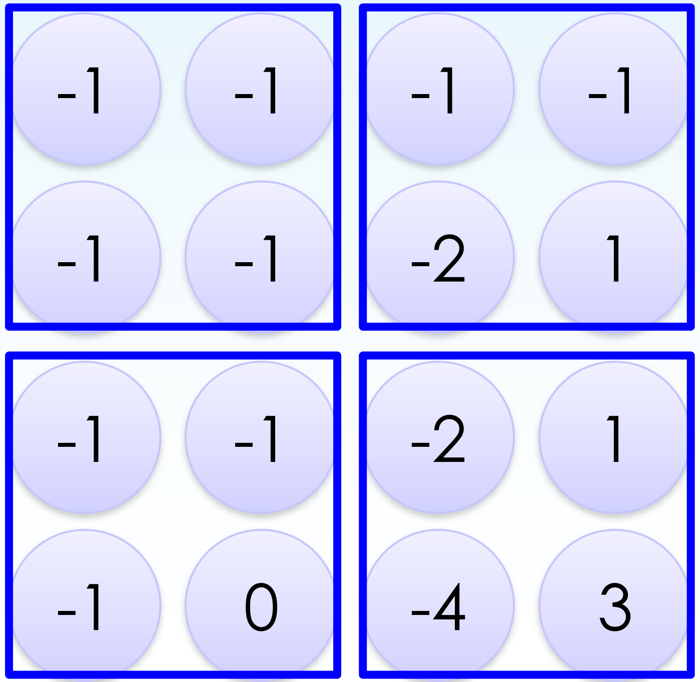
Filter 1

-1	1	-1
-1	1	-1
-1	1	-1

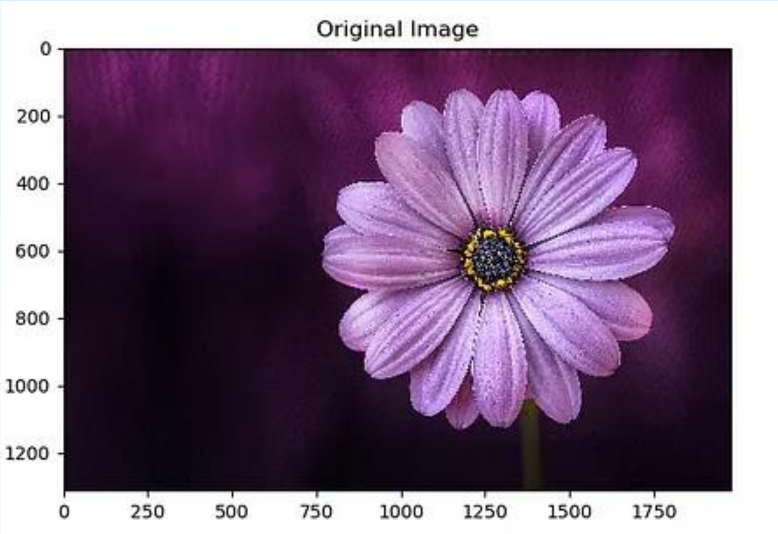
Filter 2

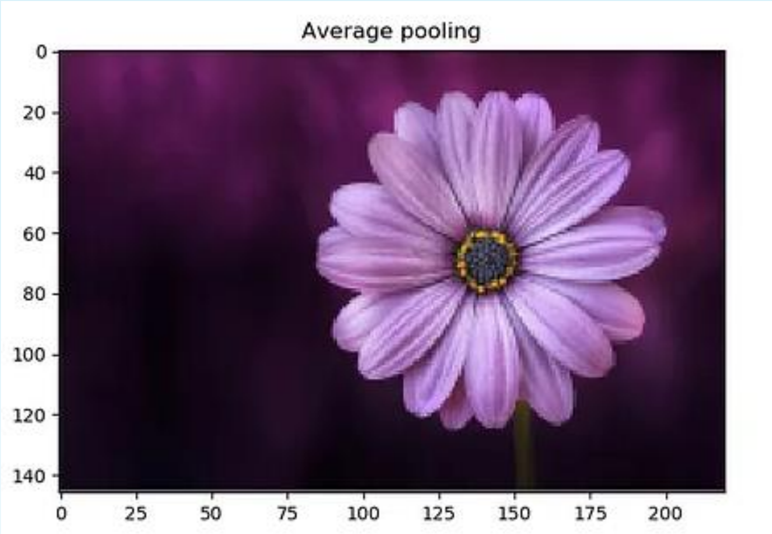


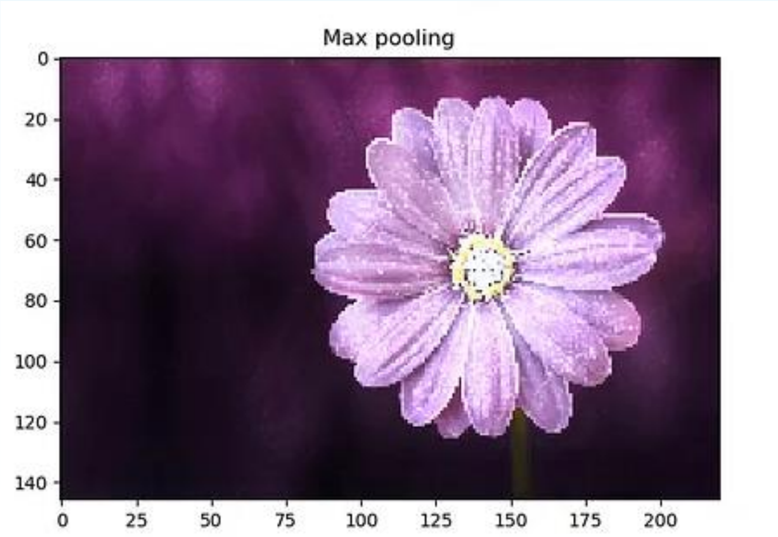
3	-1	-3	-1
-3	1	0	-3
-3	-3	0	1
3	-2	-2	-1

