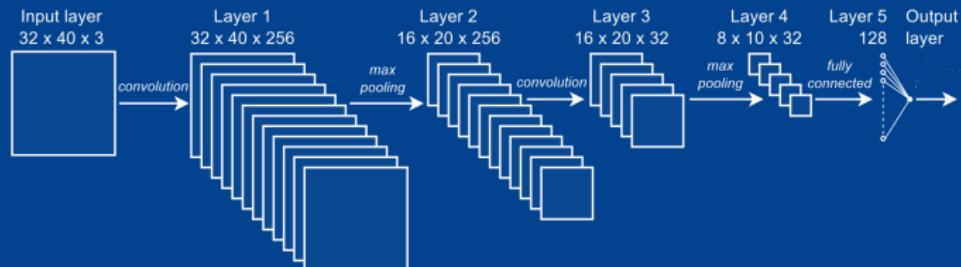




LESSON 09: Deep Learning + CNN's + Hardware

CARSTEN EIE FRIGAARD

SPRING 2021



"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E." — Mitchell (1997).

L09: DL + CNN's + HW, Agenda

- ▶ Spørge-minutter..
- ▶ Admin:
 - ▶ O3 opgavesæt afsluttet, Deadline 14/4, 18:00:
 - ▶ Opgave fra L07: `capacity_under_overfitting.ipynb`
 - ▶ Opgave fra L07: `generalization_error.ipynb`
 - ▶ Opgave fra L08: `regulizers.ipynb`
 - ▶ Opgave fra L08: `gridsearch.ipynb`
 - ▶ **'Cheat'-review** af O3: før aflevering, se beskrivelse i O1+3
'cheat'-review. Resume skal denne gang inkluderes i O3 afleveringen.
 - ▶ Ændringer i "Almindelige formelle journal krav":
 - ▶ NY: Ingen direkte genbrug af overskrifter ...
 - ▶ NY: Skriv kun på eet sprog igennem hele journalen..

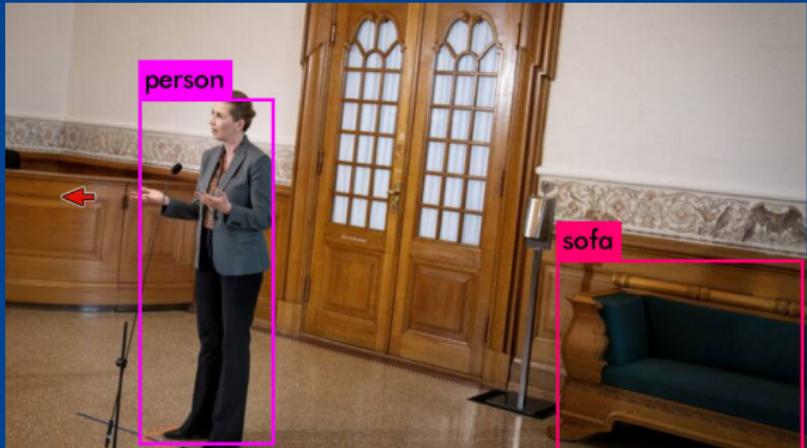
▶ Convolutional Neural Networks (CNN's),

- ▶ and Deep-learning (DL).

▶ Hardware for Machine Learning,

- ▶ (CPUs), GPUs, TPUs, Exotic Hardware.

CONVOLUTIONAL NEURAL NETWORKS



Deep-learning (DL)

Artificial Intelligence:

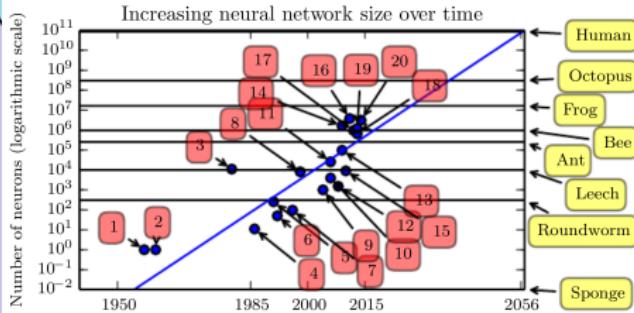
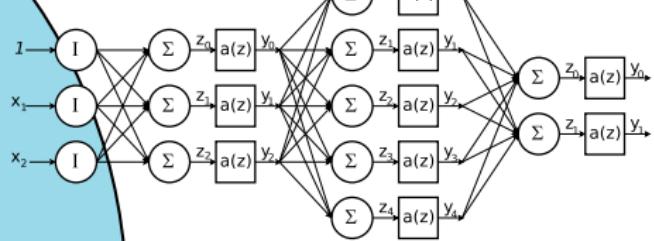
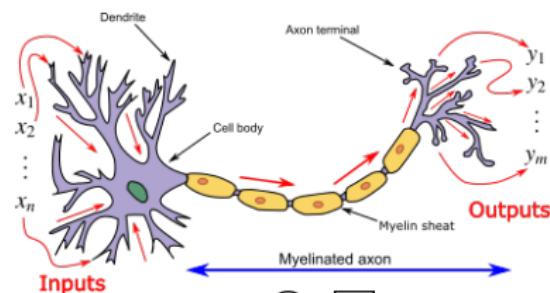
Mimicking the intelligence or behavioural pattern of humans or any other living entity.

Machine Learning:

A technique by which a computer can "learn" from data, without using a complex set of different rules. This approach is mainly based on training a model from datasets.

Deep Learning:

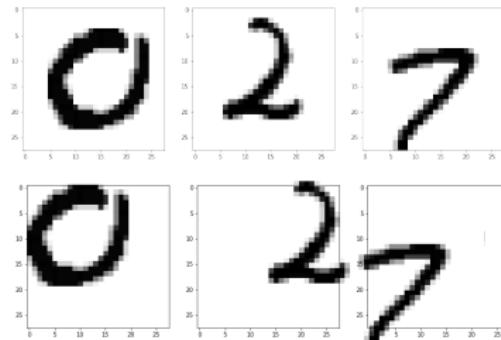
A technique to perform machine learning inspired by our brain's own network of neurons.



Definition of DL=??

Preprocessing/Feature-extraction + Machine Learning

Can your ML model handle simple image translation (rotation and scaling)?



..no problem for Your visual cortex, right?

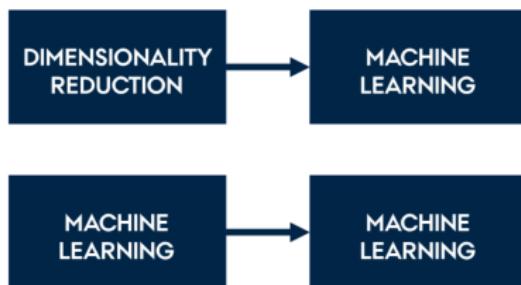
Introducing the Convolutional Neural Networks: trade off

- ▶ preprocessing/feature-extraction + classification with:
- ▶ feature learning (CNN kernels) + classification (fully connected NN)

Convolutional Neural Networks

Feature Extraction vs. Feature Learning

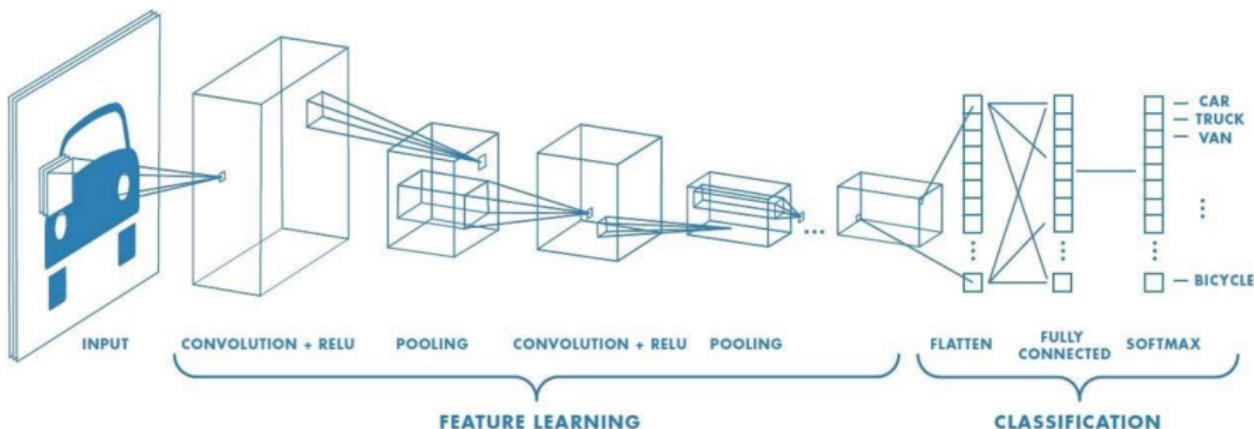
- ▶ Smart filtering:
the distinction between dimensionality reduction and machine learning blurs..



- ▶ Fundamental problem of filtering:
What is noise, what is signal?

Convolutional Neural Networks

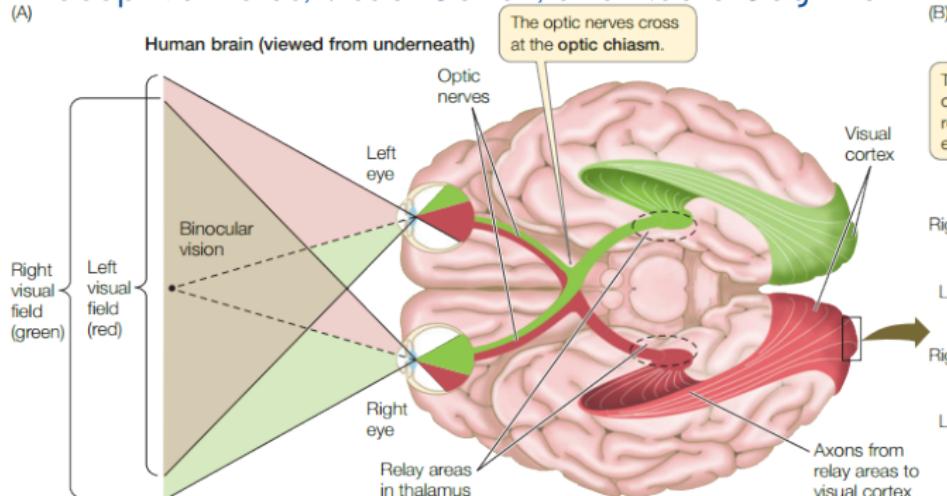
Feature Extraction vs. Feature Learning



Convolutional Neural Networks

Human Eye and Brain Image Processing: Receptive Fields, Visual Cortex, and Neuro Cognition

(A)



(B)

