

Postdoc position

Title: Information extraction from scanned images for consumption analysis on the out-of-home market

Period: May 1, 2020 to October 30, 2021 (18 months)

Contact: A. Belaïd, LORIA, READ Team, France, abdel.belaid@loria.fr

The project

This is an AMI (Appel à Manifestation d'Intérêt Economie Numérique) project in collaboration with the VAZEE Company (Strasbourg, France) and LORIA laboratory (Nancy, France).

The objective

The aim is to create a set of digital tools for food brands and out-of-home catering professionals, enabling promotion to end consumers and analysis of consumption in the out-of-home market. From a simple photograph of their out-of-home consumption receipt, the user acquires loyalty points to be exchanged for partner offers. The system to be implemented comprises several stages: acquisition of the image of the receipt from his smartphone, location of the text lines, and information extraction.

The literature has for some years been proposing solutions for certain activity sectors, such as drinking places, but these solutions are too restrictive to the types of information to be extracted, or even in the processing of languages. We can mention Mobishop [1] which is a participative detection application which facilitates the sharing of product prices between participants, OCRdroid [2] which is based on the optical character recognition engine Tesseract [3] and seeks to improve problems of misalignment and insufficient light, Receiptlog [4] to identify users' specialized purchasing preferences in order to predict their future behavior, or finally the application of Dikici and Saraclar [5] which uses linguistic benchmarks to correct rolling texts that are featured in newsletters.

In addition, this year, on the sidelines of the international conference ICDAR (International Conference on Document Analysis and Recognition), a group for the evaluation of receipts called SROIE (Scanned Receipts OCR and key information extraction) is created [6]. The report produced at the end of this competition indicates: "although the tasks of localization and recognition of the text seem relatively easy to tackle, it is interesting to observe the variety of ideas and approaches proposed for information extraction. Based on the communications results, we believe there is still room to improve information retrieval performance." We have made the same observation and offer to the VAZEE Company a more efficient solution for its receipts.

The subject

Inspired by the recent successes recorded by deep learning models and convinced by the contribution of linguistics in this case, we propose a neural solution using language models. For the detection of text lines, we would like to adapt the EAST method (Efficient and Accurate Scene Text Detector) [7] and the Connectionist Text Proposal Network (CTPN) [8] which have been proposed for the text detection in natural images.

EAST uses a single neural network to predict a word or text at line level. It can detect text in an arbitrary orientation with quadrilateral shapes. In 2017, this algorithm surpassed advanced methods. It consists of a fully convolutional network with a merge state with non-maximum suppression.

CTPN explores rich context information from an input image, making it a powerful tool for detecting text in different ticket formats. The CTPN structure is similar to Faster R-CNN, but with the addition of the LSTM layer. The network model mainly consists of three parts: the extraction of characteristics by VGG16, the bidirectional LSTM and the regression of the selection framework.

Both the language and the linguistic aspects will be integrated into these networks. This will avoid classic systems that require finding the language first, then applying the concomitant rules to find the information accordingly. A large number of annotated document samples would be required to complete this step.

The information to be extracted corresponds to a word or a group of words united in an assembly that is called "context". The chosen network will have to learn these assemblies and predict them during the recognition phase. For their representation, we can either use a vector representation like Word2Vec or a neighborhood graph, like GAN or GraphGAN [9].

References

- [1] Sehgal, S., Kanhere, S. S. and Chou, C. T., "Mobishop: Using mobile phones for sharing consumer pricing information," Proc. Conference on Distributed Computing in Sensor Systems, Santorini, Greece (2008).
- [2] Joshi, A., Zhang M., Kadmawala, R., Dantu, K., Poduri, S. and Sukhatme, G., "OCRdroid : A Framework to Digitize Text Using Mobile Phone," Proc. ICST International Conference on Mobile Computing, Applications, and Services, (2009).
- [3] Smith, R., "An overview of the Tesseract OCR engine," Proc. International Conference on Document Analysis and Recognition, Brazil, (2007).
- [4] Tokunaga, S., Matsumoto, S. and Nakamura, M., "Receiptlog: A consumer-oriented lifelog service for storing and reviewing daily receipts", IEICE Technical Report, vol. 111, no.107, 23–28 (2011).
- [5] Dikici, E., and Saraçlar, M., "Sliding text recognition in broadcast news," Proc. IEEE 16th SignalProcessing and Communications Applications Conference (SIU), (2008).
- [6] Z. Huang, K. Chen, J. He, X. Bai, D. Karatzas, S. Lu and C. V. Jawahar, ICDAR2019 Competition on Scanned Receipt OCR and Information Extraction, ICDAR 2019, pp. 1516-1520.
- [7] Xinyu Zhou, Cong Yao, He Wen, Yuzhi Wang, Shuchang Zhou, Weiran He, Jiajun Liang, EAST: An Efficient and Accurate Scene Text Detector, CVPR 2017, pp. 5551-5560.
- [8] Tian, Zhi, Weilin Huang, Tong He, Pan He, and Yu Qiao, Detecting Text in Natural Image with Connectionist Text Proposal Network, ECCV (2016).
- [9] Hongwei Wang, Jia Wang, Jialin Wang, Miao Zhao, Weinan Zhang, Fuzheng Zhang, Xing Xie, Minyi Guo, GraphGAN: Graph Representation Learning with Generative Adversarial Nets, The 32nd AAAI Conference on Artificial Intelligence (AAAI 2018), arXiv:1711.08267.