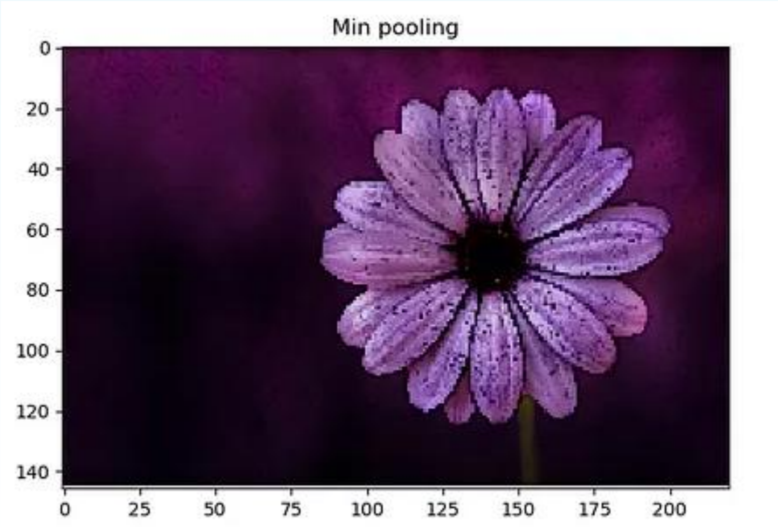


-1	-1	-1	-1
-1	-1	-2	1
-1	-1	-2	1
-1	0	-4	3

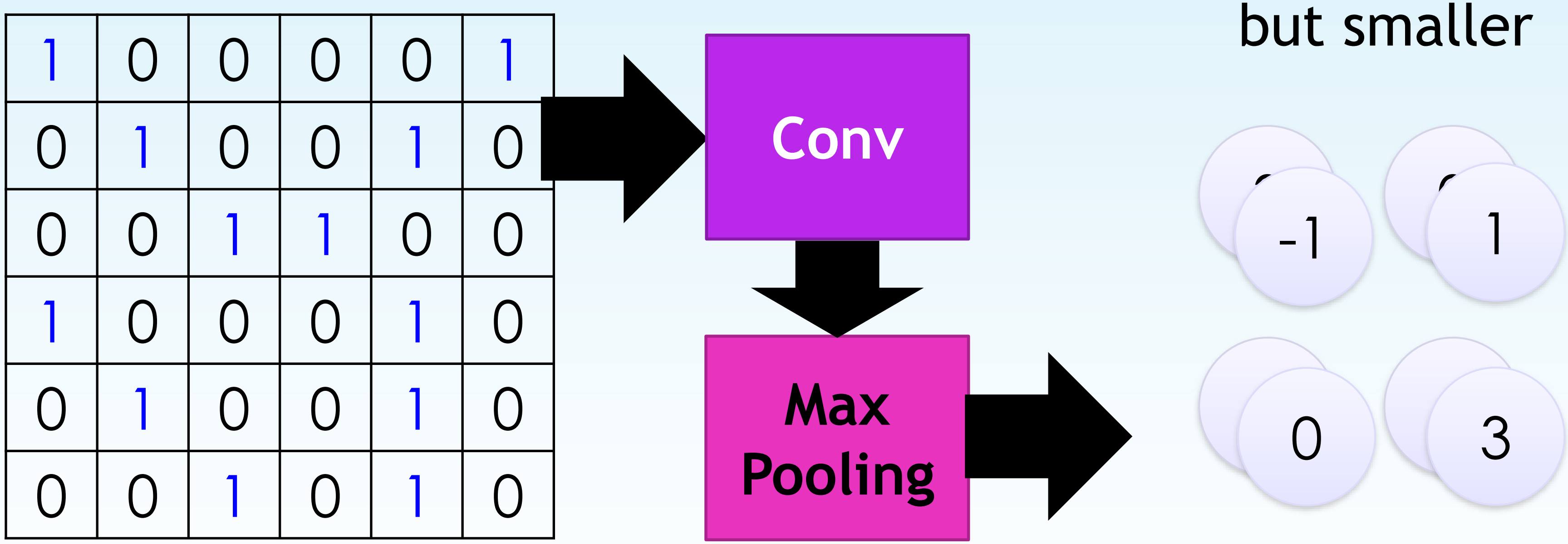

Original Image
0
200
400
600
800
1000
1200
0
250
500
750
1000
1250
1500
1750


Average pooling
0
20
40
60
80
100
120
140
0
25
50
75
100
125
150
175
200


Max pooling
0
20
40
60
80
100
120
140
0
25
50
75
100
125
150
175
200


Min pooling
0
20
40
60
80
100
120
140
0
25
50
75
100
125
150
175
200

Max Pooling



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