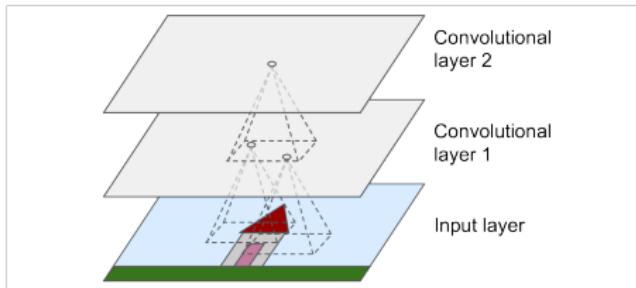
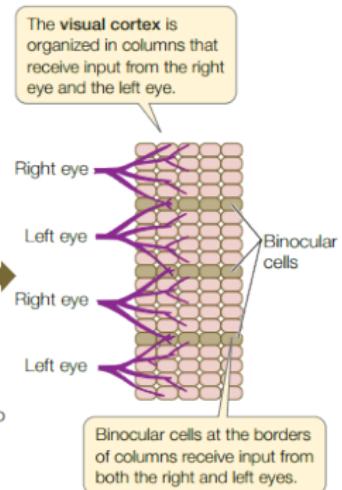


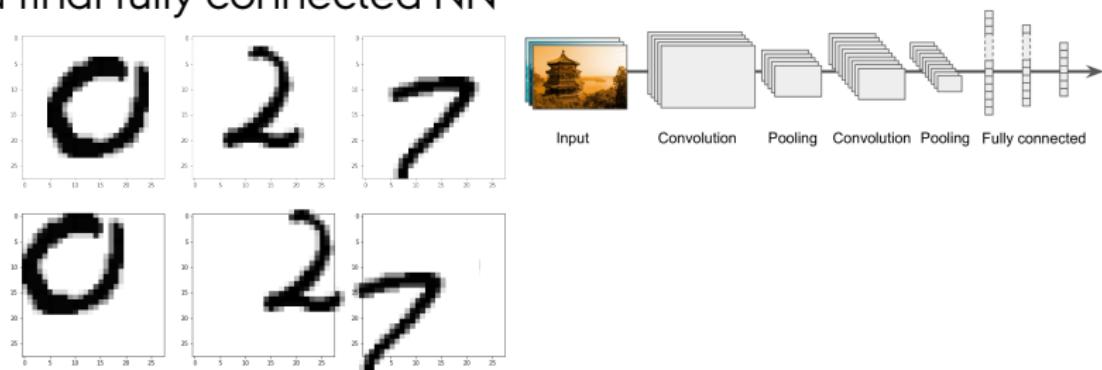
Figure 14-2. CNN layers with rectangular local receptive fields

Feature Learning + Machine Learning

CNN principle

Translation, rotation, and scaling invariant

- ▶ automatic image **feature extraction** that is **feature learning** via
 - ▶ 'convolutional', 'pooling' CNN layers, then a final fully connected NN



..no problem for Your CNN ML model, right?

Feature Learning + Machine Learning

CNN principle: Convolutional Layer (via CNN kernels)

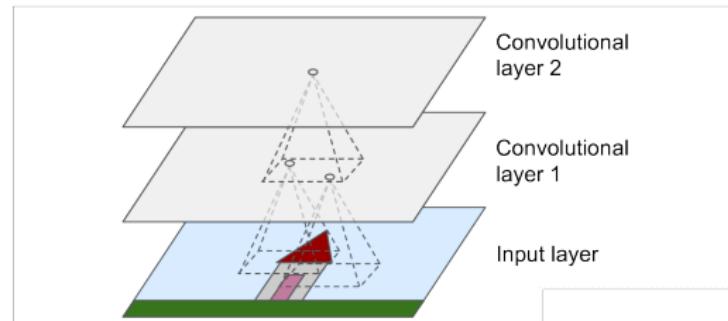


Figure 14-2. CNN layers with rectangular local receptive fields

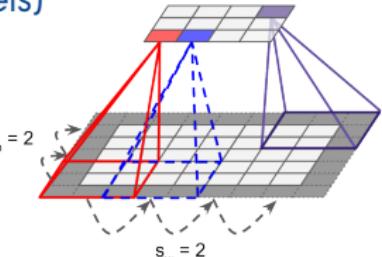
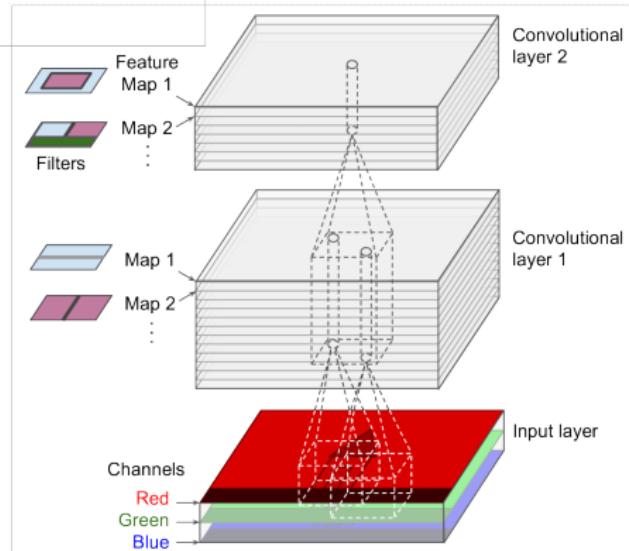


Figure 14-4. Reducing dimensionality using a stride of 2



Feature map 1 => with kernel 1

Feature map 2 => with kernel 2

Figure 14-6. Convolution layers with multiple feature maps, and images with three color channels

CNN Kernels

2D Convolution with a Kernel: Principle

0	0	0	0	0	0	0	0
0	60	113	56	139	85	0	0
0	73	121	54	84	128	0	0
0	131	99	70	129	127	0	0
0	80	57	115	69	134	0	0
0	104	126	123	95	130	0	0
0	0	0	0	0	0	0	0

Kernel

0	-1	0
-1	5	-1
0	-1	0

114				

$$5*60 + (-1)*113 + (-1)*73 = 114$$

CNN Kernels

2D Convolution with a Kernel: Different Kernels

-0.2	0.0	0.5
1.0	0.3	-0.6
0.0	0.0	0.8

0.0	0.0	0.0
0.8	-0.5	0.8
0.0	-0.2	0.0

0.4	0.2	-0.2
-0.8	0.0	0.8
0.0	-0.5	0.2



CNN Kernels in 3D

3D Convolution with a Kernel: Principle

0	0	0	0	0	0	...
0	156	155	156	158	158	...
0	153	154	157	159	159	...
0	149	151	155	158	159	...
0	146	146	149	153	158	...
0	145	143	148	148	158	...
...

Input Channel #1 (Red)

0	0	0	0	0	0	...
0	167	166	167	169	169	...
0	164	165	168	170	170	...
0	160	162	166	169	170	...
0	156	156	159	163	168	...
0	155	153	153	158	168	...
...

Input Channel #2 (Green)

0	0	0	0	0	0	...
0	163	162	163	165	165	...
0	160	161	164	166	166	...
0	156	158	162	165	166	...
0	155	155	158	162	167	...
0	154	152	152	157	167	...
...

Input Channel #3 (Blue)

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

Kernel Channel #1



308

+

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

Kernel Channel #2



-498

0	1	1
0	1	0
1	-1	1

Kernel Channel #3



164

+



Bias = 1

-25				...
				...
				...
				...
...

Output

CNN Pooling

The Principle

Max Pooling

29	15	28	184
0	100	70	38
12	12	7	2
12	12	45	6

2 x 2
pool size

100	184
12	45

Average Pooling

31	15	28	184
0	100	70	38
12	12	7	2
12	12	45	6

2 x 2
pool size

36	80
12	15

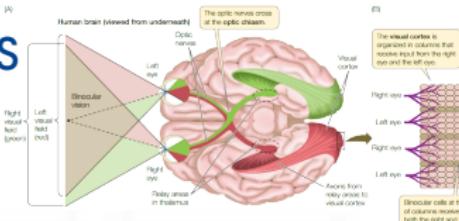
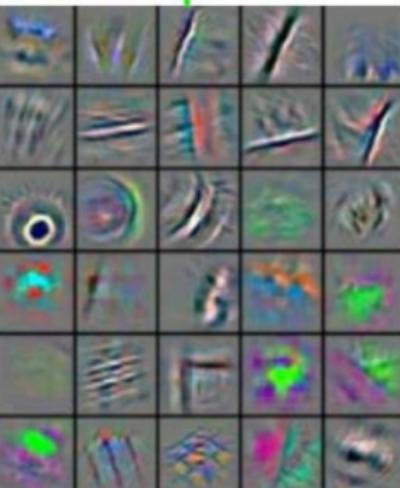
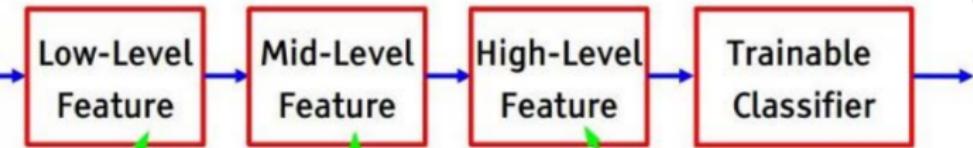
CNN Pooling

On real data (effectively subsampling)



