No Padding Please: Efficient Neural Handwriting Recognition.

Gideon Maillette de Buy Wenniger*, Lambert Schomaker*, Andy Way**

*Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, *University of Groningen*, Groningen, The Netherlands

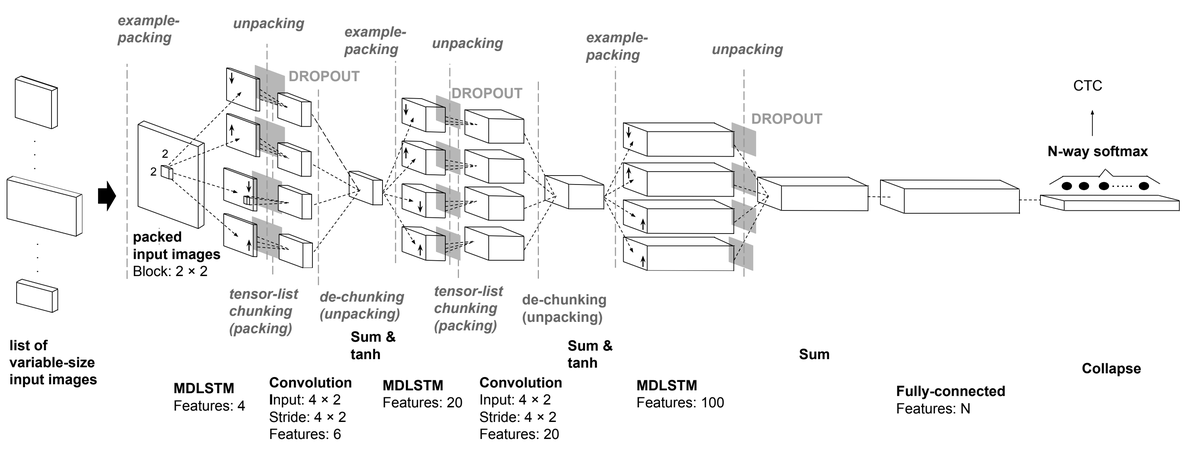
**The Adapt Centre, *Dublin City University*, Dublin, Ireland

gemdbw@gmail.com, l.r.b.schomaker@rug.nl, andy.way@adaptcentre.ie

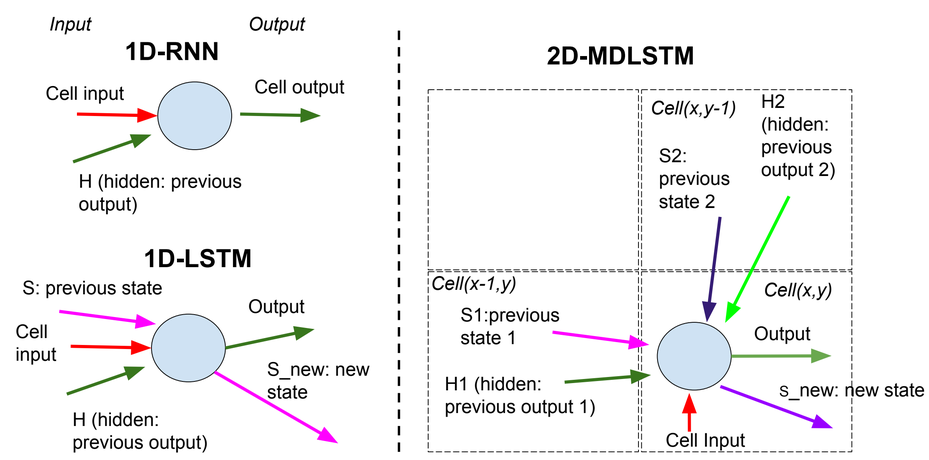
1. Introduction

Models with multi-dimensional long short-term memory (MDLSTM) layers have achieved state-of-the art results on handwritten text recognition tasks. Multi-directional MDLSTM layers have an unbeaten ability to capture the complete context in all directions, but this limits the possibilities for parallelization. We develop methods to create efficient MDLSTM-based models for NHR:
1) Example-packing: a new method that eliminates computational waste resulting from padding, 2) A technique to optimize parallelization using convolutions with grouping, 3) A method for parallelization across GPUs for variable-length example batches.

