Multi-column Deep Neural Networks for Image Classification

Duy Hung Nguyen

TECHNISCHE UNIVERSITÄT DARMSTADT

Gliederung

TECHNISCHE UNIVERSITÄT DARMSTADT

1. Einleitung

2. Multi-Column Deep Neural Networks (MCDNN) Architektur

    2.1. Deep Convolutional Neural Network (DNN)
    2.2. MCDNN

3. Experimente

    3.1. MNIST
    3.2. NIST SD 19
    3.3. Chinese characters
    3.4. NORB
    3.5. Traffic signs
    3.6. CIFAR 10

4. Fazit

Einleitung

TECHNISCHE UNIVERSITÄT DARMSTADT

- MCDNN Architektur verbessert die Erkennung von handgeschriebenen Ziffern, Buchstaben, Bildern oder Verkehrsschildern.

- DNNs entwickelt ihre vollständige Potenzial wenn sie breit und tief sind.
  - breit : viele Karten pro Schicht
  - tief : viele Schichten

- Um riesige DNNs zum trainieren, führen wir sie auf schnelle GPUs aus.

- Die Kombination von mehreren Spalten DNN in eine Multi-Spalte DNN (MCDNN) sinkt die Fehlerrate von 30-40 %

Einleitung

TECHNISCHE UNIVERSITÄT DARMSTADT

- 4 x GTX 580 1.5GB RAM
- >6 TFLOPS (maximum theoretical speed)
- 50-100x speed-up compared with a single threaded CPU version of the CNN program (one day on GPU instead of two months on CPU)

Deep Convolutional Neural Network (DNN)

TECHNISCHE UNIVERSITÄT DARMSTADT

- Vorgestellt von Fukushima (80), verbessert von LeCun u.a.(98), verfeinert und vereinfacht von Riesenhuber u.a.(99), Simard u.a.(03), Behnke (03), Ciresan u.a.(11).

- DNN ist von der Neocognition inspiriert :
  - tief (hunderte von Karten pro Schicht)
  - viele Schichte (6 – 10) aus nicht-linearen Neuronen übereinander gestapelt.

- Neocognition: Das Netzwerk ist selbstorganisiert durch “Lernen ohne Lehrer” , erwirbt die Fähigkeit, die Reizmuster zu erkennen auf der Grundlage der geometrischen Ähnlichkeit (Gestalt) der Formen, ohne von ihren Positionen betroffen.

Deep Convolutional Neural Network (DNN)

TECHNISCHE UNIVERSITÄT DARMSTADT

- Vielschichtige DNNs sind vor zwei Jahrzehnten schwer mit Gradientenabstiegs zu trainieren
- Aktuelle CPU ist mehr als 60000 mal schneller.
- Sorgfältige konzipierte Code für massiv parallele GPU : zusätzliche Bechleunigung Faktor 50-100 über Seriencode für Standard-Computern.