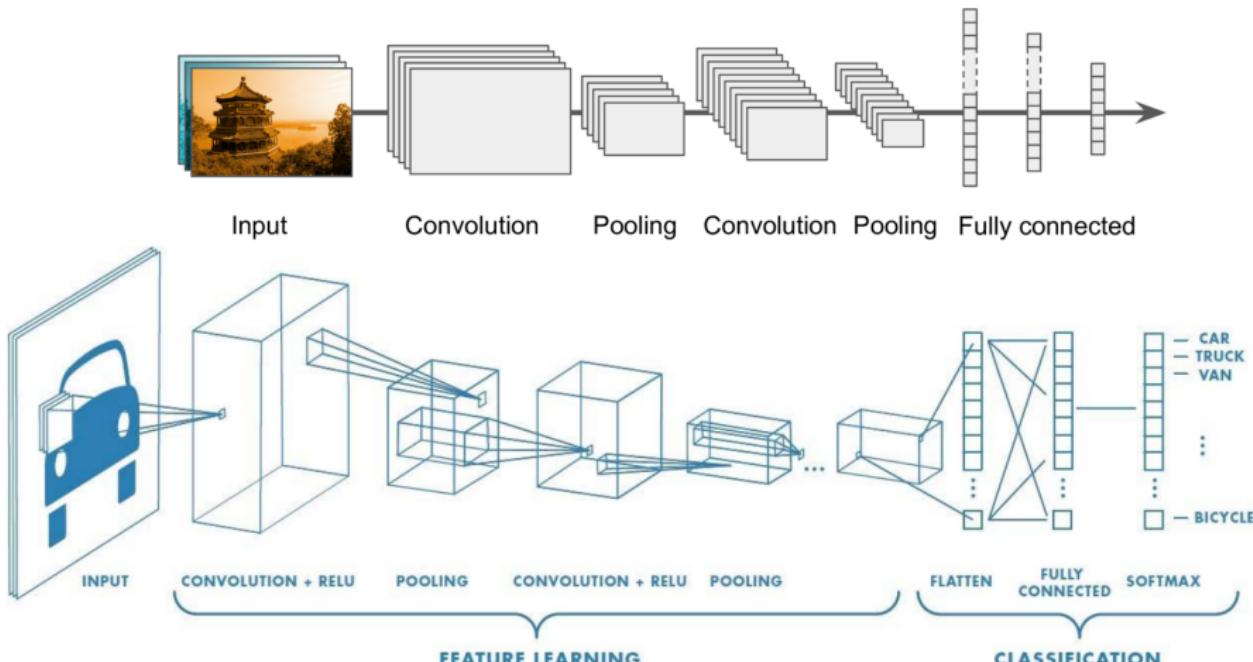
Convolutional Neural Networks

Low-, Mid-, and High-Level Feature Extraction



Convolutional Neural Networks

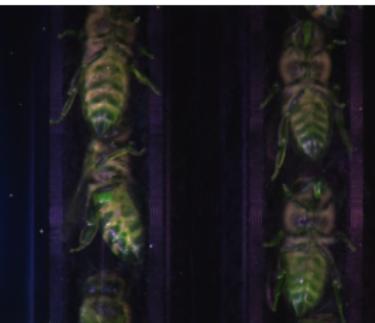
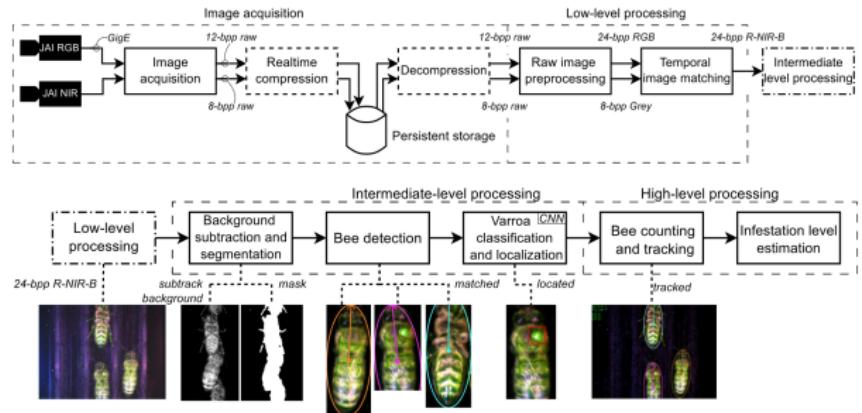
Stacking It All Up..



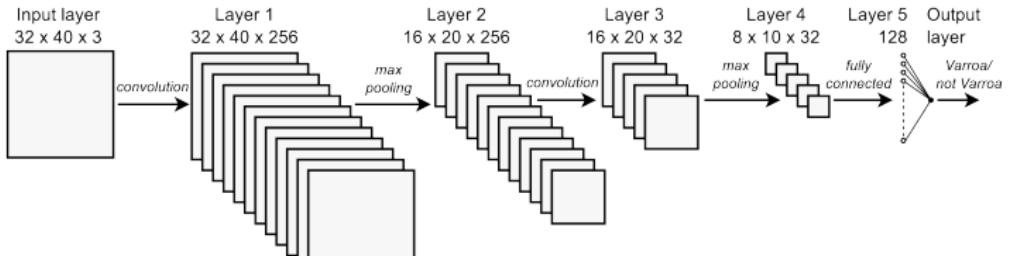
CNN's in Practice

Varroa mite detector—with a CNN somewhere in the pipeline

Image processing pipeline



CNN architecture



A computer vision system to monitor the infestation level of Varroa
detectors in a honeybee colony

Nicolaus^a*, Christian De Paoli^b, Peter Hug^b, Michael Nieder^b, Michael Möbius^b,
Pia Röger^c

^a Institute of Robotics, Mechatronics and Production Engineering, University of Karlsruhe, Karlsruhe, Germany; ^b Institute of Animal Ecology and Conservation Biology, University of Regensburg, Regensburg, Germany; ^c Institute of Zoology, University of Regensburg, Regensburg, Germany

* To whom all correspondence should be addressed. E-mail: nico@ira.uka.de

Received 20 August 2007; accepted 20 April 2008; published online 20 May 2008

© 2008 American Society of Agronomy, Inc. All rights reserved. *Communications in Soil Science and Plant Analysis*, Vol. 39, No. 10, pp. 2005–2010, 2008

ISSN: 0010-236X print/1542-0483 online

DOI: 10.1002/agro.20787

© 2008 American Society of Agronomy, Inc.

Published online in *Communications in Soil Science and Plant Analysis* on May 20, 2008.

Manuscript submitted June 1, 2007; accepted April 2, 2008.

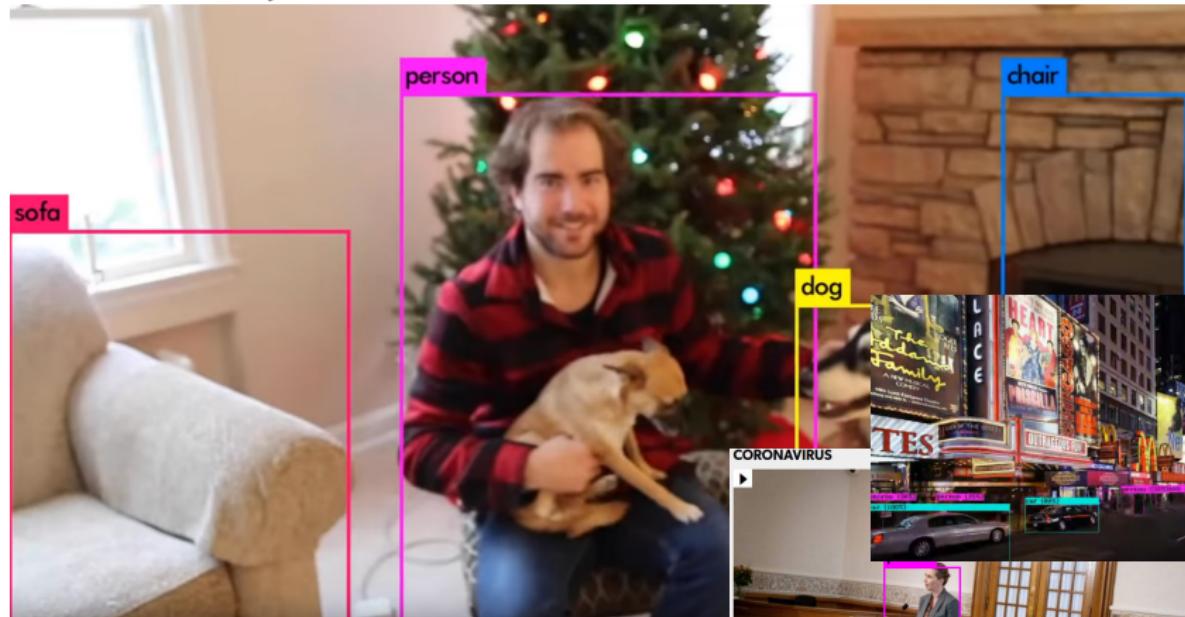
Editorial handling, Brian Loeffler; Reviewer distribution: Ann McElroy, Multi-sensor Biomonitoring

