Institut
Mines-Télécom

Recognition and information extraction in multi-lingual documents with Recurrent Neural Networks and Deep Neural Networks

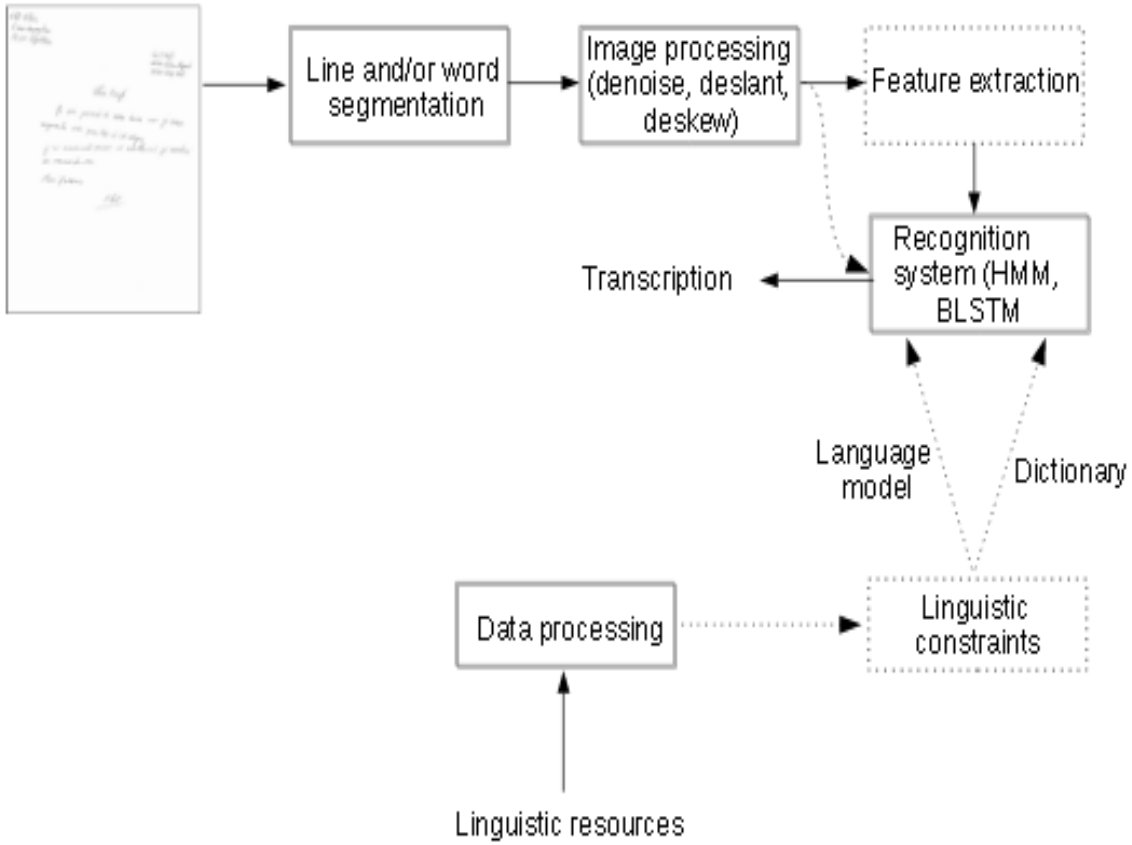
PhD Candidate: Bogdan-Ionut Cirstea

Supervisors: Laurence Likforman-Sulem (Télécom Paris-Tech), Emmanuèle Grosicki (DGA)

