Multi-task Learning for Low-Resolution License Plate Recognition

Gabriel Resende Gonçalves[1]([✉]), Matheus Alves Diniz[1], Rayson Laroca[2], David Menotti[2], and William Robson Schwartz[1]

[1] Smart Sense Laboratory, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil
{gabrielrg,matheusad,william}@dcc.ufmg.br
[2] Department of Informatics, Universidade Federal do Paraná, Curitiba, Paraná, Brazil
{rblsantos,menotti}@inf.ufpr.br

Abstract. License plate recognition is an important task applied to a myriad of important scenarios. Even though there are several methods for performing license plate recognition, our approach is designed to work not only on high resolution license plates but also when the license plate characters are not recognizable by humans. Early approaches divided the task into several subtasks that are executed in sequence. However, since each task has its own accuracy, the errors of each are propagated to the next step. This is critical in the last two steps of the pipeline known as segmentation and recognition of the characters. Thus, we employ a technique to perform these two steps at once. The approach is based on a multi-task network where each task represents the recognition of an entire license plate character. We do not address the license plate detection problem in this paper. We also propose the use of a so called generative model for data augmentation of low-resolution images simulating images as if they were acquired farther away from where they actually are. We are able to achieve very promising results with improvements of more than 30% points of accuracy on images with multiple resolutions and a character recognition accuracy on low-resolution images higher than 87%.

Keywords: Multi-task learning · Low-resolution · Deep learning · CNN

1 Introduction

Nowadays, recognize vehicles using their license plate is a well-known task and it is performed by many companiengs from different segments. This task is called *Automatic License Plate Recognition (ALPR)* and it was widely explored by researchers in the last two decades. Hence, there are multiple works proposing techniques to execute this task and improve its results [3,15].
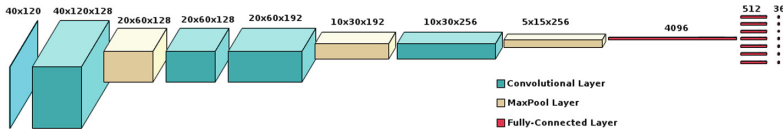
Fig. 1. Architecture of the proposed multi-task CNN to recognize license plate images.

Early approaches divide the license plate recognition into multiple subtasks and execute them in sequence [2]. These subtasks normally are (i) license plate detection; (ii) character segmentation; and (iii) optical character recognition (OCR). According to Gonçalves et al. [3], this has an important drawback since all error are propagated to the next step through the entire ALPR pipeline. Therefore, these approaches might have a large error rate at the end even if each subtask has good performance when evaluated separately. Moreover, cascading multiple approaches can be very time-consuming which is not desirable for real scenarios systems. Therefore, current approaches commonly try to perform ALPR without explicit executing all aforementioned subtasks [1,3,10].

Despite researches focus on proposing an end-to-end ALPR system [3,10,11, 16], there are many papers in the literature proposing techniques to solve only a few tasks of all steps described earlier. In this paper, we address two steps of the license plate recognition that we consider the most important and sensitive in terms of accuracy among all five: segmentation and recognition of the license plate characters. The former is usually approached by the researchers using algorithms that are not learning-based such pixel counting [19], mathematical morphology [13] and template matching [17]. The latter is a step known as optical character recognition (OCR) and is a wide-explored topic on the computer vision research community. Thus, many techniques provide accurate recognition rates on multiples challenging datasets [15,22].

In this work, we propose to perform license plate recognition in low-resolution images, focusing on segmenting and recognizing all characters holistically, avoiding the need to explicit segment the license plates (which is very hard on low-resolution images). Our experiments are carried out using a dataset with Brazilian license plates. Nonetheless, the approach can be further fine-tuned to work with any license plate standards. Our main concerns are to avoid the need to have high-resolution images and to avoid performing multiple tasks in sequence. We focus on a system to handle low-resolution images to reduce the cost of ALPR systems and improve its feasibility on companies or government departments that do not have a large budget to invest on high quality cameras. In addition, this work might be employed to forensics sciences, which have to handle poor quality images captured from a crime scene.

