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

Vasiliki Tassopoulou, Giorgos Retsinas, Petros Maragos
tassopoulouvasiliki@gmail.com, gretsinas@central.ntua.gr, maragos@cs.ntua.gr

Motivation

Problem : Line Level Handwritten Text Recognition

Sequence of Visual Features → Sequence of Characters drawn from an alphabet

So far HTR has been handled as : Single Task

In this paper we handle the HTR as : Multitask

We formulate the Multitask Scheme as a way to **integrate language domain knowledge**. In HTR, the domain knowledge was integraded so far in the postprocessing step, the decoding via Statistical Language Models, either character level or word level. In our case, we integrate implicitly into the training procedure the n-gram character level information, since we "force" the model to learn except from unigrams, n-grams. As a result we obtain the visual n-gram probability.

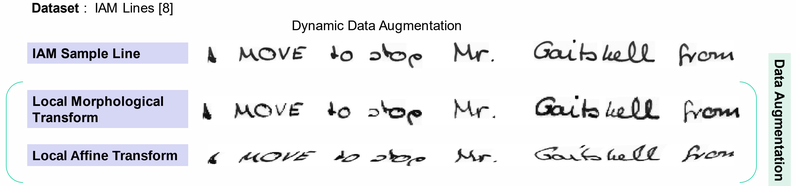
Goals :

- Enable language domain knowledge integration via Multitask Scheme
- Explore fine-to-coarse n-gram granulatities
- Explore Multitask Architectures
- Perform unigram level decoding in the inference. No computational burden in the decoding process

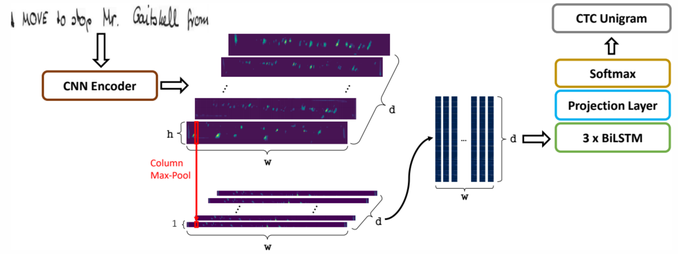
Baseline and Data Augmentation

Dataset : IAM Lines [8]

Dynamic Data Augmentation



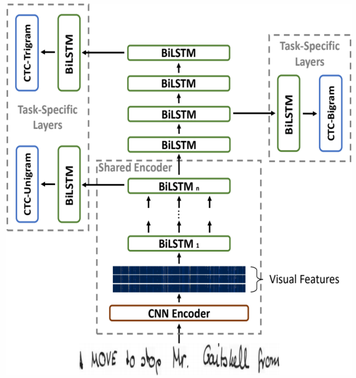
Baseline Architecture



Model Architectures

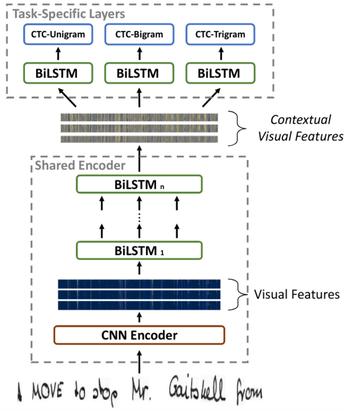
Hiearchical Multitask Architecture

Each n-gram builds its task specific layer upon previous n-gram layers



Block Multitask Architecture

All n-gram task-specific layers share the same encoder



Target Units Selection

Target Units	Word Decomposition
Unigram	b-e-t-t-e-r
All Bigrams	be-et-tt-te-er
Partial Bigrams	be-tt-er
All Trigrams	bet-ett-tte-ter
Partial Trigrams	bet-tte-ter

Example of a single alignmemt for word "better". In the case where a subset of all possible trigrams is selected as target units, the missing trigram in every word is substitued with the blank character "-" . The same applies to bigrams.

Training

CTC Objective :
Y : groundtruth text
A: set of all possible alignments of Y
P(Y|X) = Σ P(a| X)

Intuition : Sum up the probability to have the all possible alignments.
Consider them all when Maximimizing

Multitask CTC Loss : Composed of unigram, bigram and Trigram CTC losses.

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

Inference

B : Mapping that removes the repeating concecutive characters and Then the "blank" characters. Converts an alignment to sentence.

Greedy

$$y_{dec} = B(\arg\max_x \prod_{t=1}^{T} P(x_t|\boldsymbol{X}))$$

CTC Decoding + Character LM

$$_{dec} = B(\arg\max_x \sum(\prod_{t=1}^{T} P(x_t|\boldsymbol{X}) \cdot P_{LM}(x_t|y_{dec}(t-1))))$$

Evaluation and Conclusion

Evaluation Metrics : Word Error Rate, Character Error Rate

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**

- In all the above experiments we utilize only the unigram posteriors in the inference so as to keep the computational cost of decoding as low as possible
- There is no substantial difference in recognition performance between the Hierarchical and Block Architecture. Thus we focus on the BLock Mutitask architecture with Unigram and Bigram CTC levels
- Comparing our Single-task architecture with the Block Multiatsk we observe the improvement in both WER and CER in the greedy decoding where no explicit language knowledge was utilized. This result indicates that the using unigrams and bigrams (character level) in a multitask-scheme improves the internal learned representations and leads to better performance metrics

Aknowledgement

The work of Prof. Petros Maragos was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the "First Call for H.F.R.I. Research Projects to support Faculty members and Researchers and the procurement of high-cost research equipment grant" (Project: "SL-ReDu", Project Number: 2456).

References

[1] Vu Pham and Christopher Kermorvant and J´er´ome 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\