Reviewers: Ann McElroy, Brian Loeffler, Reviewer 1, Reviewer 2, Reviewer 3, Reviewer 4, Reviewer 5, Reviewer 6, Reviewer 7, Reviewer 8, Reviewer 9, Reviewer 10, Reviewer 11, Reviewer 12, Reviewer 13, Reviewer 14, Reviewer 15, Reviewer 16, Reviewer 17, Reviewer 18, Reviewer 19, Reviewer 20, Reviewer 21, Reviewer 22, Reviewer 23, Reviewer 24, Reviewer 25, Reviewer 26, Reviewer 27, Reviewer 28, Reviewer 29, Reviewer 30, Reviewer 31, Reviewer 32, Reviewer 33, Reviewer 34, Reviewer 35, Reviewer 36, Reviewer 37, Reviewer 38, Reviewer 39, Reviewer 40, Reviewer 41, Reviewer 42, Reviewer 43, Reviewer 44, Reviewer 45, Reviewer 46, Reviewer 47, Reviewer 48, Reviewer 49, Reviewer 50, Reviewer 51, Reviewer 52, Reviewer 53, Reviewer 54, Reviewer 55, Reviewer 56, Reviewer 57, Reviewer 58, Reviewer 59, Reviewer 60, Reviewer 61, Reviewer 62, Reviewer 63, Reviewer 64, Reviewer 65, Reviewer 66, Reviewer 67, Reviewer 68, Reviewer 69, Reviewer 70, Reviewer 71, Reviewer 72, Reviewer 73, Reviewer 74, Reviewer 75, Reviewer 76, Reviewer 77, Reviewer 78, Reviewer 79, Reviewer 80, Reviewer 81, Reviewer 82, Reviewer 83, Reviewer 84, Reviewer 85, Reviewer 86, Reviewer 87, Reviewer 88, Reviewer 89, Reviewer 90, Reviewer 91, Reviewer 92, Reviewer 93, Reviewer 94, Reviewer 95, Reviewer 96, Reviewer 97, Reviewer 98, Reviewer 99, Reviewer 100, Reviewer 101, Reviewer 102, Reviewer 103, Reviewer 104, Reviewer 105, Reviewer 106, Reviewer 107, Reviewer 108, Reviewer 109, Reviewer 110, Reviewer 111, Reviewer 112, Reviewer 113, Reviewer 114, Reviewer 115, Reviewer 116, Reviewer 117, Reviewer 118, Reviewer 119, Reviewer 120, Reviewer 121, Reviewer 122, Reviewer 123, Reviewer 124, Reviewer 125, Reviewer 126, Reviewer 127, Reviewer 128, Reviewer 129, Reviewer 130, Reviewer 131, Reviewer 132, Reviewer 133, Reviewer 134, Reviewer 135, Reviewer 136, Reviewer 137, Reviewer 138, Reviewer 139, Reviewer 140, Reviewer 141, Reviewer 142, Reviewer 143, Reviewer 144, Reviewer 145, Reviewer 146, Reviewer 147, Reviewer 148, Reviewer 149, Reviewer 150, Reviewer 151, Reviewer 152, Reviewer 153, Reviewer 154, Reviewer 155, Reviewer 156, Reviewer 157, Reviewer 158, Reviewer 159, Reviewer 160, Reviewer 161, Reviewer 162, Reviewer 163, Reviewer 164, Reviewer 165, Reviewer 166, Reviewer 167, Reviewer 168, Reviewer 169, Reviewer 170, Reviewer 171, Reviewer 172, Reviewer 173, Reviewer 174, Reviewer 175, Reviewer 176, Reviewer 177, Reviewer 178, Reviewer 179, Reviewer 180, Reviewer 181, Reviewer 182, Reviewer 183, Reviewer 184, Reviewer 185, Reviewer 186, Reviewer 187, Reviewer 188, Reviewer 189, Reviewer 190, Reviewer 191, Reviewer 192, Reviewer 193, Reviewer 194, Reviewer 195, Reviewer 196, Reviewer 197, Reviewer 198, Reviewer 199, Reviewer 200, Reviewer 201, Reviewer 202, Reviewer 203, Reviewer 204, Reviewer 205, Reviewer 206, Reviewer 207, Reviewer 208, Reviewer 209, Reviewer 210, Reviewer 211, Reviewer 212, Reviewer 213, Reviewer 214, Reviewer 215, Reviewer 216, Reviewer 217, Reviewer 218, Reviewer 219, Reviewer 220, Reviewer 221, Reviewer 222, Reviewer 223, Reviewer 224, Reviewer 225, Reviewer 226, Reviewer 227, Reviewer 228, Reviewer 229, Reviewer 230, Reviewer 231, Reviewer 232, Reviewer 233, Reviewer 234, Reviewer 235, Reviewer 236, Reviewer 237, Reviewer 238, Reviewer 239, Reviewer 240, Reviewer 241, Reviewer 242, Reviewer 243, Reviewer 244, Reviewer 245, Reviewer 246, Reviewer 247, Reviewer 248, Reviewer 249, Reviewer 250, Reviewer 251, Reviewer 252, Reviewer 253, Reviewer 254, Reviewer 255, Reviewer 256, Reviewer 257, Reviewer 258, Reviewer 259, Reviewer 260, Reviewer 261, Reviewer 262, Reviewer 263, Reviewer 264, Reviewer 265, Reviewer 266, Reviewer 267, Reviewer 268, Reviewer 269, Reviewer 270, Reviewer 271, Reviewer 272, Reviewer 273, Reviewer 274, Reviewer 275, Reviewer 276, Reviewer 277, Reviewer 278, Reviewer 279, Reviewer 280, Reviewer 281, Reviewer 282, Reviewer 283, Reviewer 284, Reviewer 285, Reviewer 286, Reviewer 287, Reviewer 288, Reviewer 289, Reviewer 290, Reviewer 291, Reviewer 292, Reviewer 293, Reviewer 294, Reviewer 295, Reviewer 296, Reviewer 297, Reviewer 298, Reviewer 299, Reviewer 300, Reviewer 301, Reviewer 302, Reviewer 303, Reviewer 304, Reviewer 305, Reviewer 306, Reviewer 307, Reviewer 308, Reviewer 309, Reviewer 310, Reviewer 311, Reviewer 312, Reviewer 313, Reviewer 314, Reviewer 315, Reviewer 316, Reviewer 317, Reviewer 318, Reviewer 319, Reviewer 320, Reviewer 321, Reviewer 322, Reviewer 323, Reviewer 324, Reviewer 325, Reviewer 326, Reviewer 327, Reviewer 328, Reviewer 329, Reviewer 330, Reviewer 331, Reviewer 332, Reviewer 333, Reviewer 334, Reviewer 335, Reviewer 336, Reviewer 337, Reviewer 338, Reviewer 339, Reviewer 340, Reviewer 341, Reviewer 342, Reviewer 343, Reviewer 344, Reviewer 345, Reviewer 346, Reviewer 347, Reviewer 348, Reviewer 349, Reviewer 350, Reviewer 351, Reviewer 352, Reviewer 353, Reviewer 354, Reviewer 355, Reviewer 356, Reviewer 357, Reviewer 358, Reviewer 359, Reviewer 360, Reviewer 361, Reviewer 362, Reviewer 363, Reviewer 364, Reviewer 365, Reviewer 366, Reviewer 367, Reviewer 368, Reviewer 369, Reviewer 370, Reviewer 371, Reviewer 372, Reviewer 373, Reviewer 374, Reviewer 375, Reviewer 376, Reviewer 377, Reviewer 378, Reviewer 379, Reviewer 380, Reviewer 381, Reviewer 382, Reviewer 383, Reviewer 384, Reviewer 385, Reviewer 386, Reviewer 387, Reviewer 388, Reviewer 389, Reviewer 390, Reviewer 391, Reviewer 392, Reviewer 393, Reviewer 394, Reviewer 395, Reviewer 396, Reviewer 397, Reviewer 398, Reviewer 399, Reviewer 400, Reviewer 401, Reviewer 402, Reviewer 403, Reviewer 404, Reviewer 405, Reviewer 406, Reviewer 407, Reviewer 408, Reviewer 409, Reviewer 410, Reviewer 411, Reviewer 412, Reviewer 413, Reviewer 414, Reviewer 415, Reviewer 416, Reviewer 417, Reviewer 418, Reviewer 419, Reviewer 420, Reviewer 421, Reviewer 422, Reviewer 423, Reviewer 424, Reviewer 425, Reviewer 426, Reviewer 427, Reviewer 428, Reviewer 429, Reviewer 430, Reviewer 431, Reviewer 432, Reviewer 433, Reviewer 434, Reviewer 435, Reviewer 436, Reviewer 437, Reviewer 438, Reviewer 439, Reviewer 440, Reviewer 441, Reviewer 442, Reviewer 443, Reviewer 444, Reviewer 445, Reviewer 446, Reviewer 447, Reviewer 448, Reviewer 449, Reviewer 450, Reviewer 451, Reviewer 452, Reviewer 453, Reviewer 454, Reviewer 455, Reviewer 456, Reviewer 457, Reviewer 458, Reviewer 459, Reviewer 460, Reviewer 461, Reviewer 462, Reviewer 463, Reviewer 464, Reviewer 465, Reviewer 466, Reviewer 467, Reviewer 468, Reviewer 469, Reviewer 470, Reviewer 471, Reviewer 472, Reviewer 473, Reviewer 474, Reviewer 475, Reviewer 476, Reviewer 477, Reviewer 478, Reviewer 479, Reviewer 480, Reviewer 481, Reviewer 482, Reviewer 483, Reviewer 484, Reviewer 485, Reviewer 486, Reviewer 487, Reviewer 488, Reviewer 489, Reviewer 490, Reviewer 491, Reviewer 492, Reviewer 493, Reviewer 494, Reviewer 495, Reviewer 496, Reviewer 497, Reviewer 498, Reviewer 499, Reviewer 500, Reviewer 501, Reviewer 502, Reviewer 503, Reviewer 504, Reviewer 505, Reviewer 506, Reviewer 507, Reviewer 508, Reviewer 509, Reviewer 510, Reviewer 511, Reviewer 512, Reviewer 513, Reviewer 514, Reviewer 515, Reviewer 516, Reviewer 517, Reviewer 518, Reviewer 519, Reviewer 520, Reviewer 521, Reviewer 522, Reviewer 523, Reviewer 524, Reviewer 525, Reviewer 526, Reviewer 527, Reviewer 528, Reviewer 529, Reviewer 530, Reviewer 531, Reviewer 532, Reviewer 533, Reviewer 534, Reviewer 535, Reviewer 536, Reviewer 537, Reviewer 538, Reviewer 539, Reviewer 540, Reviewer 541, Reviewer 542, Reviewer 543, Reviewer 544, Reviewer 545, Reviewer 546, Reviewer 547, Reviewer 548, Reviewer 549, Reviewer 550, Reviewer 551, Reviewer 552, Reviewer 553, Reviewer 554, Reviewer 555, Reviewer 556, Reviewer 557, Reviewer 558, Reviewer 559, Reviewer 560, Reviewer 561, Reviewer 562, Reviewer 563, Reviewer 564, Reviewer 565, Reviewer 566, Reviewer 567, Reviewer 568, Reviewer 569, Reviewer 570, Reviewer 571, Reviewer 572, Reviewer 573, Reviewer 574, Reviewer 575, Reviewer 576, Reviewer 577, Reviewer 578, Reviewer 579, Reviewer 580, Reviewer 581, Reviewer 582, Reviewer 583, Reviewer 584, Reviewer 585, Reviewer 586, Reviewer 587, Reviewer 588, Reviewer 589, Reviewer 590, Reviewer 591, Reviewer 592, Reviewer 593, Reviewer 594, Reviewer 595, Reviewer 596, Reviewer 597, Reviewer 598, Reviewer 599, Reviewer 600, Reviewer 601, Reviewer 602, Reviewer 603, Reviewer 604, Reviewer 605, Reviewer 606, Reviewer 607, Reviewer 608, Reviewer 609, Reviewer 610, Reviewer 611, Reviewer 612, Reviewer 613, Reviewer 614, Reviewer 615, Reviewer 616, Reviewer 617, Reviewer 618, Reviewer 619, Reviewer 620, Reviewer 621, Reviewer 622, Reviewer 623, Reviewer 624, Reviewer 625, Reviewer 626, Reviewer 627, Reviewer 628, Reviewer 629, Reviewer 630, Reviewer 631, Reviewer 632, Reviewer 633, Reviewer 634, Reviewer 635, Reviewer 636, Reviewer 637, Reviewer 638, Reviewer 639, Reviewer 640, Reviewer 641, Reviewer 642, Reviewer 643, Reviewer 644, Reviewer 645, Reviewer 646, Reviewer 647, Reviewer 648, Reviewer 649, Reviewer 650, Reviewer 651, Reviewer 652, Reviewer 653, Reviewer 654, Reviewer 655, Reviewer 656, Reviewer 657, Reviewer 658, Reviewer 659, Reviewer 660, Reviewer 661, Reviewer 662, Reviewer 663, Reviewer 664, Reviewer 665, Reviewer 666, Reviewer 667, Reviewer 668, Reviewer 669, Reviewer 670, Reviewer 671, Reviewer 672, Reviewer 673, Reviewer 674, Reviewer 675, Reviewer 676, Reviewer 677, Reviewer 678, Reviewer 679, Reviewer 680, Reviewer 681, Reviewer 682, Reviewer 683, Reviewer 684, Reviewer 685, Reviewer 686, Reviewer 687, Reviewer 688, Reviewer 689, Reviewer 690, Reviewer 691, Reviewer 692, Reviewer 693, Reviewer 694, Reviewer 695, Reviewer 696, Reviewer 697, Reviewer 698, Reviewer 699, Reviewer 700, Reviewer 701, Reviewer 702, Reviewer 703, Reviewer 704, Reviewer 705, Reviewer 706, Reviewer 707, Reviewer 708, Reviewer 709, Reviewer 710, Reviewer 711, Reviewer 712, Reviewer 713, Reviewer 714, Reviewer 715, Reviewer 716, Reviewer 717, Reviewer 718, Reviewer 719, Reviewer 720, Reviewer 721, Reviewer 722, Reviewer 723, Reviewer 724, Reviewer 725, Reviewer 726, Reviewer 727, Reviewer 728, Reviewer 729, Reviewer 730, Reviewer 731, Reviewer 732, Reviewer 733, Reviewer 734, Reviewer 735, Reviewer 736, Reviewer 737, Reviewer 738, Reviewer 739, Reviewer 740, Reviewer 741, Reviewer 742, Reviewer 743, Reviewer 744, Reviewer 745, Reviewer 746, Reviewer 747, Reviewer 748, Reviewer 749, Reviewer 750, Reviewer 751, Reviewer 752, Reviewer 753, Reviewer 754, Reviewer 755, Reviewer 756, Reviewer 757, Reviewer 758, Reviewer 759, Reviewer 760, Reviewer 761, Reviewer 762, Reviewer 763, Reviewer 764, Reviewer 765, Reviewer 766, Reviewer 767, Reviewer 768, Review

CNN's in Practice

YOLOv2+3+4 (You Only Look Once)

Real-time object detection, demo..