Since many works have provided very promising results in computer vision problems using deep learning techniques [7,8,10] and multi-task learning [12,14, 23,24], we decided to employ a deep convolutional network with multiple tasks, in which, each task of our network represents the recognition of one character. This idea was already proved to work before [1,20]. In this case, the character

is not explicitly located using our approach as the network only outputs the predicted characters. In addition, since deep learning networks require a large amount of data to learn, we employ two data augmentation techniques to increase the number of training samples, where we are able to train our network using only $2,520$ original license plate samples (later increased to $1,200,000$ samples with the data augmentation). Finally, we also design a network to generate new low-resolution samples to train our proposed multi-task network. There are two main contributions in this work: (i) a multi-task CNN model to segment and recognize low-resolution license plates characters; (ii) a deep generative network trained to create synthetic low-resolution images.

We evaluate our CNN model on two sets containing high-resolution and low-resolution images. Both sets were sampled from the *SSIG-ALPR* dataset [3]. The multi-task network is able to recognize 83.3% and 40.3% of the high and low-resolution sets, respectively. These results represent an improvement of 34.9 and 39.4 percentage points when compared to the best baselines. Moreover, our generative network brought an improvement of 4.9 p.p. in the experiments with the low-resolution images.

2   Proposed Approach

In this section, we detail our techniques. First, we present the architecture of the network used to simultaneously perform segmentation and recognition on low-resolution license plate images. Then, we describe the data augmentation techniques specially designed for low-resolution images.

Our approach consists on a deep convolutional network that receives a detected license plate and outputs the characters predicted considering the multiple tasks. Multi-task networks hypothesize that it is possible to improve the robustness of the network by learning a joint representation that is useful to describe more than one task on the same image [23]. In our case, our hypothesis is that the segmentation of a license plate character is dependent from the segmentation of the remaining characters since most characters are located sequentially in the same line. Moreover, the OCR also depends on how accurate the character is segmented. Hence, we train our deep network to segment and recognize all license plate characters at once to achieve better results than using two separated techniques (i.e., a for segmentation and a for OCR). This removes the need for cascading two approaches and bypasses the aforementioned error propagation. Even though this multi-task technique was already employed in previous approaches in literature [1,3], they do not focus on recognizing low-resolution license plate images.

It is common sense that deep learning techniques demand a great amount of training samples to learn discriminative and representative features to achieve promising results. Thereby, in addition to the proposed CNN, we also design a technique to permute characters aiming at performing data augmentation. Furthermore, since our focus is to increase the robustness of the network to handle low-resolution images, we need to train it with such images. However,

Table 1. Hyperparameters of the CNN trainable layer (upper table) and non-trainable (lower table) layers.

Layer	# Filters/Neurons	Filter size
Conv1	128	{5, 5}
Conv2	128	{3, 3}
Conv3	192	{3, 3}
Conv4	256	{3, 3}
Shared FC	512	–
Non-shared FCs	512	–
Output layers	36	–
Layer	Size	Stride
Input layer	{120, 40}	–
MaxPool layers (1, 2 and 3)	{2, 2}	{2, 2}

instead of only down-sampling the image with standard pixel-based techniques, we create a generative network that creates synthetic license plate images as if they were acquired farther away from where they actually were.

2.1 Multi-task CNN

As stated earlier, we propose the use of a convolutional neural network with multiple tasks to simultaneously perform segmentation and recognition of license plate characters. Our architecture is designed to implicitly perform the segmentation whereas the recognition are explicitly returned, one character per task. The network is designed to perform the character segmentation in the shared convolutional layers and recognize them on each task afterwards. The architecture is composed by five weighted shared layers: four convolutional layers and one fully-connected layer. There are also three shared max-pooling layers that do not contain any weights to train. Moreover, each task contains additional two fully-connected layers. We also placed dropout layers in between fully-connected layers to prevent overfitting. The input layer receives images of size 120×40 pixels and each output layers contains 36 neurons representing each one of the possible characters. This architecture is illustrated in Fig. 1.

The hyperparameters of the network are described in Table 1. Note that the non-shared layers are replicated for the each task, therefore, we have seven tasks with two layers each.

