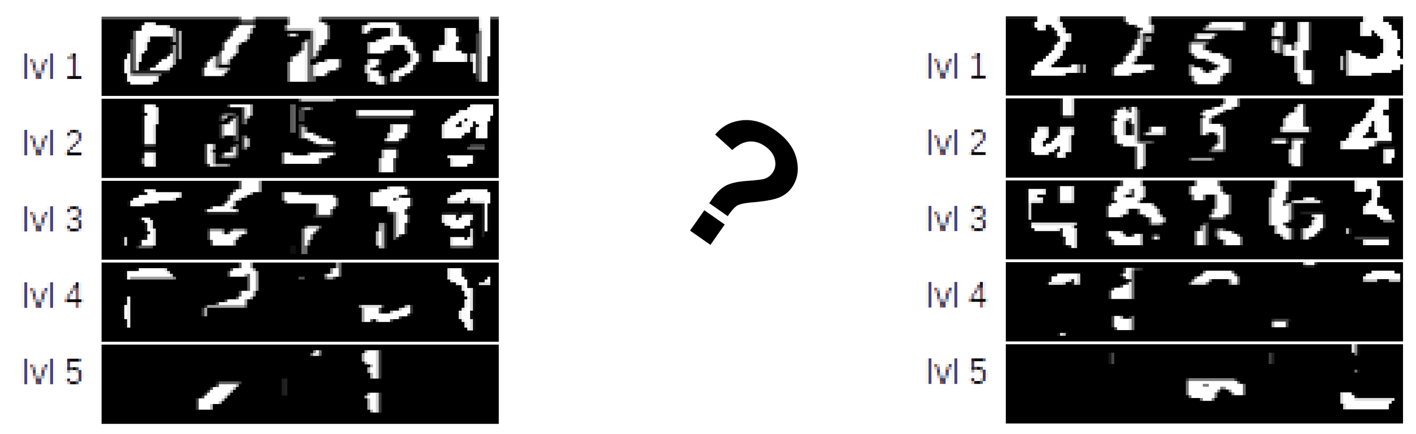
CLASSIFICATION OF SEVERELY OCCLUDED IMAGE SEQUENCES VIA CONVOLUTIONAL RECURRENT NEURAL NETWORKS

BINGHAMTON UNIVERSITY
State University of New York

Jian Zheng, Yifan Wang, Xiaonan Zhang, Xiaohua Li
Dept. of ECE, Binghamton University



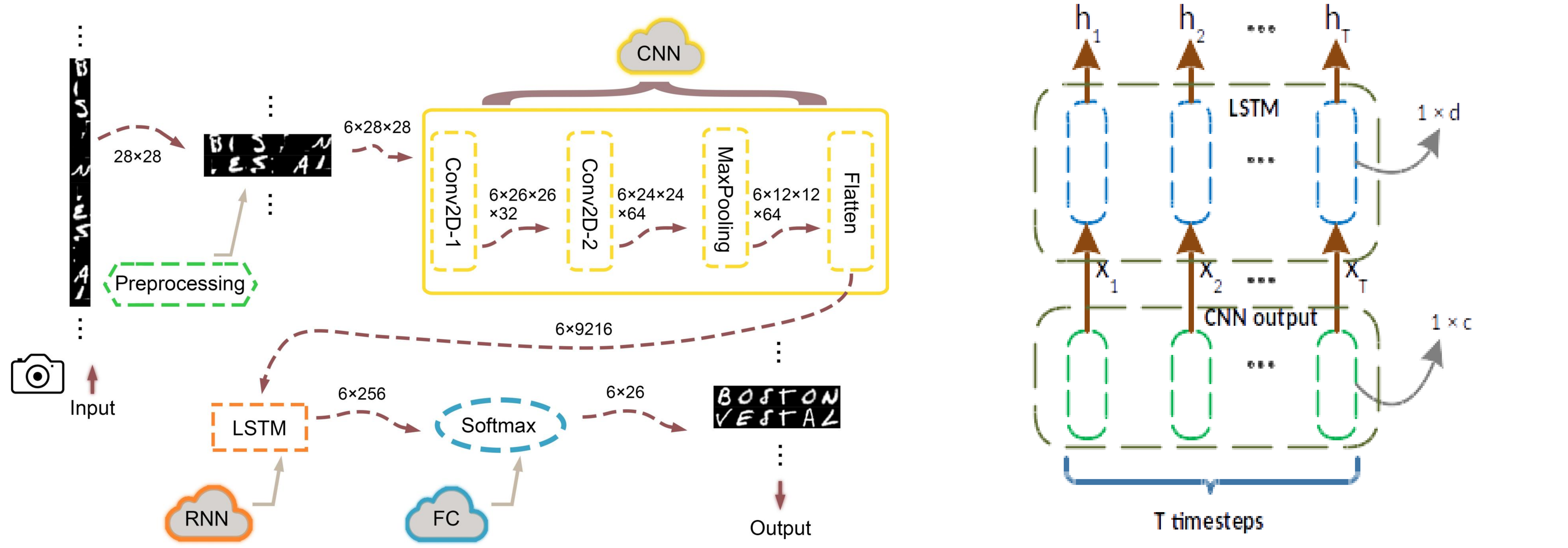
Is the machine able to identify severely occluded images? How?