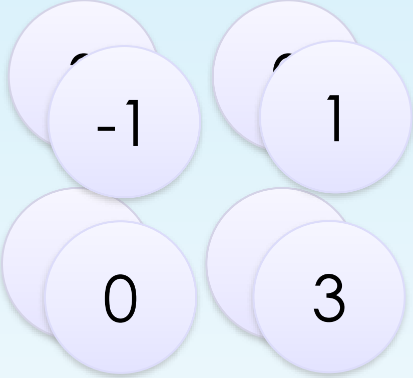
6 x 6 image

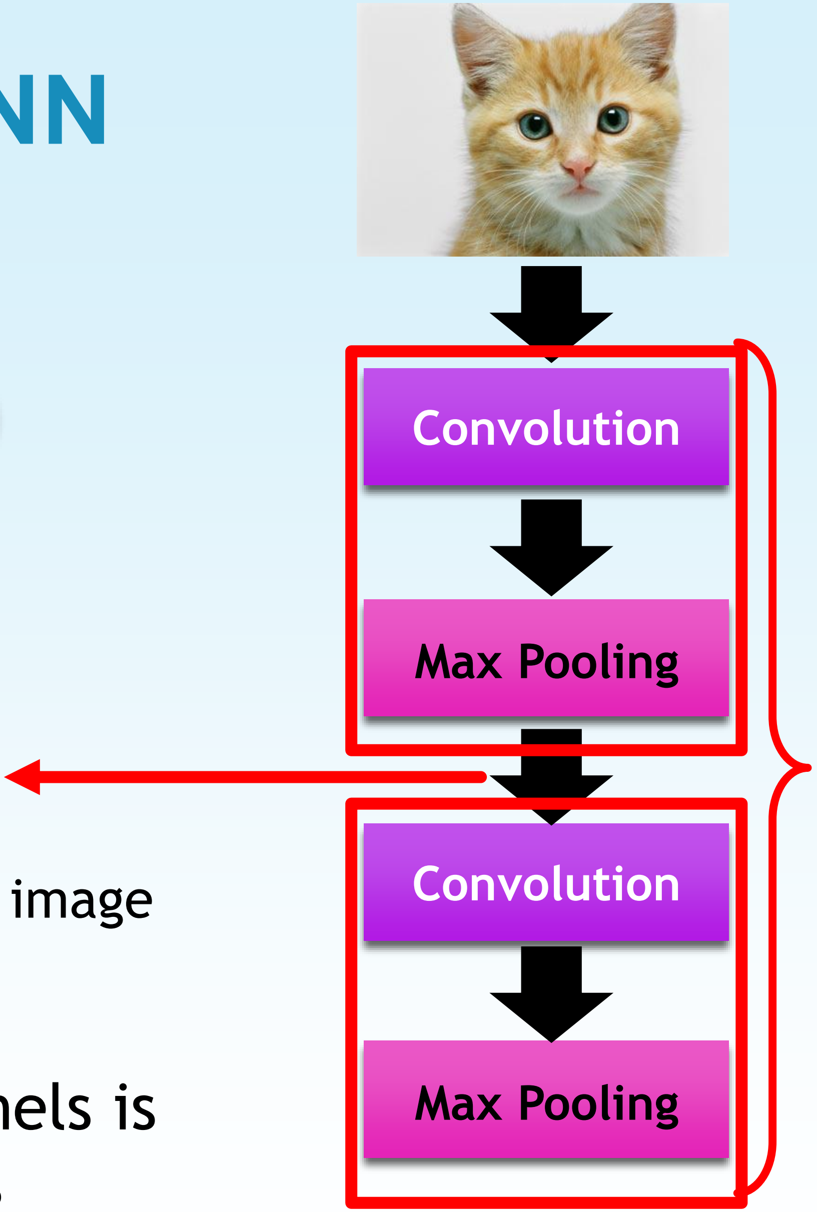
New image
but smaller

2 x 2 image

Each filter
is a channel

The whole CNN





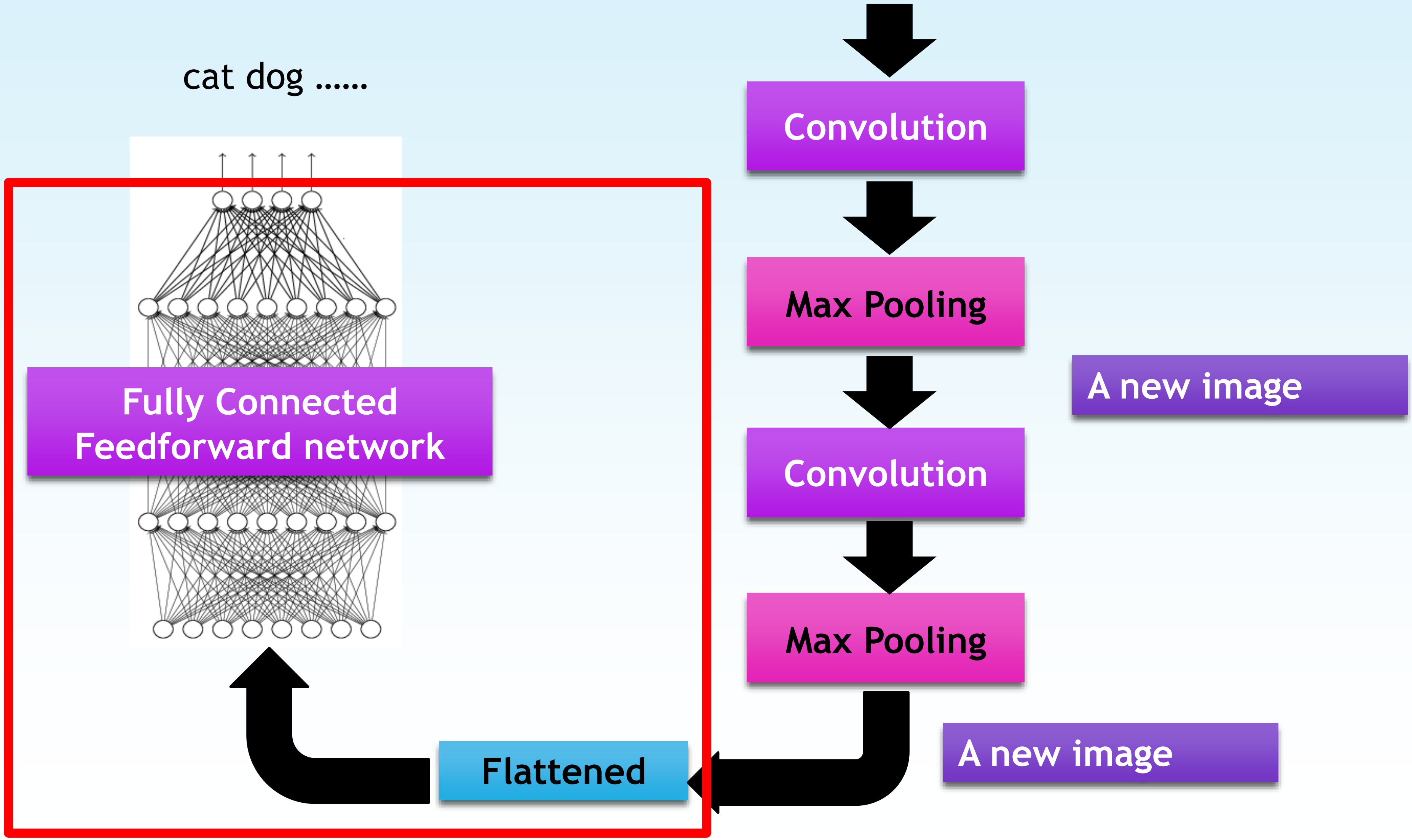
A new image

Smaller than the original image

The number of channels is the number of filters

Can repeat many times

The whole CNN



cat dog ……

Convolution

Max Pooling

A new image

Convolution

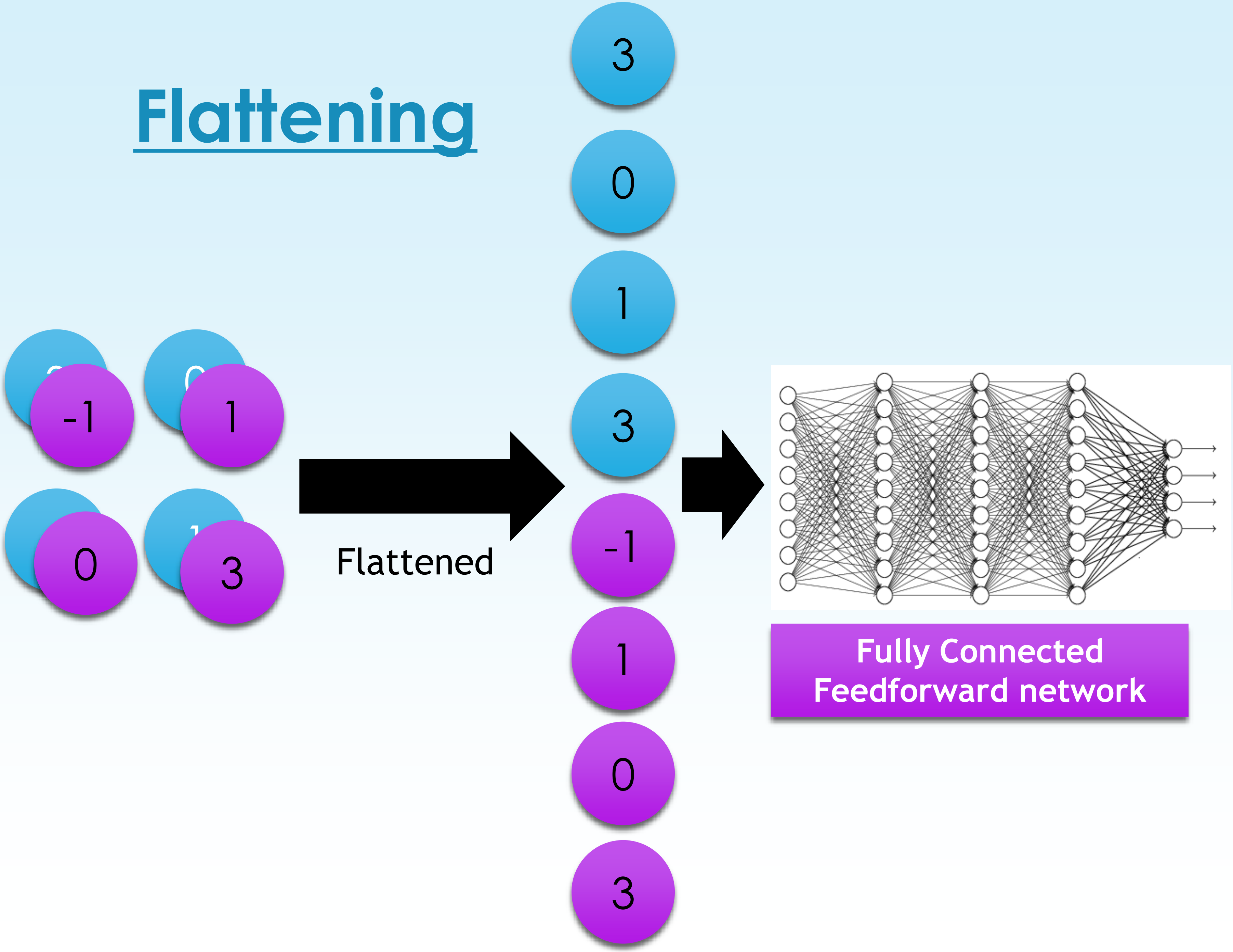