[<https://holm.info/yolodemo>]

[<https://github.com/AlexeyAB/darknet>]

Mette Frederiksen: Uklogt og uholdbart at vi

CNN's in Practice

YOLOv2+3+4 (You Only Look Once)

```
[~] Terminal
o obj/art.o obj/tag.o obj/cifar.o obj/go.o obj/rnn.o obj/segmenter.o obj/regressor.o obj/classifier.o obj/coco.o
obj/yolo.o obj/detector.o obj/nightmare.o obj/instance-segmenter.o obj/darknet.o libdarknet.a -o darknet -lm -lpthread
-L/opt/opencv/opencv4/lib -lopencv_core -lopencv_imgproc -lopencv_imgcodecs -lopencv_videoio -lopencv_highgui -ltaff -lstdc++ libdarknet.a
layer    filters   size      input           output
  0 conv     32  3 x 3 / 1  256 x 256 x  3  ->  256 x 256 x 32  0.113
  1 max      2 x 2 / 2  256 x 256 x 32  ->  128 x 128 x 32
  2 conv     64  3 x 3 / 1  128 x 128 x 32  ->  128 x 128 x 64  0.604
  3 max      2 x 2 / 2  128 x 128 x 64  ->  64 x 64 x 64
  4 conv    128  3 x 3 / 1  64 x 64 x 64  ->  64 x 64 x 128  0.604
  5 conv     64  1 x 1 / 1  64 x 64 x 128 ->  64 x 64 x 64  0.067
  6 conv    128  3 x 3 / 1  64 x 64 x 64  ->  64 x 64 x 128  0.604
  7 max      2 x 2 / 2  64 x 64 x 128 ->  32 x 32 x 128
  8 conv    256  3 x 3 / 1  32 x 32 x 128 ->  32 x 32 x 256  0.604
  9 conv    128  1 x 1 / 1  32 x 32 x 256 ->  32 x 32 x 128  0.067
 10 conv   256  3 x 3 / 1  32 x 32 x 128 ->  32 x 32 x 256  0.604
 11 max      2 x 2 / 2  32 x 32 x 256 ->  16 x 16 x 256
 12 conv   512  3 x 3 / 1  16 x 16 x 256 ->  16 x 16 x 512  0.604
 13 conv   256  1 x 1 / 1  16 x 16 x 512 ->  16 x 16 x 256  0.067
 14 conv   512  3 x 3 / 1  16 x 16 x 256 ->  16 x 16 x 512  0.604
 15 conv   256  1 x 1 / 1  16 x 16 x 512 ->  16 x 16 x 256  0.067
 16 conv   512  3 x 3 / 1  16 x 16 x 256 ->  16 x 16 x 512  0.604
 17 max      2 x 2 / 2  16 x 16 x 512 ->  8 x 8 x 512
 18 conv   1024 3 x 3 / 1  8 x 8 x 512 ->  8 x 8 x 1024  0.604
 19 conv   512  1 x 1 / 1  8 x 8 x 1024 ->  8 x 8 x 512  0.067
 20 conv   1024 3 x 3 / 1  8 x 8 x 512 ->  8 x 8 x 1024  0.604
 21 conv   512  1 x 1 / 1  8 x 8 x 1024 ->  8 x 8 x 512  0.067
 22 conv   1024 3 x 3 / 1  8 x 8 x 512 ->  8 x 8 x 1024  0.604
 23 conv   1000 1 x 1 / 1  8 x 8 x 1024 ->  8 x 8 x 1000  0.131
 24 avg          8 x 8 x 1000 ->  1000
 25 softmax
Loading weights from darknet19.weights...Done!
test.jpg: Predicted in 1.965434 seconds.
64.68%: mountain bike
16.00%: web site
12.87%: bicycle-built-for-two
1.05%: crash helmet
0.86%: alp
[ce]f@leno:~/textmal/darknet$
```

De danske cykellyttere har aldrig været bedre, siger cykelekspert Brian Nygaard efter Kasper Asgreens sejr i Flandern Rundt.



Kasper Asgreen slog Mathieu van der Poel i Flandern Rundt og vandt den sjette danske sejr på årets World Tour. Foto: AP/Bild ID: Stockom. © Scarpido



CNN's in Practice

The LeNET-5 Architecture

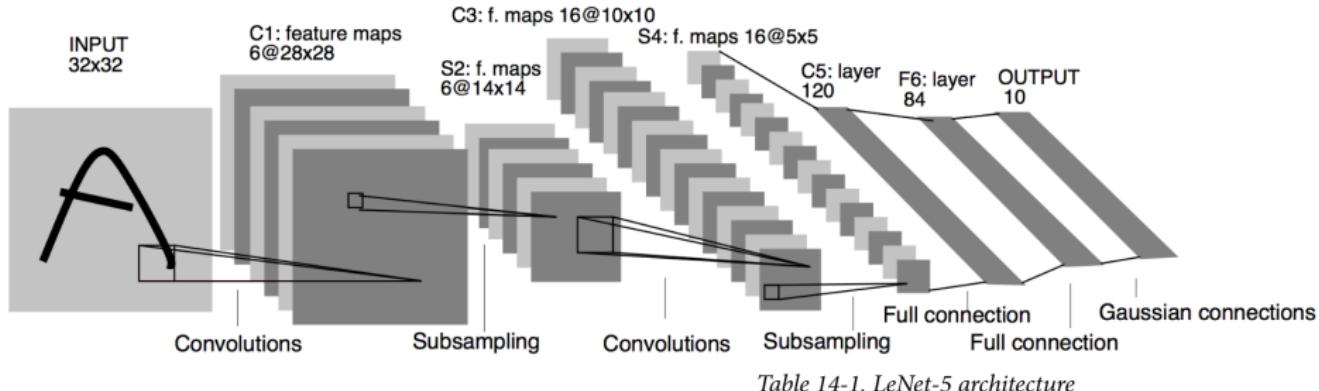


Table 14-1. LeNet-5 architecture

Other famous CNN-architectures:

- ▶ AlexNet,
- ▶ GoogLeNet (inception),
- ▶ ResNet (152 layers,
skip-connections),
- ▶ VGGNet,
- ▶ Inception-v4 (GoogLeNet + ResNet),
- ▶ ...

Layer	Type	Maps	Size	Kernel size	Stride	Activation
Out	Fully Connected	–	10	–	–	RBF
F6	Fully Connected	–	84	–	–	tanh
C5	Convolution	120	1 × 1	5 × 5	1	tanh
S4	Avg Pooling	16	5 × 5	2 × 2	2	tanh
C3	Convolution	16	10 × 10	5 × 5	1	tanh
S2	Avg Pooling	6	14 × 14	2 × 2	2	tanh
C1	Convolution	6	28 × 28	5 × 5	1	tanh
In	Input	1	32 × 32	–	–	–

CNN's in Practice

A LeNET-5 'Like' Architecture

```
In [9]: 1 import keras
2 from keras import layers
3
4 model = keras.Sequential()
5
6 model.add(layers.Conv2D(filters=6, kernel_size=(3, 3),
7                         activation='relu', input_shape=(32,32,1)))
8 model.add(layers.AveragePooling2D())
9
10 model.add(layers.Conv2D(filters=16, kernel_size=(3, 3), activation='relu'))
11 model.add(layers.AveragePooling2D())
12
13 model.add(layers.Flatten())
14
15 model.add(layers.Dense(units=120, activation='relu'))
16 model.add(layers.Dense(units=84, activation='relu'))
17 model.add(layers.Dense(units=10, activation='softmax'))
18
19 model.summary()
```

Layer (type)	Output Shape	Param #
conv2d_5 (Conv2D)	(None, 30, 30, 6)	60
average_pooling2d_5 (Average)	(None, 15, 15, 6)	0
conv2d_6 (Conv2D)	(None, 13, 13, 16)	880
average_pooling2d_6 (Average)	(None, 6, 6, 16)	0
flatten_3 (Flatten)	(None, 576)	0
dense_7 (Dense)	(None, 120)	69240
dense_8 (Dense)	(None, 84)	10164
dense_9 (Dense)	(None, 10)	850
=====		
Total params: 81,194		
Trainable params: 81,194		
Non-trainable params: 0		

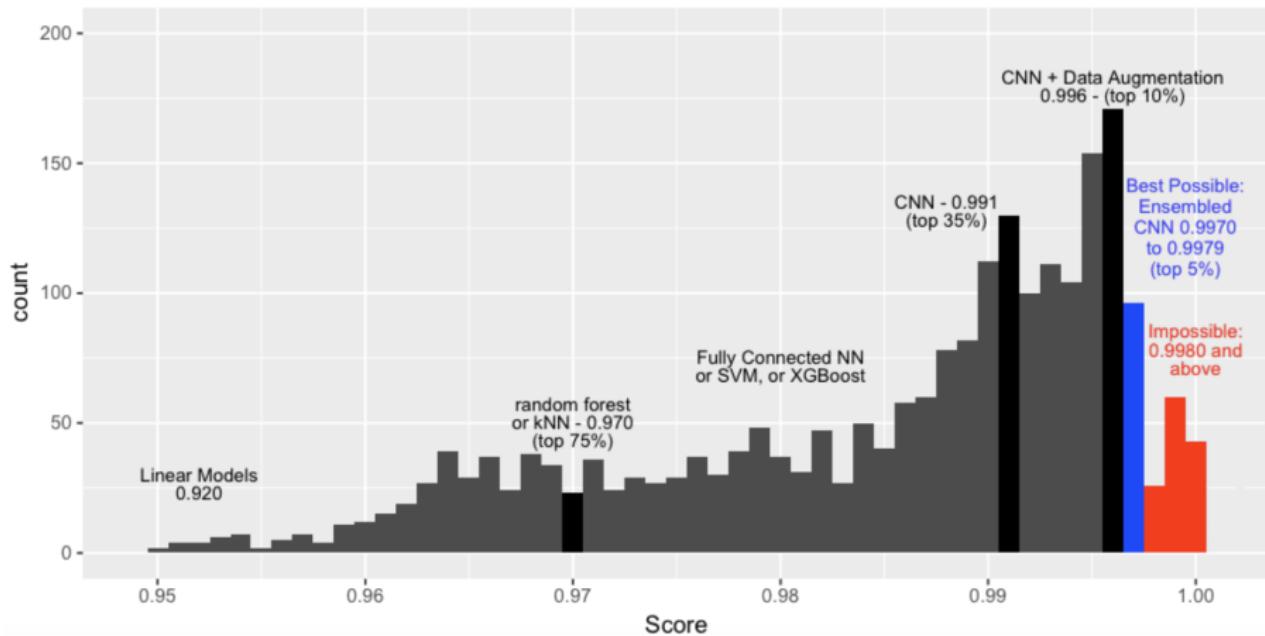
Table 14-1. LeNet-5 architecture

CNN's in Practice

The LeNET-5 Architecture on MNIST

Histogram of Kaggle MNIST

public leaderboard scores, July 15 2018



- ▶ using pre-trained models => Transferred Learning,
- ▶ object detection,
- ▶ semantic segmentation,
- ▶ time series => RNN's, ...