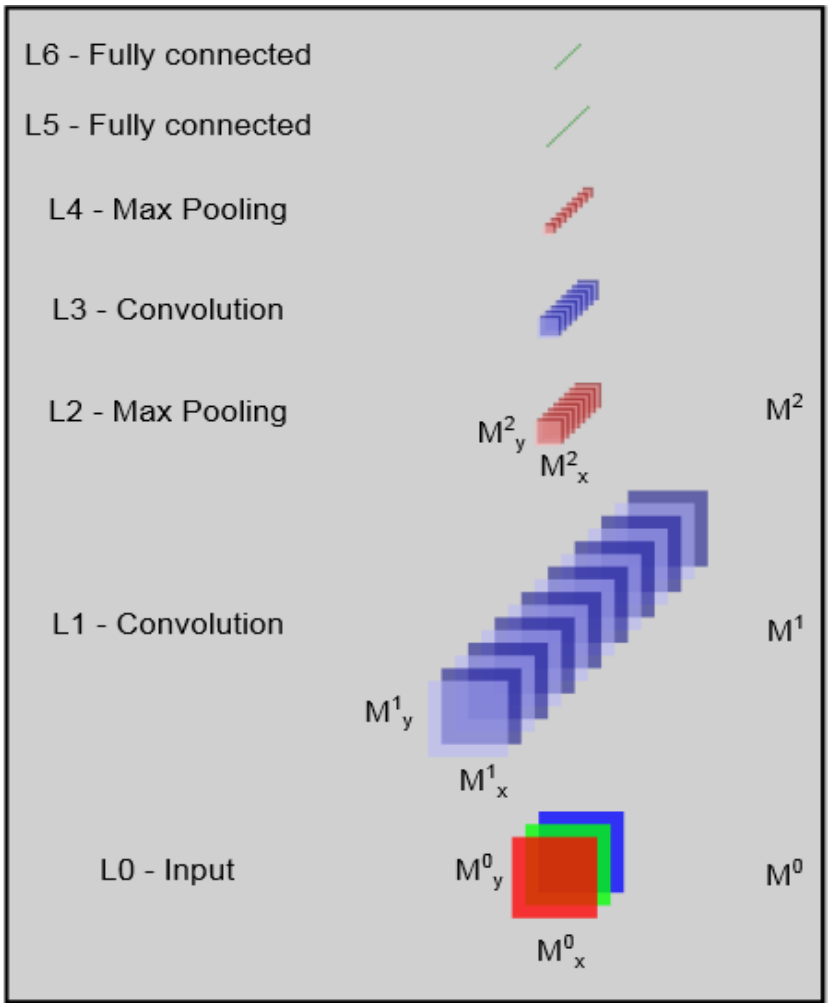
DNN Achitektur

TECHNISCHE UNIVERSITÄT DARMSTADT

- DNN haben 2-dimensionale Schichten von Winner-take-all Neuronen mit überlappenden Rezeptiven Felder, deren Gewichte geteilt werden.
- Eine einfache max Pooling-Technik bestimmt gewinnen Neuronen durch die Aufteilung die Schichten in quadratische Regionen der lokalen Hemmung, die aktivsten Neuron der jeweiligen Region ist Gewinner.
- Die Gewinner stellen eine kleinere, downsampling-Schicht mit geringerer Auflösung, versorgen der nächsten Schicht in der Hierarchie.
- Dieser Ergebnis ist inspiriert von Hubel und Wiesel (mit einfache und komplexe Zelle)

