CRN: End-to-end convolutional recurrent network structure applied to vehicle classification

Mohamed Ilyes Lakhal, Sergio Escalera, Hakan Cevikalp
m.i.lakhal@qmul.ac.uk,{sergio.escalera.guerrero, hakan.cevikalp}@gmail.com

1. Introduction

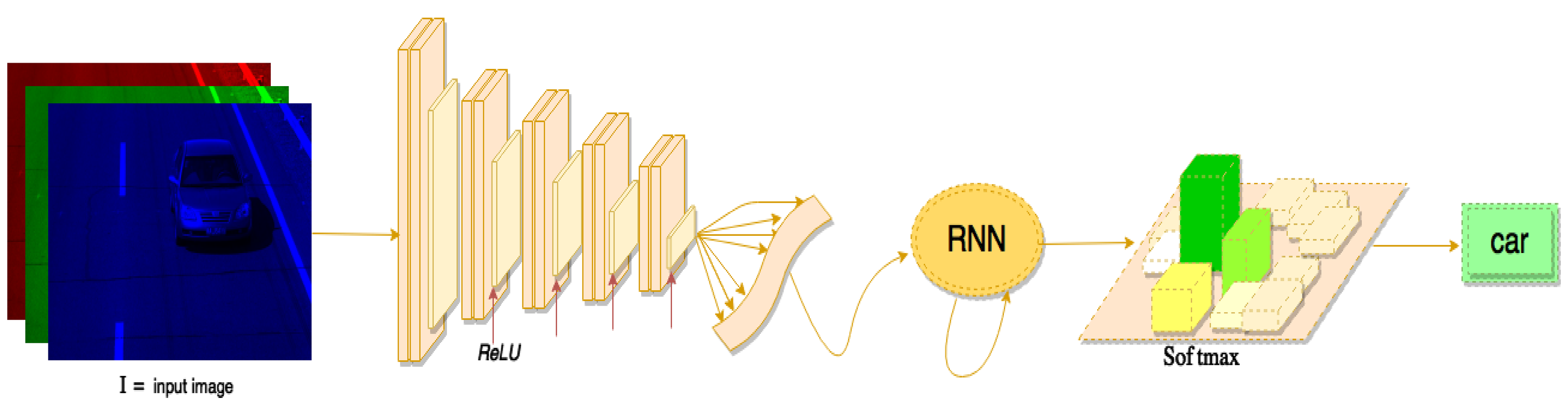
- Traffic Surveillance Cameras (TSC)
  - essential for an Intelligent Traffic System
  - captures images of passing vehicles and other objects [1]
- **Problem statement**
  - classify different types of vehicles from images
- **Challenges**
  - changes in illumination, scale, surface color of vehicles and viewpoint

2. Proposed Convolutional Recurrent Network (CRN)

- Takes advantage of the high level feature from CNN and the flexibility of RNN into a joint model
- Merges two deep learning models into a single structure:
  - CNN part learns discriminative features from input data
  - the output of the CNN is directly fed to an RNN that is used as classifier
- End-to-end feature learning model

3. Implementation

- ESOGU model
  - Extracts intermediate features from CNN
  - Flattens the CNN features to form a single descriptor
  - RNN is used as a classifier instead of the softmax layer



- Model architecture

Module	Layers	Output Size
CNN	Conv	[224 × 224]
	Conv	[224 × 224]
	Pool	[112 × 112]
	ConvBlock$_1$	[56 × 56]
	ConvBlock$_2$	[28 × 28]
	ConvBlock$_3$	[14 × 14]
	ConvBlock$_4$	[7 × 7]
	FC$_1$	[25088]
	FC$_2$	[4096]
	FC$_3$	[4096]
RNN	LSTM	$nb_{classes}$

ESOGU

ESOGU_fc

4. Experimental results

- Datasets

Dataset	Resolution	No-samples	No-classes
Road vehicle [4]	-	2,427	2
BIT-Vehicle [5]	1600x1200 or 1920x1080	900	6
MIO-TCD [6]	-	786,702	11