2.2 Data Augmentation

To improve the network robustness, we employed two techniques to increase the number of samples to train the network. The first technique is proposed by Gonçalves et al. [3] and is responsible to generate new artificial images with
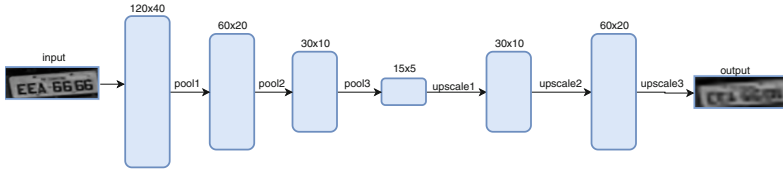
Fig. 2. Architecture of the DGM. The first and last layers contain a single filter. All other layers contains 64 filters.

license plate that were not initially on the dataset. The strategy consists on modifying the license plate images by changing (or permuting) the order of its characters. The authors claim that it helps the network to create an association of the character position with the correspondent task. This strategy was also successfully used in Laroca et al. [9].

Deep Generative Model. The second technique is a Deep Generative Model used to create new low-resolution license plates and is illustrated in Fig. 2. This model (hereinafter called DGM) is trained using the original images that we already had. Instead of using Generative Adversarial Networks [6], we created a network very similar to a variational autoencoder [21]. The goal is to generate low-resolution images from high-resolution ones. We could simply downscale the images in the dataset, but we believe that this does not emulate the actual behavior of low-resolution license plate that contains noises arising usually from long distance captures or by low-resolution cameras. Our model contains six convolution layers where the first three are followed by a max pooling layer and the last three are followed by a upsampling layer, similar to a generic convolutional autoencoder. The network is trained with pairs of images of the same vehicle license plate: one high-resolution image captured close to the camera used as input and a low-resolution image captured far from the camera used as output.

We performed the model training as follows. We recorded a video with multiple vehicles and track them throughout the video. Then, we manually chose a frame where the license plate is recognizable by a human and a frame where the license plate is not recognizable by a human. Afterwards, we had multiple pairs of high-resolution and low-resolution images from the same license plate, which have been use to train the network.

3   Experimental Results

In this section, we present the experiments carried out to evaluate the approach to segment and recognize license plate characters simultaneously. First, in Sect. 3.1, we present the datasets and baselines and then we describe and discuss the results achieved (Sect. 3.2).

Fig. 3. Training sample on top-left; validation sample on top-right; low-resolution and high-resolutions testing samples on bottom-left and bottom-right, respectively.

3.1   Evaluation Protocol

We created an evaluation protocol to measure the effectiveness of our approach. The protocol establishes three datasets, one for each task: training, testing and validation; and also a comparison scheme with four baselines. The images from the first and second dataset were sampled from the *SSIG-ALPR* dataset proposed in Gonçalves et al. [3].

The *first dataset* is used for training. It was originally composed from 2,520 images. However, we applied two data augmentation algorithms described in the previous section to increase the number of samples. We generate 800,000 images using the permutation technique and created other 400,000 using the DGM. Therefore, we presented 1,200,000 samples of license plate images to learn the model on the proposed network architecture. The license plate images have average size of 94×35 pixels. The images in the *second dataset* were used to test our proposed approach and are divided into two partitions: high-resolution and low-resolution. The high-resolution one contains 2,360 license plate images having average size of 104×39 pixels and the low-resolution one contains 820 images with average size of 49×17 pixels. The images of the first and second datasets were acquired in Brazil on two different places of the UFMG campus. The *third dataset*, called *SSIG-SegPlate*, was proposed by Gonçalves et al. [5] and contains 2,000 images of multiple resolutions and was only used for validation purposes. Figure 3 shows some examples from the three described datasets.