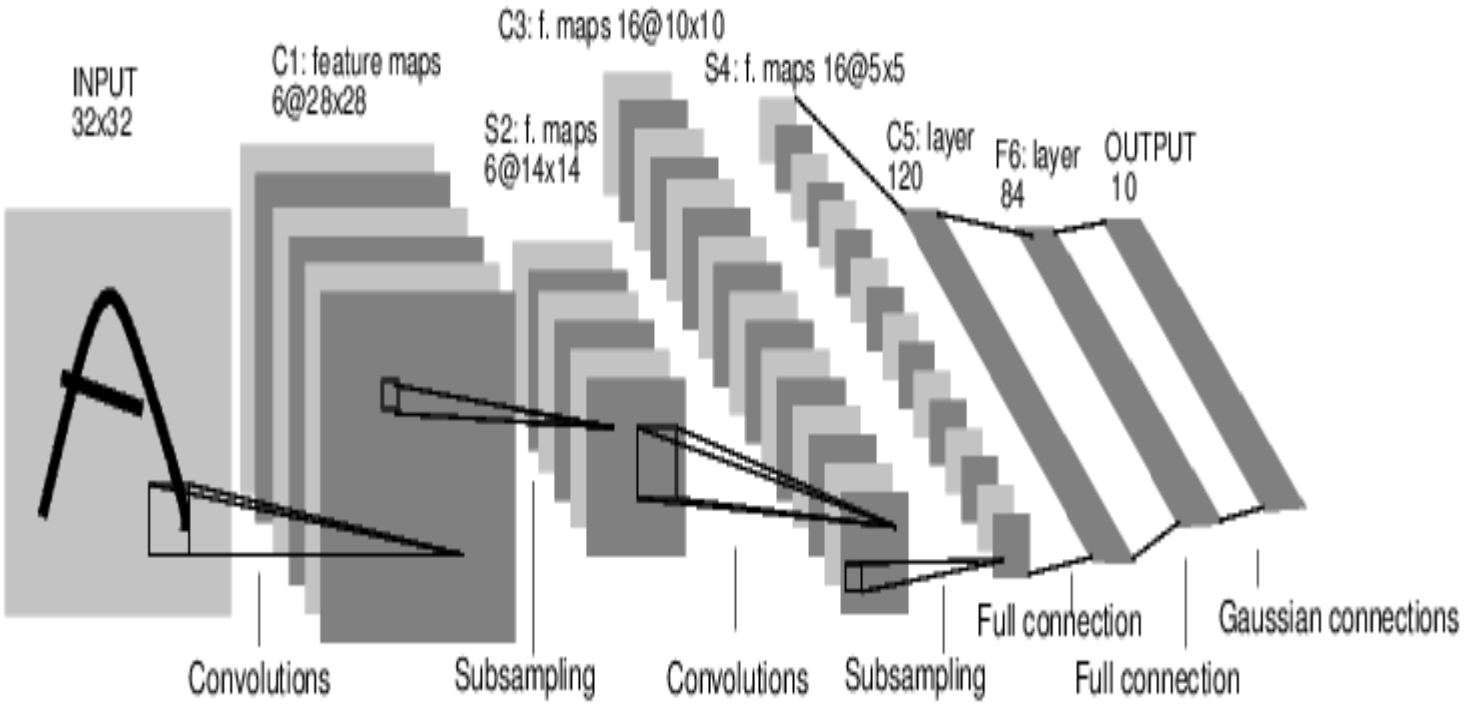
Handwriting recognition

- **Relevance**
  - Automatic document processing – bank checks, postal envelopes, handwritten mails, bills, forms
  - From isolated characters to words and pre-segmented lines of handwritten text
- **Difficulty**
  - Difficult conditions: multi-lingual documents, large vocabulary and variety of styles, ancient degraded documents
- **Historically significant dataset – MNIST (isolated handwritten digits)**
- **State of the art results with first deep learning models**
  - Convolutional neural networks (CNN) – LeNet5 (LeCun, 1998) on MNIST
  - Long short term memory (LSTM) recurrent neural networks (RNN) – MDLSTM (Graves, 2009) for handwriting sequences recognition

Example of a classical handwriting recognition system

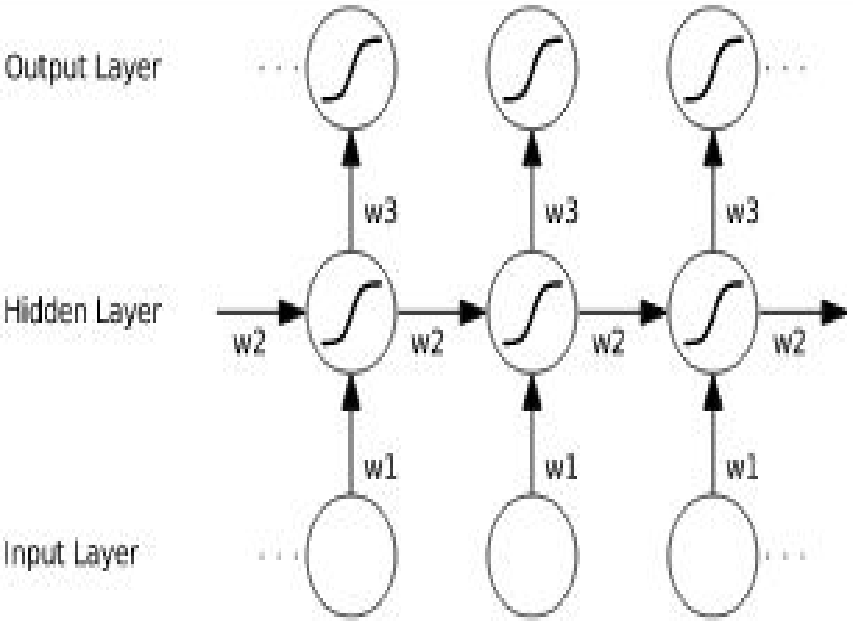


LeNet-5 Convolutional neural network (CNN) (LeCun, 1998)