- Metrics

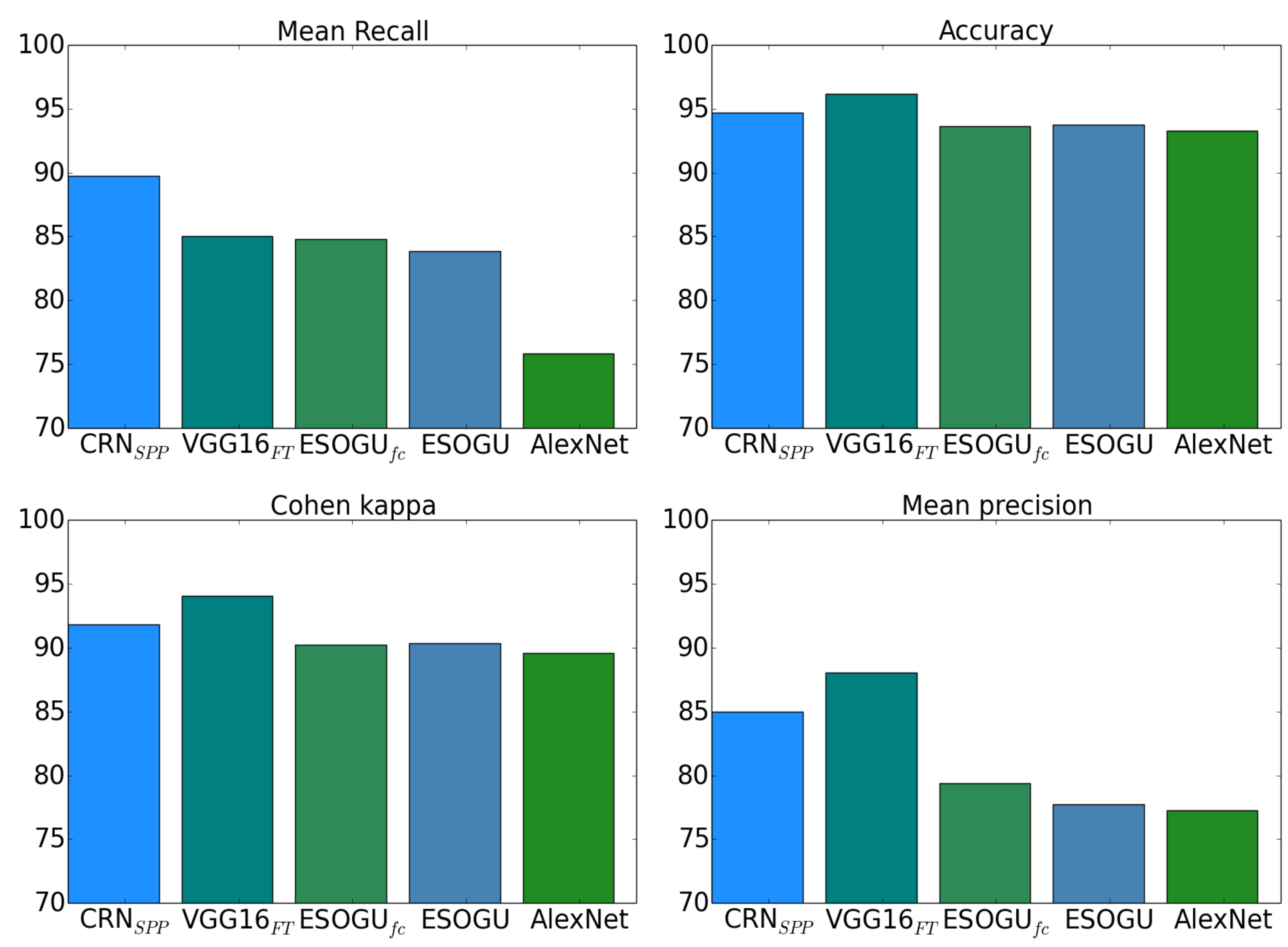
$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad \kappa = (p_o - p_e)/(1-p_e)$$

$$Prec_i = \frac{TP}{TP+FP} \qquad Rec_i = \frac{TP}{TP+FN}$$

$$MeanPrecision = \sum_{i=1}^{C} Prec_i \qquad MeanRecall = \sum_{i=1}^{C} Rec_i$$

p_0 : Empirical probability agreement, p_e : Expected agreement
TP: True positives, FP: False positives, TN: True negatives, FN: False negatives

- Comparison with other state-of-the-art
  - **VGG16** [2]
  - **AlexNet** [3]
  - **CRN_spp** (ours) [follow-up work]
  - **ESOGU_fc** (ours)
  - **ESOGU** (ours)



Test Results on MIO-TCD dataset

5. Conclusion

- Feature learning part is done using CNN
- Recurrent neural network is used as a classifier
- Model trainable in an end-to-end fashion
- Adaptable to other scenarios like multi-label image classification

References

[1] Yong Tang, Congzhe Zhang, Renshu Gu, Peng Li, and Bin Yang. Vehicle detection and recognition for intelligent traffic surveillance system. Multimedia Tools and Applications, 2017.
[2 K. Simonyan and A. Zisserman. Very deep convolutional networksfor large-scale image recognition. ICLR, 2015.
[3] Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutionalneural networks. Advances in Neural Information Processing Systems, 2012.
[4] Y. Zhou, H. Nejati, T. T. Do, N. M. Cheung, and L. Cheah. Image-based vehicle analysis using deep neu- ral network: A systematic study. IEEE DSP, 2016.
[5] Z. Dong, M. Pei, Y. He, T. Liu, Y. Dong, and Y. Jia. Vehicle type classification using unsupervised convolutional neural network. ICPR 2014.
[6] http://tcd.miovision.com/challenge/dataset/

Acknowledgment

This work has been partially supported by the Spanish project TIN2016-74946-P (MINECO/FEDER, UE) and CERCA Programme / Generalitat de Catalunya.