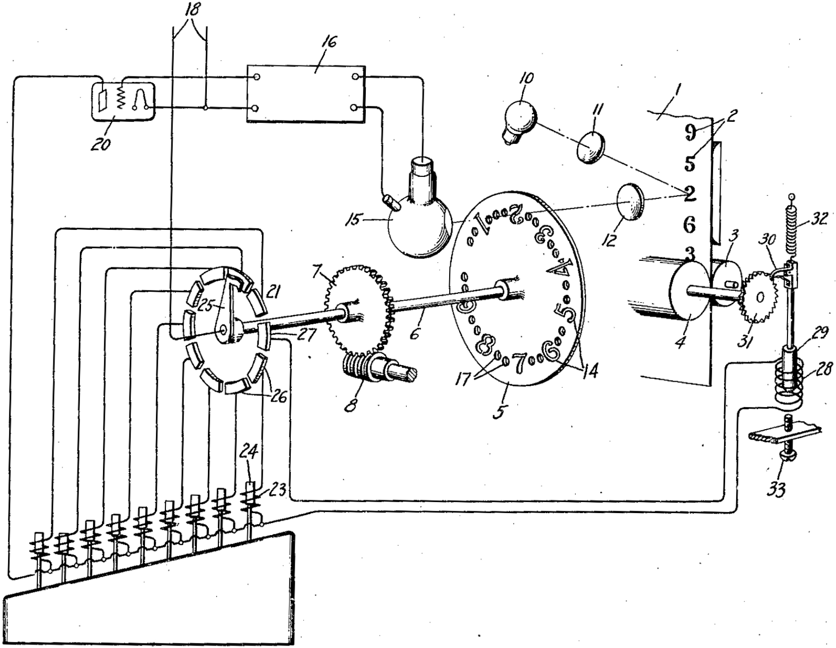
Spatially-Sparse Convolutional Neural Networks

Ben Graham

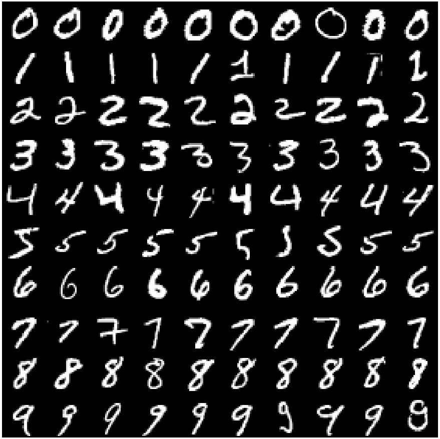
Department of Statistics
University of Warwick*

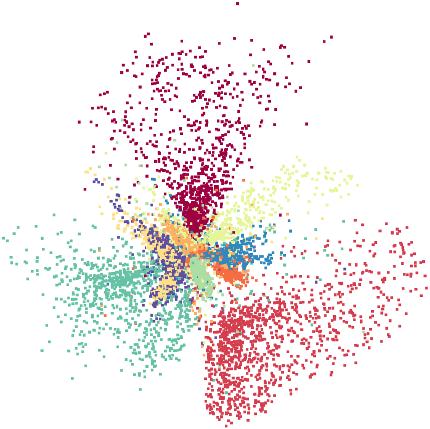
31 May 2016

Motivation (Paul Handel, 1931)

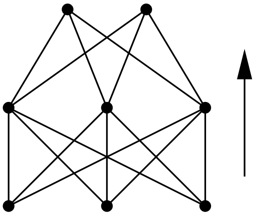


Motivation: Unsupervised learning for MNIST





Artificial neural networks



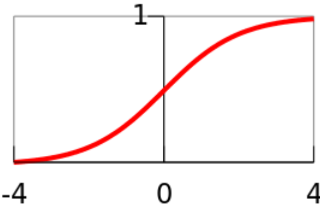
Directed weighted graph
For each node:
output=$\sigma(b+\sum_i w(i)\mathrm{input}(i))$

For classification, the final layer is weighted to give a probability distribution.

