Convolutional Neural Networks - CNN

The Power of Deep Learning for Image Recognition

February 6, 2023

Primoz Konda
pk@business.aau.dk

Aalborg University Business School

AALBORG UNIVERSITY
DENMARK

Agenda

Wider context

Computer vision

CNN model
- Convolution layer
- Downsampling
- Flattening

R-CNN

Coding example

Discussion

References

Top growing technologies in 2022

According to the latest ifiCLAIMS report[1], the fastest growing technology was:

- AUTONOMOUS DRIVING
- modern EV cars have 20+ sensors and cameras
- 9th place: Machine learning

[1] https://www.ificlaims.com/rankings-tech-growth-2022.htm

Why good vision is important?

By far the most important organs of sense are our eyes.

What per cent of all impressions do we perceive by means of our sight?

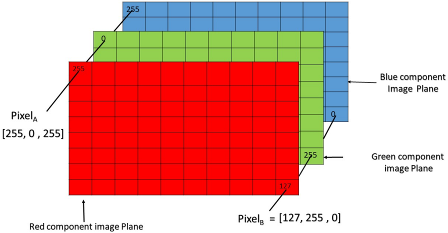
- 80%

As vision plays a crucial role in our daily lives, CNNs have made it possible for machines to mimic human vision and process images effectively, improving efficiency and accuracy in various fields.

How computers "see"?

0	2	15	0	0	11	10	0	0	0	0	9	9	0	0	0
0	0	0	4	60	157	236	255	255	177	95	61	32	0	0	29
0	10	16	119	238	255	244	245	243	250	249	255	222	103	10	0
0	14	170	255	255	244	254	255	253	245	255	249	253	251	124	1
2	98	255	228	255	251	254	211	141	116	122	215	251	238	255	49
13	217	243	255	155	33	226	52	2	0	10	13	232	255	255	36
16	229	252	254	49	12	0	0	7	7	0	70	237	252	235	62
6	141	245	255	212	25	11	9	3	0	115	236	243	255	137	0
0	87	252	250	248	215	60	0	1	121	252	255	248	144	6	0
0	13	113	255	255	245	255	182	181	248	252	242	208	36	0	19
1	0	5	117	251	255	241	255	247	255	241	162	17	0	7	0
0	0	0	4	58	251	255	246	254	253	255	120	11	0	1	0
0	0	4	97	255	255	255	248	252	255	244	255	182	10	0	4
0	22	206	252	246	251	241	100	24	113	255	245	255	194	9	0
0	111	255	242	255	158	24	0	0	6	39	255	232	230	56	0
0	218	251	250	137	7	11	0	0	0	2	62	255	250	125	3
0	173	255	255	101	9	20	0	13	3	13	182	251	245	61	0
0	107	251	241	255	230	98	55	19	118	217	248	253	255	52	4
0	18	146	250	255	247	255	255	255	249	255	240	255	129	0	5
0	0	23	113	215	255	250	248	255	255	248	248	118	14	12	0
0	0	6	1	0	52	153	233	255	252	147	37	0	0	4	1
0	0	5	5	0	0	0	0	0	14	1	0	6	6	0	0

How computers "see"?



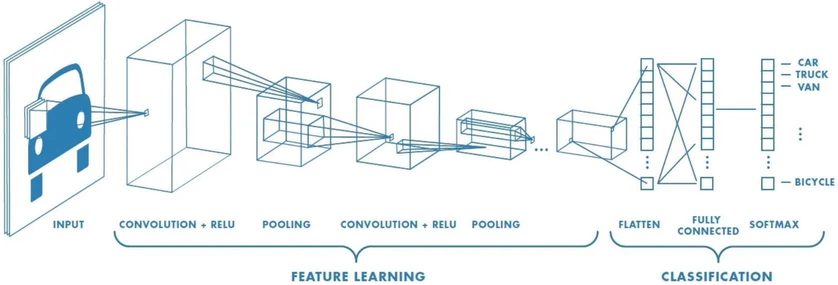
Pixel of an RGB image are formed from the corresponding pixel of the three component images

How humans recognize objects?

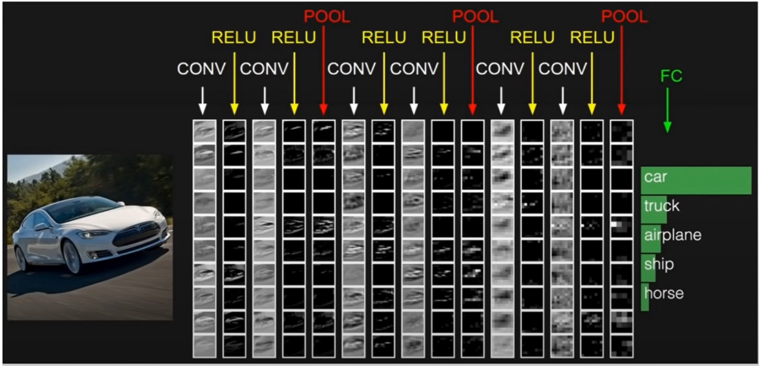
How do we recognize a car?

