Enhancing Handwritten Text Recognition with N-gram Sequence Decomposition and Multitask Learning

Vasiliki Tassopoulou, Giorgos Retsinas, Petros Maragkos

National Technical University of Athens

Problem Definition

- **Problem:** Line Level Handwritten Text Recognition
- Sequence of Visual Features → Sequence of Target Units (Characters, Words, Subwords)

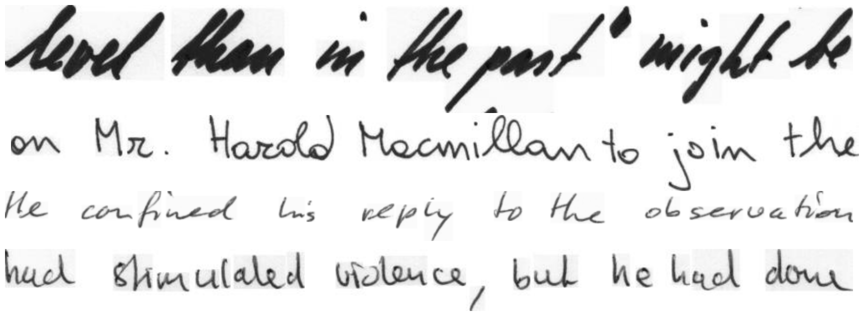


Figure 1: IAM Dataset Line Samples

Motivation

- HTR is handled so far as a Single Task
- A Model is optimized to learn alignments of target units of one level. For example character level unigrams.
- Language Knowledge is integrated explicitly in the decoding step

In this work we want :

- Integrate Domain Knowledge during training
- Leverage the advantages of Multitask Learning

Baseline Model

Architecture

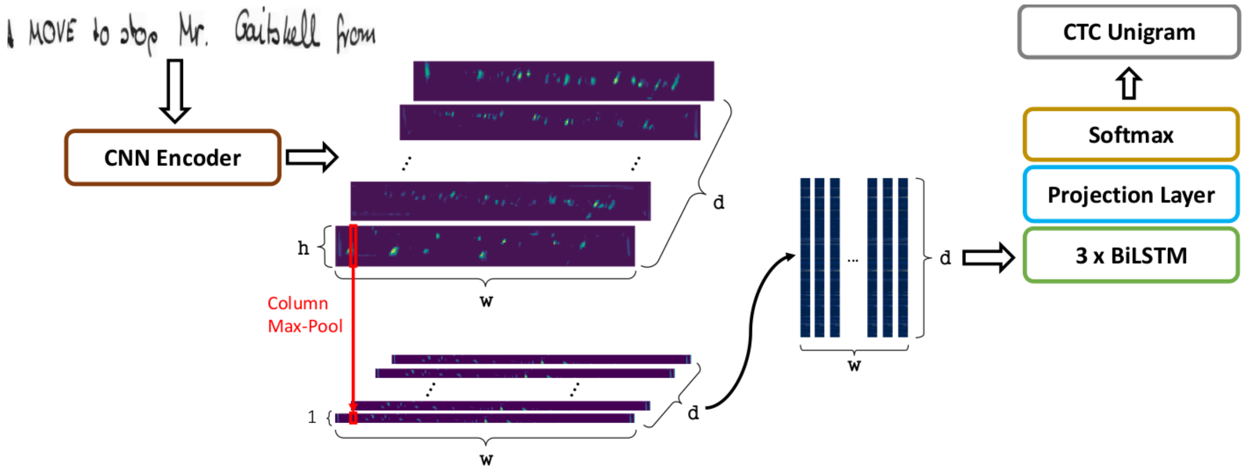
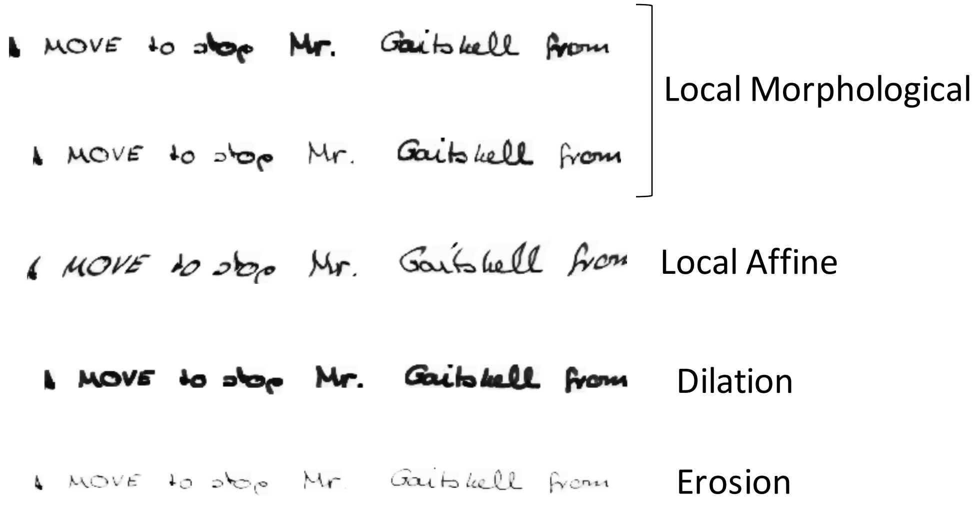