Recurrent neural network (RNN)

[from (Graves, 2013a)]



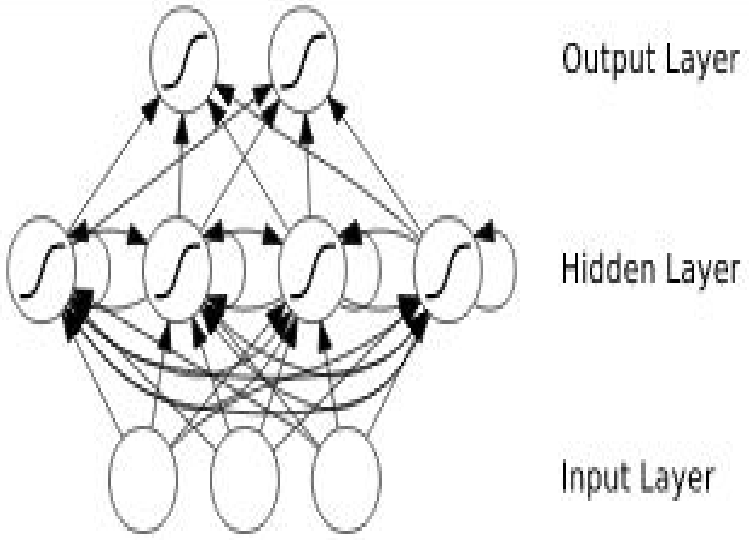