Qd MNIST Search Quest II

Fra tidligere semester:

Highest score = 0.97369,

→ ITMALGpr28:

```
1 model=KNeighborsClassifier(  
2     algorithm='ball_tree',  
3     n_neighbors=3,  
4     p=4,  
5     weights='distance')
```

12/10 11:13> "ITMALGrp17: best: dat mnist, score=0.97206,
model=KNeighborsClassifier(leaf_size=40, p=4)"
08/10 22:56> "ITMALGrp09: dat=mnist, score=0.96973,
model=KNeighborsClassifier(leaf_size=10, n_neighbors=3, weights='distance')"
08/10 22:04> "ITMALGrp09: dat=mnist, score=0.93550,
model=KNeighborsClassifier(leaf_size=10, n_neighbors=3, weights='distance')"
08/10 10:02> "08/10 10:01 Is from 'ITMALGrp28'"
08/10 10:01> "best: dat mnist, score=0.97369,
model=KNeighborsClassifier(algorithm='ball_tree', n_neighbors=4, weights='distance')"
08/10 08:53> "ITMALGrp17: dat=mnist, score=0.47090, model=KNeighborsClassifier(algorithm='auto',
n_neighbors=3, weights='distance', p=2, metric='minkowski', metric_params=None, n_jobs=None),
score=0.97206"
07/10 21:43> "ITMALGrp25: KNeighborsClassifier(algorithm='ball_tree', leaf_size=30,
metric='minkowski', metric_params=None, n_jobs=None, n_neighbors=5, p=2, weights='uniform'),
score=0.96771<2857142857"
07/10 18:53> "ITMALGrp21: dat=mnist, score=0.96878,
model=KNeighborsClassifier(algorithm='ball_tree', n_neighbors=3)"
06/10 22:54> "ITMALGrp08: dat=mnist, score=0.96973,
model=KNeighborsClassifier(algorithm='ball_tree', leaf_size=10, n_neighbors=3, weights='distance')"
29/09 13:40> "ITMALGrp05: dat=mnist , score=0.97257,
model=KNeighborsClassifier(n_neighbors=4,p=2,weights='distance')"
24/09 13:55> "ITMALGrp02: dat=mnist, score=0.96810, model=RandomForestClassifier(n_estimators=1000)*

Training time?

```
12/10 11:13> "ITMALGrp17: score=0.97206, model=KNeighborsClassifier...  
08/10 22:56> "ITMALGrp09: score=0.96973, model=KNeighborsClassifier...  
08/10 22:04> "ITMALGrp09: score=0.93550, model=KNeighborsClassifier...  
08/10 10:02> "08/10 10:01 Is from 'ITMALGrp28'"  
08/10 10:01> " score=0.97369, model=KNeighborsClassifier...  
08/10 08:53> "ITMALGrp17: score=0.47090, model=KNeighborsClassifier...  
07/10 21:43> "ITMALGrp25: score=0.9677, model=KNeighborsClassifier...  
07/10 18:53> "ITMALGrp21: score=0.96878, model=KNeighborsClassifier...  
06/10 22:54> "ITMALGrp08: score=0.96973, model=KNeighborsClassifier...  
29/09 13:40> "ITMALGrp05: score=0.97257, model=KNeighborsClassifier...  
24/09 13:55> "ITMALGrp02: score=0.96810, model=RandomForestClassifier...
```

HARDWARE FOR MACHINE LEARNING



Hardware for Machine Learning

Methods and Terminology (SKIP most except μ, η)

Objective:

Why optimize using 'application specific' hardware?

▶ Effectiveness:

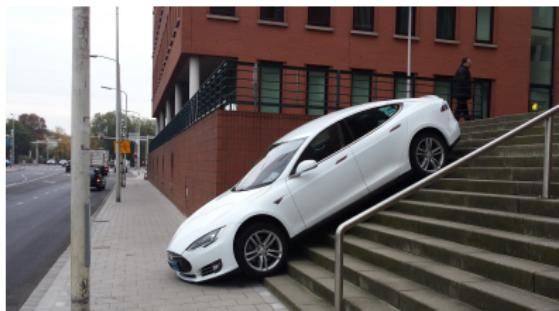
- ▶ cost of purchasing/operating systems,
 $\mu = \text{FLOPS}/\$$
- $\eta = \text{FLOPS}/\text{Watt}$
- ▶ cut-down developer waiting time,
- ▶ make modelling iterations fast (say minutes).

▶ Big-data:

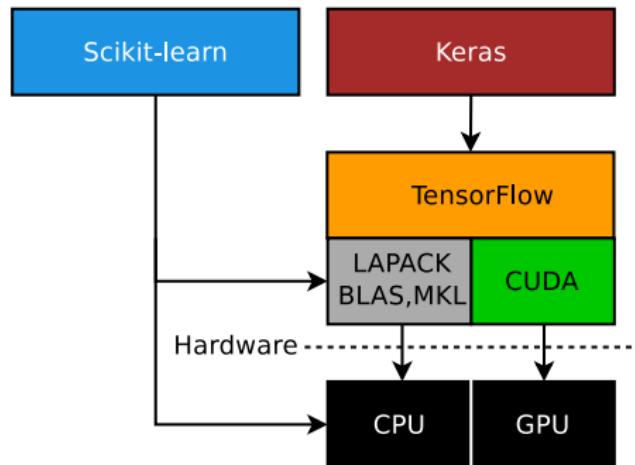
- ▶ enable training on x-large data.

▶ Real-time constraints:

- ▶ inference (on visual data) in real-time,
- ▶ low-power constraints.



RESUMÉ: Keras and Tensorflow



GP-GPU: General-Purpose Graphics Processing Unit...or just **GPU**.

CUDA: Compute Unified Device Architecture, API for SIMD/SIMT on GPU,

CPUs

Build Tensorflow from source (SKIP)

- ▶ for specific architecture, say ARM,
- ▶ or for HPC optimization for all CPU feature
- > lscpu

```
Architecture: x86_64
Model name:  Intel(R) Core(TM) i7-6600U CPU @ 2.60GHz
Flags:          fpu vme de pse tsc msr pae mce cx8 apic sep mtrr
               pge dts acpi mmx fxsr fma .sse sse2..sse4_1sse4_2..
```



Using Docker and pulling TF from GIT + a lot of scripting!

```
> git clone https://github.com/tensorflow/tensorflow
> git checkout 1.12
(lots of scripting and pain..)
> bazel build -copt=-mfma -copt=-msse4.2 tensorflow
```

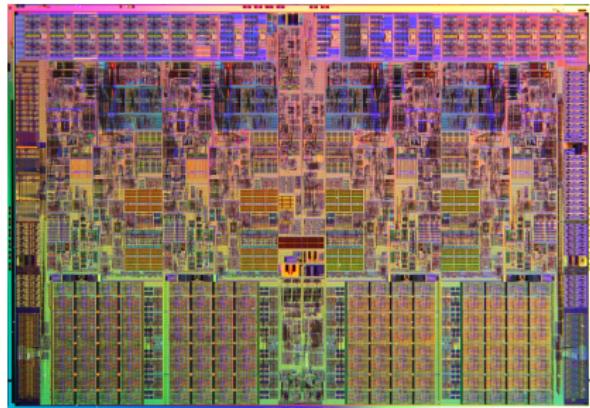
Howtos:

<https://www.tensorflow.org/install/source>

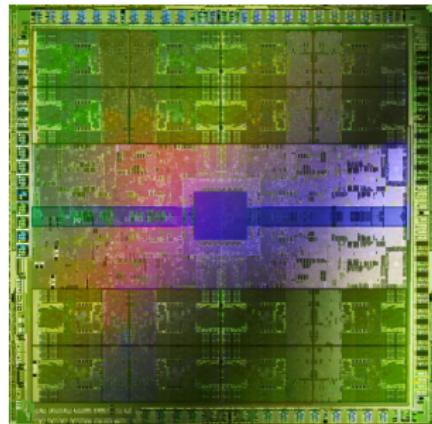
<https://www.pugetsystems.com/labs/hpc/Build-TensorFlow-CPU-with-MKL-and-Anaconda-Python-3-6-using-a-Docker-Container-1133>

<https://www.pugetsystems.com/labs/hpc/Build-TensorFlow-GPU-with-CUDA-9-1-MKL-and-Anaconda-Python-3-6-using-a-Docker-Container-1134>

CPUs vs GPUs



Nehalem CPU
die size: $\sim 700 \text{ mm}^2$
transistors: $\sim 2.3 \cdot 10^9$



Fermi GPU
die size: 520 mm^2
transistors: $\sim 3 \cdot 10^9$

Nvidia arch. transistor counts

Pascal	$\sim 15 \cdot 10^9$
Turing	$\sim 19 \cdot 10^9$
Volta	$\sim 21 \cdot 10^9$
Ampere	$\sim 28 \cdot 10^9$ (GPU 3090, 8nm)

CPUs vs GPUs

So many transistors, but how many for the ALUs/FPUs?