Max Pooling

A new image

Flattened

Fully Connected Feedforward network

Flattening

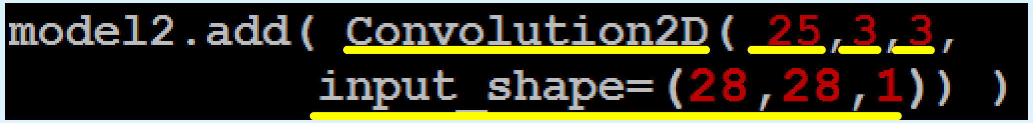


Flattened

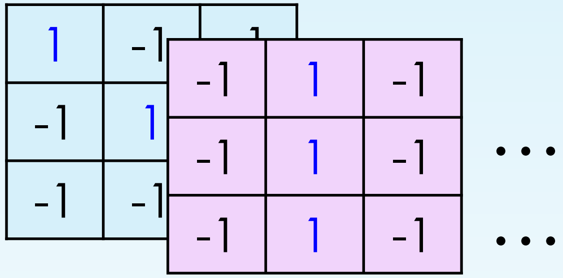
Fully Connected Feedforward network

CNN in Keras

Only modified the ***network structure*** and ***input format (vector -> 3-D tensor)***



```
model2.add( Convolution2D( 25,3,3,
             input_shape=(28,28,1)) )
```

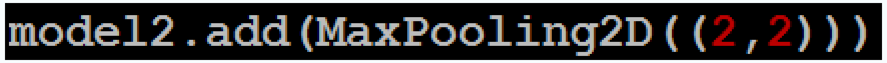


There are 25 3x3 filters.

