Automatic Structured Text Reading for License Plates and Utility Meters

Marçal Rusiñol[1,2]
marcal@cvc.uab.es
Lluís Gómez[1,2]
lgomez@cvc.uab.es
Adriaan Landman[2]
adriaan@allread.ai
Miguel Silva-Constenla[2]
miguel@allread.ai
Dimosthenis Karatzas[1,2]
dimos@cvc.uab.es

[1] Computer Vision Center, Dept. Ciències de la Computació
Edifici O, Campus UAB
08193 Barcelona, Spain.

[2] AllRead Machine Learning Technologies
Pier01, Pl. de Pau Vila 1
08019 Barcelona, Spain.

Abstract

Reading text in images has attracted interest from computer vision researchers for many years. Our technology focuses on the extraction of structured text – such as serial numbers, machine readings, product codes, etc. – so that it is able to center its attention just on the relevant textual elements. It is conceived to work in an end-to-end fashion, bypassing any explicit text segmentation stage. In this paper we present two different industrial use cases where we have applied our automatic structured text reading technology. In the first one, we demonstrate an outstanding performance when reading license plates compared to the current state of the art. In the second one, we present results on our solution for reading utility meters. The technology is commercialized by a recently created spin-off company, and both solutions are at different stages of integration with final clients.

1 Introduction

Text is omnipresent in man-made environments, and conveys important, semantic information about the world around us. Everyday tasks require humans to read and act upon all sorts of textual information. In industry, text is habitually used to transmit structured information such as numeric values, serial numbers, machine readings, product codes, etc. Such kind of textual information is not easy to read by human operators. The human reading process is based on word-level processing of the textual content, and any structured textual content requiring a character-by-character interpretation is not natural for humans to read. As such humans often err when reading such content. The fact that such structured text is habitually presented in complex scenarios, and that a single reading mistake can have important consequences, aggravates the situation. When it comes to structured text, machine-based automated reading software can be the solution, but instead of a generic OCR a different approach is needed, adaptable and specialized to each reading domain.

Although OCR engines, able to convert document images to electronic text, have been around for several decades, we have seen in the last years a revolution in this field powered by the latest advancements in deep learning. Nowadays, state of the art methods are able to locate and read text not only in scanned documents but also in natural scenes "in the wild" [3, 4, 7]. However, scene text recognizers are often designed to work with a closed vocabulary, and thus tend to perform badly when dealing with structured text that cannot be explicitly incorporated in any lexicon, such as price tags, utility meter readings, license plates, serial numbers, etc. Our proposed solution is not yet another OCR, in the sense that we do not want it to read all the text within an image, but to focus only on a specific textual pattern. Our recognition models are carefully tailored to work in specific industrial scenarios. By training *ad-hoc* models, we guarantee an important recognition accuracy, but also, we are able to extract just the text that is deemed important for the user, ignoring the rest of non-relevant textual information appearing in the image.

Another noteworthy aspect of our proposed approach, is that it is able to directly process incoming images and output their corresponding readings, without any explicit text segmentation step. State-of-the-art scene text extraction methods typically localize first text regions in the image and then feed them to a recognition process. Any localization errors affect the subsequent text recognition step. In order to overcome this, over-segmentation strategies are usually adopted, e.g. [2, 4]. Our proposed solution on the other hand, avoids any explicit localization step, and allows instead end-to-end recognition of structured text in images.

Our technology was born in the *Computer Vision Center*, and is being commercialized today by a new spin-off company, *AllRead Machine Learning Technologies*. The first commercial use cases that have been rolled out are focused on the digitization of manual processes that involve reading specific textual information, such as utility meters, measurement instruments or printed identification codes. In this paper, after a brief presentation of our solution, we will present the results obtained first on a public dataset for license plates, and then on one of our industrial use cases for reading utility meters.

2 Reading Structured Text

Our proposed solution is based on a convolutional neural network that processes images and outputs the desired structured texts in a single shot. The network is composed of a convolutional backbone followed by stacking several independent fully connected layers, each one with a Softmax activation producing a probability distribution over the n possible classes for each expected symbol of the final reading. The n outputs of the network are then treated as a typical classification output and trained using the Cross-Entropy loss function.

Intuitively for the network being able to read in an end-to-end manner it has to take into account global information extracted over the entire image to make the individual symbol predictions. The initial convolutional layers extract visual features of the whole image, while the fully connected layers specialize in predicting the output probabilities for each symbol.

This end-to-end model has several benefits over traditional methods of robust reading. First, the network can be trained with full images, without any explicit segmentation, and directly optimizes the end-to-end reading performance. Second, the particular design of the network allows for real time reading speeds while achieving high reading accuracy.

3 License Plate Reading

Automatic license plate recognition has been addressed within the computer vision community for many years. It is particularly interesting for security and surveillance applications. In order to assess the performance of our solution and compare its accuracy against the state of the art, we decided to use as benchmarking test the recent CCPD dataset of Chinese license plates [12]. The dataset is composed of 250,000 images with ground truth annotations of the license plate position and transcription. Each Chinese license plate number is comprised of a Chinese character, a letter, and five letters or numbers. The CCPD dataset comes from several test sets with images having different problematics (see Fig. 1) such as blur, distance, illumination, weather conditions, etc.

