Deep learning in computer vision

Mihai Teletin
Faculty of Mathematics and Computer Science
Babeș-Bolyai University
mihai.teletin@cs.ubbcluj.ro

Abstract

Computer vision is a domain that includes challenging tasks that might be tackled using deep learning. In the latest period, convolutional neural networks were successfully applied in order to address such a task. In this paper, we present two research directions that were carried out during our PhD research. Namely, we are going to present a lightweight solution for solving an image classification problem, fruit recognition. Afterward, we are going to demonstrate a deep learning based preprocessing method based on document detection using a convolutional neural network.

Keywords: Deep learning, Computer vision, Fruits recognition, Document deskewing

Domain: computer science
Section: Postdoc

An important subdiscipline of ML that has emerged from the study of artificial neural networks is deep learning. The field has recorded notable success in various fields while it managed to improve the state of the art performance.

Computer vision is a well known computer science field for which deep learning techniques were proposed and obtained state of the art performance. Problems like image classification [1] and object detection [4] received particular attention from deep learning researchers. Therefore, today, almost all production-ready solutions for such problems are deep learning based.

The first task discussed in this paper is an application based on lightweight deep neural networks for fruit recognition [8]. Such models were successfully applied in order to tackle these tasks and were showcased on the fruits-360 dataset [6]. It is composed of thousands of images of fruits. At the time we performed this experiment, 81 classes of fruits (e.g. apple green, orange, guava, dates, raspberry, etc) were available. Each dataset instance is labeled using one of these categories. Moreover, for benchmarking reasons, the authors also provided a distinct testing set. The dataset is well balanced thus it is more straightforward to train a machine learning algorithm since one doesn't have to use any kind of balancing method. On the other hand, such a dataset has disadvantages because it does not express reality. One can argue that in real world, the model will commonly deal with some fruits (e.g. apples and pears) than others (e.g. dates and maracuja).

The experimented models were inspired by MobileNetV2 [7] and ShuffleNetV2 [5] approaches. Two perspectives were taken into consideration when developing the model: time and classification efficiency.

We ran the experiments described in the original paper [8] and reported the results in Table 1. 95% confidence intervals are used and the best performance is highlighted. If we compare the models trained from scratch we observe that ShuffleNet has managed to generalize better. This is to be expected since ShuffleNet is a two times simpler model. These results empirically demonstrated our assumption: on this dataset lightweight models are generalizing better. On the other hand, when using transfer learning we initialize our MobileNet with ImageNet pretrained weight. Results proved that this process helps to improve the performance. Moreover, we observe that simple augmentations (random flips) made the test accuracy more robust since the confidence intervals were shrunk. Thus, the best performing model was the MobileNet V2 based version that used ImageNet initialization and data augmentation.

Model backbone	Transfer learning	Augmentations	Test accuracy	Best Test accuracy	Worst Test accuracy
MobileNet V2	No	No	$97.3\% \pm 0.3$	98.0%	96.4%
MobileNet V2	No	Yes	$98.0\% \pm 0.2$	98.5%	97.4%
MobileNet V2	ImageNet	No	$98.6\% \pm 0.2$	98.9%	97.9%
MobileNet V2	ImageNet	Yes	$98.7\% \pm 0.1$	99.1%	98.2%
ShuffleNet V2	No	No	$97.6\% \pm 0.3$	98.2%	96.9%
ShuffleNet V2	No	Yes	$98.4\% \pm 0.1$	98.8%	98.1%

Table 1: Results of various settings. 95% CIs are provided [8].

The second task that we managed to tackle using convolutional neural networks based approaches is the preprocessing step for optical character recognition. Basically, given a picture of a cash receipt, we proposed a neural network to detect 4 key points to be used by a projective transformation [2]. Thus, in the projected space, the background was removed and the text was horizontally aligned. An important particularity of the proposed convolutional neural network is the use of an angular loss, a term proposed to take into consideration the quality of the resulted projection [2]. Since the focus was to develop a lightweight approach, we compared various experimental settings that were based on MobileNet backbones [3].

The dataset used in our experiment is described in the original paper [2]. It comprised a collection of photographies representing various types of cash receipts that were collected from different sources. All the images are of the same resolution, 1920x1080 but for decreasing the complexity of the model, all the images (and the corresponding labels) were resized to 480x270.

Model	MAE	Angular error	Hough
MobileNet $\lambda = 0$	3.66 ± 0.07	4.87 ± 0.07	1.04 ± 0.01
MobileNet $\lambda = 1$	3.37 ± 0.04	4.05 ± 0.12	0.96 ± 0.01
MobileNet $\lambda = 5$	3.38 ± 0.05	3.22 ± 0.14	0.88 ± 0.01

Table 2: Results obtained by the deskew approach [2].

The obtained model was tested against a collection of images that consisted of 700 images. The cash receipts come from different providers than those found in the training set and were designed to be very difficult. For each sample, we detect 4 key points and report the mean absolute error, the angular error, and the absolute value of the skew angle of the resulted

projection. We performed three experiments: the first version was using MSE only as loss while the other two versions were enhancing it using the angular loss function [2]. The obtained results are depicted in Table 2. The experiments were repeated 10 times for each setting for providing the 95% confidence intervals.

We presented a fruit classification system based on convolutional neural networks. We demonstrated good results on an open source dataset. Afterward, we presented a new preprocessing technique effective for making images more accessible to OCR algorithms. It was based on two steps: document detection and deskewing.

Bibliography

[1] F. Chollet, "Xception: Deep Learning with Depthwise SeparableConvolutions,"CoRR, vol. abs/1610.02357, 2016. [Online]. Available:http://arxiv.org/abs/1610.02357

[2] Lorand Dobai and Mihai Teletin. A document detection technique using convolutional neural networks for optical character recognition systems. European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, Bruges, Belgium, ESANN, pp. 547–552, 2019.

[3] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, WeijunWang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXivpreprint arXiv:1704.04861, 2017

[4] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu,and A. C. Berg, "Ssd: Single shot multibox detector," inEuropeanconference on computer vision. Springer, 2016, pp. 21–37.

[5] Ningning Ma, Xiangyu Zhang, Hai-Tao Zheng, and Jian Sun. Shufflenetv2: Practical guidelines for efficient cnn architecture design. arXiv preprintarXiv:1807.11164, 2018

[6] Horea Mureşan and Mihai Oltean. Fruit recognition from images using deep learning. Acta Universitatis Sapientiae, Informatica, 10(1):26–42, 2018.

[7] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and pattern recognition, pages 4510–4520, 2018

[8] Mihai Teletin and Lorand Dobai. Lightweight models for fruits recognition. IEEE13th International Symposium on Applied Computational Intelligence and Informatics, SACI 2019, Timisoara, Romania, pp. 69-74, 2019.