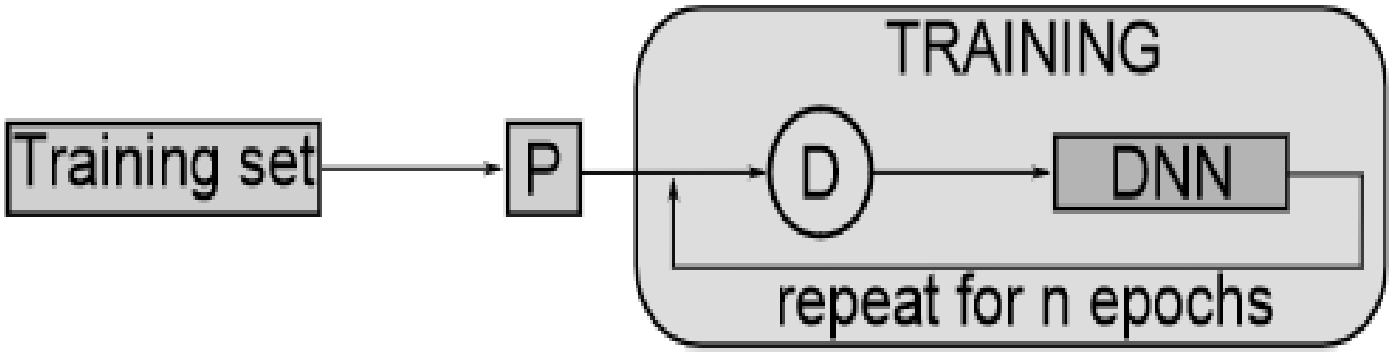
DNN Architektur

TECHNISCHE UNIVERSITÄT DARMSTADT



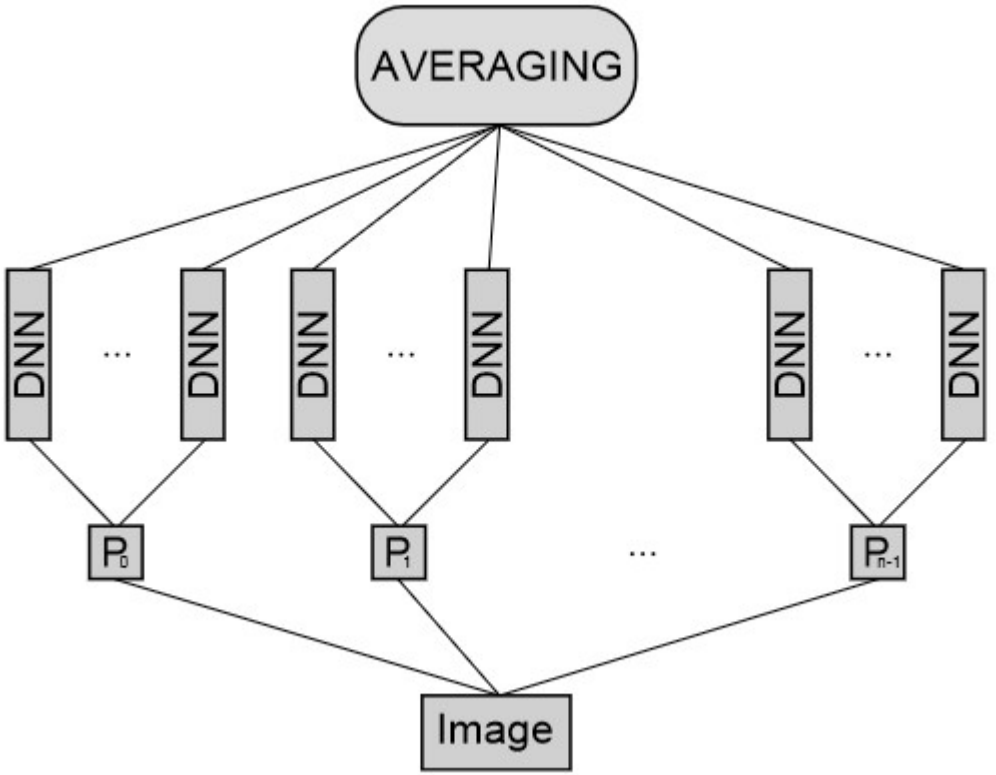
Trainieren ein DNN

TECHNISCHE UNIVERSITÄT DARMSTADT



MCDNN Architektur

TECHNISCHE UNIVERSITÄT DARMSTADT



MCDNN Architektur

TECHNISCHE UNIVERSITÄT DARMSTADT

$$y^i_{MCDNN} = \frac{1}{N} \sum_{j}^{\#columns} y^i_{DNN_j}$$

Predictions of all columns are averaged

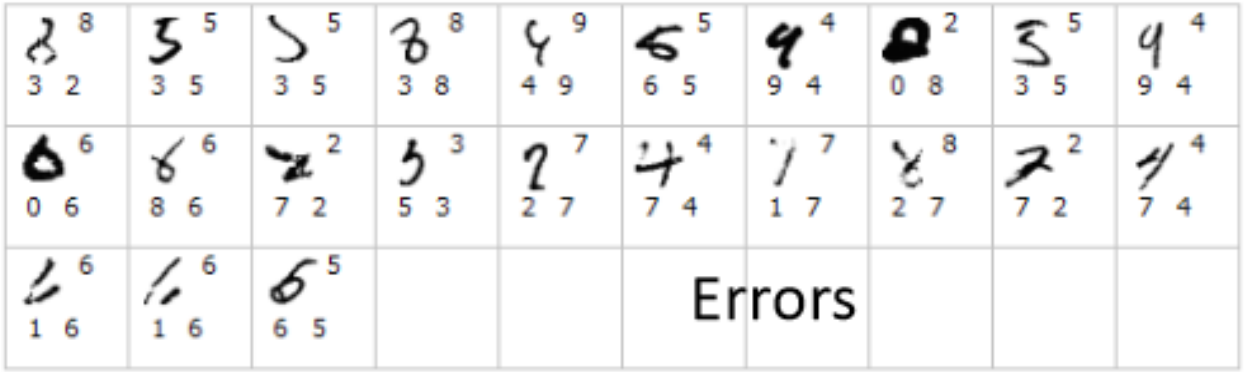
i corresponds to the i-th Class

j runs over all DNN

Experiment : MNIST

TECHNISCHE UNIVERSITÄT DARMSTADT

- Simard et al. (2003) – 0.40%, Ciresan et al. (2010) – 0.35%
- Distortions: translation, rotation, scaling, elastic
- 800 epochs of training for each DNN
- DNN architecture: 29x29-20C4-MP2-40C5-MP3-150N-10N
- 35column MCDNN: 6+1 width normalizations x 5 instances
- **0.23%** error rate (20 out of 23 digits have a correct second prediction)



Experiment: NIST SD 19

TECHNISCHE UNIVERSITÄT DARMSTADT

- Collection of ~0.8M handwritten digits and Latin characters
- Same setup as for MNIST

Data (# classes)	**MCDNN error [%]**	Previous best result [%]
All (62)	**11.63**	-
Digits (10)	**0.77**	1.88
