

# 1. Handwriting Recognition with Large Multidimensional Long Short-Term Memory Recurrent Neural Networks

Paul Voigtlaender, Patrick Doetsch, Hermann Ney

在之前的实验中，由于缺少对 GPU 的充分利用，计算能力不足，限制了 LSTM 网络的深度和宽度。这篇论文就是创造了一个高效的基于 GPU 的应用。运用多维 LSTM，然后使用了 GPU 进行并行运算，剪短了时间，然后就可以训练更深更宽的网络，达到一个更好的结果。网络结构如 Figure1 所示。

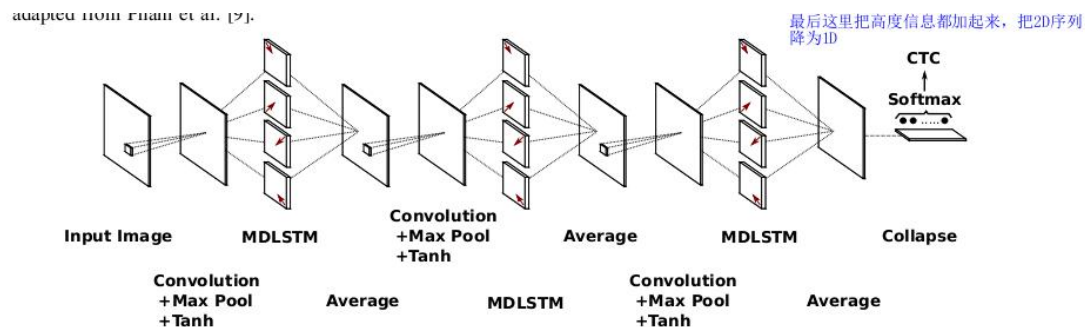


Figure1. Large Multidimensional Long Short-Term Memory Recurrent Neural Networks

另外，在获取多维信息的时候，为了减少计算的次数，使用了红角线顺序进行并行计算。如 Figure2 所示。其中(a)为最普通的计算方法，(b)为(a)的数字表示，(c)为对角线计算的数字表示。可以看出，对于一个  $N \times N$  的图像，普通的计算方法需要计算  $N^2$  次，而对角线顺序只需要  $2 \times N - 1$  次，可以大大减少计算时间。

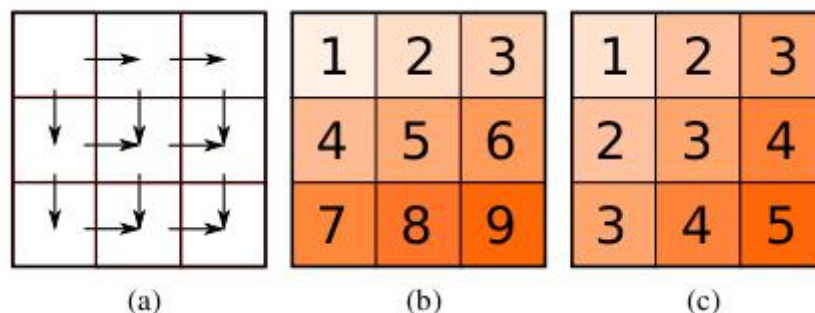


Figure2. 计算顺序

论文中也对初始值初始化的方法进行了对比，主要是对比了 Glorot initialization 跟 Normal initialization。使用不同的随机种子，对比网络的收敛性以及收敛速度，结果如 Figure 3 所示。可以看出 Glorot initialization 具有更快的收敛速度，而且对随机种子的选取的依赖性比较低。