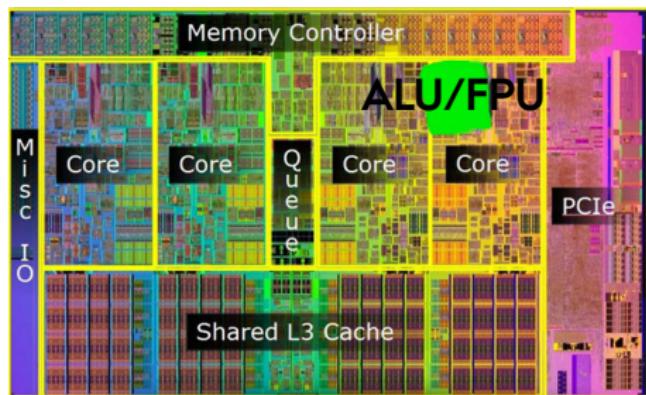
What makes GPUs such excellent HW for Machine Learning?

ALU: Arithmetic logic unit

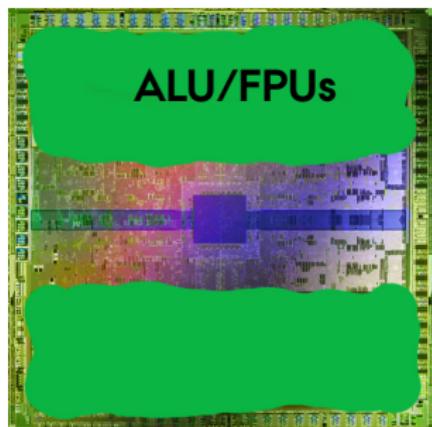
FPU: Floating-point unit

Memory Controller: six controllers on GPU,

CPU: lots of speculative execution; waste of transistors.



CPU (type?)
with one ALU/FPU marked



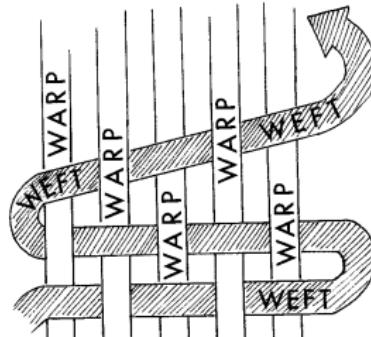
Fermi GPU
ALUs/FPUs all over

GPUs

Fundamental Problems with the GPUs Hardware (SKIP most, except WARP)

GPUs had several *Achilles heels* related to its hardware, many of them addressed in the latest Vola V100 architecture:

- ▶ coding problems graphically, now CUDA,
- ▶ no STACK, now added,
- ▶ no CACHE (or Texture only), now both L₁ and L₂ cache,
- ▶ distinct GPU memory, now UNIFIED memory,
- ▶ SIMT WARP-bunch of 32-threads, now true SIMD,



GPUs

GPU architecture: Core Design, Streaming Multiprocessors



GPUs

Core Design for a SM (Streaming Multiprocessor)

Volta SM design (new gen. GPU):

FP64/32:FPGAs
INT: ALUs
TENSOR CORES: ?

Ampere, RTX, 3090: Raytracingkerner: ?



GEFORCE RTX 3090

NVIDIA CUDA® kerner

10496

Høj CPU-hastighed (GHz)

1.70

Normal CPU-hastighed (GHz)

1.40

Raytracingkerner

ner 2. generation

Tensor Cores

3. generation

Standard hukommelseskonfiguration

Hukommelsesgrænsefladens bredde

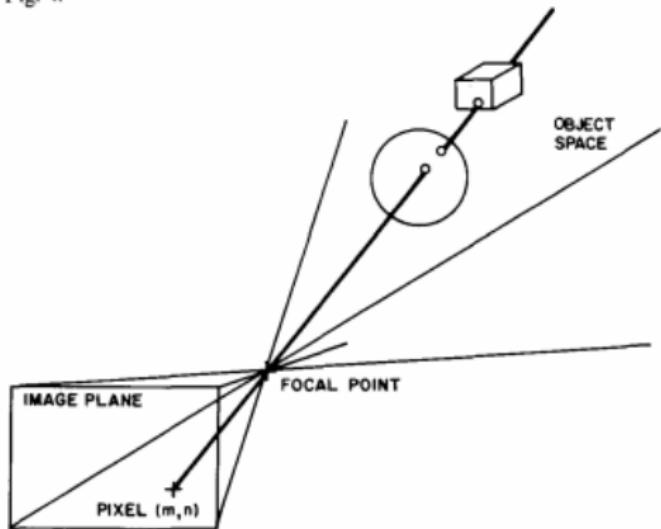
24 GB GDDR6X

GPUs

Tensor cores: Raytracing vs Rasterization on GPUs

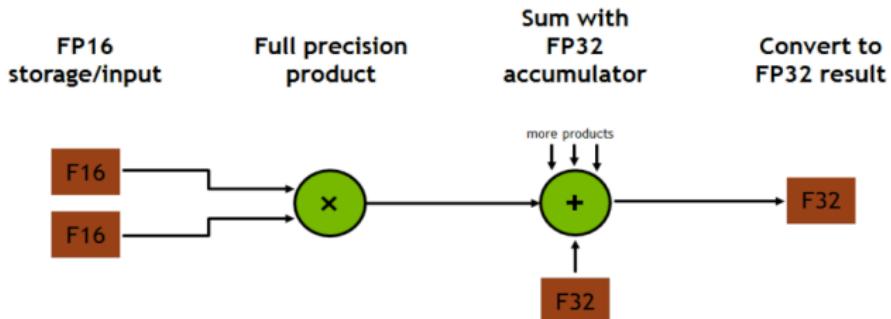


Fig. 4.



GPUs

Tensor Cores
(SKIP most)



$$D = \left(\begin{array}{cccc} A_{0,0} & A_{0,1} & A_{0,2} & A_{0,3} \\ A_{1,0} & A_{1,1} & A_{1,2} & A_{1,3} \\ A_{2,0} & A_{2,1} & A_{2,2} & A_{2,3} \\ A_{3,0} & A_{3,1} & A_{3,2} & A_{3,3} \end{array} \right) \left(\begin{array}{cccc} B_{0,0} & B_{0,1} & B_{0,2} & B_{0,3} \\ B_{1,0} & B_{1,1} & B_{1,2} & B_{1,3} \\ B_{2,0} & B_{2,1} & B_{2,2} & B_{2,3} \\ B_{3,0} & B_{3,1} & B_{3,2} & B_{3,3} \end{array} \right) + \left(\begin{array}{cccc} C_{0,0} & C_{0,1} & C_{0,2} & C_{0,3} \\ C_{1,0} & C_{1,1} & C_{1,2} & C_{1,3} \\ C_{2,0} & C_{2,1} & C_{2,2} & C_{2,3} \\ C_{3,0} & C_{3,1} & C_{3,2} & C_{3,3} \end{array} \right)$$

FP16 or FP32 FP16 FP16 or FP32

Tesla V100's Tensor Cores deliver up to 125 Tensor TFLOPS for training and inference applications.

[volta-architecture-whitepaper.pdf]

Analog: DSD: half-adder, full-adder, ripple-carry-adder with n-bit multiplier based on adders (scales n-bit²)..

HPC Top500

Best High-Performance Computer

[<https://www.top500.org/list/2007/06/>]

[https://en.wikipedia.org/wiki/List_of_Nvidia_graphics_processing_units]



HOME	LISTS ▾	STATISTICS ▾	RESOURCES
------	---------	--------------	-----------

Home » Lists » TOP500 » June 2007

JUNE 2007

Rank	System	Cores	Rmax [TFlop/s]	Rpeak [TFlop/s]	Power (kW)
1	BlueGene/L - eServer Blue Gene Solution, IBM DOE/NNSA/LLNL United States	131,072	280.6	367.0	1,433
2	Jaguar - Cray XT4/XT3, Cray/HPE DOE/SC/Oak Ridge National Laboratory United States	23,016	101.7	119.3	
3	Red Storm - Sandia/ Cray Red Storm, Opteron 2.4 GHz dual core, Cray/HPE NNSA/Sandia National Laboratories United States	26,544	101.4	127.4	

GPUs

When is the GPU faster than the CPU for NN?

GPU slower for CPU for a three-layer NN + MNIST, why?

- ▶ GPU needs a reasonable amount of trainable parameters + data to beat the CPU!

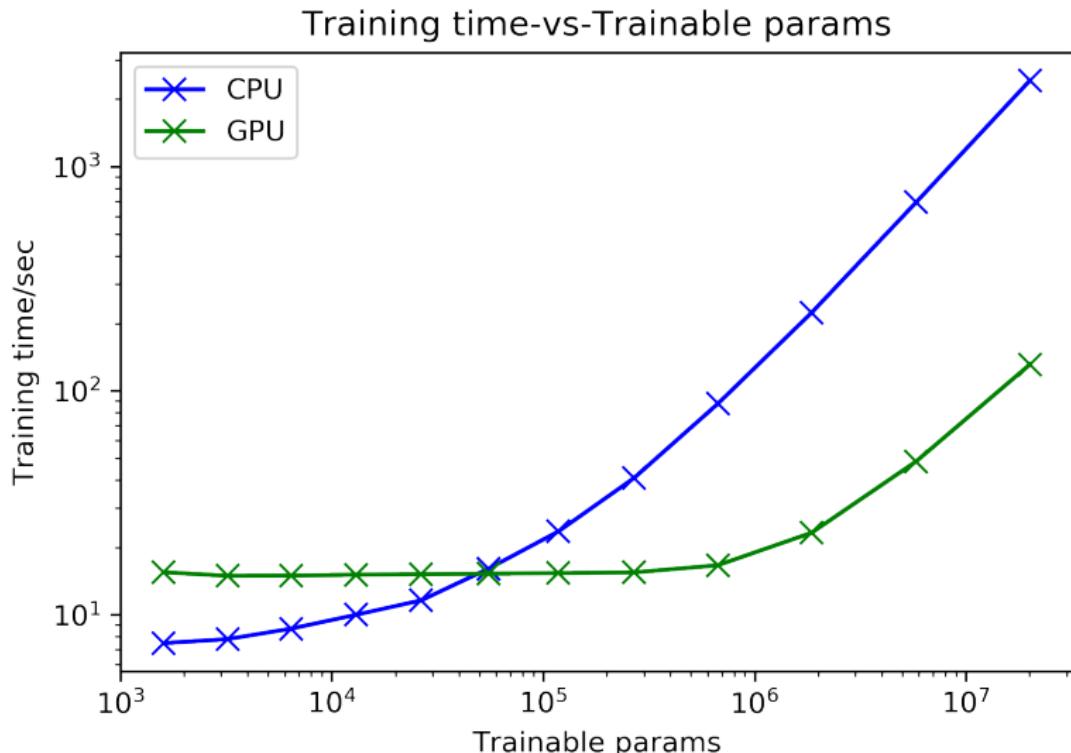
`model.summary():`

```
1 # i=12, n=4096
2 #
3 # Layer (type)          Output Shape       Param #
4 # =====
5 # dense_46 (Dense)     (None, 4096)        3215360
6 #
7 # dropout_31 (Dropout) (None, 4096)        0
8 #
9 # dense_47 (Dense)     (None, 4096)        16781312
10 #
11 # dropout_32 (Dropout) (None, 4096)        0
12 #
13 # dense_48 (Dense)     (None, 10)          40970
14 # =====
15 # Total params: 20,037,642
16 # Trainable params: 20,037,642
17 # Non-trainable params: 0
```