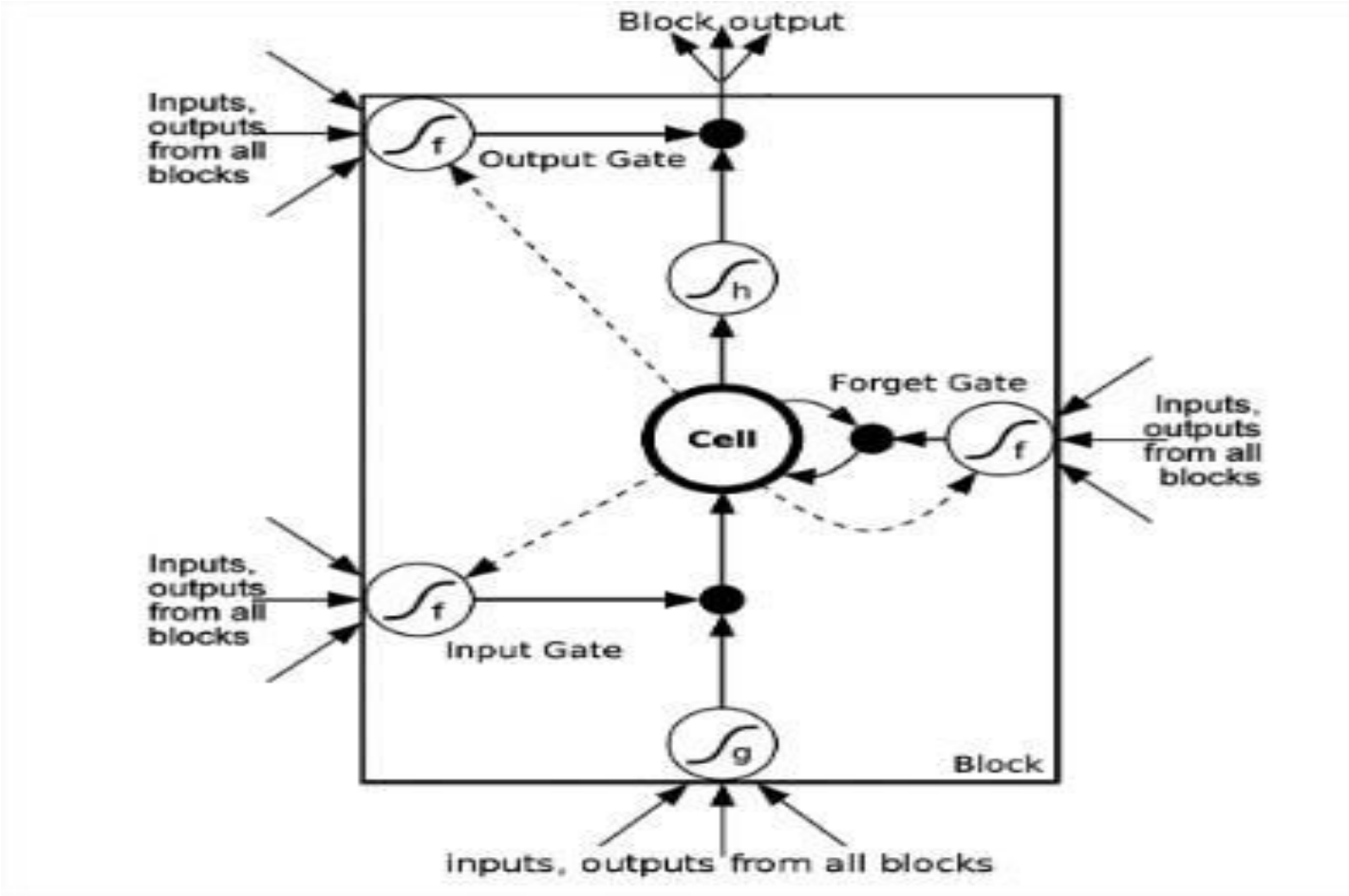
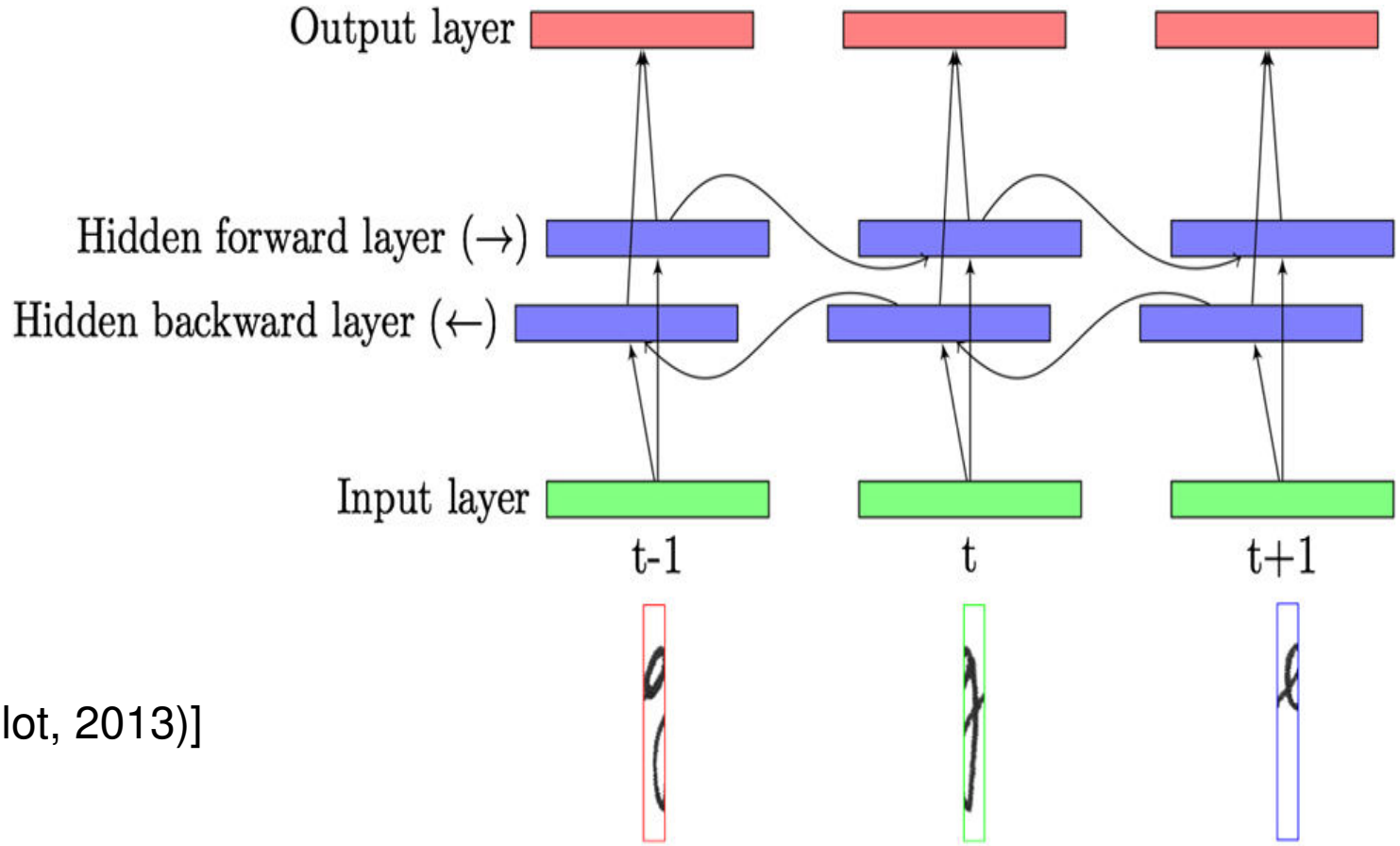