Letters (52)	**21.01**	30.91
Letters (26) - merged	**7.37**	13.00
Merged (37)	**7.99**	-
Uppercase (26)	**1.83**	6.44
Lowercase (26)	**7.47**	13.27

Experiment : Chinesische Charakter

TECHNISCHE UNIVERSITÄT DARMSTADT

- 300 writers, 3755 classes, 1.2 Million characters, 3GB
- One week of training on GPU (corresponding to ~14 months on CPU)
- DNN: 48x48-100C3-MP2-200C2-MP2-300C2-MP2-400C2-MP2-500N-3755N
- First method which works directly on images
- Offline task: **6.5%** error vs. 10.01% (Liu et al. 2010)
- Online task: **5.61%** vs. 7.61% (Liu et al. 2010)
- **1th** place at ICDAR 2011 competition

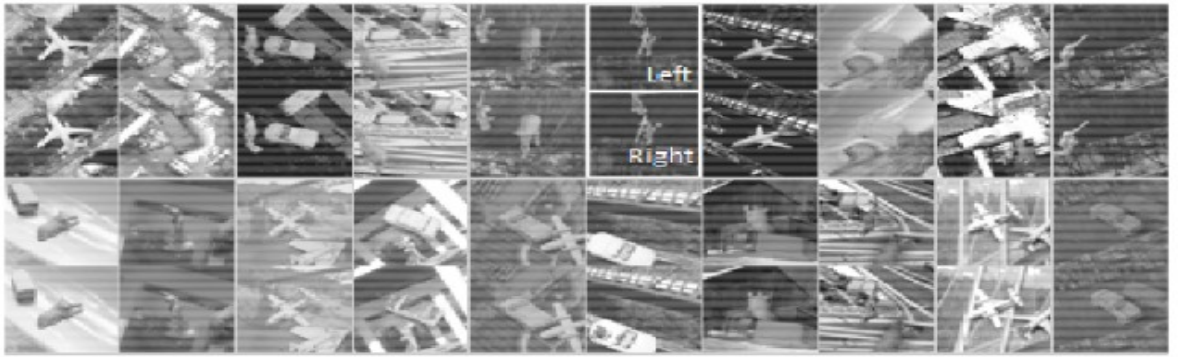
Experiment : NORB

TECHNISCHE UNIVERSITÄT DARMSTADT

- 3D object recognition from stereo images
- Training set: 291600 48x48 stereo images
- 5 classes with 10 instances: 5 instances for training and 5 for testing
- Challenging dataset, only 5 instances/class, some instances from test set are completely different than the one from training set
- DNN: 2x48x48-100C5-MP2-100C5-MP2-100C4-MP2-300N-100N-6N
- Distortions: translation (max 15%), rotation (max 15°), scaling (max 15%)