We compare our approach with four other techniques used as baselines. The first baseline contains two deep convolutional networks in cascade, in which the first was trained to segment the license plate characters and the second one to recognize them. We train these networks using our train set. We decided to utilize this baseline to demonstrate our hypothesis that training a single CNN to perform both tasks might avoid (and reduce) the error propagation previously mentioned. The second baseline is the approach proposed by Silva and Jung [18] that contains an end-to-end vehicle identification pipeline with two deep network; currently, this approach achieves state-of-the-art results in the Brazilian license plates recognition. Since we are assuming the license plate is already detected, we utilized only their recognition network. The third baseline is a hand-crafted
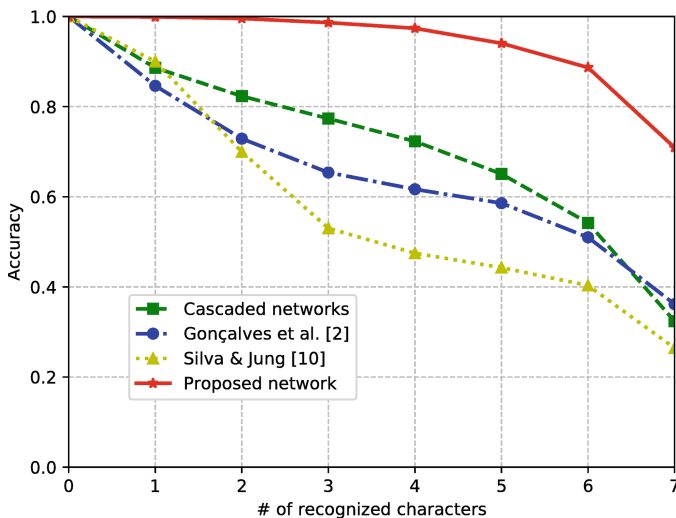
Fig. 4. Accuracy of all baselines as a function of the number of characters in the license plate that were correctly predicted.

Table 2. Recognition rates achieved by the proposed approach compared to baselines.

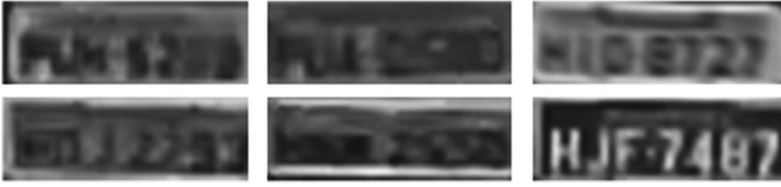| Approach	High-resolution	Low-resolution
Cascade networks	43.3%	0.9%
Gonçalves et al. [4]	48.3%	0.1%
Silva and Jung [18]	35.4%	0.4%
Proposed network	**83.2%**	**35.4%**

approach proposed by Gonçalves et al. [4] which employs a HOG-SVM classifier. Finally, our fourth baseline is the free version of the system called OpenALPR[1].

3.2   Results and Discussion

We perform two experiments to validate the proposed approach. The first evaluates our technique with the baselines in two different test sets, the partition with high-resolution images and the partition with low-resolution images. In this case, the samples generated by DGM were not considered. Then, the second experiment evaluates the influence of adding the examples of our generative network (DGM). We trained our multi-task network with these new samples and use it to predict the images from these two partitions of the test set once again.

The results of the first experiments are showed in Table 2. According to the results, the proposed multi-task network outperformed all baselines. If we consider only samples with low-resolution, the approaches proposed by Gonçalves

---

[1] Available at http://www.openalpr.com.

Fig. 5. Low-resolution license plates correctly recognized.

et al. [4] and by Silva and Jung [18] were able to recognize an insignificant number of license plates. This is expected since these approaches were not designed to handle low-resolution images and they also are dependent of the character segmentation, which might be a very difficult task on low-resolution characters. On the other hand, our approach does not have this problem because our network does not have to perform the segmentation of the characters explicitly. The baseline composed by two cascade models was able to achieve 43.3% on the high-resolution set and 0.9% on the low-resolution set, the latter result is very low, as the ones achieved by the other baselines. Our approach, on the other hand, was able to outperform the best baselines by 34.9 and 34.5 percentage points in high- and low-resolution images, demonstrating our hypothesis regarding the increasing of error when two steps are performed (i.e., character segmentation followed by character recognition) instead of a single step for both, as executed by our approach.

We also compared our approach with the commercial system OpenALPR. However, we were not able to use only the segmentation and the recognition steps of this technique. Therefore, we only evaluate the images in which the license plate was correctly detected by the OpenALPR system. The system was able to recognize 86.3% of high-resolution license plates with a full pipeline. Analyzing only the samples where the license plates were detected by OpenALPR, our method achieved 87.1% of accuracy on the high-resolution set, outperforming, therefore, the OpenALPR. Regarding the low-resolution license plates, the OpenALPR was able to detect only six of 820 license plates and was not capable of recognizing any of them. On the other hand, our approach was able to recognize correctly three of the six low-resolution license plates detected by the OpenALPR system, giving an edge for our approach on this set.

