

Contents

Abstract	v
Resumen	vii
Resum	ix
General Index	xi
List of Figures	xvi
List of Tables	xxi
1 Introduction	1
1.1 Motivation	1
1.2 Tasks and tools	3
1.3 End-to-end: The Handwriting Text Recognition pipeline	5
1.4 Objectives	6
1.5 Document structure	8
2 Related work	13
2.1 Document Image Binarization	14
2.1.1 Taxonomy	15

2.1.2 Performance and efficiency	21
2.1.3 Evaluation	21
2.2 Text Line Extraction	22
2.2.1 Bottom-up strategies and text aggregation	23
2.2.2 Projection profiles	24
2.2.3 Blurring and smearing	25
2.2.4 Line seams	26
2.2.5 Text Reference Lines and baseline detection	26
2.2.6 Touching components	27
2.2.7 Text Line as sequences	29
2.2.8 Historical Documents	30
2.2.9 Machine Learning	31
2.3 Text Line Preprocessing	32
2.3.1 Tracking Text Reference Lines	33
2.3.2 Skew and Slant correction	33
2.3.3 Size Normalization	35
2.4 Handwriting Text Recognition	36
2.4.1 Optical Modeling	37
2.4.2 Feature extraction	40
2.4.3 Trends for HWR	40
2.5 Summary	41
3 Neural Networks	43
3.1 ANN basics	44
3.2 Training the networks	46
3.2.1 Stopping criteria	48
3.3 Architectures	49
3.4 Regularization	53
3.5 Deep learning	56
3.6 Hyper-parameters tuning	57
3.7 One-to-one labeling	58
3.7.1 Sliding window CNN and MLP	59
3.7.2 MultiDirectional Long Short Term Memories	61
3.8 Summary	62

4 Image Cleaning and Enhancement	63
4.1 Supervised Methods	65
4.2 Cleaning and enhancement as a probability pixel estimation problem.	67
4.3 Ground Truth Generation.	68
4.3.1 Denoising supervision.	68
4.3.2 Noising methods	70
4.4 Measures and evaluation	70
4.4.1 Subjective evaluation	71
4.4.2 Objective evaluation.	71
4.5 Connectionist methods	74
4.5.1 Multilayer Perceptron	75
4.5.2 Multilayer Perceptron with additional features	76
4.5.3 Convolutional Neural Networks	76
4.5.4 MultiDirectional Long Short Term Memories	78
4.5.5 Analytic cost	79
4.6 Ensembles.	82
4.7 Experiments and results	84
4.7.1 Insights of CNN for Document Image Binarization.	84
4.7.2 Comparison of connectionist methods	88
4.7.3 Results.	90
4.8 Discussion.	95
4.9 Summary	100
5 Text Line Extraction	101
5.1 Ground-truth formats	104
5.2 Evaluation and metrics	104
5.2.1 TLE metrics discussion	112
5.3 Text Reference Lines.	114
5.4 Text reference line estimation by Local Extrema Points	117
5.4.1 Local Extrema Points	118
5.4.2 Interest Points	119
5.5 Text Segments from Local Extrema Interest Points	122
5.5.1 Merge of Text Segments	126

5.5.2 Text Segments split (Touching Components)	128
5.6 Text Segment Aggregation Problem.	133
5.6.1 Combination and Optimization Problem	133
5.6.2 Two by two segments joiner	137
5.7 Text Line Extraction from MBA estimation	145
5.7.1 Text Block Extraction	146
5.7.2 Main Body Area pixel classification	147
5.7.3 Text Line Segmentation by Watershed Transformation	148
5.7.4 Post-processing.	149
5.8 Experimental setup and Results	151
5.9 Layout Analysis	153
5.10 Discussion	156
5.11 Summary.	157
6 Text Line Normalization	159
6.1 Skew, slant and slope correction	161
6.2 Text Line Height Normalization	162
6.2.1 Column resizing model	163
6.3 HMM column model	164
6.3.1 Statistical framework	164
6.3.2 Column HMM/ANN Modeling	165
6.3.3 Adding the rest of Text Reference Lines	166
6.4 A combined Convolutional Neural Network and Dynamic Programming approach	167
6.4.1 Pixel probability map	169
6.4.2 MBA edge detection	170
6.4.3 Reference lines extraction by Dynamic Programming	170
6.5 Experimental Setup	171
6.5.1 HMM/ANN column model setup.	171
6.5.2 Combined CNN and Dynamic Programming.	172
6.6 Results	173
6.7 Discussion.	174
6.8 Summary	175

7	Decoding from Scratch	177
7.1	Previous feature extraction and receptive field	178
7.2	Deep MLPs for extracting features	181
7.3	CNNs for preprocessing	181
7.3.1	Using known architectures	182
7.3.2	Adhoc networks dealing with the singularity of HWR	182
7.3.3	Cell feature extraction by convolutions.	183
7.3.4	1D convolutions	185
7.4	Experimental setup.	187
7.4.1	Using raw input and deep MLPs.	187
7.4.2	Using CNNs for preprocessing	188
7.5	Results	188
7.5.1	Recognition results	188
7.5.2	Kernel and map visualization.	191
7.5.3	Comparison with other approaches	195
7.6	Summary	195
8	F-Measure as Neural Network optimization function	199
8.1	Error-backpropagation with F-Measure	200
8.2	Experimental Setup and Results	202
8.3	conclusions	204
9	Conclusions and Future Work	207
9.1	What has been accomplished	207
9.2	Half way	210
9.3	What should have been done	212
9.4	Derived publications	213
9.5	Future work	215
	Appendices	217
A	Appendix Description of the APRIL-ANN toolkit	219

B Appendix Hybrid HMM/ANN Modeling	221
B.1 Handwriting Text Recognition (HWR)	221
B.2 Optical models by Hybrid HMM/ANN	222
B.3 Language Modeling	223
B.4 HMM/ANN engine description (Baseline).	225
B.4.1 N-gram model	225
C Appendix Document Image Corpora	227
C.1 Document Image Binarization	227
C.1.1 DIBCO	227
C.1.2 Saint Gall	228
C.1.3 Noisy Office.	230
C.1.4 Corpora Information.	231
C.2 IAM Historical Document Database (IAM-HistDB)	231
C.2.1 Saint Gall	233
C.2.2 Parzival	233
C.3 IAM Offline Database	234
C.3.1 Interest Points, Text Reference Lines and MBA ground-truth	234
D Appendix Text Line ground truth formats	235
E Appendix Text Reference Lines (soft) ground truth generation	251
E.1 Interest Point (IP) supervision	251
E.2 Main Body Area supervision.	257
E.3 Case of study. Bootstrap	258
References	263