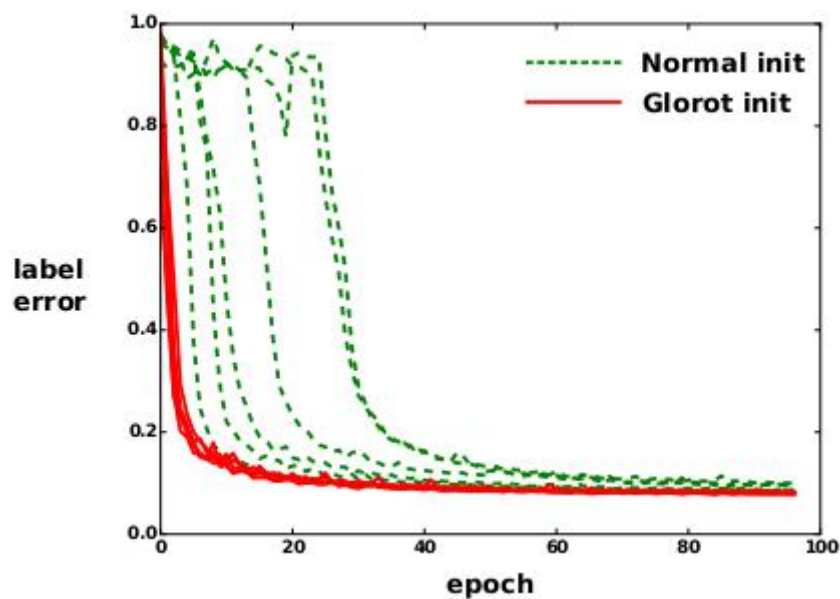


Figure 3.初始化对比

网络结果如 Figure 4 所示，与其他实验相比有较大的提升，应该是 IAM 文本行识别中，当前准确率最高的结果。

System	WER[%]		CER[%]	
	dev	eval	dev	eval
Our system	7.1	9.3	2.4	3.5
Doetsch et al. [2]	8.4	12.2	2.5	4.7
Voigtlaender et al. [30]	8.7	12.7	2.6	4.8
Pham et al. [9]	11.2	13.6	3.7	5.1

Figure4 .实验结果

## 2. Improvements in RWTH's system for off-line handwriting recognition

Michał Kozielski, Patrick Doetsch, Hermann Ney

主要方法：PCA 降维特征，HMM structure consists of 2 states，MLLR (THE MAXIMUM LIKELIHOOD LINEAR REGRESSION) 用于 **writer adpatation**，使用混合语言模型（贝叶斯判断准则）

网络大概结构如 Figure5 所示。

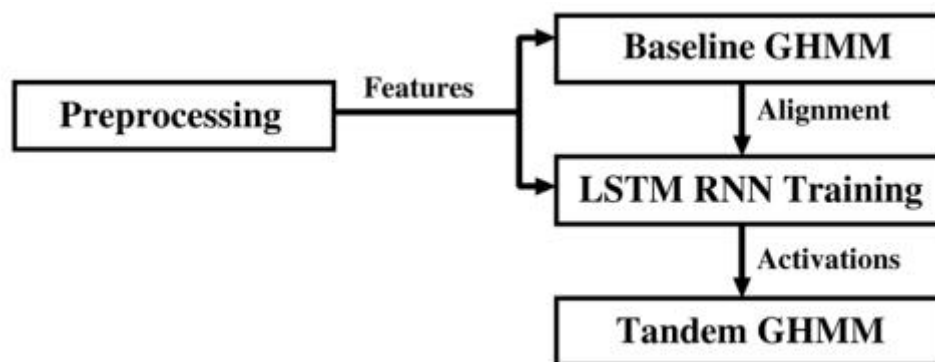


Figure5 .网络基本结构

IAM 数据结果如 Figure6 所示。

Systems	Voc.	WER [%]		CER [%]	
		Dev.	Eval	Dev.	Eval
Our system	50k	9.5	13.3	2.7	5.1
España et al. [25]	50k	19.0	22.4	-	9.8
Toselli et al. [26]	9k	-	25.8	-	-
Graves et al. [12]	20k	-	25.9	-	18.2
Bertolami et al. [27]	20k	26.8	32.8	-	-

Figure6 .IAM 数据集测试结果

### 3. Fast and robust training of recurrent neural networks for offline handwriting recognition

Patrick Doetsch, Michal Kozielski and Hermann Ney

这个论文主要是对 RWTH（论文 SEQUENCE. DISCRIMINATIVE TRAINING OF RECURRENT NEURAL NETWORKS）的改良，使用了 LSTM-RNN；使用了 sequence chunk 进行并行加速（把序列分割成等长块，见 Figure 7）；使用了 proposed gate scaling method，改变激活函数形状（见 Figure 8），在计算复杂度几乎没有改变的情况下提高了准确率（见 Figure9）。

proposed gate scaling method 有效的原因，主要是记忆细胞里面的中心单元是线性的，这样就是说是无边界的，但是输入门跟输出门接收来自记忆细胞的输入是通过激活函数的，这样他们的输入在输出空间里是有界的。而 proposed gate scaling method 就是改变这个从无界到有界空间的转换的程度。