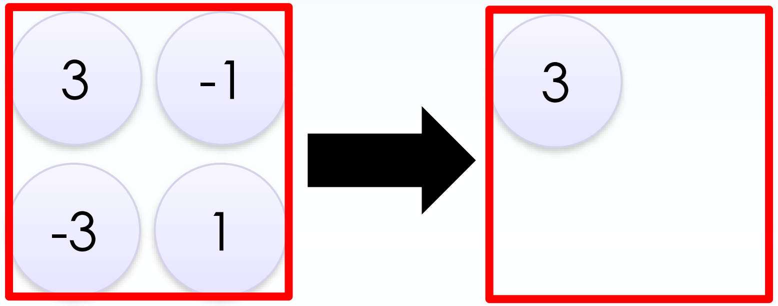
Input_shape = (28 , 28 , 1)

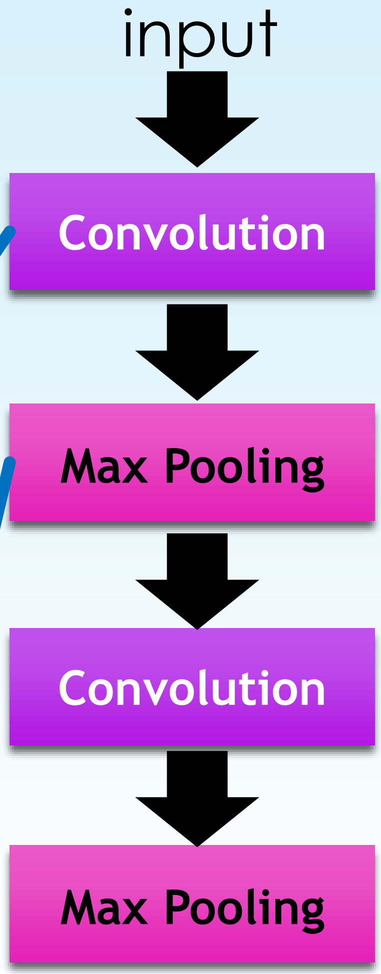
28 x 28 pixels

1: black/white, 3: RGB



```
model2.add(MaxPooling2D((2,2)))
```



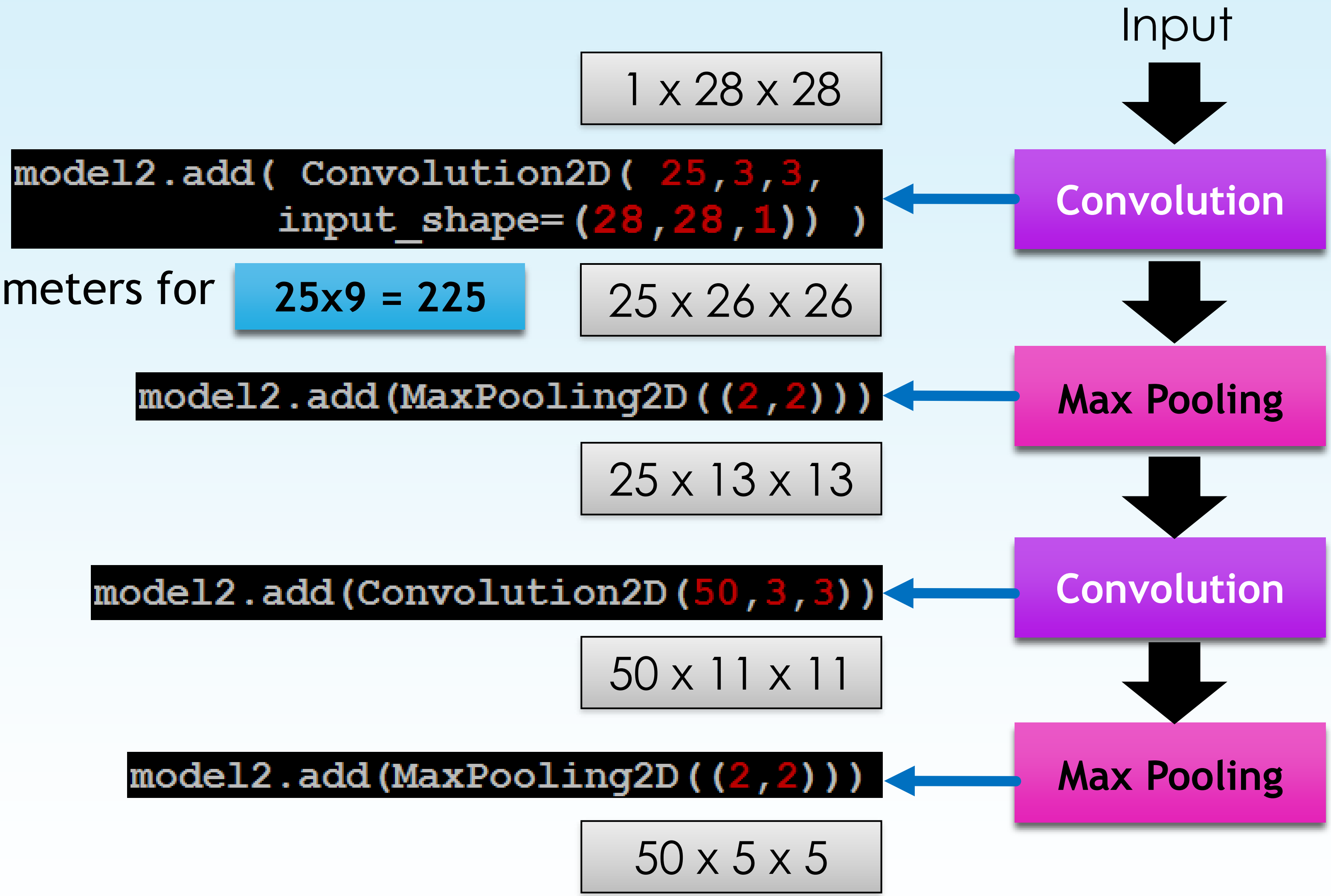


CNN in Keras

Only modified the ***network structure*** and ***input format (vector -> 3-D array)***

How many parameters for each filter?

