Bundeskriminalamt

FAU
BKA

Friedrich-Alexander-Universität
Erlangen-Nürnberg

Forensic License Plate Recognition with Compression-Informed Transformers

Denise Moussa[1,2], Anatol Maier[2], Andreas Spruck[3], Jürgen Seiler[3], Christian Riess[2]

[1]Federal Criminal Police Office (BKA), Germany
[2]IT Security Infrastructures Lab, Computer Science, Friedrich-Alexander-Universität Erlangen-Nürnberg
[3]Multimedia Communications and Signal Processing, Electrical Engineering, Friedrich-Alexander-Universität Erlangen-Nürnberg

Motivation: Driving Forward Forensic License Plate Recognition

Challenges for Image Forensics in Daily Police Work

- Low-cost cameras often introduce high compression and low resolution
- Classic tools for image enhancement may fail for very low quality footage

Our Proposals

We effectively improve the recognition rate of (especially very low quality) license plate images by:

1. Improving the Neural Network Architecture
   - Sequence-to-sequence approach based on the Transformer [1] model
   - Higher performance while needing less parameters
2. Exploiting Image Quality Information
   - Feed estimated compression strength to network

   ⇒ The lower the image quality, the higher the additional benefit of this side information

Methods

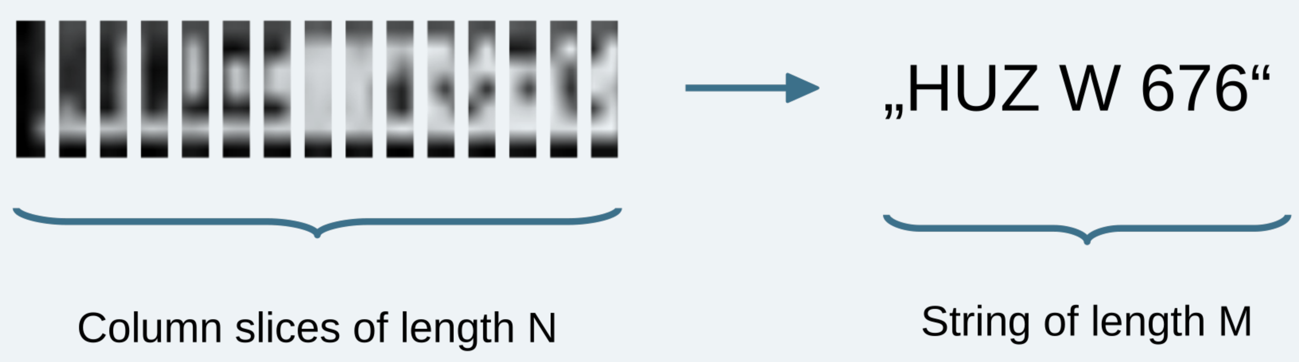
1: Estimating the JPEG Compression Strength

Image analysis

1. Extract/estimate the quantization matrix from the image
2. Regress it to the closest standard $\mathrm{QF} \in [1, 100]$ of libjpeg [2]