Figure 3.3: **A recurrent neural network.**

Long short term memory (LSTM) block



Bidirectional LSTM (BLSTM) for recognizing handwriting sequences



[from (Morillot, 2013)]

Hierarchical subsampling RNNs (Graves, 2013a)

combination of subsampling layers and multi-dimensional LSTM

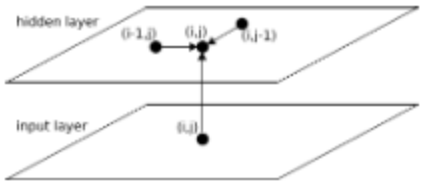
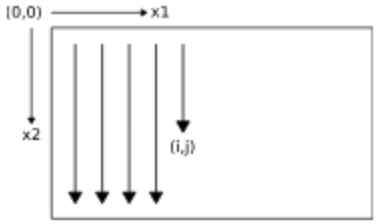
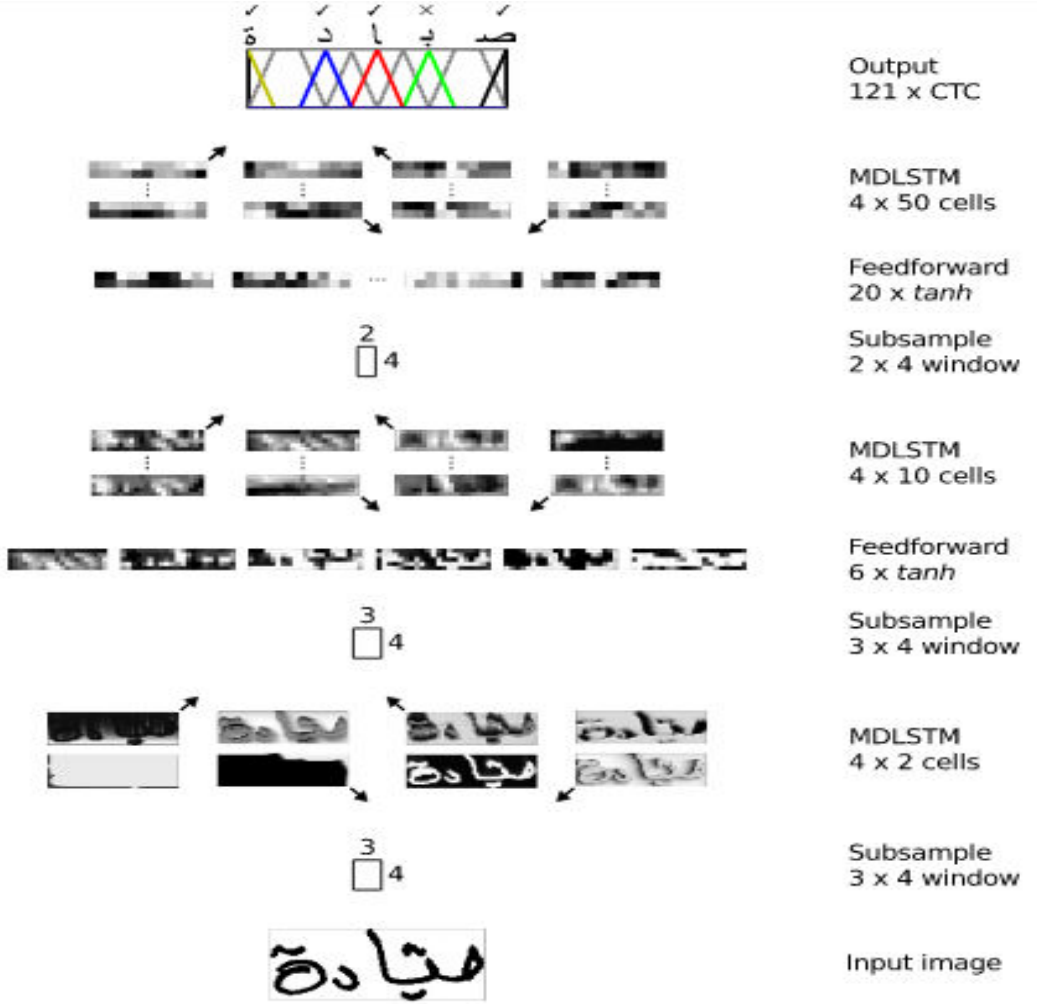