Is this electric car?

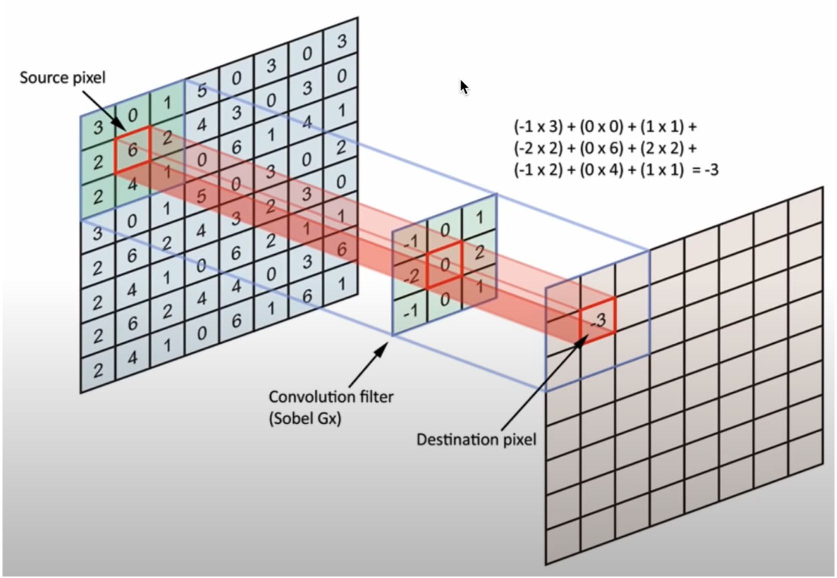
CNN architecture



Visual process



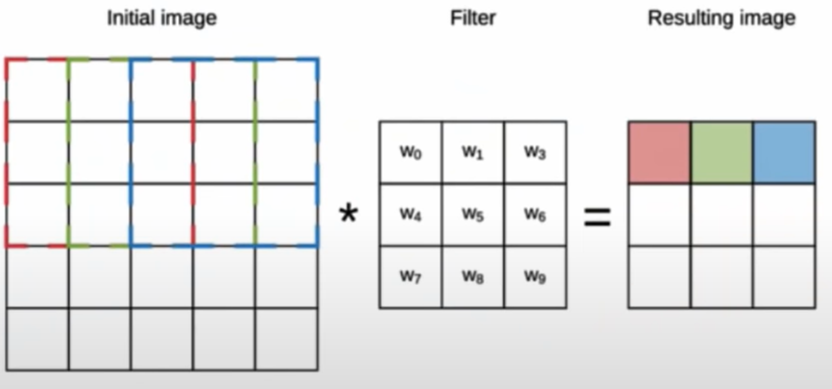
Convolution layer



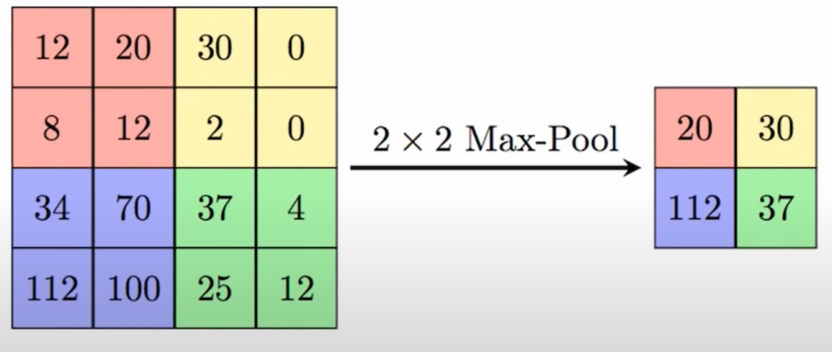
Convolution filter

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)	$\frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$	