⇒ Scalar quality surrogate value for any JPEG image

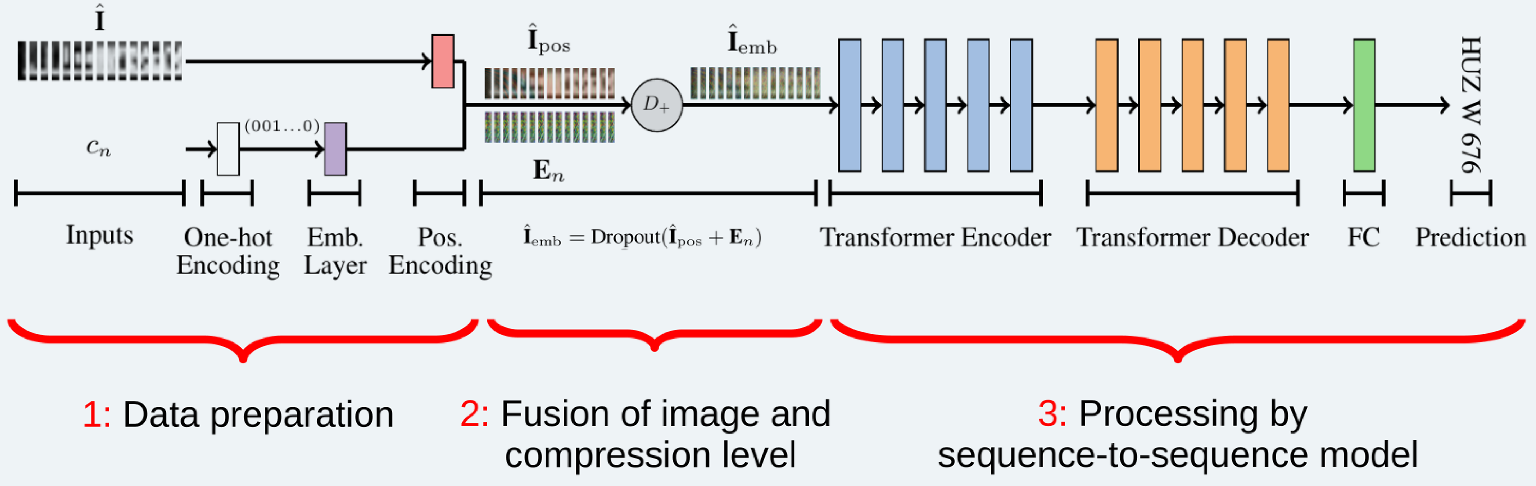
2: Sequence Based License Plate Recognition



⇒ Enables for predicting captions of arbitrary length

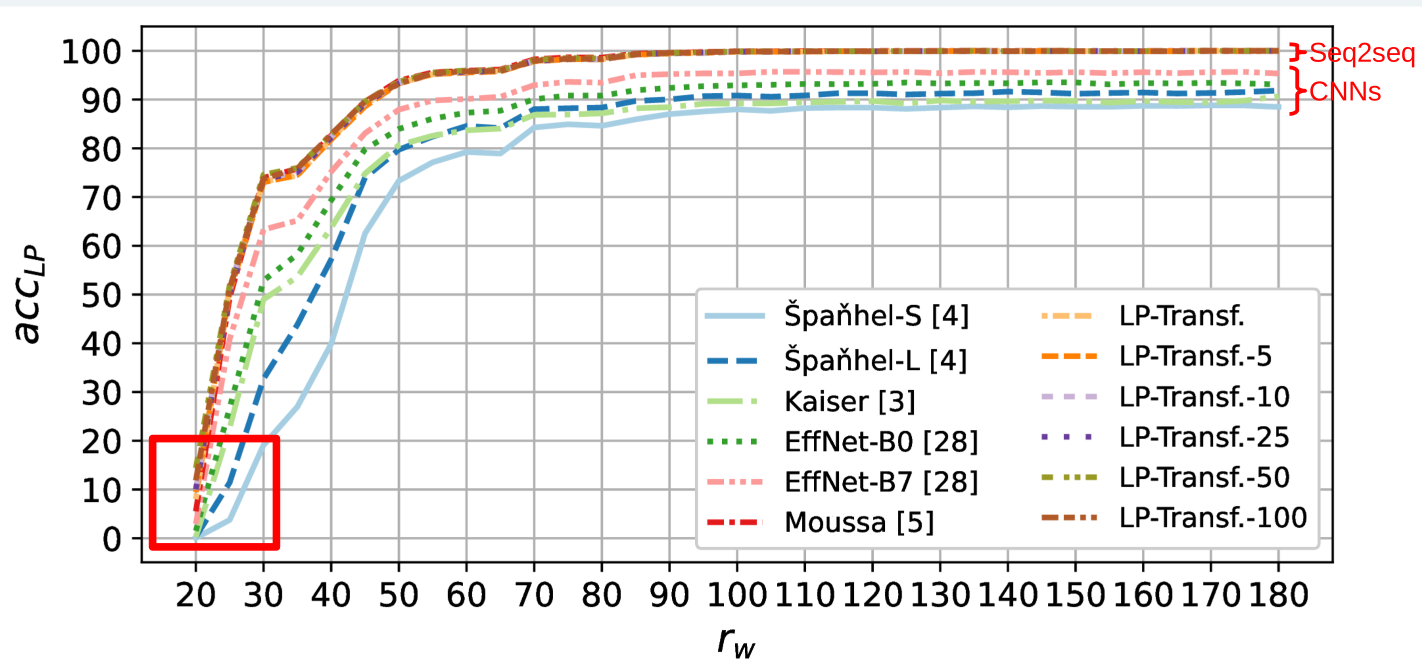
⇒ Exploits sequential context information from license plates