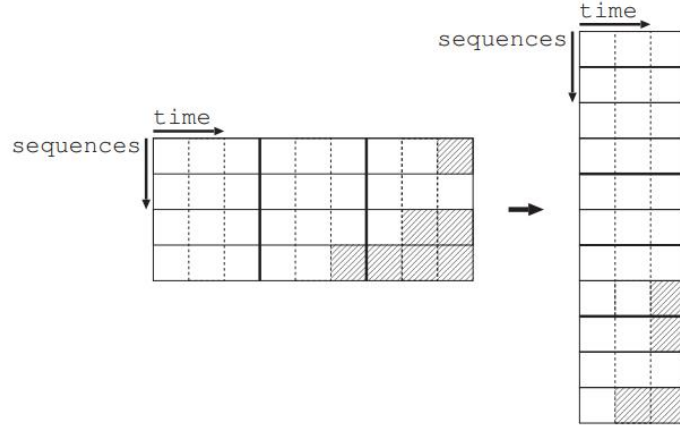


Figure7 .sequence chunk

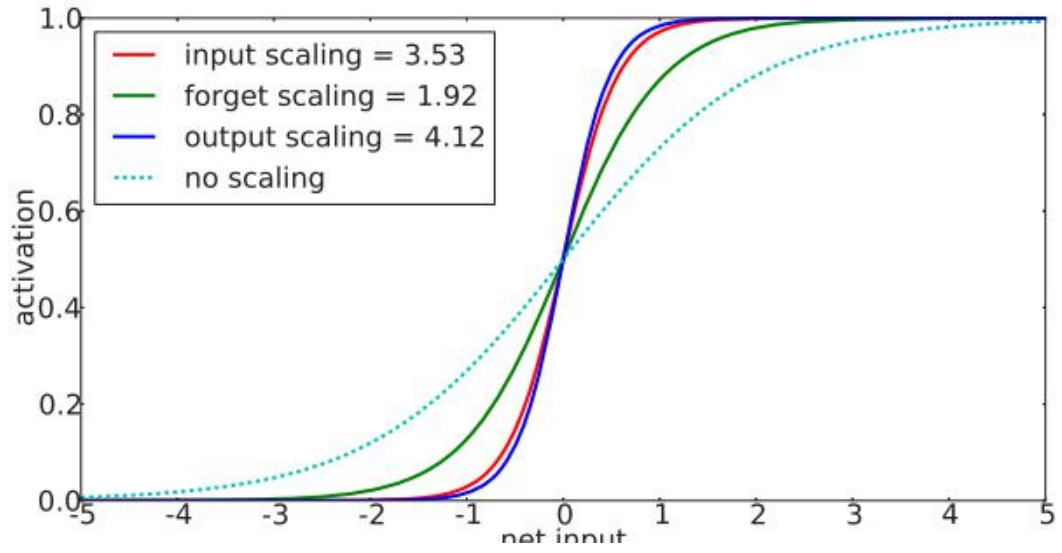


Figure 8.proposed gate scaling method

TABLE IV. COMPARISON OF THE PROPOSED SYSTEM TO RESULTS REPORTED BY OTHER GROUPS ON THE IAM DATABASE.

Systems	Voc.	WER [%]		CER [%]	
		Dev.	Eval	Dev.	Eval
RWTH	50k	8.4	12.2	2.5	4.7
Kozielski et al. [14]	50k	9.5	13.3	2.7	5.1
Boquera et al. [22]	50k	19.0	22.4	-	9.8
Dreuw et al. [23]	50k	22.7	32.9	7.7	12.4
Bertolami et al. [24]	20k	26.8	32.8	-	-