Convolution layer



Downsampling

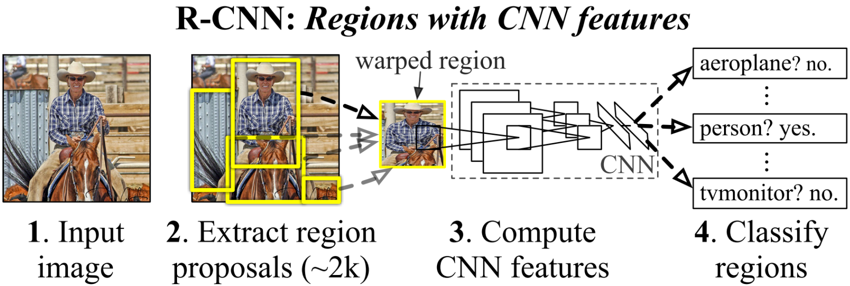


Fully connected layers

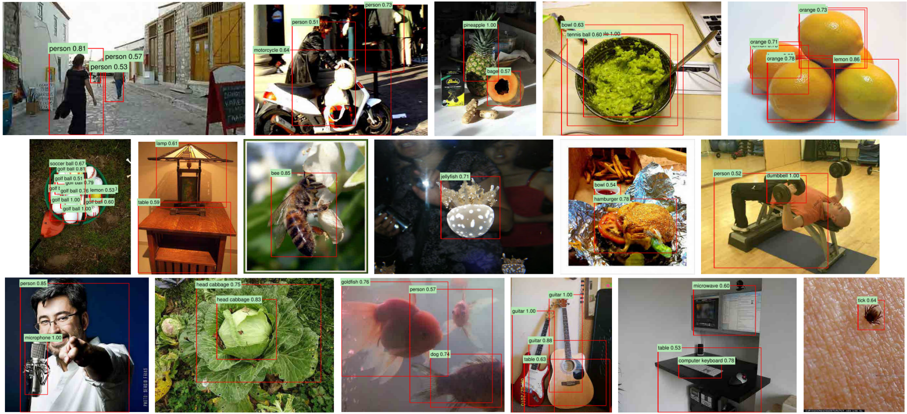
First fully connected layers take the output of the previous layers, “flattens” them and turns them into a single vector that can be an input for the next stage.

Fully connected output layer gives the final probabilities for each label.

R-CNN


R-CNN: *Regions with CNN features*
warped region
aeroplane? no.
person? yes.
tvmonitor? no.
CNN
1. Input image
2. Extract region proposals (~2k)
3. Compute CNN features
4. Classify regions

R-CNN



person 0.81
person 0.57
person 0.53
person 0.51
person 0.73
motorcycle 0.64
pineapple 1.00
bagel 0.57
bowl 0.63
tennis ball 0.60
orange 0.73
orange 0.71
orange 0.78
lemon 0.86
soccer ball 0.67
golf ball 0.81
golf ball 0.51
golf ball 0.76
lemon 0.53
golf ball 1.00
golf ball 0.60
golf ball 1.00
lamp 0.61
table 0.59
bee 0.85
jellyfish 0.71
bowl 0.54
hamburger 0.78
person 0.52
dumbbell 1.00
person 0.85
microphone 1.00
head cabbage 0.75
head cabbage 0.83
goldfish 0.76
person 0.57
dog 0.74
guitar 1.00
guitar 1.00
guitar 0.88
table 0.63
microwave 0.60
table 0.53
computer keyboard 0.78
tick 0.64

Dataset CIFAR-10

- 60.000 images
- RGB (3 channels)
- 32x32 pixel images
- 10 classes (5000 train 1000 test for each class)

Dataset CIFAR-10

Results

Model Name	Accuracy of the network:
e6, b4, r00.1	51.69%
e8, b4, r00.1	54.70%
e10, b4, r00.1	58.57%
e10, b4, r00.3	62.41%

Results

Class:	Accuracy:
Accuracy of the network:	62.41 %
Accuracy of plane:	54.2 %
Accuracy of car:	74.4 %
Accuracy of bird:	57.3 %
Accuracy of cat:	54.3 %
Accuracy of deer:	53.6 %
Accuracy of dog:	43.5 %
Accuracy of frog:	78.7 %
Accuracy of horse:	61.3 %
Accuracy of ship:	80.9 %
Accuracy of truck:	65.9 %

Model optimization

- Hyper-parameter tuning: learning rate, batch size, epochs, number of filters in each convolutional layer
- Data augmentation: increase dataset size by rotating, flipping, scaling
- Transfer learning: using a pre-trained models and fine-tune with your dataset
- Regularization: dropout, L1/L2 regularization, and early stopping
- Change architecture: find the optimal network architecture for a given problem (use academic articles)
- ...

Business applications

what, how, why, and for who

References

- O'Shea, K., Nash, R. (2015). **An introduction to convolutional neural networks.** arXiv preprint arXiv:1511.08458.
- Albawi, S., Mohammed, T. A., Al-Zawi, S. (2017). **Understanding of a convolutional neural network.** In 2017 international conference on engineering and technology (ICET) (pp. 1-6). Ieee.
- Ajit, A., Acharya, K., Samanta, A. (2020, February). **A review of convolutional neural networks.** In 2020 international conference on emerging trends in information technology and engineering (ic-ETITE) (pp. 1-5). IEEE.
- Girshick, R., Donahue, J., Darrell, T., Malik, J. (2013). **Rich feature hierarchies for accurate object detection and semantic segmentation.** arXiv e-prints. arXiv preprint arXiv:1311.2524, 396.