Figure 1: Sample images from the CCPD [12] dataset

Method	Test Set						
	Base	DB	FN	Rotate	Tilt	Weather	Challenge
Cascade [11] + HC [10]	69.7	67.2	69.7	0.1	3.1	52.3	30.9
SSD300 [6] + HC [10]	98.3	96.6	**95.9**	88.4	91.5	87.3	83.8
YOLO9000 [8] + HC [10]	98.1	96.0	88.2	84.5	88.5	87.0	80.5
Faster-RCNN [9] + HC [10]	97.2	94.4	90.9	82.9	87.3	85.5	76.3
TE2E [5]	97.8	94.8	94.5	87.9	92.1	86.8	81.2
RPNet [12]	98.5	96.9	94.3	90.8	92.5	87.9	**85.1**
Ours	**99.0**	**97.7**	94.7	**97.5**	**98.0**	**96.6**	82.6

Table 1: End-to-end recognition on the test subsets of the CCPD Dataset. DB and FN denote the Dark/Bright and Far/Near test subsets respectively. HC denotes Holistic-CNN [10].

We present in Table 1 the reached end-to-end recognition accuracy in comparison with the different state-of-the-art baselines presented in [12]. Our proposed solution outperforms the rest of approaches in most of the test sets. It is worth to note its notable improvement in scenarios in which we have rotation, tilt or weather conditions. In such test sets methods that first try to segment the license plate for a latter recognition tend to be damaged, whereas since our method bypasses such segmentation, our performance is maintained.

4 Utility Meters Reading

A 2014 European Commission report [1] on the deployment of smart metering estimated that close to 200 million smart meters for electricity and 45 million for gas will be rolled out in the EU by 2020. This represents a potential investment of €45 billion. However such estimates are still far from complete roll-out. In order to automate the process of consumption reading, utility companies allow users to submit a legally binding photograph of their meters. In such cases, to automatically extract the meter reading entails important economical benefits. Our solution has been implemented by a utility company and is currently processing 8,000 images per week. Such images are captured and sent by the end users and there are over forty different meter models represented in them (see Fig 2).

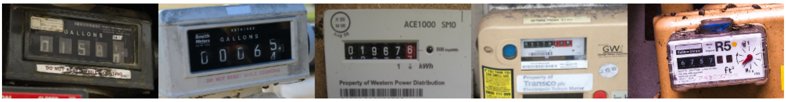

Figure 2: Examples of utility meters. Public domain images similar to the ones in our private dataset shown for illustrative purposes.

We present below, one of our latest benchmark studies carried out with over a million annotated images. Although most of the meter models register the reading with five digits, some less frequent meters have either four or six digits. We show in Table 2 the different distribution of this dataset, where we can appreciate how unbalanced the distribution of meter models is. Since reading errors might be critical when it comes to bill incorrect amounts,

Set	4 digits	5 digits	6 digits	Total
Train	30,189	836,589	7,238	874,016
Test	7,603	209,073	1,829	21,8505
Total	37,792	1,045,662	9,067	1,092,521

Table 2: Utility meters dataset statistics.

Coverage	4 digits	5 digits	6 digits	All
100	91.80	94.49	85.57	94.33
95	96.06	98.17	89.42	98.03
90	99.48	99.37	92.78	99.42
75	99.92	99.89	99.19	99.91

Table 3: End-to-end recognition rates at different test dataset coverage points.

we set up a rejection strategy for the images that present a confidence score below a certain threshold. We present in Table 3 the obtained results at different coverage points. From this table we can see that when forwarding a 10% of the images to manual inspection, the remaining 90% of the images are automatically processed with an accuracy beyond 99%.

5 Conclusions

In this paper we have presented two industrial use cases where we have applied our automatic structured text reading technology. Our solution outperforms the current state-of-the-art methods on license plate recognition, by bypassing any explicit text segmentation step. We also presented the utility meter use case where our solution has been implemented. It currently processes 8,000 images per week with an accuracy beyond 99% when applying a rejection strategy.

References

[1] European commission. The role of DSOs in a smart grid environment. https://ec.europa.eu/energy/sites/ener/files/documents/20140423_dso_smartgrid.pdf, 2014.

[2] L. Gómez and D. Karatzas. Textproposals: A text-specific selective search algorithm for word spotting in the wild. *Pattern Recognition*, 70:60–74, October 2017.

[3] Lluís Gómez, Andrés Mafla, Marçal Rusinol, and Dimosthenis Karatzas. Single shot scene text retrieval. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 700–715, 2018.

[4] Max Jaderberg, Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Reading text in the wild with convolutional neural networks. *International Journal of Computer Vision*, 116(1):1–20, 2016.

[5] Hui Li, Peng Wang, and Chunhua Shen. Toward end-to-end car license plate detection and recognition with deep neural networks. *IEEE Transactions on Intelligent Transportation Systems*, 20(3):1126–1136, 2018.

[6] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pages 21–37. Springer, 2016.

[7] Lukáš Neumann and Jiří Matas. Real-time scene text localization and recognition. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3538–3545. IEEE, 2012.

[8] Joseph Redmon and Ali Farhadi. Yolo9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7263–7271, 2017.

[9] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015.

[10] Jakub Špaňhel, Jakub Sochor, Roman Juránek, Adam Herout, Lukáš Maršík, and Pavel Zemčík. Holistic recognition of low quality license plates by cnn using track annotated data. In *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–6. IEEE, 2017.

[11] Shen-Zheng Wang and Hsi-Jian Lee. A cascade framework for a real-time statistical plate recognition system. *IEEE Transactions on Information Forensics and Security*, 2 (2):267–282, 2007.

[12] Zhenbo Xu, Wei Yang, Ajin Meng, Nanxue Lu, Huan Huang, Changchun Ying, and Liusheng Huang. Towards end-to-end license plate detection and recognition: A large dataset and baseline. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 255–271, 2018.