input$\in \mathbb{R}^a$

$$\begin{aligned}\text{hidden1} &= \sigma(\mathrm{input}\cdot W_1 + B_1) \in \mathbb{R}^{\mathrm{b}}\\ \text{hidden2} &= \sigma(\mathrm{hidden1}\cdot W_2 + B_2) \in \mathbb{R}^{\mathrm{c}}\\ \text{hidden3} &= \sigma(\mathrm{hidden2}\cdot W_3 + B_3) \in \mathbb{R}^{\mathrm{d}}\end{aligned}$$

output$= \mathrm{softmax}(\mathrm{hidden3}\cdot W_4 + B_4) \in \mathbb{R}^{\mathrm{e}}$

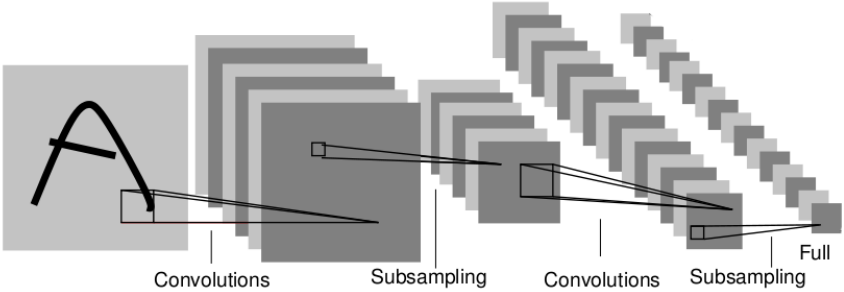


#Parameters $(a+1)\times b+(b+1)\times c+(c+1)\times d+(d+1)\times e$

ConvNets

- 1998: LeNet-5 (LeCun et al)

$$32 \times 32 \text{ input: } 6\text{C}5 - \text{MP}2 - 16\text{C}5 - \text{MP}2 - 120\text{C}5 - \text{output}$$



Convolutional Neural Network Software

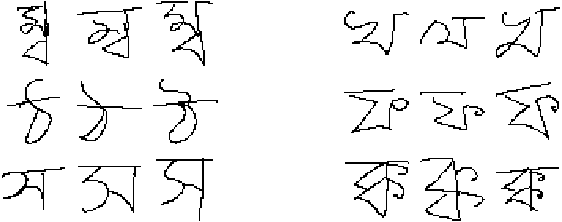
- cuda-convnet
- Torch7
- Theano: Keras, Lasagne, PyLearn2
- OverFeat
- Caffe
- CxxNet
- CuDNN
- MatConvNet
- TensorFlow
- Marvin
- **SparseConvNet (2012-)**

Why Spatially Sparse ConvNets?

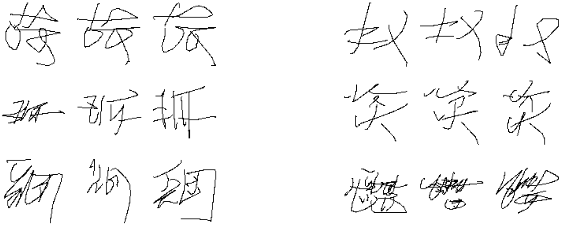
- Research keyword bingo
- Online Handwriting
  - 1d manifold in 2d space
  - computation should scale linearly with rendering scale
  - CVJK - fine detail
- Data augmentation
- Differences in input size
- Curse of dimensionality - higher dimensional spaces are sparser
- 3D space - includes 2+1 dimensional space time
- 4D space - includes 3+1 dimensional space time
- 5D space ??

Handwriting recognition

- 183 Assamese characters, 36 training exemplars per class



- 3755 Chinese CASIA-OLHWDB1.1 characters, 240 exemplars/class



Network structure = Prior information

$$\overline{o} \qquad \overset{\rule{1em}{0.4pt}}{o} \qquad \underline{o}$$

Data augmentation

- Assamese at scale 2^ℓ

DeepCNet(ℓ, 30)	None	Translations	Affine
$\ell = 3$	51.4%	38.3%	35.2%
$\ell = 4$	26.8%	9.29%	7.47%
$\ell = 5$	18.2%	5.28%	2.73%
$\ell = 6$	14.1%	4.61%	1.76%
$\ell = 7$	**13.8%**	**4.07%**	**1.70%**

Sparsity

- Sparse matrix multiplication has high overheads
  - loss of Strassen's algorithm
  - inefficient memory access
- Spatially-Sparse ConvNets are relatively efficient
  - Contiguous memory access
  - Small filters and pooling regions preserve sparsity
  - Small filters work well with FMP