THE 16TH INTERNATIONAL CONFERENCE ON FRONTIERS IN HANDWRITING RECOGNITION

CHARACTER AND TEXT RECOGNITION OF KHMER HISTORICAL PALM LEAF MANUSCRIPTS

Dona Valy, Michel Verleysen, Sophea Chhun, and Jean-Christophe Burie

ICHFR2018

August 5-8, 2018

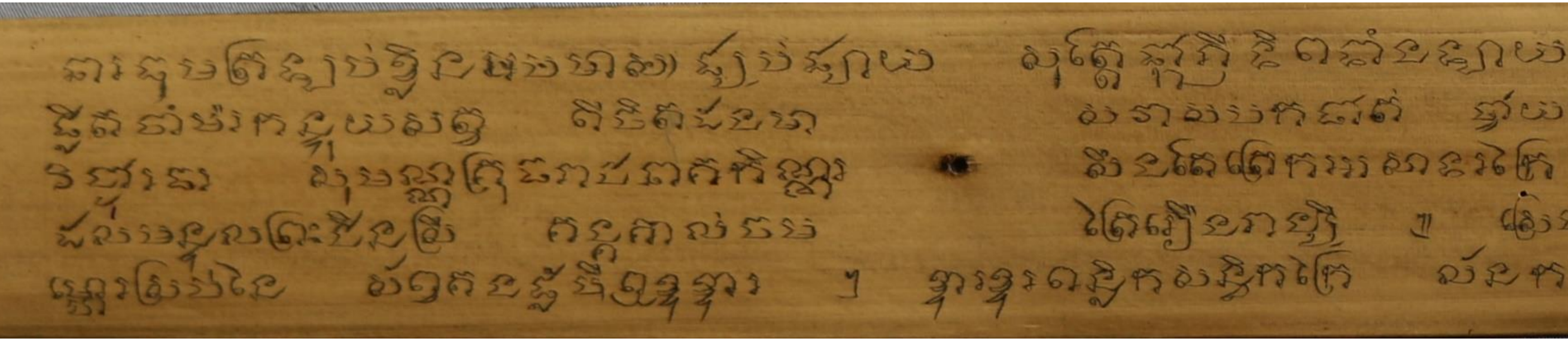
Overview

- Khmer Palm Leaf Manuscripts
- Task 1: Isolated Character Classification
- Task 2: Word/Text Recognition
- Conclusion

KHMER PALM LEAF MANUSCRIPTS

Introduction

- Palm Leaf Manuscripts or Sleuk Rith in Khmer
  - [**Sleuk:** leaf] + [**Rith:** to bind/tie together]

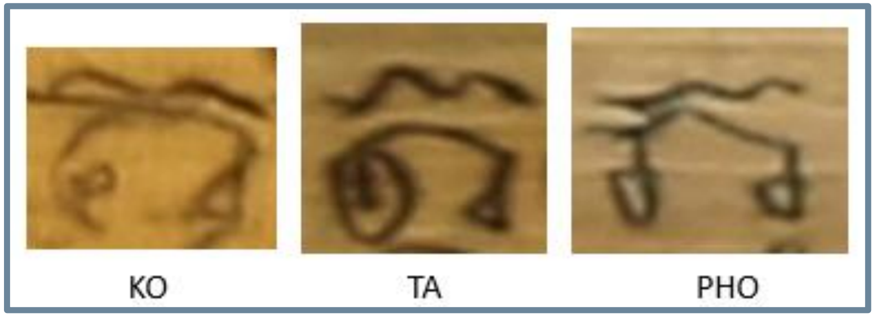


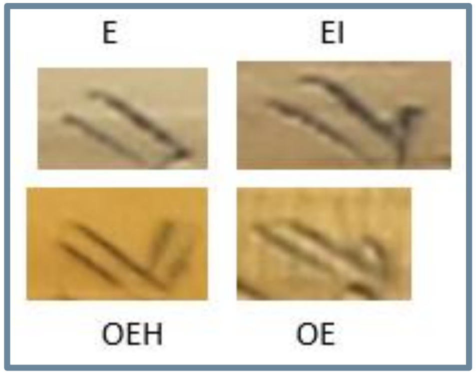
Challenges

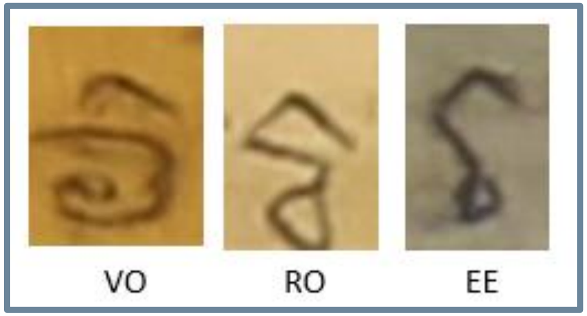
- Degradations and defects

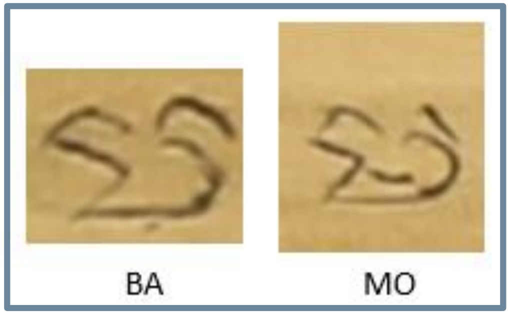
Challenges

- Ambiguity of certain characters
  - Khmer alphabet (more or less 70 symbols)
  - Similarity between characters