Motivations

- Most deep learning tasks are conducted with high quality data sources;
- Their performance may not be reliable in noisy or occluded data sets;
- Distortions such as blur, noise, occlusion, etc., will degrade image classification performance;
- Severe occlusions, which block major or key areas of the images, are especially detrimental;
- Classification of severely occluded images is highly needed in many real applications such as self-driving.

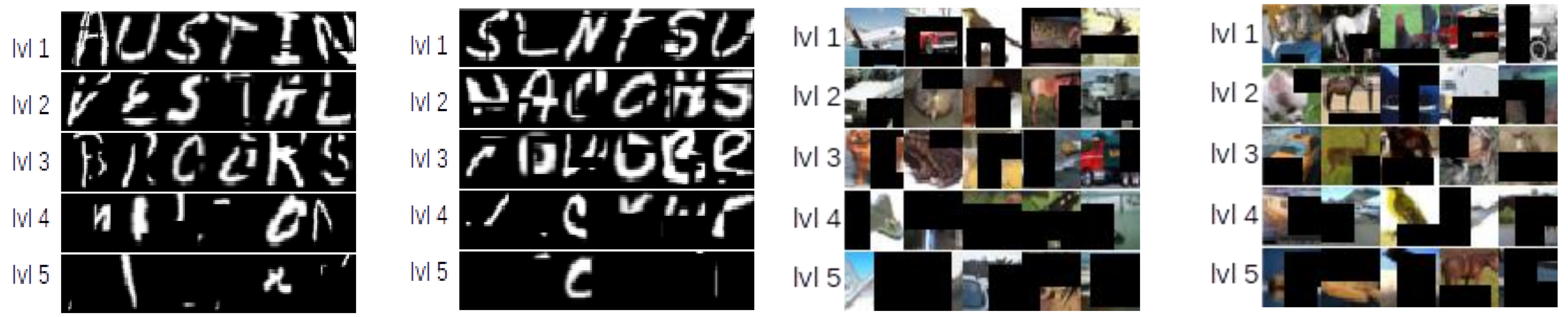
Contributions

- We propose to apply convolutional recurrent neural network (CRNN) for classifying severely occluded image sequences;
- We create three occluded image datasets based on MNIST [1], EMNIST [2] and CIFAR-10 [3], with which we conduct both machine learning experiments and human learning experiments;
- Experiment results show that the proposed CRNN outperforms both conventional methods and most human experimenters.



1. *Data preprocessing*: convert input images into image sequences with fixed length *T*;
2. *Image embedding:* learn image representations for each input image and combine the feature maps of the *T* images;
3. *LSTM for image sequences:* model the spatial contextual dependency among image sequences;
4. *Classification:* classify image sequences based on the output of LSTM unit.

Examples of occluded image sequences



(a) City names (b) Random character sequences (c) CIFAR-10 image sequences with regular patterns (d) Random CIFAR-10 image sequences

References

[1] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
[2] Gregory Cohen, Saeed Afshar, Jonathan Tapson, and Andr e´ van Schaik, "Emnist: an extension of mnist to handwritten letters," arXiv preprint arXiv:1702.05373, 2017.
[3] L. Wan, M. Zeiler, S. Zhang, Y. LeCun, and R. Fergus, "Regularization of neural networks using dropconnect," *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, 2013, pp. 1058–1066.