Figure 2: The backbone baseline architecture for Single Task Unigram Level HTR HTR

Dynamic Data Augmentation



Connectionist Temporal Classification – Objective Function

X, Y pair of image and transcript
A : set of aligments such that Y = B(A)
B : mapping between an alignment **a** and transcript Y.

$$P(Y|X) = \sum_{a \in A} P(a|X)$$

Alignment Example

Y : hello
Aligments : { h---e-l-lll—oooo, hel-l—o, h-e----l---l-o}

Multitask Architectures

Models

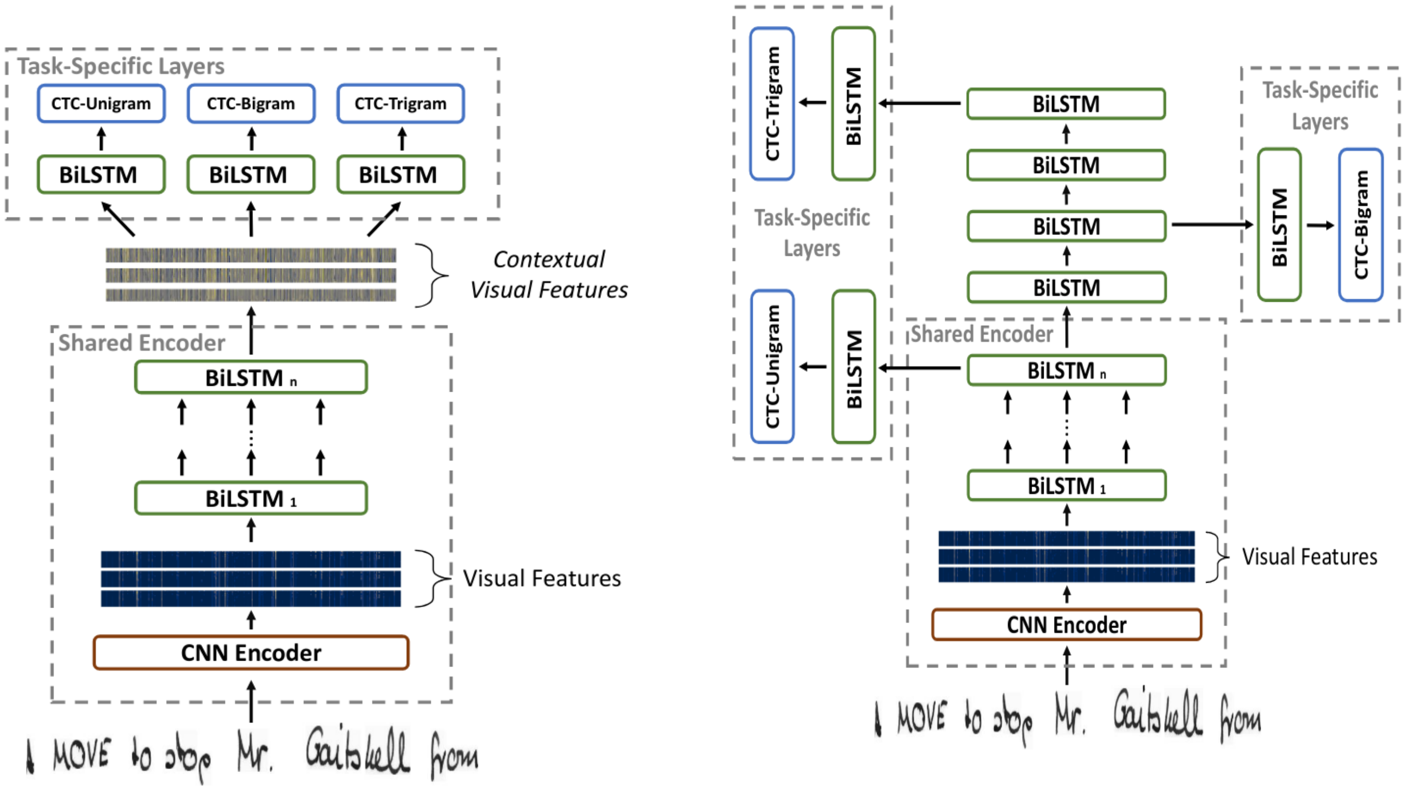


Figure 3: (a) Block Multitask Architecture (b) Hierarchical Multitask Architecture

