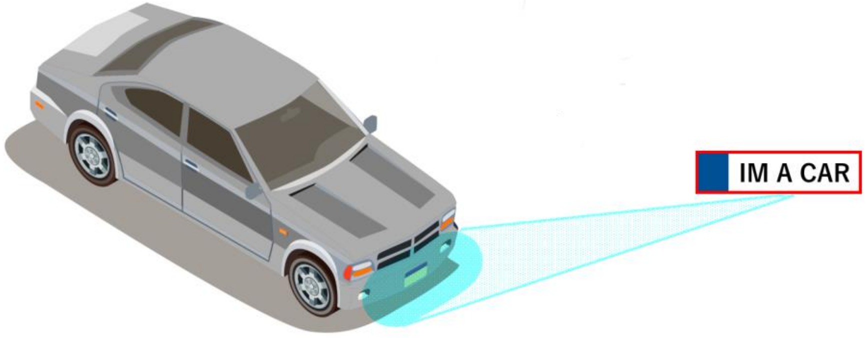
12th International Young Scientists Conference on Computational Science

CIS Multilingual License Plate Detection and Recognition Based on Convolutional and Transformer Neural Networks

Irek Saitov and Andrey Filchenkov
ITMO university

Relevance

Applicability:
- automatic collection of money on toll roads;
- traffic safety monitoring;
- access control to closed areas.



Difficulties:
- size, shape, color, font and placement of symbols variations;
- different languages and special symbols;
- video or photo qualities.

Standard steps for license plate recognition:
1. License plate detection (bounding box / segmentation).
2. Character segmentation.
3. Optical character recognition (OCR).

Related works

Nomeroff Net – open-source framework for license plates recognition.

- License plate detection model based on YOLOv5 convolutional neural network;
- Licence plate OCR model based on GRU modules.

Main goal

Goal: Develop the main components of the ALPR system: license plate detection model and multilingual license plate recognition model for the CIS countries.

Tasks:

1. Develop license plate detection model

2. Develop multilingual license plate recognition model for the CIS countries.

3. Perform a comparative analysis with Nomeroff Net approach and draw conclusions based on the results of experiments of the developed models.

Detection model

YOLO[1] application: object detection, classification, semantic segmentation, object tracking, and pose estimation.

YOLOv5 improvements:
- Adaptive anchor box selection process “autoanchor”.
- Mosaic augmentation.

YOLOv8 improvements:
- New structural convolution block.
- Anchor-free detection system.
- Corrected use of mosaic augmentation.

1. YOLO by Ultralytics. URL: https://github.com/ultralytics/ultralytics

OCR model

TrOCR[1]

Training: synthetic Wikipedia corpus of raw text images.

Encoder: visual transformer
- DEiT;
- BEiT.

Decoder: text transformer
- MiniLM;
- RoBERTa.

Pretrained versions:
1. First-stage.
2. Second-stage:
   a) handwritten
   b) printed