Challenges

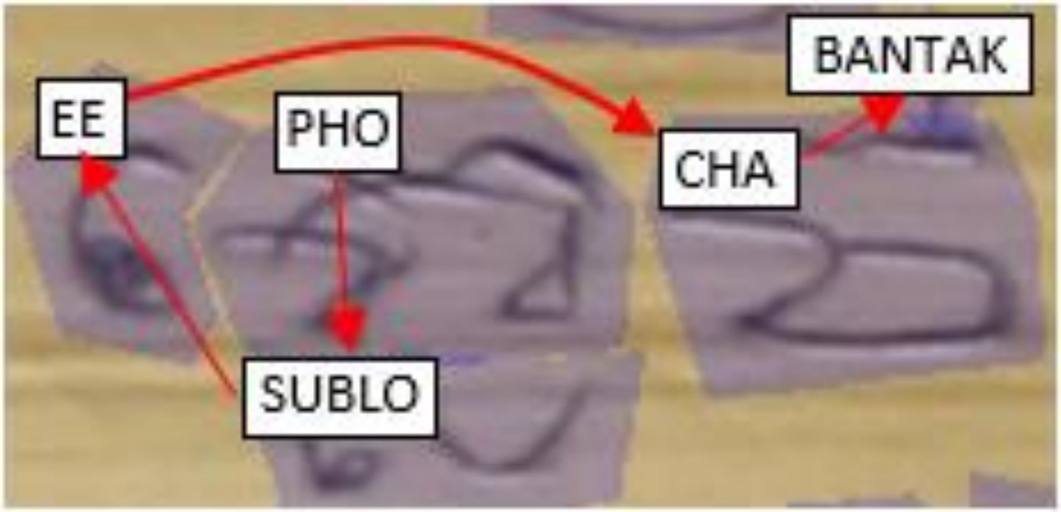
- Sequential order of characters composing a word
  - Khmer alphabet (more or less 70 symbols)
  - Irregularity of how characters are combined into words

SA-SUBDA-AEU-NGO

SleukRith Set

- A collection of annotated data created from 657 pages of digitized Khmer palm leaf manuscripts
- Composed of 3 types of annotated data:
  - Character/Glyph
  - Word
  - Line

Annotating a word



Available at https://github.com/donavaly/SleukRith-Set

SleukRith Set

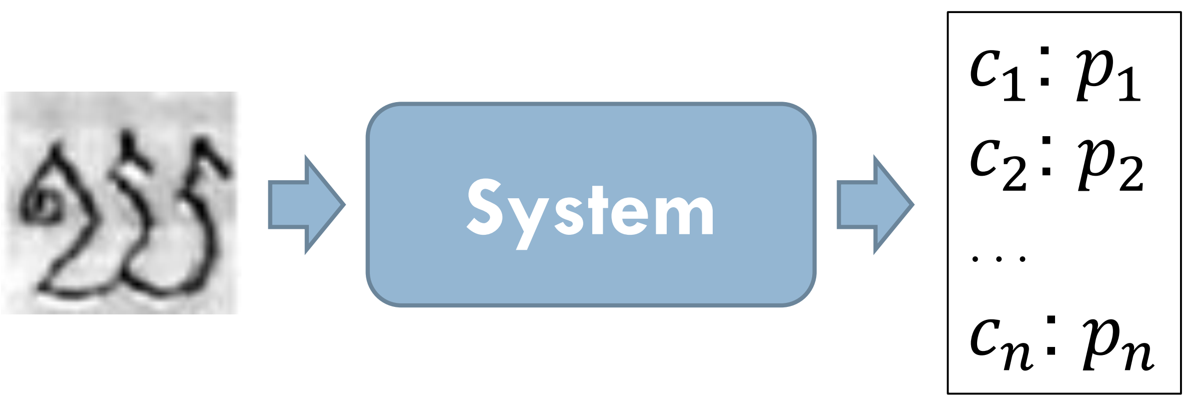
- Statistics of SleukRith Set

Data	Quantity
Annotated Characters/Glyphs	301,626
Annotated Words	73,359
Text Lines	3,245

- Character and word image patches

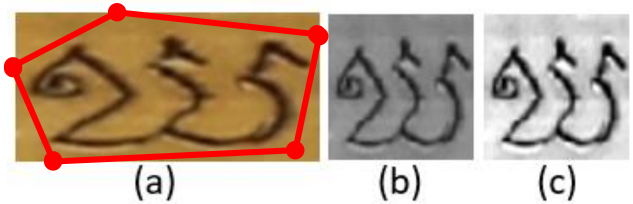
Available at https://github.com/donavaly/SleukRith-Set