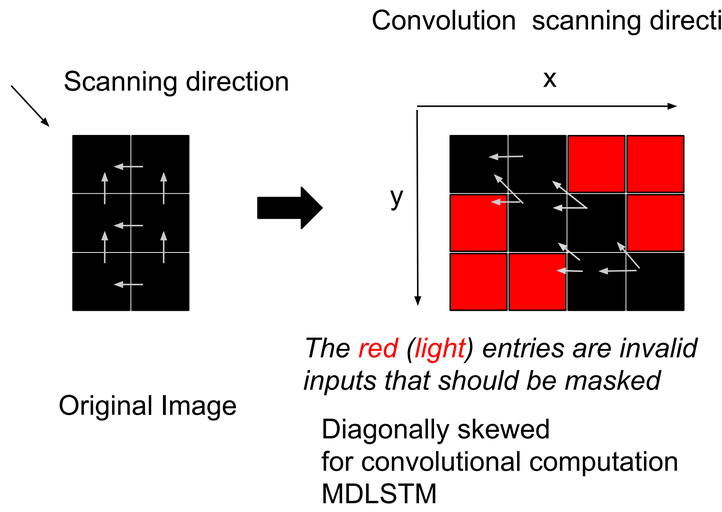
2. Model Structure



3. What are 2D-MDLSTMs?

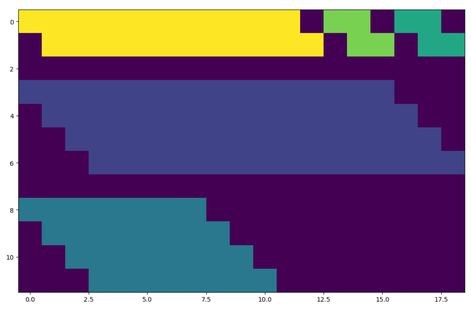


4. Efficient MDLSTM computation by convolution

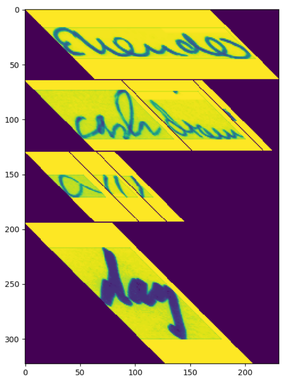


5. Example packing

Packing example (schematic)



Word-strips (skewed) packing



6. Example packing - details

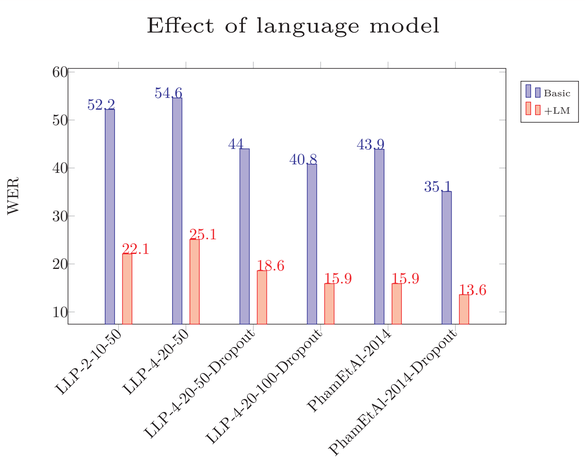
- Every row is filled greedily up to the maximum width
- Examples within a row must share same height, but different rows are allowed to have different heights
- Packing/unpacking done in pairs:
  - packing: receives a list and outputs a tensor
  - unpacking receives a tensor and ouptus a list
- Packing done before every MDLSTM layer, unpacking after it
- Major gains especially in word-based handwriting recognition setting (due to large variance in word lengths)

7. Recognition Performance

Comparison to literature results on IAM lines.

	validation		test	
System	WER	CER	WER	CER
Leaky LP Cell [4,20,100], dropout	33.9	9.8	40.8	12.9
+ Vocabulary and LM	15.5	6.4	15.9	6.6
Pham et.al (2014), no dropout	36.5	10.4	43.9	14.4
+ Vocabulary and LM	12.1	4.2	15.9	6.3
Pham et.al (2014), dropout	27.3	7.4	35.1	10.8
+ Vocabulary and LM	11.2	3.7	13.6	5.1
Voigtlaender et.al (2016)	7.1	2.4	9.3	3.5

Effect of language model



8. Speedup of example packing

Preparation of batch examples	batch size	time per epoch (HH:MM:SS)	examples per second	max GPU1 memory use (MB)	max GPU2 memory use (MB)
IAM lines					
batch-padding	8	07:24:06	0.243	10824	10675
example-packing	12	05:04:45	0.355	10694	10780
IAM words					
batch-padding	20	06:26:48	2.38	11074	11144
example-packing	200	00:58:22.	16.1	10827	10849

Memory and time usage for models with and without example packing, with batch sizes chosen the maximal possible given the observed maximum GPU memory usage.

9. Conclusions

- Example packing gives major speedup, factor 6.6 on words, on top of speedup by other techniques
- Performance system comparable to similar state-of-the-art systems
- Makes it feasible to use computationally expensive MDLSTMs within standard deep learning frameworks

10. References

[1] Vu Pham, Théodore Bluche, Christopher Kermorvant, and Jérôme Louradour. Dropout improves recurrent neural networks for handwriting recognition. *Arxiv https://arxiv.org/abs/1312.4569v2)*, pages 1–6, 2016.

[2] Gundram Leifert, Tobias Strauß, Tobias Grüning, Welf Wustlich, and Roger Labahn. Cells in multidimensional recurrent neural networks. *Arxiv https://arxiv.org/abs/1412.2620)*, pages 1–35, 2014.

[3] Paul Voigtlaender, Patrick Doetsch, and Hermann Ney. Handwriting recognition with large multi-dimensional long short-term memory recurrent neural networks. *In proceedings of the International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pages 228–233, 2016.

Source Code

`https://github.com/gwenniger/multi-hare` (To be released soon after ICDAR)

This research has been supported by the ADAPT Centre for Digital Content Technology which is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 713567.