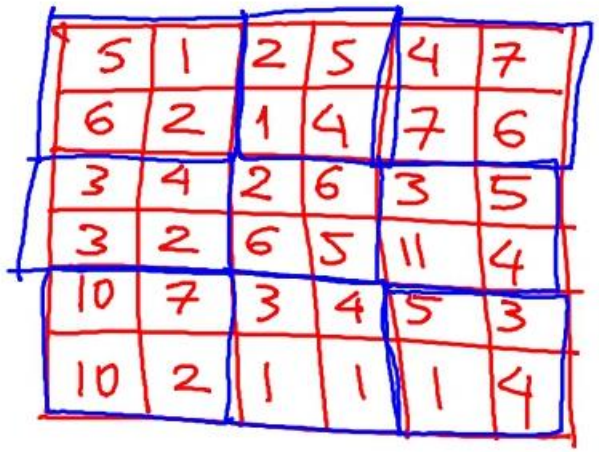
Why deep networks work

1. GPUs
2. Big data
   - Available.
   - Data augmentation.
3. Activation functions (ReLU's)
4. Dropout method.
5. Pooling.



6	5	7
4	6	11
10	4	5

Max pooling

3,5	3	6
..	.	. ..
.	...	. .

Average pooling.

Convolution + ReLU + pooling.

1	6	-3	2	5
-6	-1	4	1	5
7	-5	7	3	1
3	2	-4	5	2
1	-5	2	-3	6

Input

2	1
-1	-2

Kernel

Convolution:

$$\begin{bmatrix} 1 & 6 \\ -6 & -1 \end{bmatrix} * \begin{bmatrix} 2 & 1 \\ -1 & -2 \end{bmatrix} = 1 \cdot 2 + 6 \cdot 1 + (-6) \cdot (-1) + (-1)(-2) = 16$$

$$\begin{bmatrix} 6 & -3 \\ -1 & 4 \end{bmatrix} * \begin{bmatrix} 2 & 1 \\ -1 & -2 \end{bmatrix} = 12 - 3 + 1 - 8 = 2$$

Result

16	2	-10	-2
-10	-7	-4	3
2	3	11	6
17	1	1	-1

ReLU →

16	2	0	0
0	0	0	3
2	3	11	6
17	1	1	0

Max Pooling

16	2
17	11

(output)

Average pooling

4,5	0,5
5,75	4,5

(Output)

On MNIST database.

① 1 hidden layer of 100 neurons (fully connected) (sigmoid a.f.) 97,8%.

② ① + 1 conv. layer + max pooling. 98,78%. (sigmoid a.f.)

③ ② but with 2 conv. layers + pooling. 99,06%.

④ ③ + ReLU + R_2 regularisation 99,23%.

(5) (4) + augmented (shifted) patterns 99,37%

(6) (5) + dropout + 2 layers in the classifier part 99,6%

(7) (6) + Ensembling. 99,7%