TASK1: ISOLATED CHARACTER CLASSIFICATION



Isolated Character Dataset

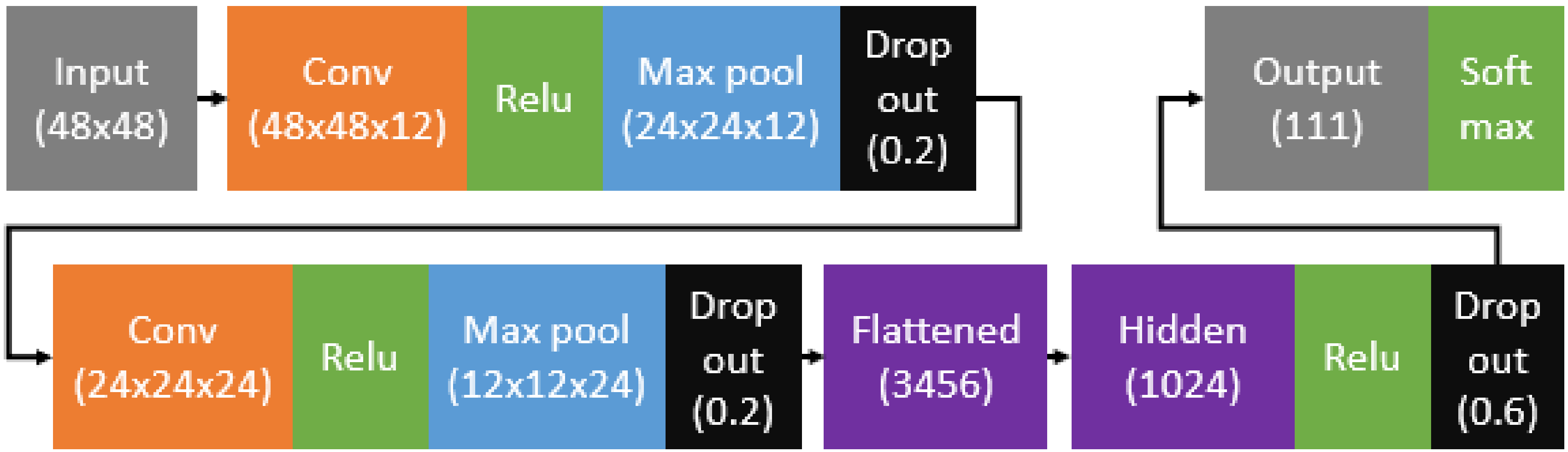
- Data normalization



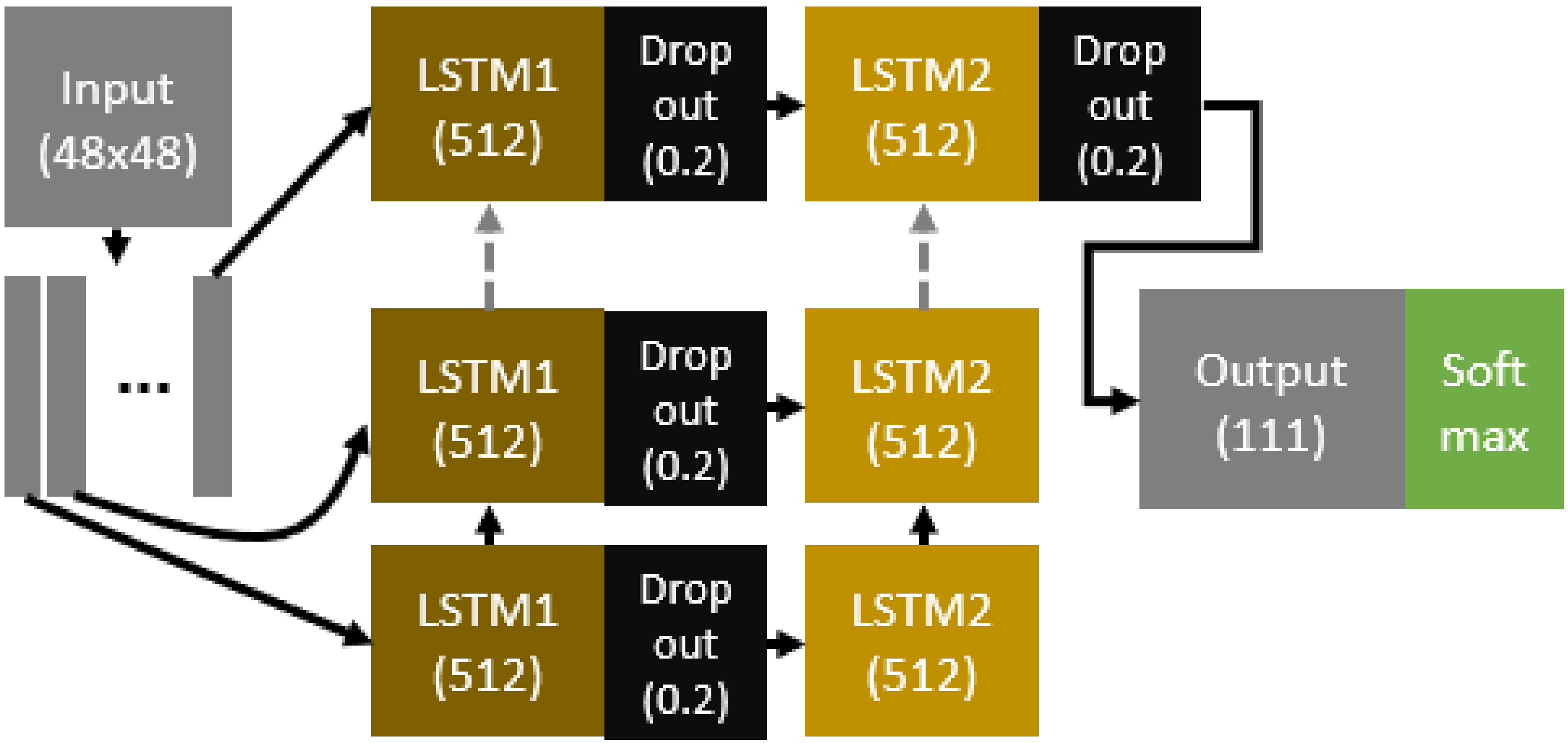
(a). Original image, (b). Gray scaled and resized to 48x48, (c). Normalized