25x9 = 225



```
model2.add( Convolution2D( 25,3,3,
            input_shape=(28,28,1)) )
```

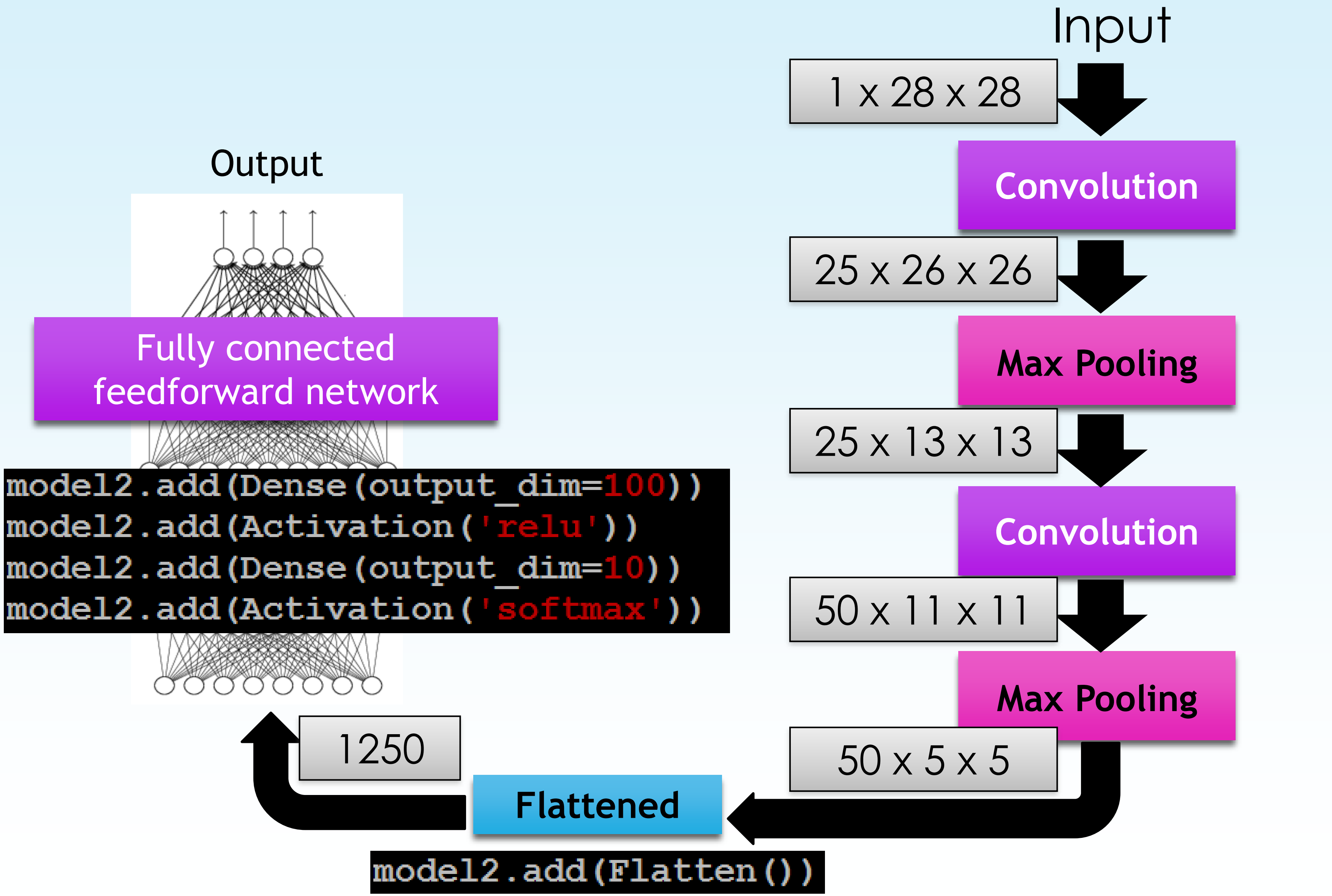
```
model2.add(MaxPooling2D((2,2)))
```

```
model2.add(Convolution2D(50,3,3))
```

```
model2.add(MaxPooling2D((2,2)))
```

CNN in Keras

Only modified the ***network structure*** and ***input format (vector -> 3-D array)***

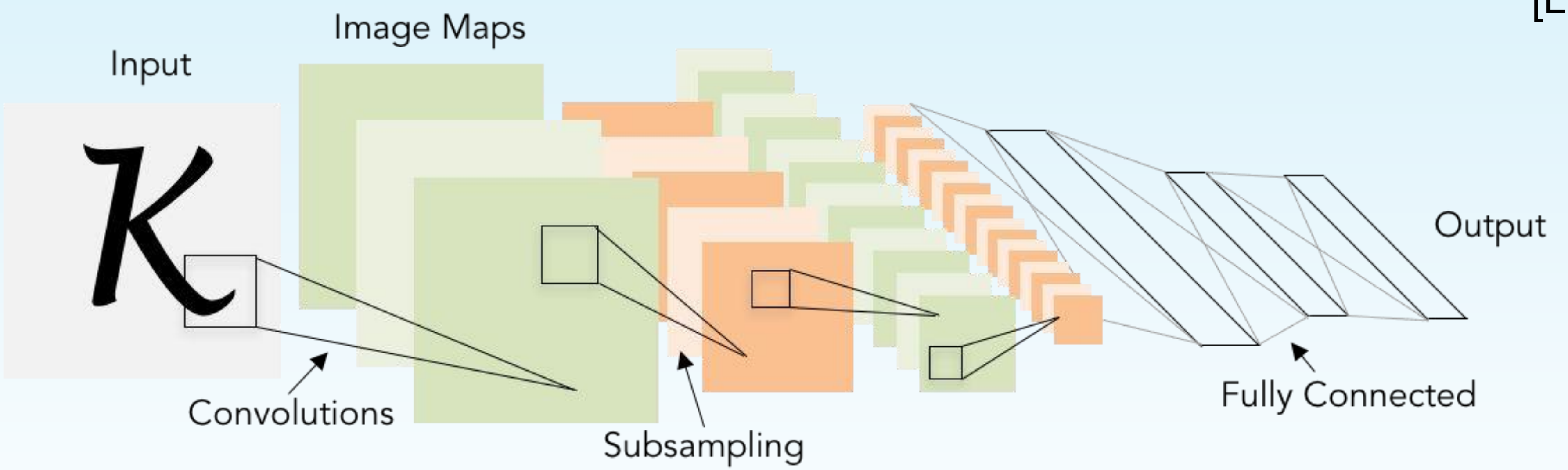


```
model2.add(Dense(output_dim=100))
model2.add(Activation('relu'))
model2.add(Dense(output_dim=10))
model2.add(Activation('softmax'))
```

```
model2.add(Flatten())
```