Figure 8.1: **MDRNN forward pass**

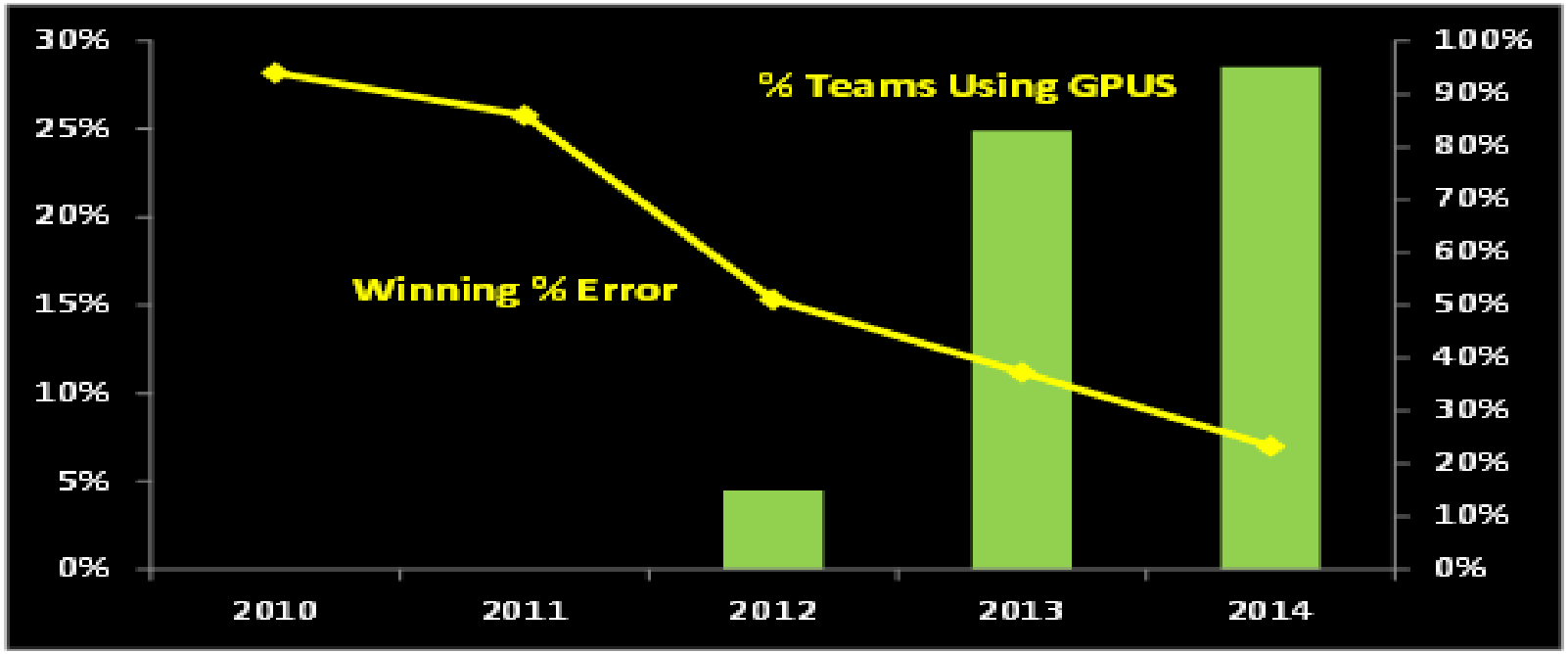


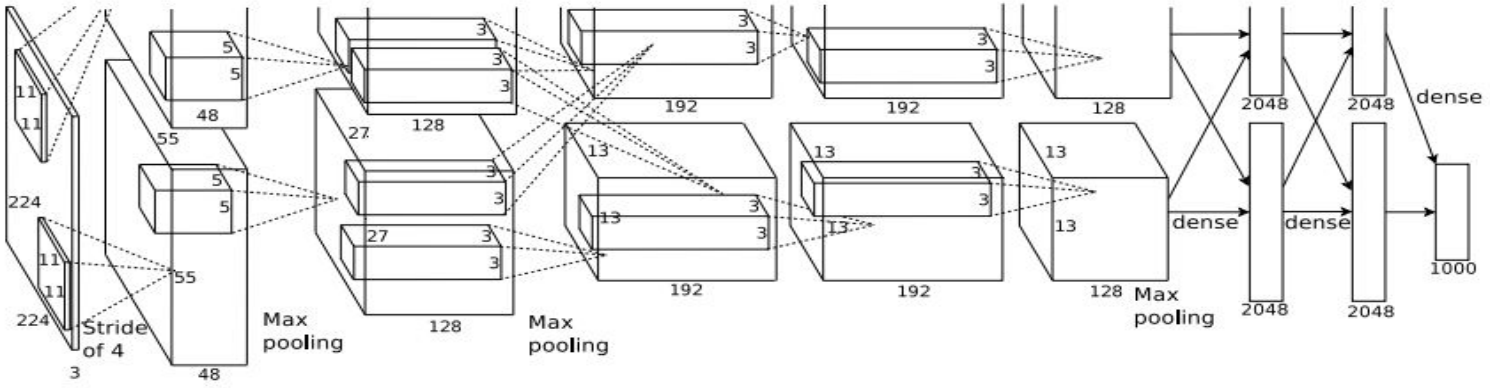


Handwriting recognition state of the art – NIST 2013 OpenHaRT

Team	Image preproce-ssing	Hand - designed features	Optical model	1-WER accuracy
A2iA	No	No	MDLSTM	0.799