- Dataset:
  - Train: ~113k
  - Test: ~91k
  - Number of classes: 111

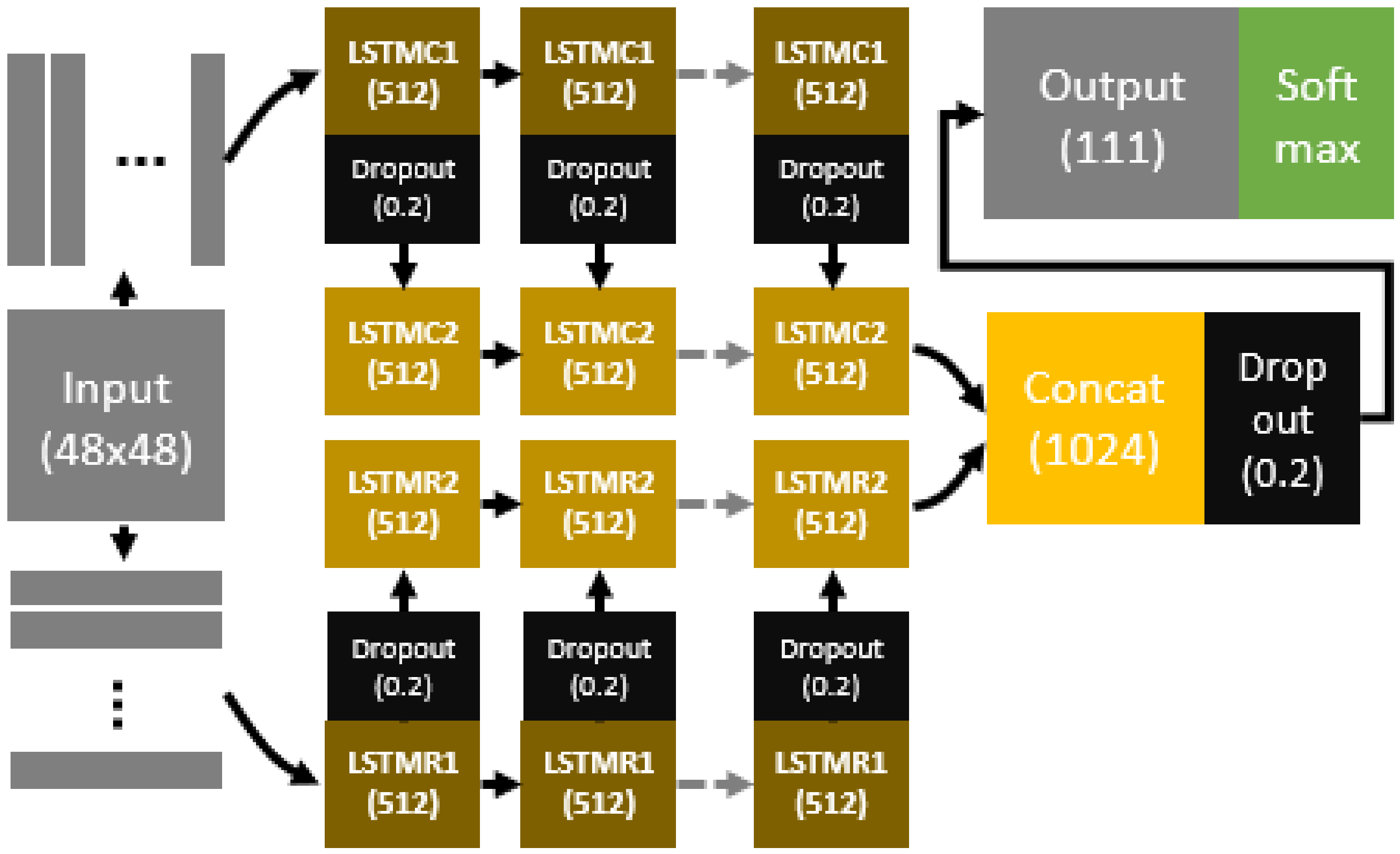
Network 1.1: CNN



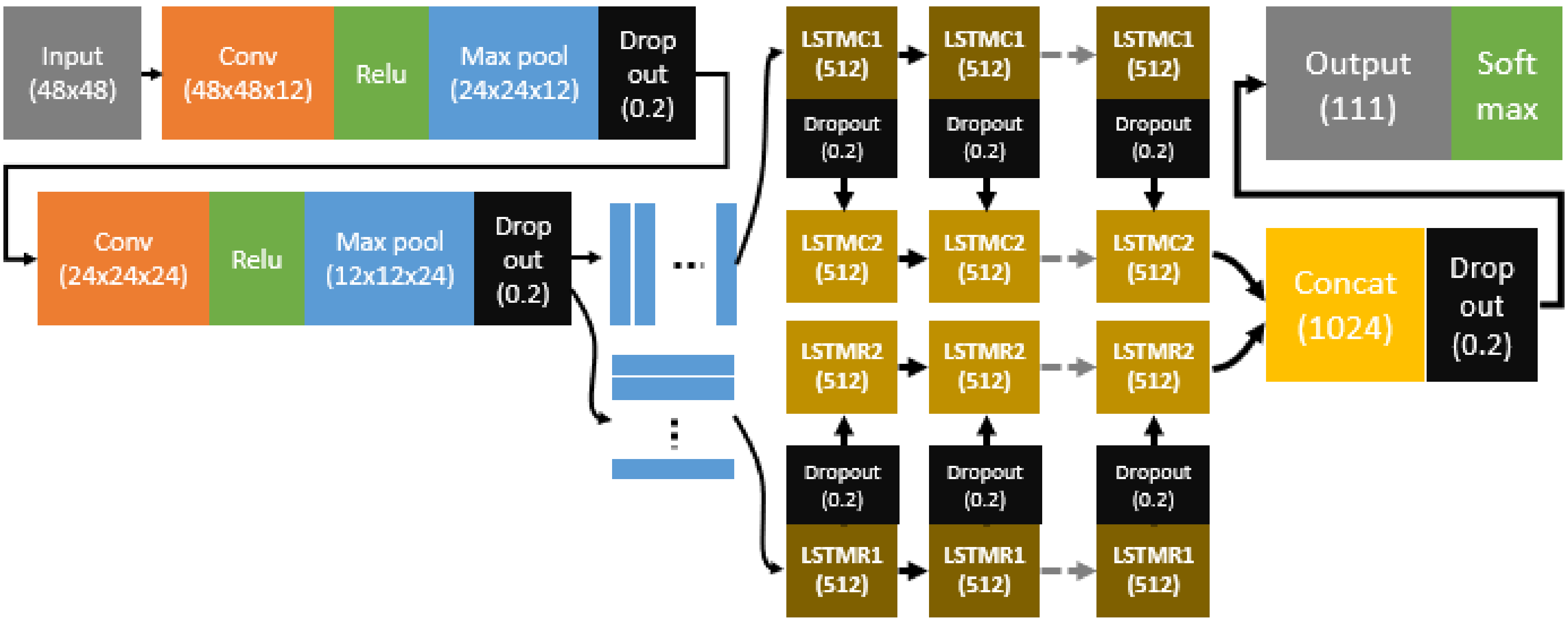
Network 1.2: Column LSTM



Network 1.3: Row-Column LSTM



Network 1.4: CNN-LSTM

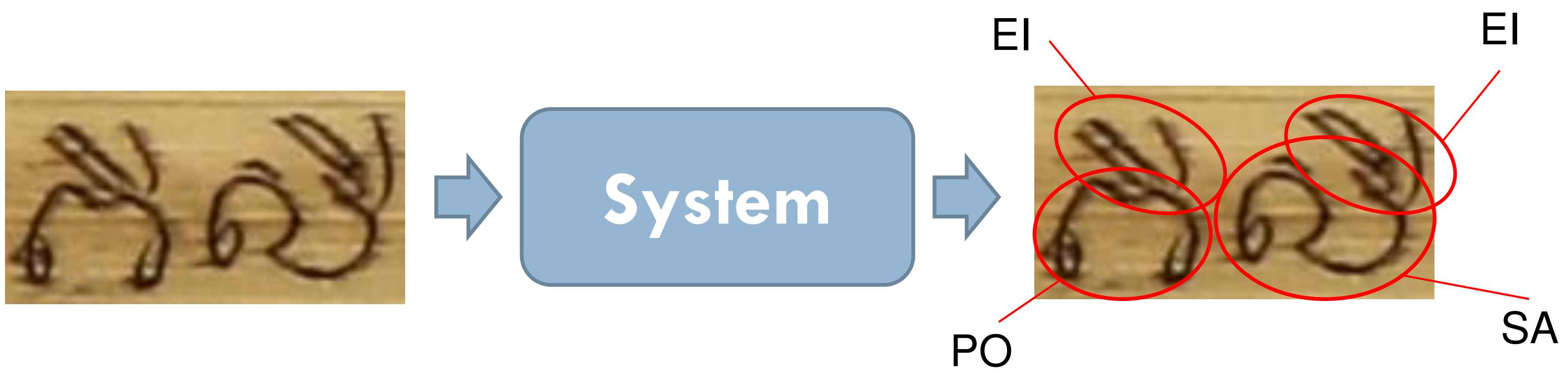


Experiments and Results

- Training configurations:
  - Batch size: 300
  - Samples are reshuffled after each epoch