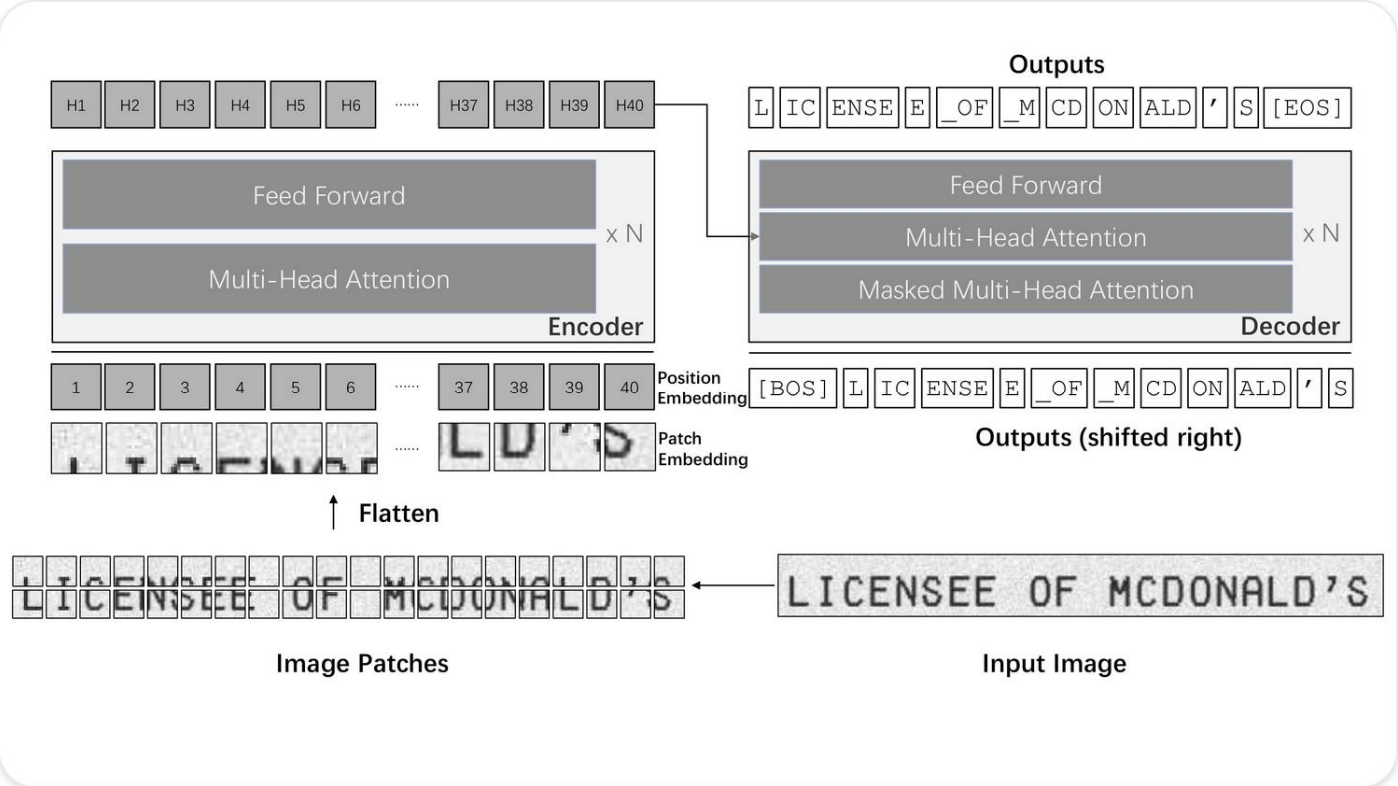


Fig. 1: The architecture of TrOCR, where an encoder-decoder model is designed with a pre-trained image Transformer as the encoder and a pre-trained text Transformer as the decoder.

1. Li, M., Lv, T., Chen, J., Cui, L., Lu, Y., Florencio, D., Zhang, C., Li, Z., Wei, F., 2021. Trocr: Transformer-based optical character recognition with pre-trained models. arXiv preprint arXiv:2109.10282 .

Data

License plate detection data (1/2)

The detection dataset has 11384 images.

Data split:
- train: 7968 (70%)
- validation: 1708 (15%)
- test: 1708 (15%).

Preprocessing:
- Image resizing: 640x640

Augmentation:
- scaling;
- color space adjustment;
- mosaic transforms.

Fig. 1. Vehicles examples with license plates from Nomeroff Net dataset. Photos taken under different environments. In many samples, the camera viewpoint transforms the perspective projection of license plates in the image

Data

Optical character recognition data (2/2)

Preprocessing:

- Image resizing:
  - TrOCR fine-tuning: 384x384
  - Scratch visual transformers: 224x224
- Square operation (for some experiments)

Augmentation:

- Random choice from next list:
  - random rotation (-10 to 10 degrees),
  - gaussian blurring,
  - image dilation,
  - image erosion,
  - downscaling,
  - underlining,
  - keeping the original image.

Table 1. Number of samples according to the split sets for each available license plate OCR dataset from Nomeroff Net.

Country	Short name	Train	Valid	Test
Kazakhstan	Kz	8642	1001	279
Moldova	Md	10531	789	849
Armenia	Am	11524	565	620
Georgia	Ge	24986	738	2777
Kyrgyzstan	Kg	29204	1645	1068
ex USSR	Su	35310	1874	1618
Europe	Eu	42738	1359	1531
Russia	Ru	49382	4893	2845
Ukraine	Ua	121721	2146	2304

Fig. 1. License plates examples from Kazakhstan, Georgia, Armenia, and Kyrgyzstan. Each country has an original design, where the font, position and the order of characters differ

Metrics

Object detection metric:

$$mAP@50 = \frac{1}{n}\sum_{k=1}^{n} AP_k\,, where\ TP\ if\ IoU > 0.5 \tag{1}$$

$$AP = \sum (r_{n+1} - r_n) p_{interp}(r_{n+1}) \tag{2}$$

$$p_{interp}(r_{n+1}) = \max_{\tilde{r} \geq r_{n+1}} p(\tilde{r}) \tag{3}$$

OCR metric:

$$CER = \frac{I + D + S}{N}, \tag{4}$$

where I num of insertions, D is a number of deletions, S is a number of substitutions, N is a number of characters in the prediction.