Training set size	Errors for 4 runs [%]				Mean [%]
First 2 folds	4.49	4.71	4.82	4.85	4.72±0.16
	4-net MCDNN error: 3.57%				
All 10 folds	3.32	3.18	3.73	3.36	3.40±0.23
	4-net MCDNN error: **2.70%**				

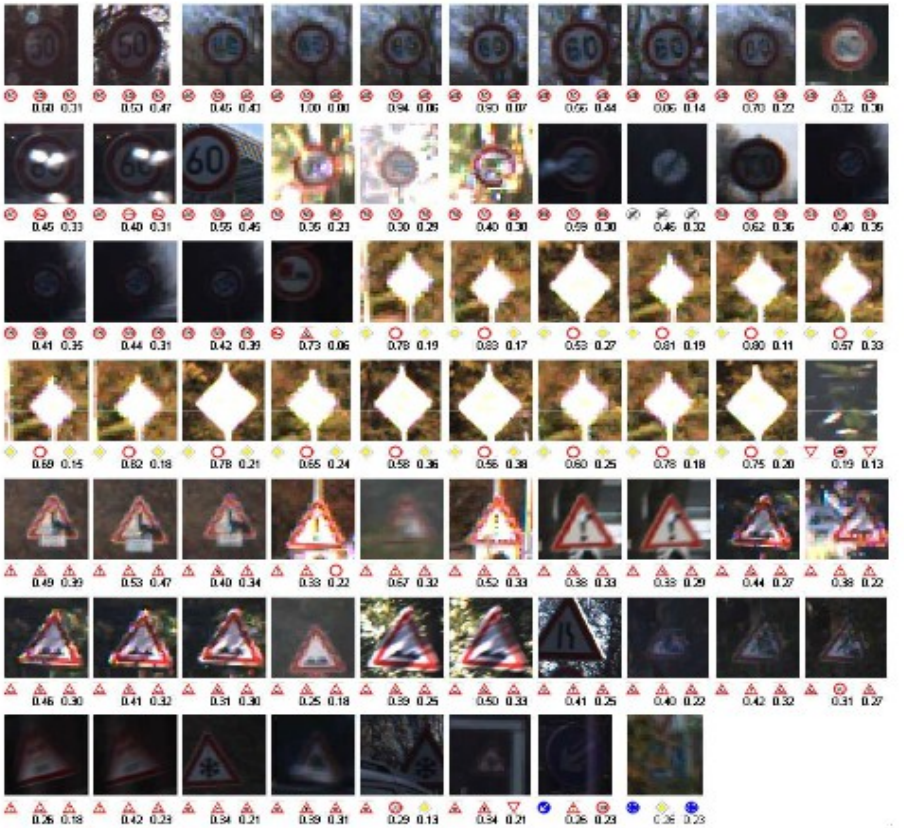
Previous lowest error: 5.00% (Coates et al. 2011)



Experiment : Traffic signs

TECHNISCHE UNIVERSITÄT DARMSTADT

- 26640 training and 12569 testing images, 43 classes
- 25 column MCDNN: 4+1 normalization x 5 instances
- DNN: 3x48x48-100C7-MP2-150C4-150MP2-250C4-250MP2-300N-43N
- **1th** place at Traffic Sign competition at IJCNN 2011
- Error rate: **0.54%**; outperforms the second best algorithm by a factor of three