RWTH	Yes	Yes	BLSTM + HMM	0.762
CITLAB	Yes	No	MDRNN	0.737
UBV	Yes	Yes	Bernoulli HMM	0.706
UOB-TPT	Yes	Yes	BLSTM + HMM	0.520
LITIS	Yes	Yes	HMM	0.224

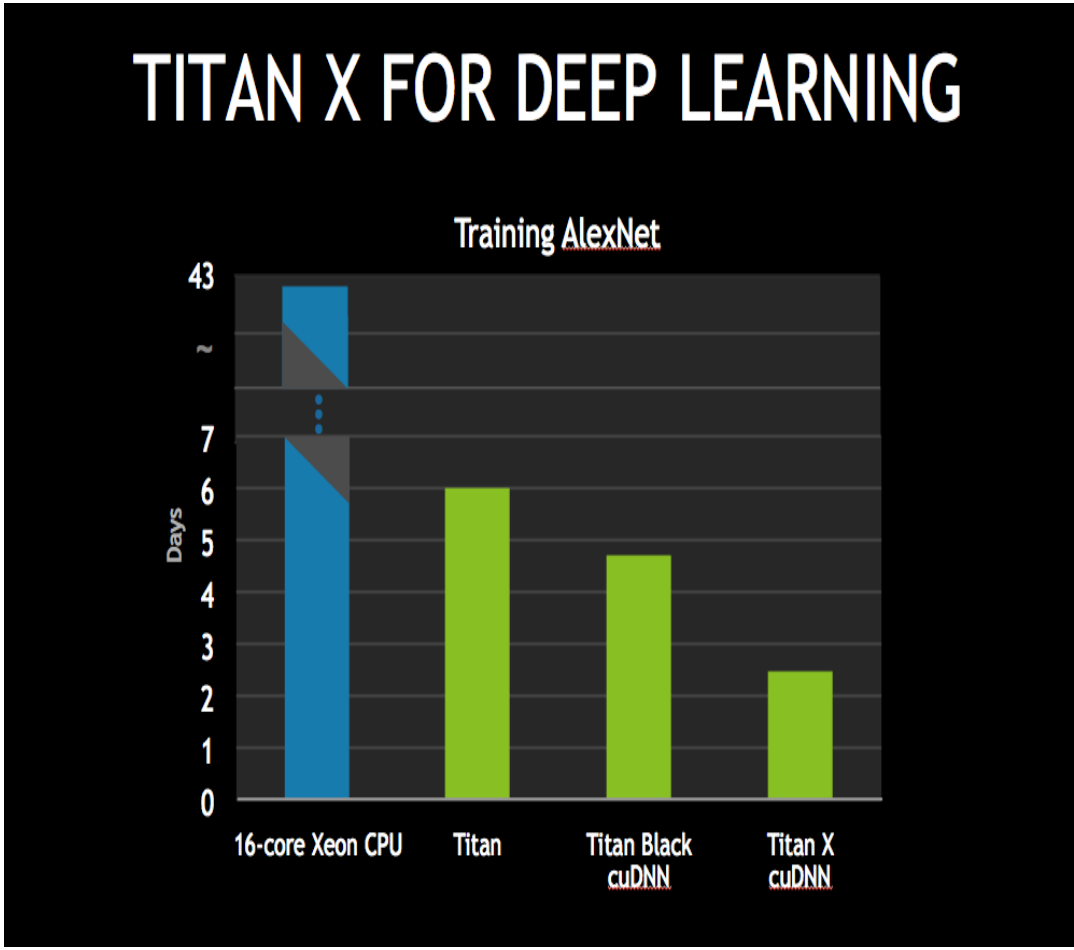
ImageNet Large Scale Visual Recognition Challenge – (Krizhevsky, 2012), [2]