  - Stop condition:
    - average loss does not improve after $N = 10$ consecutive tests
    - each test is done for every 50 iterations
- Results: top-k error rate

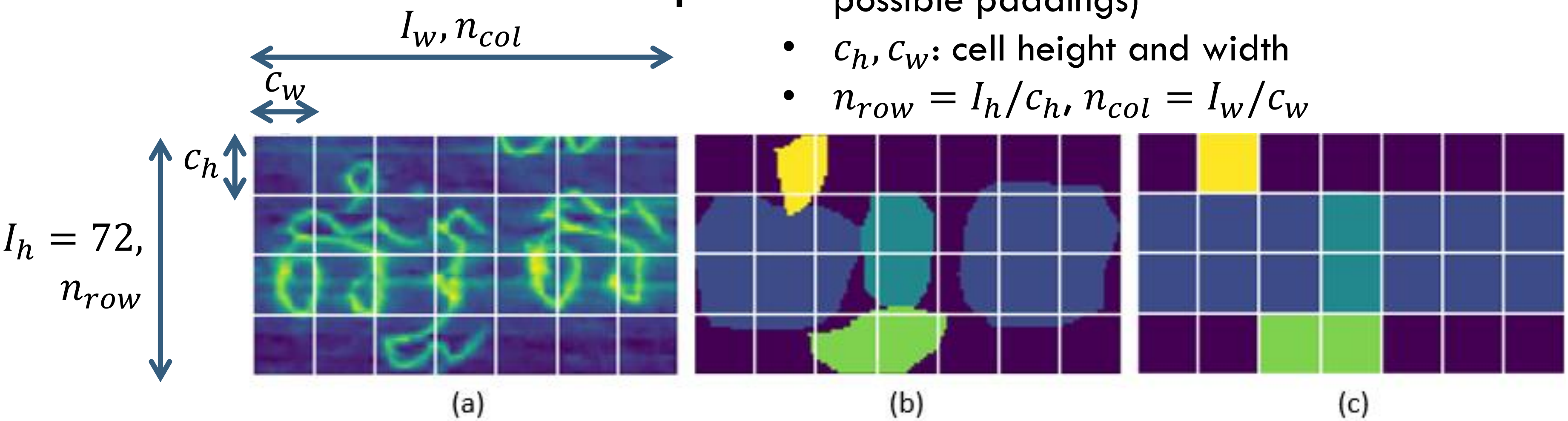
Architecture	*Error Rate (%)*	
	Top 5	*Top 1*
Network 1.1: CNN	0.65	6.29
Network 1.2: Column LSTM	1.05	8.49
Network 1.3: Row-Column LSTM	0.82	7.00
Network 1.4: Conv-LSTM	**0.46**	**5.01**