CNN Architectures

LeNet-5
[LeCun et al., 1998]

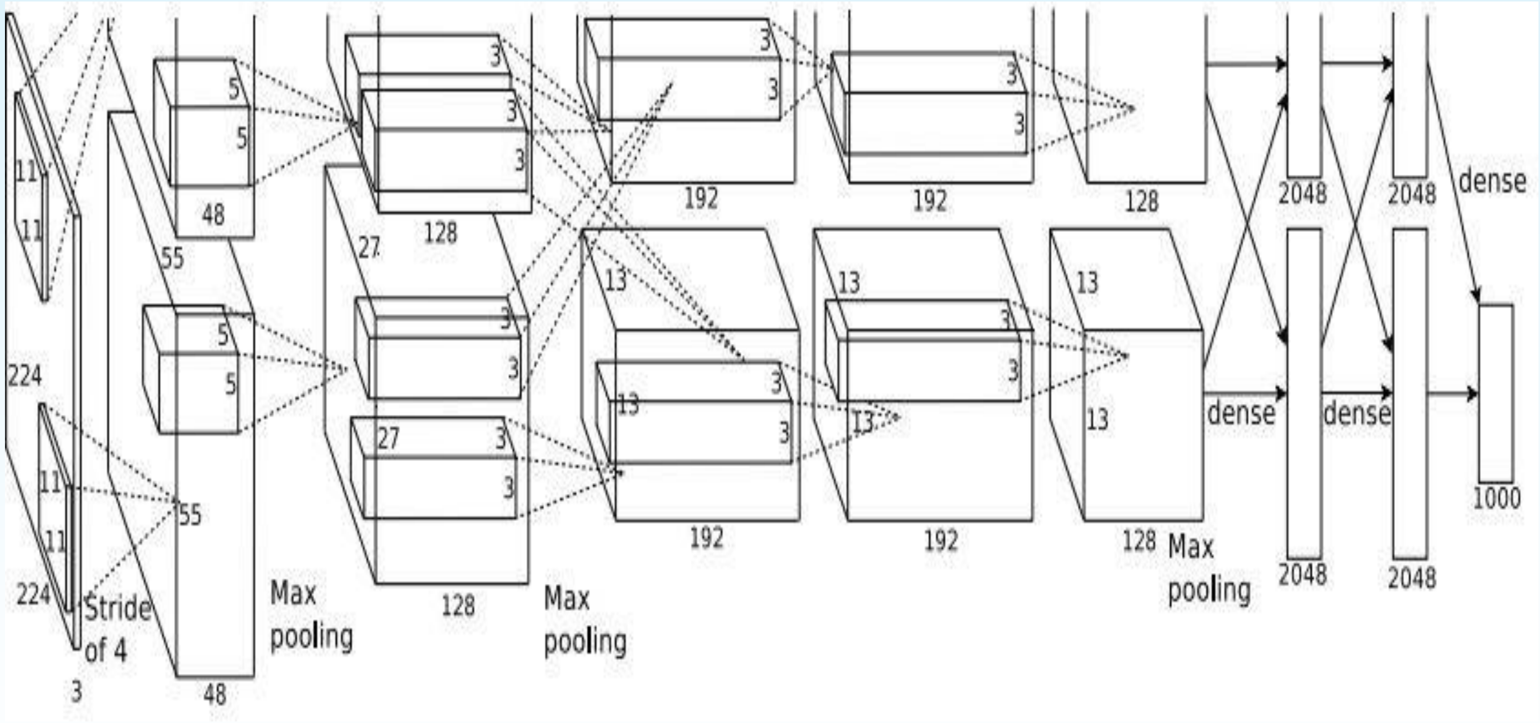


- Conv filters were 5x5, applied at stride 1
- Subsampling (Pooling) layers were 2x2 applied at stride 2
- i.e. architecture is [CONV-POOL-CONV-POOL-FC-FC]

AlexNet

[Krizhevsky et al. 2012]

Architecture:
CONV1
MAX POOL1
NORM1
CONV2
MAX POOL2
NORM2
CONV3
CONV4
CONV5
Max POOL3
FC6
FC7
FC8



What is ImageNet DataSet

- It is a large dataset of annotated photographs.
- This dataset consists of about 14 million images, more than 21000 groups or classes and more than 1 million images that have bounding box annotation.

ImageNet large scale visual recognition Challenge (ILSVRC)

- ImageNet large scale visual recognition Challenge for short ILSVRC. The goal of this challenge is to train a model that can correctly classify an image into a class out of 1000 separate object categories.

ImageNet Large Scale Visual Recognition Challenge (ILSVRC) winners

- AlexNet
- GoogLeNet (e.g. InceptionN),
- VGGNet (e.g. VGG16 or VGG19),
- Residual Network (e.g. ResNetN)