The need for GPUs [1, 2]

Training a state of the art language model		
Device	Small LSTM	Large LSTM
Intel Core i7 2.6 GHZ	156 minutes	9351 minutes
NVIDIA GTX 980	30 minutes	1006 minutes
Speedup	5.2x	9.29x



Limiting factors for applying DL to handwriting recognition (hypotheses)

- **Less accessible and smaller datasets than in computer vision**
  - Hand-designed features and simpler methods have often worked well
- **No public MDLSTM implementation trainable on GPUs**
- **Sequential models less accessible than feedforward ones from the most popular deep learning frameworks (Torch, Theano, Caffe)**
- **Not enough / adapted computational power – from (Morillot , 2013):**
  - BLSTM – WER: 16.1%; total training time: 351 hours
  - HMM – WER: 22.4%; total training time: 21 hours
  - Intel Xeon E5620 2.4 Ghz, 8 threads
  - RNNlib (no GPU support)

What we plan to do (I)

- **Combine deep neural networks and recurrent neural networks in robust systems, trained end-to-end, with the same loss function**

- **Build bigger / more flexible / more complex models**
  - Variants of activation functions
  - Architecture variants
  - Better learning algorithms and optimization (especially for recurrent neural networks)
  - Hyperparameter learning (including architecture)
  - Attention models
  - Encoder/decoder approaches
  - Introduce document context

What we plan to do (II)

- **Use more data**
  - Transfer learning, multi-task learning, domain adaptation: single characters, multiple alphabets, multiple languages, synthetic data (e.g. CAPTCHAS)
  - Generate data - models that can learn to generate handwriting sequences, variational approaches

[from (Graves, 2013b)]

- **All the approaches above can benefit significantly from more computational power**

Tools we are using

- **Torch**
  - Fast computationally and quite easy to develop in / extend
  - GPU and CPU support
  - Used in both industrial (Facebook, Google, Google DeepMind) and academic settings (NYU)
- **Theano**
  - Automatic differentiation
- **Caffe**
  - Probably fastest for CNNs
- **RNNlib**
  - Framework in which handwriting recognition models have previously been developed at TPT
  - No GPU support
  - Port code to (some of) the frameworks above

References

[1] http://devblogs.nvidia.com/parallelforall/understanding-natural-language-deep-neural-networks-using-torch/

[2] http://www.slideshare.net/NVIDIA/gtc2015-final-published

(Graves, 2009) Alex Graves, Jurgen Schmidhuber, 'Offline Handwriting Recognition with Multidimensional Recurrent Neural Networks'

(Graves, 2013a) Alex Graves, 'Supervised Sequence Labeling with Recurrent Neural Networks'

(Graves, 2013b) Alex Graves, 'Generating sequences with recurrent neural networks'

(LeCun, 1998) Yann LeCun, Léon Bottou, Yoshua Bengio and Patrick Haffner, 'Gradient-Based Learning Applied to Document Recognition'

(Krizhevsky, 2012) Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, 'ImageNet Classification with Deep Convolutional Neural Networks'

(Morillot, 2013) Olivier Morillot, Laurence Likforman-Sulem, Emmanuèlle Grosicki, 'New baseline correction algorithm for text-line recognition with bidirectional recurrent neural networks'

Questions?