TASK2: WORD/TEXT RECOGNITION



Annotated Word Dataset

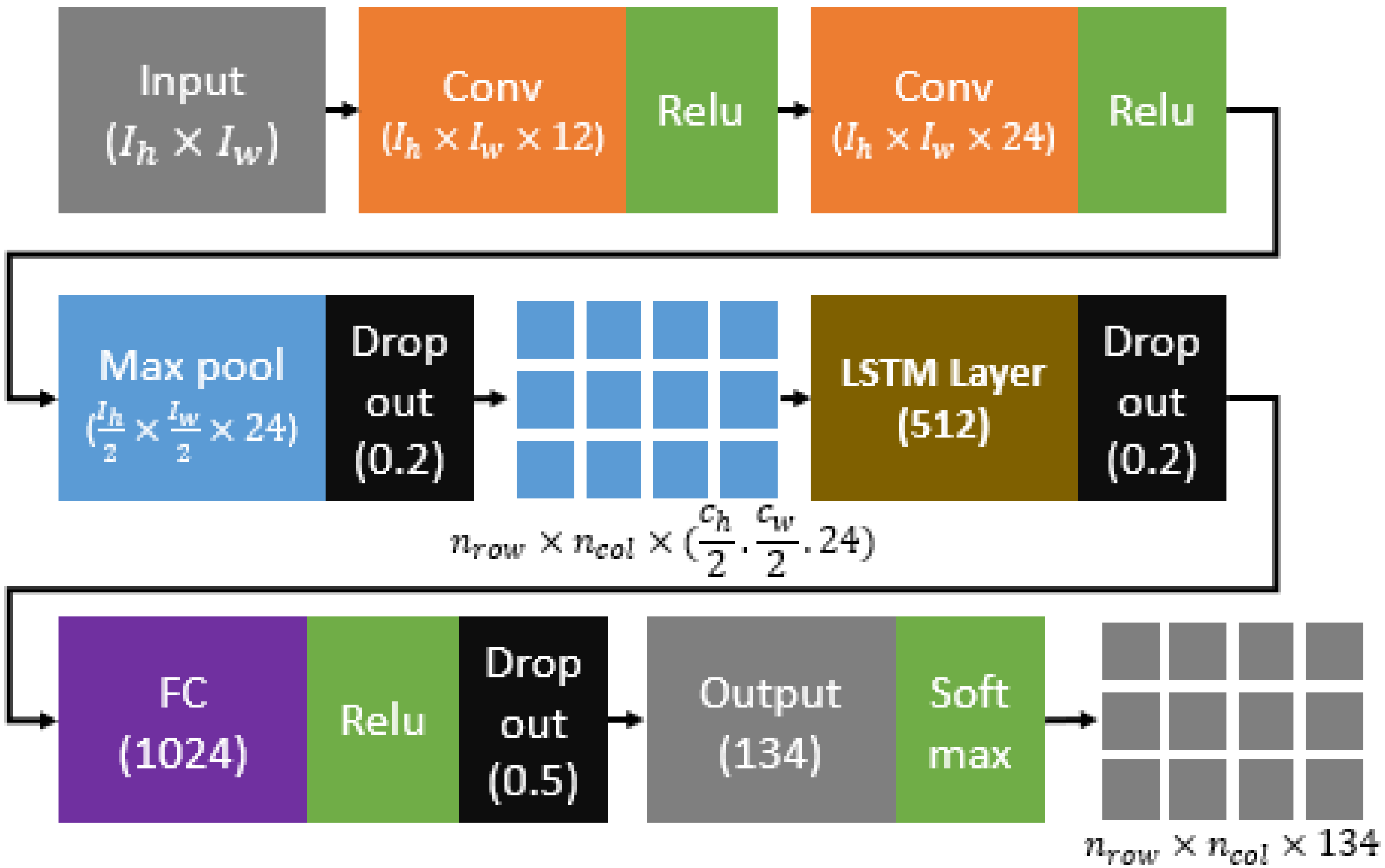
- Character-Class Map
  - I_h, I_w: height and width of the image (after possible paddings)
  - c_h, c_w: cell height and width
  - $n_{row} = I_h / c_h$, $n_{col} = I_w / c_w$



(a). Original word image patch, (b). Annotated character information in the word: polygon boundaries of all characters, (c). Character-class map

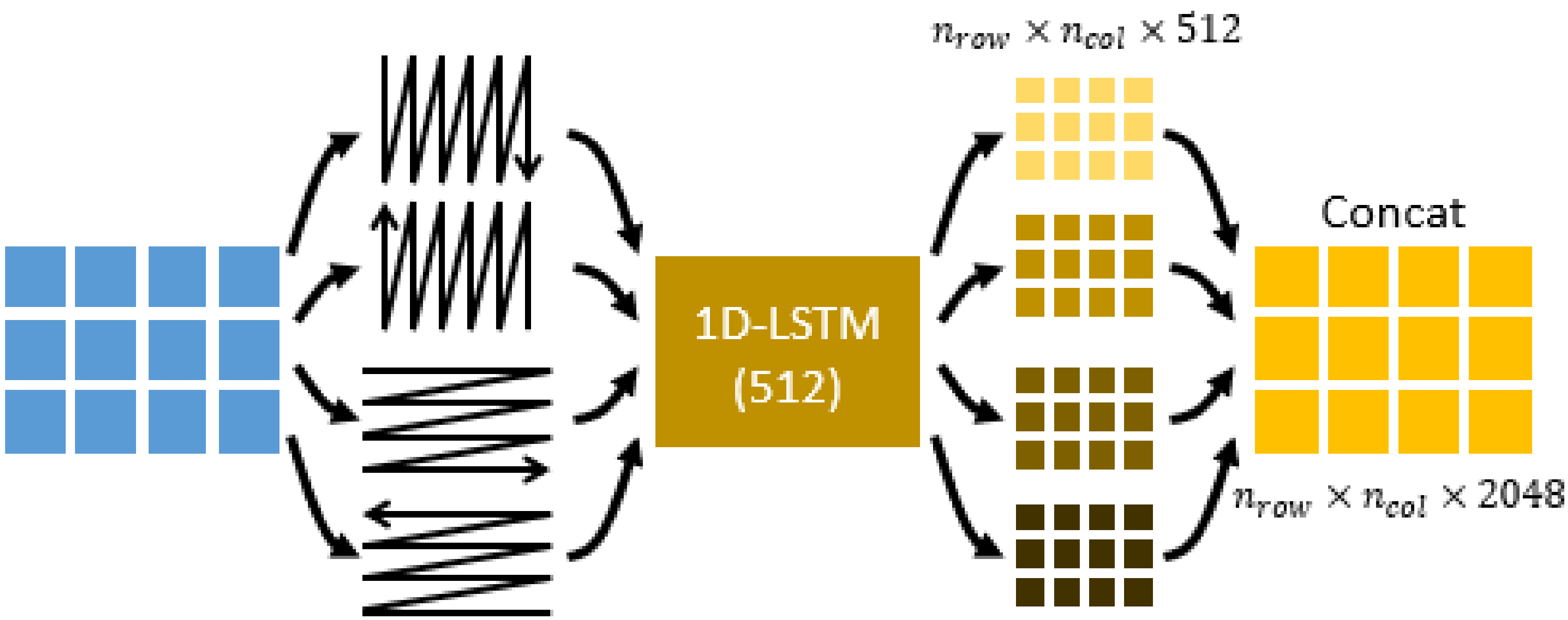
- Dataset:
  - Train: ~16k
  - Test: ~8k
  - Number of character-classes: 134 (including 1 token class for background or blank space)