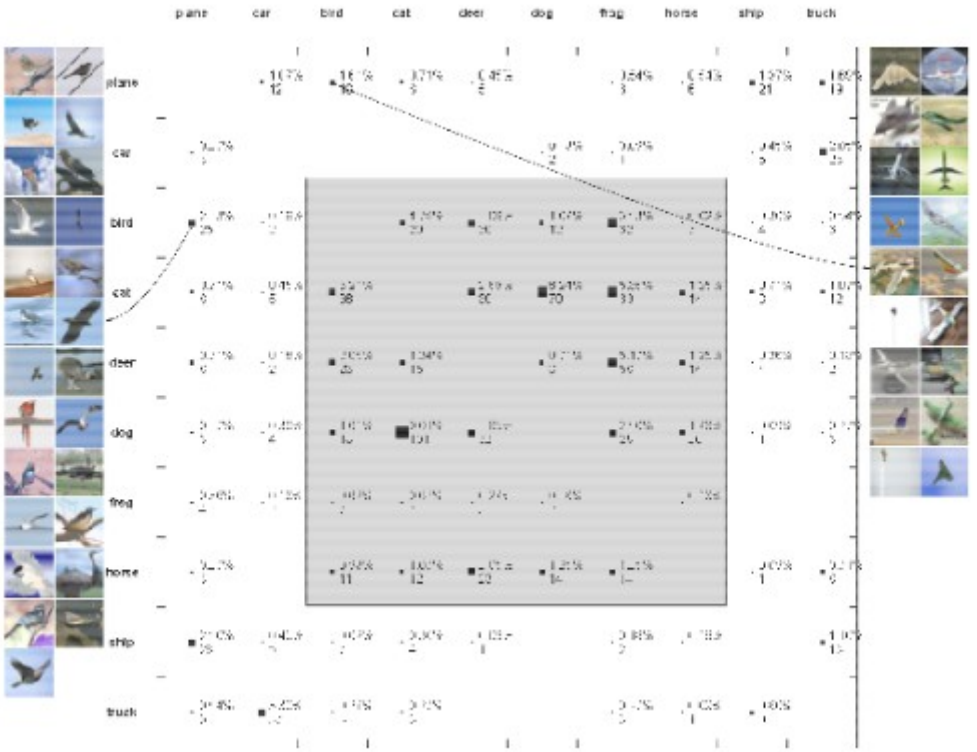
Experiment : CIFAR 10

TECHNISCHE UNIVERSITÄT DARMSTADT

- Small, 32x32 pixels color images, complex backgrounds
- DNN: 3x32x32-300C3-MP2-300C2-MP2-300C3-MP2-300C2-MP2-300N-100N-10N
- Affine distortions

preprocessing	Errors for 8 runs [%]				Mean [%]
Yes	16.47	19.20	19.72	20.31	18.93±1.69
No	15.63	15.85	16.13	16.05	15.91±0.22
	8-net average error: 17.42±1.96%				
	8-net MCDNN error: **11.21%**				

Previous lowest error: 18.50% (Coates et al. 2011)



Fazit

TECHNISCHE UNIVERSITÄT DARMSTADT

- Mensch-Wettbewerbsergebnisse werden auf weit verbreitete Computer-Vision-Benchmarks angegeben.
- MCDNN verbessert die state-of-the-art von 30-80%.
- Vollständig überwacht, verwendet keine zusätzlichen unmarkierten Datenquelle.
- Keine Notwendigkeit, handgearbeiteten Merkmale extrahieren, arbeitet an Roh-Pixel-Bilder
- Robust (kleinste Fehlerraten) und schnell (103-105 Bilder / s) für die sofortige industrielle Anwendungen

Fazit

TECHNISCHE UNIVERSITÄT DARMSTADT

Dataset	Best result of others [%]	MCDNN error [%]	Relative improvement [%]
MNIST	0.39	0.23	41
NIST SD 19			30-80
HWDB1.0 on.	7.61	5.61	26
HWDB1.0 off.	10.01	6.50	35
CIFAR 10	18.50	11.21	39
Traffic signs	1.69	0.54	72
NORB j.-c.	5.00	2.70	46