3: Combining Both Aspects in One Architecture

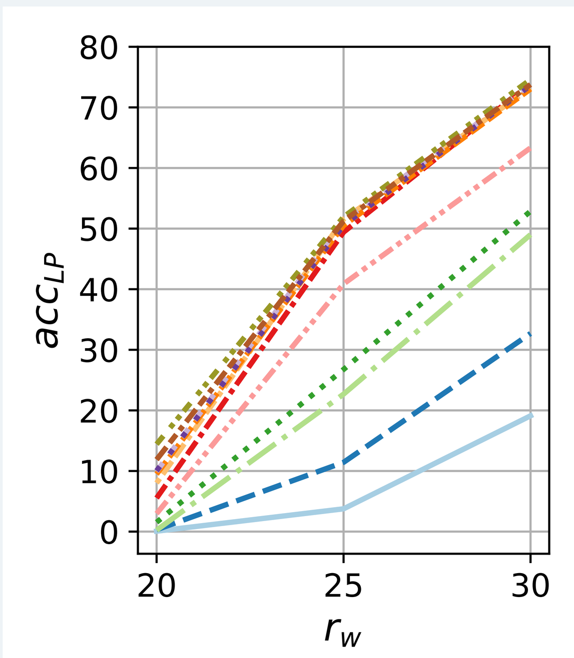


1. Send compression level c_n to an embedding layer and project to dimensions of input image $\hat{\mathbf{I}}$
2. Fuse input image $\hat{\mathbf{I}}$ and compression level c_n
3. Feed result $\hat{\mathbf{I}}_{emb}$ to Transformer sequence-to-sequence network to yield the final prediction of the license plate caption

Experiments



(a) Accuracy per license plate ($\mathrm{acc_{lp}}$) for resolutions in pixel width $r_w \in [20, 180]$. Compression levels $\mathrm{QF} \in [1, 100]$ are included for each r_w.



(b) Close-up of $r_w \in [20, 30]$

Method	$\mathrm{acc_{lp}}$	CER
Moussa [5]	5.53%	0.3496
LP-Transf.	8.02%	0.3254
LP-Transf.-5	9.48%	0.3131
LP-Transf.-10	9.94%	0.3046
LP-Transf.-25	10.04%	0.3046
LP-Transf.-50	**14.43%**	**0.2848**
LP-Transf.-100	11.88%	0.2990

(c) Lowest resolution $r_w = 20$

Seq2seq methods, thus our LP-Transf.-X model and the related Convolutional Recurrent Neural Network (CRNN) approach, surpass the Convolutional Neural Network (CNN) classifier methods. For very low resolutions (red rectangle), our models have an significant advantage. Fig. 1b shows an enlargement of the region.

Close-up of low resolution region. Our models' advantage over the CRNN baseline (red line) increases with decreasing image quality.

For very low resolutions, our models' benefit is most apparent. Distinguishing 50 compression quality levels performs best. This may be a good trade-off between quality granularity (half precision) and optimization effort for the compression embedding layer.

References

[1] A. Vaswani et al. "Attention Is All You Need". In: *Advances in Neural Information Processing Systems*. 2017, pp. 5998–6008.

[2] D. Cozzolino and L. Verdoliva. "Noiseprint: a CNN-Based Camera Model Fingerprint". In: *IEEE Transactions on Information Forensics and Security* 15 (2019), pp. 144–159.

www.fau.de

denise.moussa@fau.de