General Architecture



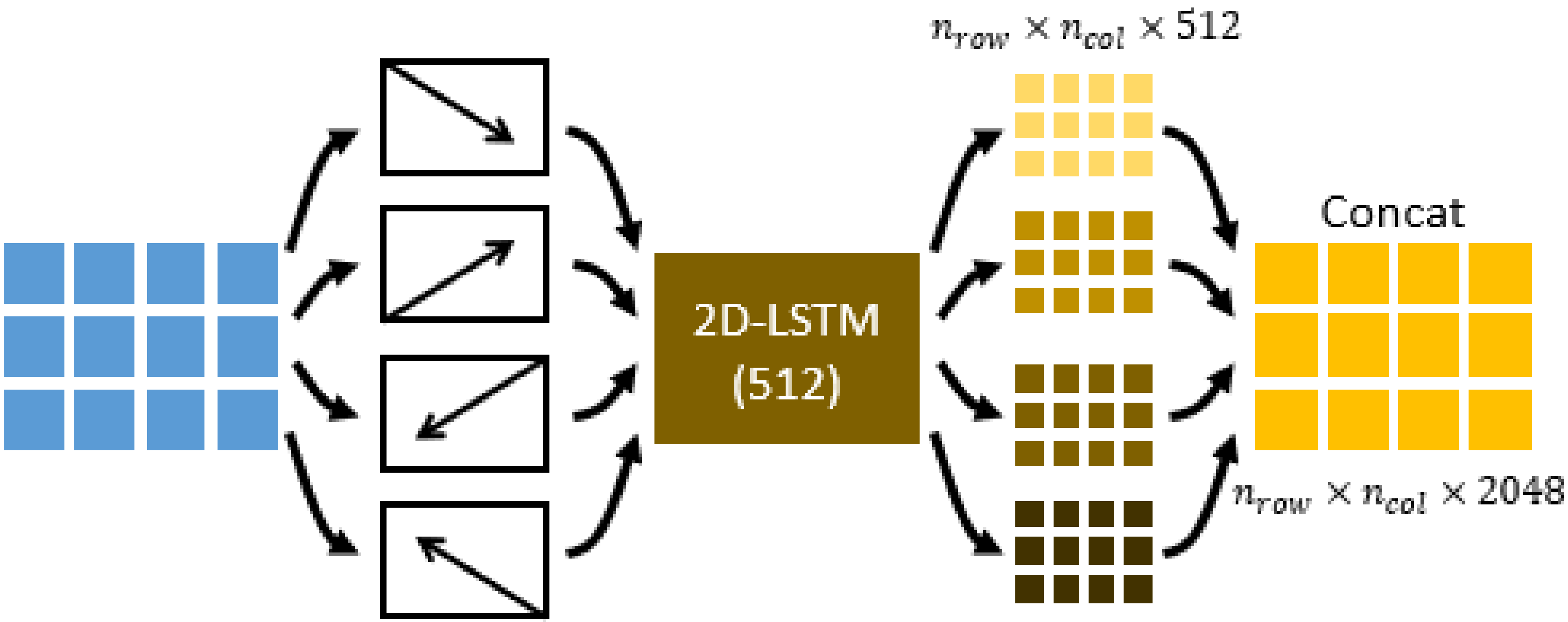
Network 2.1: 1D-LSTM

- LSTM Layer of Network 2.1



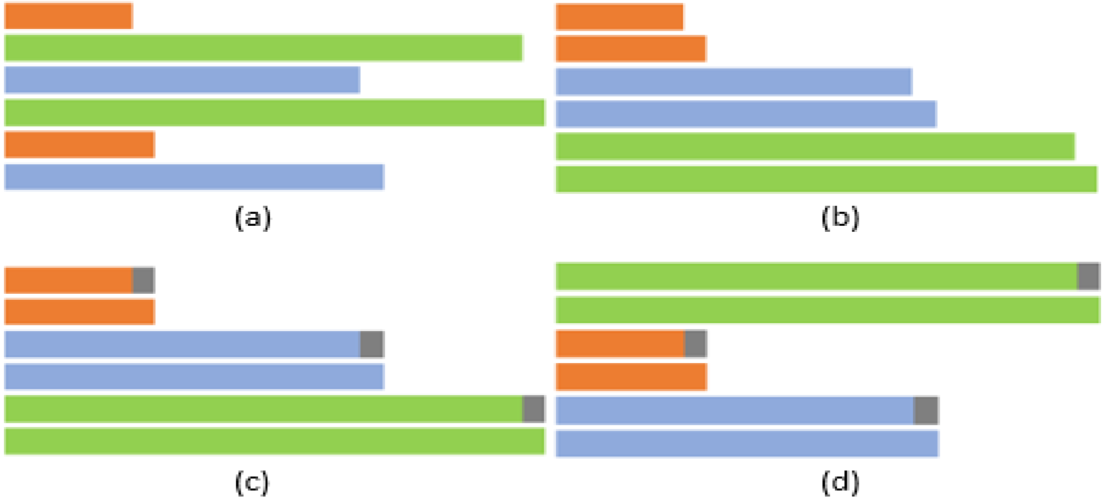
Network 2.2: 2D-LSTM

- LSTM Layer of Network 2.2



Experiments

- Training configurations:
  - Batch size: 30
  - Samples are sorted and batched according to their width