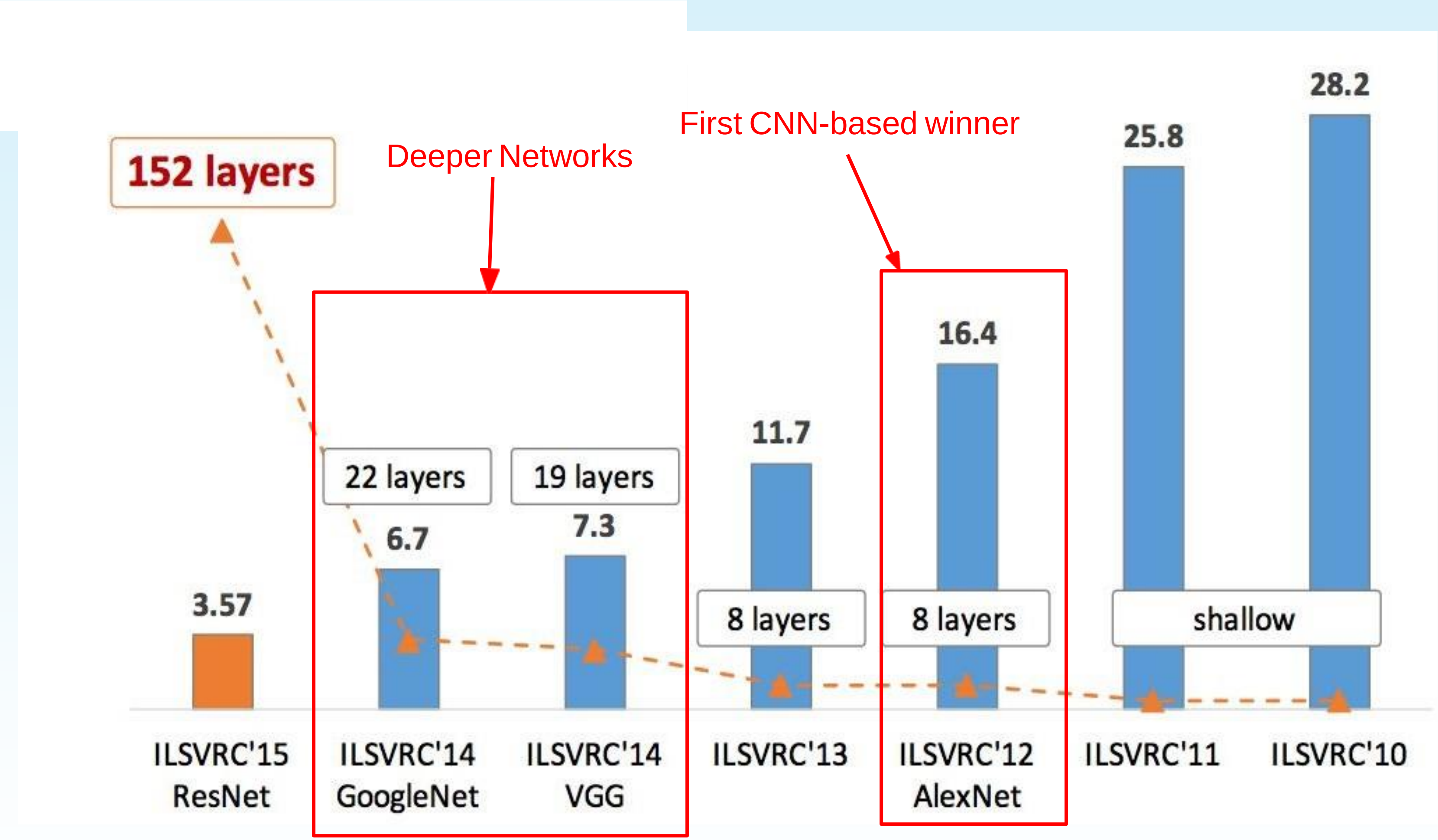
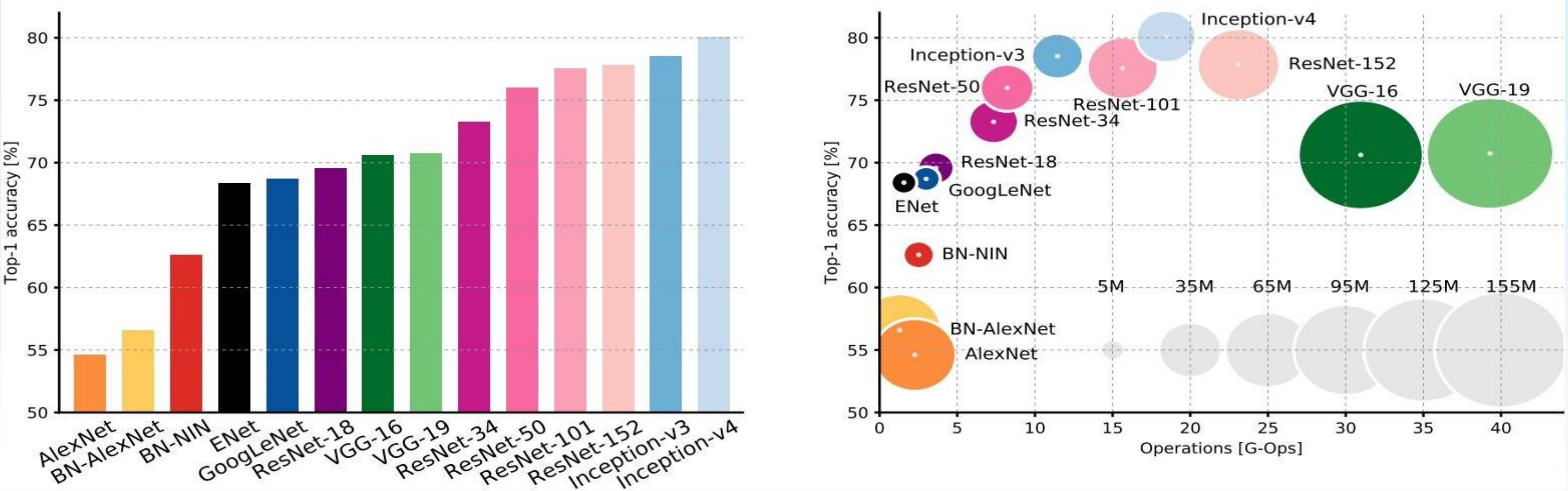


Fig: ILSVRC winners (with training error % vs Years)

Figure copyright Kaiming He, 2016. Reproduced with permission.



An Analysis of Deep Neural Network Models for Practical Applications, 2017.

Figures copyright Alfredo Canziani, Adam Paszke, Eugenio Culurciello, 2017. Reproduced with permission.

VGG: Highest memory, most operations.

GoogLeNet: Most efficient.

AlexNet: Smaller compute, still memory heavy, lower accuracy.

ResNet: Moderate efficiency depending on model, highest accuracy

Transfer Learning Model

- Transfer learning is a technique whereby a neural network model is first trained on a problem similar to the problem that is being solved.

- One or more layers from the trained model are then used in a new model hence, transfer learning is a method of reusing a pre-trained model knowledge for another task.

- There are perhaps a dozen or more top-performing models for image recognition that can be used. AlexNet, ResNet, VGG, and Inception etc. are some of the CNN based transfer learning model.

Thank You