GPUs

Actual test on the GPU-server

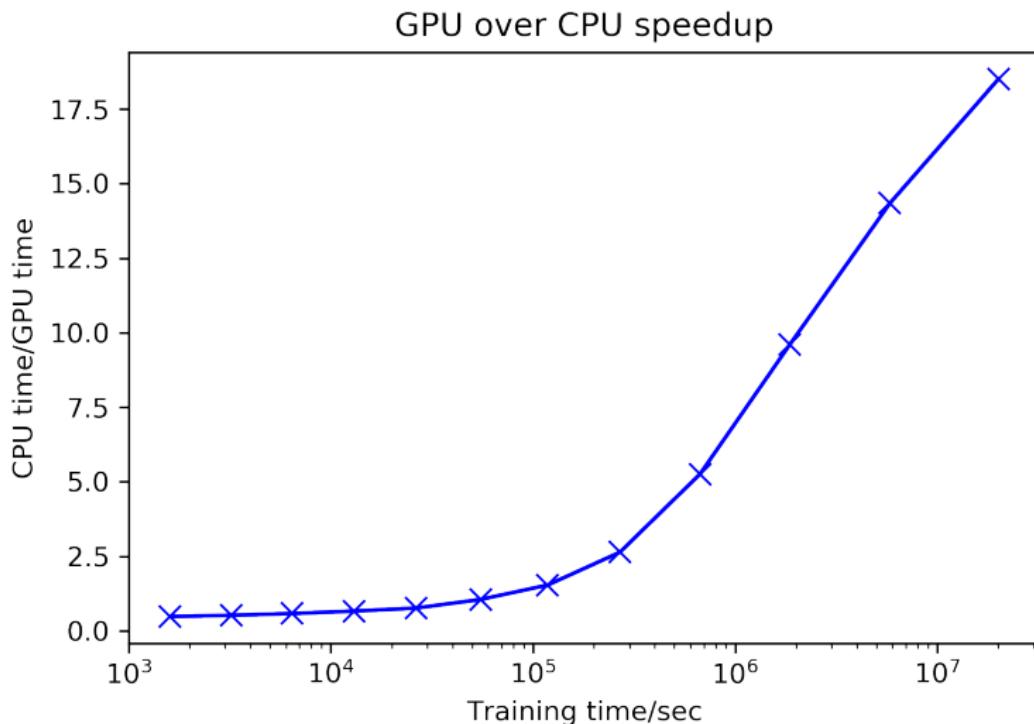
CPU vs GPU on MNIST for a three layer NN with dropout...



GPUs

Actual test on the GPU-server

CPU vs GPU on MNIST for a three layer NN with dropout...



TPUs

Tensor Processing Units

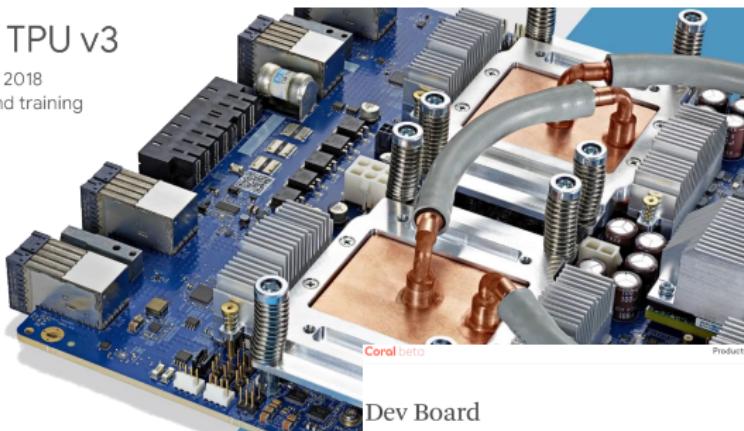
Custom ASICs by Google



Cloud TPU v3

Launched in 2018

Inference and training



Dev Board

A development board to quickly prototype on-device ML products. Scale from prototype to production with a removable system-on-module (SoM).

→ Datasheet
→ Get started guide

\$149.99

Buy



TPU v2

Launched in 2015

Inference only

Launched in 2017
Inference and training

Dev board with Google Edge TPU ML accelerator coprocessor

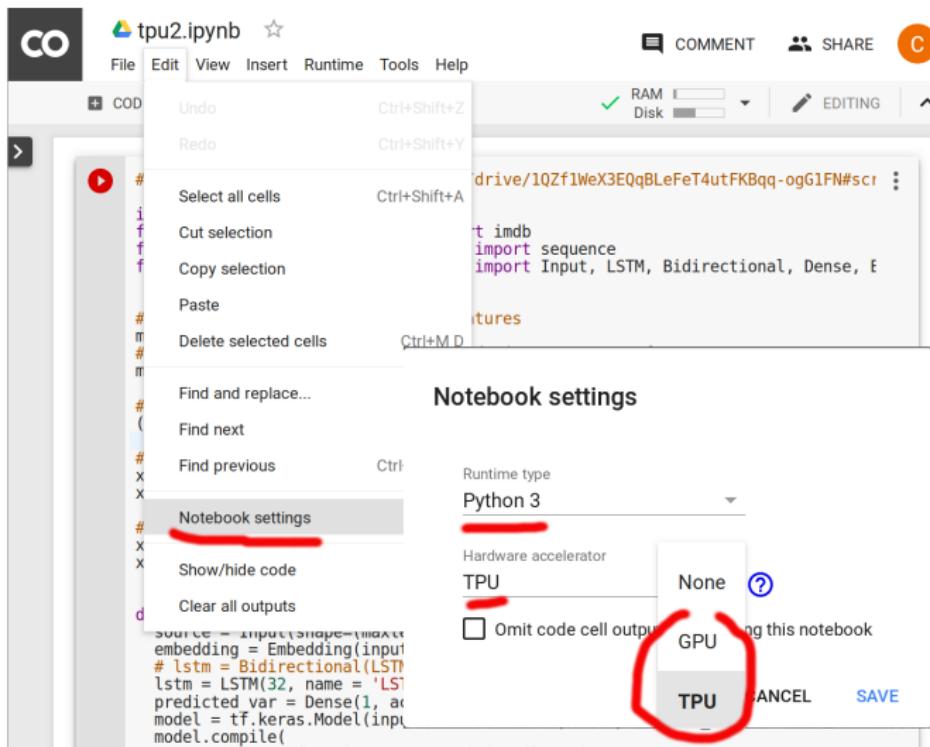
[\[https://coral.withgoogle.com/products/dev-board/\]](https://coral.withgoogle.com/products/dev-board/)



TPUs

Access to TPUs (SKIP)

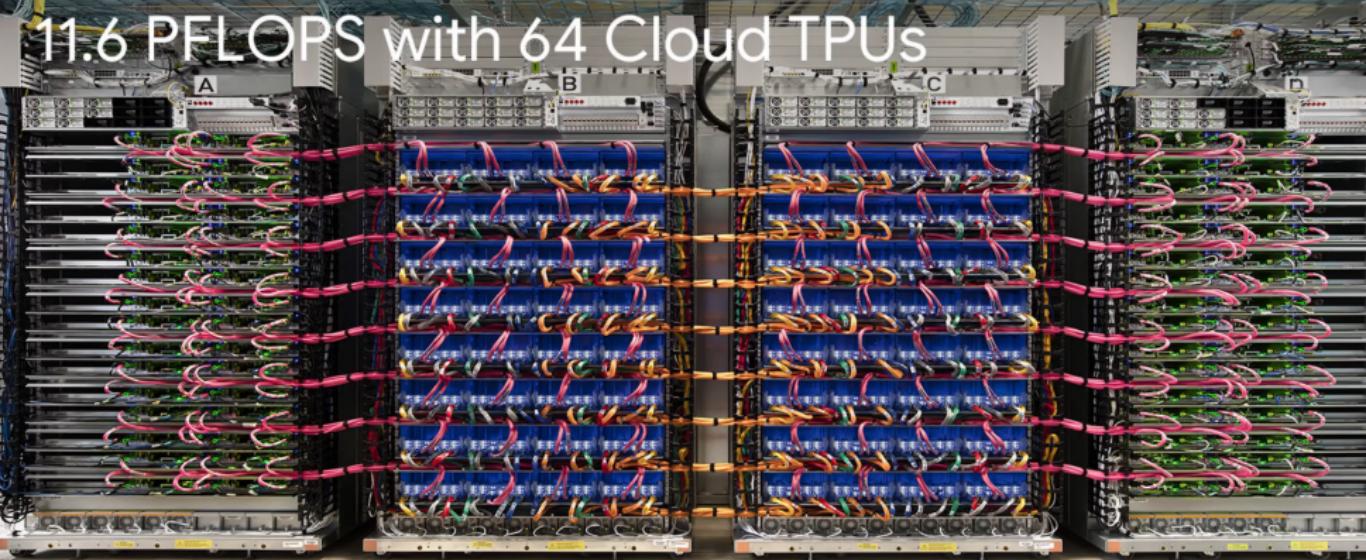
Free Jupyter Notebook environment with access to TPUs:
<https://colab.research.google.com>



TPUs

TPU Cloud

TPU v2 Pod: Google's HPC cluster for ML
11.6 PFLOPS with 64 Cloud TPUs



[<https://storage.googleapis.com/nexttpu/index.html>]

TPUs vs GPUs

Performance, TPUs vs GPUs, who wins? (SKIP most)

- ▶ Huge advantage for TPU performance-per-watt,
- ▶ Colab performance:
inconclusive (TPU part does not work yet),
- ▶ TPU only for inference?

	K80 2012	TPU 2015	P40 2016
Inferences/Sec <10ms latency	1/13 TH	1X	2X
Training TOPS	6 FP32	NA	12 FP32
Inference TOPS	6 FP32	90 INT8	48 INT8
On-chip Memory	16MB	24 MB	11 MB
Power	300W	75W	250W
Bandwidth	320 GB/S	34 GB/S	350 GB/S

[<https://www.extremetech.com/computing/247403-nvidia-claims-pascal-gpus-challenge-googles-tensorflow-tpu-updated-benchmarks>]

Exotic Hardware

- ▶ Intel Phi multicore CPU, 64 i386 cores:



- ▶ Raspberry PIs + Intel Movidius stick



OpenCV



TensorFlow



python™



RaspberryPi

- ▶ FPGAs