(a). Initial sample order
(b). Sort by the width of each sample
(c). Pad each sample to the maximum width in the batch
(d). Shuffle batch order

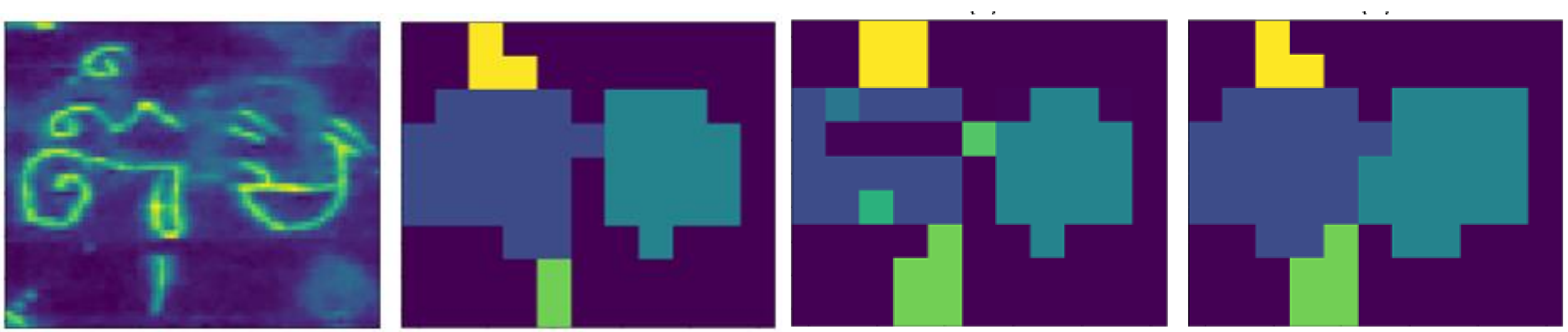
  - Stop condition:
    - average loss does not improve after $N = 30$ consecutive tests
    - each test is done for every 50 iterations

Results

- Measurement
  - Top-k error rate: average error rate of all cells in the predicted character-class map

Architecture	***Error Rate (%)***	
	Top 5	***Top 1***
Network 2.1: 1D-LSTM	8.46	32.01
Network 2.2: 2D-LSTM	**2.40**	**20.49**

(a). Original word image
(b). Ground truth character-class map
(c). Result predicted by Network 2.1
(d). Result predicted by Network 2.2



(a) (b) (c) (d)

CONCLUSION

Conclusion

- We present different approaches for two tasks on medium size datasets constructed from Khmer palm leaf manuscripts :
  - Isolated character classification
  - Word/text recognition
- The predicted character-class map from Task 2 can be used further to generate the final transcription of the word image
  - CTC and/or encoder-decoder mechanism

Thank you for your attention!

References

[1] W. Swaileh, J. Lerouge and T. Paquet, "A Unified French/English syllabic model for handwriting recognition," in 15th International Conference on Frontiers in Handwriting Recognition, 2016.

[2] T. M. Breuel, "High Performance Text Recognition using a Hybrid Convolutional-LSTM Implementation," in 14th IAPR International Conference on Document Analysis and Recognition, 2017.

[3] T. Bluche and R. Messina, "Faster Segmentation-Free Handwritten Chinese Text Recognition with Character," in 15th International Conference on Frontiers in Handwriting Recognition, 2016.

[4] X. Yang, D. He, Z. Zhou, D. Kifer and C. L. Giles, "Improving Offline Handwritten Chinese Character," in 14th IAPR International Conference on Document Analysis and Recognition, 2017.

[5] M. T. Pavez and S. A. Mahoud, "Offline Arabic handwritten text recognition: a survey," ACM Computing Surveys (CSUR), vol. 45, no. 2, p. 23, 2013.

[6] A. Khémiri, A. K. Echi, A. Belaïd and M. Elloumi, "A System for off-line Arabic Handwritten Word Recognition based on Bayesian," in 15th International Conference on Frontiers in Handwriting Recognition, 2016.

[7] A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks.," in Advances in neural information processing systems, 2012.

[8] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in arXiv preprint arXiv:1409.1556, 2014.

[9] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in IEEE conference on computer vision and pattern recognition, 2016.

References

[10] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in arXiv preprint arXiv:1406.1078, 2014.

[11] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9, no. 8, pp. 1735-1780, 1997.

[12] A. Graves and J. Schmidhuber, "Offline handwriting recognition with multidimensional recurrent neural networks," Advances in neural information processing systems, pp. 545-552, 2009.

[13] A. Graves, S. Fernández, F. Gomez and J. Schmidhuber, "Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks," in 23rd international conference on Machine learning, 2006.

[14] D. Valy, M. Verleysen, S. Chhun and J.-C. Burie, "A New Khmer Palm Leaf Manuscript Dataset for Document Analysis and Recognition - SleukRith Set," in 4th International Workshop on Historical Document Imaging and Processing (HIP), 2017.

[15] M. W. A. Kesiman, D. Valy, J.-C. Burie, E. Paulaus, M. Suryani, S. Hadi, M. Verleysen, S. Chhun and J.-M. Ogier, "Benchmarking of Document Image Analysis Tasks for Palm Leaf Manuscripts from Southeast Asia," Journal of Imaging, vol. 4, no. 2, p. 43, 2018.

[16] Y.-C. Wu, F. Yin, Z. Chen and C.-L. Liu, "Handwritten Chinese Text Recognition Using Separable Multi-Dimensional Recurrent Neural Network," in 14th IAPR International Conference on Document Analysis and Recognition, 2017.

[17] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in arXiv preprint arXiv:1412.6980, 2014.