Figure9 .results on the IAM database

## 4. Dropout improves Recurrent Neural Networks for Handwriting Recognition

Vu Pham , Theodore Bluche , Christopher Kermorvant, and Jerome Louradour

这个论文的内容就是如标题的意思了，就是说明 Dropout 在 RNN（LSTM）中的作用。文章为了避免失去 RNN 保持信息的特性，在进行 dropout 的时候只应用到 feedforward 连接上，而没有用到递归连接上（如 Figure10 所示），此外，还 drop 了卷积层输入。并且保持了用 tanh 函数作为 LSTM 单元的激活函数和 sigmoid 作为门的激活函数。整个网络如 Figure11 所示。

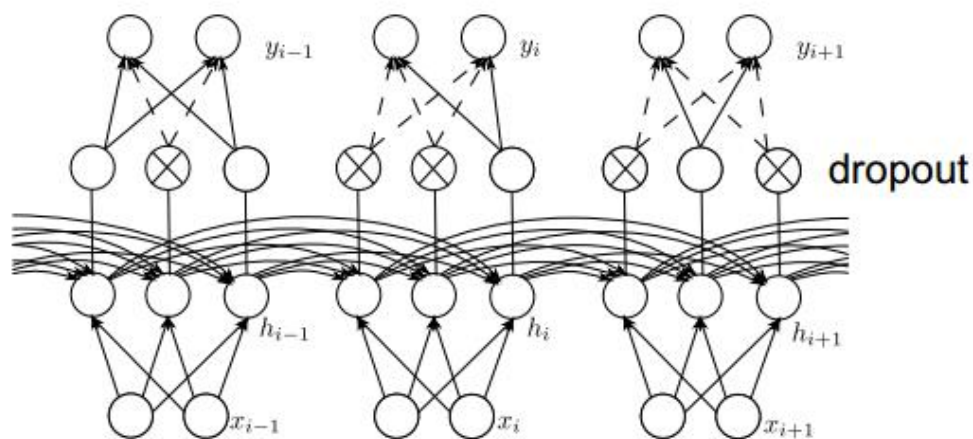


Figure10 .Dropout 示意图

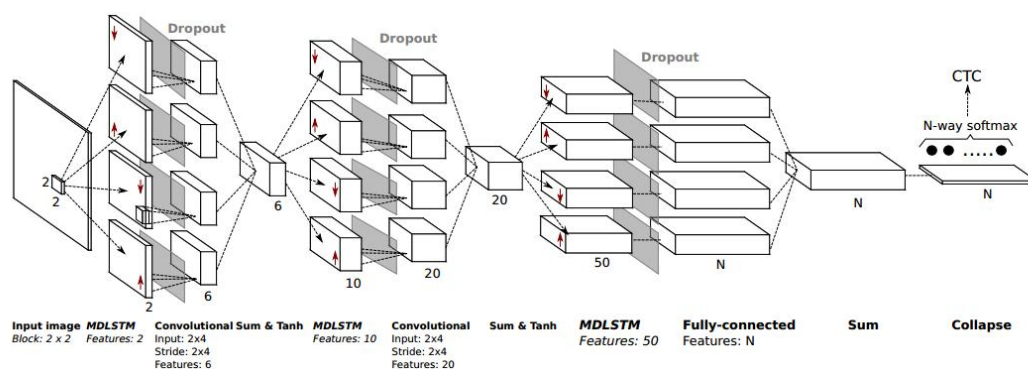


Figure11 . 基本网络结构

通过加入 Dropout，可以看出在 IAM 数据集上的准确率得到一定的提升，如 Figure 12 所示。

TABLE V  
RESULTS ON IAM

	Valid.		Eval.	
	WER	CER	WER	CER
MDLSTM-RNN	36.5	10.4	43.9	14.4
+ dropout	27.3	7.4	35.1	10.8
+ Vocab&LM	12.1	4.2	15.9	6.3
+ dropout	11.2	3.7	13.6	<b>5.1</b>
Kozielski et al. [29]	<b>9.5</b>	<b>2.7</b>	<b>13.3</b>	<b>5.1</b>
Kozielski et al. [29]	11.9	3.2	-	-
Espana et al. [31]	19.0	-	22.4	9.8
Graves et al. [32]	-	-	25.9	18.2
Bertolami et al. [33]	26.8	-	32.8	-
Dreuw et al. [34]	22.7	7.7	32.9	12.4

Figure12 .results on the IAM database