Multitask CTC Loss

y_u : Unigram posterior
y_b : Bigram posterior
y_c : Trigram posterior

$$
\begin{aligned}
L(\{y_u, y_b, y_t\}, \text{"better"}) = {} & L_{CTC}^{unigrams}(y_u, b\text{-}e\text{-}t\text{-}t\text{-}e\text{-}r) \\
& + L_{CTC}^{bigrams}(y_b, be\text{-}et\text{-}te\text{-}er) \\
& + L_{CTC}^{trigrams}(y_t, bet\text{-}ett\text{-}tte\text{-}ter)
\end{aligned}
$$

Experiments

We experiment :

- Multitask Architectures
- Target Unit Selection : Fine-to-coarse granularities
    - Unigrams + Bigrams
    - Unigrams + Bigrams + Trigrams
    - Unigrams+Bigrams+Trigrams+Fourgrams

All Bigrams
Most frequent Trigrams
Most frequent Fourgrams

Evaluation and Results

N-Grams	WER %	CER %
Single-Task		
Pham *et al.* [1]	35.10	10.80
Puigcerver *et al.* [2]	20.20	6.20
Castro *et al.* [3]	24.00	6.64
Michael *et al.* [4]	-	5.24
1-gram (ours)	19.10	5.60
Hierarchical MT		
1-grams + 2-grams	17.72	5.21
1-grams + 2-grams + 3-grams	17.70	5.37
1-grams + 2-grams + 3-grams + 4-grams	**17.68**	5.29
Block MT		
1-grams + 2-grams	17.96	5.28
1-grams + 2-grams + 3-grams	17.90	5.30
1-grams + 2-grams + 3-grams + 4-grams	**17.68**	**5.18**

Architecture	WER %	CER %
CTC Greedy Decoding		
Single-Task	19.10	5.60
BMT	17.68	5.18
CTC BeamSearch 4-Gram CharLM		
Single-Task	18.14	5.64
BMT	16.72	5.28
CTC BeamSearch 4-Gram WordLM		
Single-Task	14.81	4.60
BMT	**13.62**	**4.60**

- Block Multitask and Hierarchical Multitask have close performance
- Comparing Block Multitask with Single Task model we observe the improvement in both WER and CER in the greedy decoding, utilizing in both models the unigram posteriors

Conclusion : Block Multitask Models have learned more robust hidden representations of the line images than the Single Task Model and thus leads to better WER/CER results.

References

[1] Vu Pham and Christopher Kermorvant and J´erˆome Louradour "Dropout improves Recurrent Neural Networks for Handwriting Recognition," inInternational Conference on Frontiers in Handwriting Recognition, 2014

[2] Joan Puigcerver, "Are Multidimensional Recurrent Layers Really Neces-sary for Handwritten Text Recognition?," inInternational Conference onDocument Analysis and Recognition, 2017

[3] Dayvid Castro and Byron L. D. Bezerra and Meuser Valenca "Boostingthe Deep Multidimensional Long-Short-Term Memory Network for Hand-written Recognition Systems," inInternational Conference on Frontiersin Handwriting Recognition, 2018

[4] Johannes Michael and Roger Labahn and Tobias Gr¨uning and JochenZ¨ollner "Evaluating Sequence-to-Sequence Models for Handwritten TextRecognition," inCoRR, abs/1903.07377, 2019

[5] Patrick Doetsch and Michał Kozielski and Hermann Ney"Fast andRobust Training of Recurrent Neural Networks for Offline HandwritingRecognition," inInternational Conference on Frontiers in HandwritingRecognition, 2014

[6] Paul Voigtlaender and Patrick Doetsch and Hermann Ney "HandwritingRecognition with Large Multidimensional Long Short-Term MemoryRecurrent Neural Networks," inInternational Conference on Frontiers inHandwriting Recognition, 2016

[7] Harald Scheidl and Stefan Fiel and Robert Sablatnig"Word BeamSearch: A Connectionist Temporal Classification Decoding Algorithm," inInternational Conference on Frontiers in Handwriting Recognition, 2018

[8] Urs-Viktor Marti and Horst Bunke"The IAM-database: an Englishsentence database for offline handwriting recognition," inInternationalJournal on Document Analysis and Recognition, 39-46, 2002

Thank you!

Vasiliki Tassopoulou, MEng
tassopoulouvasiliki@gmail.com
@vtassop