Experiments

1. Baseline models

- CNN only:
  - CNN
  - CNN-2-S
  - CNN-4-S
- CNN + LSTM:
  - CRNN
  - CRNN-2-S
  - CRNN-4-S

S: replace max pooling with stride of 2
2/4: # of CNN layers

2. Experiment results

Table 1: Performances (%) with regular patterns

Dataset	MNIST	EMNIST	CIFAR10
CNN-2-S	86.22	86.01	44.39
CNN-4-S	88.02	87.26	42.12
CNN	89.44	88.90	54.99
CRNN-2-S	98.27	97.95	89.11
CRNN-4-S	98.15	97.90	90.18
CRNN	98.33	98.14	90.36

Table 2: Performances (%) without regular patterns

Dataset	MNIST	EMNIST	CIFAR10
CNN-2-S	86.37	86.10	44.21
CNN-4-S	88.28	87.81	41.40
CNN	89.37	89.22	53.88
CRNN-2-S	86.68	86.66	45.64
CRNN-4-S	88.15	88.16	48.21
CRNN	87.89	87.81	50.62

Table 3: CRNN performance with regular patterns

Seq len	$T = 5$	$T = 10$	$T = 15$	$T = 20$	$T = 25$
MNIST	98.33	98.11	98.23	98.10	98.06
CIFAR	98.36	88.29	88.11	88.19	87.93
Seq len	$T = 6$	$T = 12$	$T = 18$	$T = 24$	$T = 30$
EMNIST	98.14	97.83	97.76	97.78	97.62

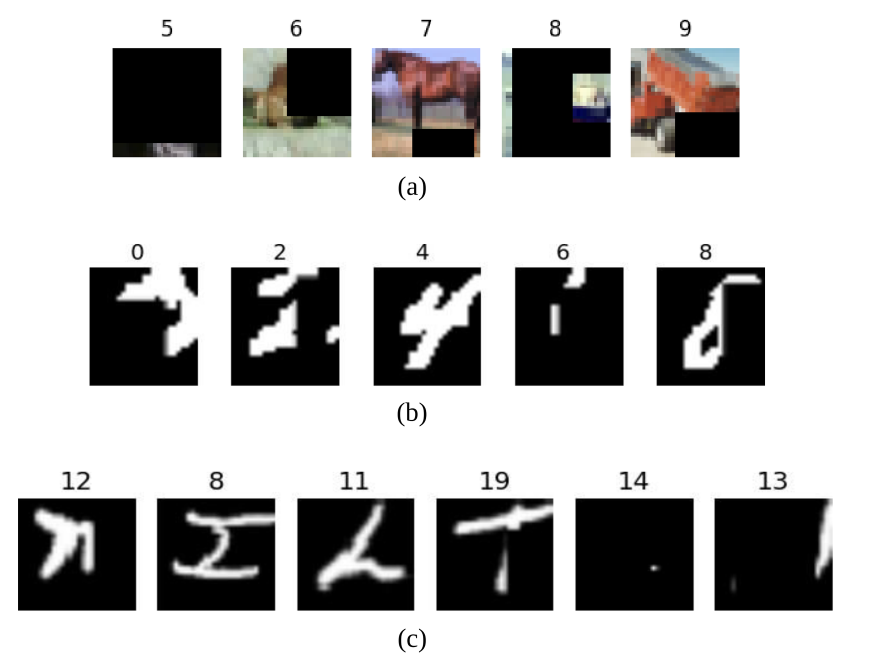
Table 4: CRNN versus human with regular patterns

Dataset	MNIST	EMNIST	CIFAR10
Non-Expert	93.96	92.52	71.11
Expert	100.00	99.26	97.78
CRNN	98.33	98.14	90.36

Table 5: CRNN versus human without regular patterns

Dataset	MNIST	EMNIST	CIFAR10
Non-Expert	75.56	71.79	50.19
Expert	81.00	78.89	71.67
CRNN	87.89	87.81	50.62

3. Three examples of CRNN classification results



(a) Correct classification of the image sequence: 5-6-7-8-9. (b) Correct classification of the digit sequence: 0-2-4-6-8. (c) Correct classification of the city name: M-I-L-T-O-N.