Experiments setup

GPU: NVIDIA TITAN RTX (24GB of RAM)

Localization model:
- Weights for YOLOv5/v8: pretrain MS COCO.
- Model size: ‘s’ small, 7.2M parameters.
- Batch size: 32.
- Input image size: 640x640.
- Epochs: 100.

OCR model:
1. TrOCR fine-tuning:
   - Model size:
     - base ($BEiT_{base}$ 12 layers + $RoBERTa_{base}$ 6 last layers, 334M parameters), batch size: 4 .
     - large ($BEiT_{large}$ 24 layers + $RoBERTa_{large}$ 12 layers, 558M parameters), batch size: 2.
   - Input image size: 384x384
   - Training: 1500 / 4000 steps
2. Scratch encoder-decoder transformer training.
   - Encoder: ViT
   - Decoder: enBERTc / deeppavlov-BERTc / geo-BERTc
   - Input image size: 224x224
   - Training: 9000 steps

Experiments

License plate detection

Fig. 1. Examples of successful license plates detections of various vehicles by trained YOLOv8 model. Recognized license plate highlighted with a green bounding box.

	mAP@50	Inference time
YOLOv5	0.951	8.7ms
YOLOv8	**0.981**	**3.3ms**

Table 1. Results by mAP@50 metric of YOLO models fine-tuning for license plate localization task. The best result achieved on the test set after 100 epoch training is shown in bold.

YSC young scientists conference

Experiments

Optical character recognition. Fine-tuning on ‘Ru’ dataset (1/2)

Table 1. Testing results of TrOCR fine-tuning on license plate recognition task by CER metric for "Ru" dataset OCR experiments. The best result is shown in bold.

Method	Step num	Model size	Pretraining	Aug	Tune	CER
Nomeroff app						**0.00050**
TrOCR	1500	base	stage1			0.00075
TrOCR	1500	base	printed			0.00084
TrOCR	1500	base	stage1	augmented		0.00063
TrOCR	4000	base	stage1	augmented		0.00485
TrOCR	1500	base	stage1		tuned	0.00121
TrOCR	1500	large	stage1			0.00134

Table 2. Vision transformer encoder-decoder scratch testing results by CER metric for ”Ru” dataset OCR experiments

Method	Encoder	Decoder	Steps	Extra params	CER
Scratch	ViT224	en-BERTc	9000	square,lower	0.00472
Scratch	ViT224	deepavlov-BERTc	9000		0.00430
Scrath	ViT224	geo-BERTc	9000		0.04516

young scientists conference

Experiments

Optical character recognition. Per country experiments (2/2)

Table 1. Results by CER metric on test data for other (Am, Eu, Kg, Kz, Su, Ge, Ua, Md) datasets OCR experiments. The best results are shown in bold

Country	Nomeroff app	TrOCR base-stage1, 1500 steps	TrOCR base-stage1, augmented, 1500 steps	TrOCR base-stage1, augmented, 1500 steps, square, lower
Kz	0.00454	0.00499	0.00136	0.00181
Md	0.00195	**0.00019**	0.00058	0.00039
Am	0.00069	0.00069	**0.00023**	0.00069
Ge	**0.00119**	0.00549	0.00281	0.00233
Kg	0.00267	**0.00069**	0.00214	0.00253
Su	0.00406	0.01094	**0.00273**	0.00361
Eu	0.00576	0.01359	**0.00490**	0.00726
Ua	0.00152	0.00211	**0.00092**	0.00211

Conclusion and future work

Conclusion.

- The proposed model can be employed as the major element of intelligent infrastructure like toll fee collection, parking management, and traffic surveillance.
- The powerful detection model shows a result of 0.983 for mAP@50 metric in the fine-tuning scenario while surpassing the baseline in terms of speed and performance.
- TrOCR model fine-tuning showed the best result on 7 out of 9 datasets (Armenia, Kazakhstan, Ukraine, Moldova, ex-USSR, Kyrgyzstan, and Europe).

Future work:

- Develop a full-fledged automatic license plate recognition (ALPR) system for the CIS countries.
    - Develop automatic OCR model selection for a specific country.
    - Develop an algorithm for normalizing perspective distortions of license plates.
- Expand the database with additional data by collecting data via Internet requests - for example, from “platesmania.com” (containing more than 40k samples for each country in question).

young scientists conference

Thanks for your attention