We also perform an evaluation of the number of characters that each method was able to recognize, combining the samples with high and low resolutions of the second dataset. According to the results showed in Fig. 4, our approach is able to recognize the entire license plate in 70.4% of the cases and predicts all but one character in 88.7% of the times. These results show an improvement of the best baselines in 34.2 (i.e., 70.4–36.2%) and 34.5 (i.e., 88.7–54.2%) percentage points in the case of zero and one misclassified characters, respectively.

The last experiment evaluates the influence of the DGM (Sect. 2.2). We trained our network using the 800,000 achieved by the character permutation and with the addition of 400,000 images from the DGM, resulting in a total of

Table 3. Accuracies achieved with and without the addition of the samples from DGM.

Approach	Accuracy	
	High-resolution	Low-resolution
Without samples from DGM	83.2%	35.4%
With samples from DGM	83.2%	40.3%

1,200,000 samples. According to the results showed in Table 3, the examples generated by the DGM were able to improve the robustness of the network on the low-resolution set in 4.9 percentage points, an improvement of 39.4 percentage points when compared to the best baseline). One also can note that the accuracy with the high-resolution was maintained, which mean that the samples generated by the DGM only improved the network generalization without compromising the results on the high-resolution test set. Figure 5 shows some low-resolution license plates correctly recognized by our approach.

4   Conclusions and Future Directions

In this paper, we introduced an approach to recognize license plates in low-resolution images. We employed a multi-task CNN to perform the segmentation and recognition of the license plate images simultaneously. We also designed a data augmentation techniques to increase the number of samples available to train our network. The DGM model was designed to simulate a license plate as if it was recorded from a long distance.

Our results demonstrated that our approach was able to recognize 83.2% and 40.3% of high- and low-resolution Brazilian license plates, respectively, being the best results on both sets among the baselines. Note that 40.3% of accuracy for license plates with seven characters stands for an accuracy of approximately 87.83% of character recognition accuracy (i.e., $0.8783^7 \approx 40.3\%$), which is a promising result for characters that are in low-resolution. The experiments also demonstrated that the use of the DGM improves our network, increasing the accuracy in the low-resolution images by 4.9 percentage points.

As future works, we intend to evaluate our network with other license plate standards and also investigate how to access the confidence of the recognition for each character in the license plate.

References

1. Dong, M., He, D., Luo, C., Liu, D., Zeng, W.: A CNN-based approach for automatic license plate recognition in the wild. In: BMVC (2017)
2. Du, S., Ibrahim, M., Shehata, M., Badawy, W.: Automatic license plate recognition (ALPR): A state-of-the-art review. TCSVT (2013)
3. Gonçalves, G., Diniz, M.A., Laroca, R., Menotti, D., Schwartz, W.R.: Real-time automatic license plate recognition through deep multi-task networks. In: SIB-GRAPI. IEEE (2018)
4. Gonçalves, G.R., Menotti, D., Schwartz, W.R.: License plate recognition based on temporal redundancy. In: ITSC (2016)
5. Gonçalves, G.R., da Silva, S.P.G., Menotti, D., Schwartz, W.R.: Benchmark for license plate character segmentation. JEI **25**, 053034 (2016)
6. Goodfellow, I., et al.: Generative adversarial nets. In: NIPS (2014)
7. Hu, C., Bai, X., Qi, L., Chen, P., Xue, G., Mei, L.: Vehicle color recognition with spatial pyramid deep learning. T-ITS **16**, 2925–2934 (2015)
8. Krizhevsky, A., Sutskever, I., Hinton, G.: Imagenet classification with deep convolutional neural networks. In: NIPS (2012)
9. Laroca, R., Barroso, V., Diniz, M.A., Gonçalves, G.R., Schwartz, W.R., Menotti, D.: Convolutional neural networks for automatic meter reading. JEI **28**, 013023 (2019)
10. Laroca, R., et al.: A robust real-time automatic license plate recognition based on the YOLO detector. In: IJCNN (2018)
11. Li, H., Wang, P., Shen, C.: Towards end-to-end car license plates detection and recognition with deep neural networks. arXiv preprint arXiv:1709.08828 (2017)
12. Moeskops, P., et al.: Deep learning for multi-task medical image segmentation in multiple modalities. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 478–486. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_55
13. Nomura, S., Yamanaka, K., Shiose, T., Kawakami, H., Katai, O.: Morphological preprocessing method to thresholding degraded word images. PRL **30**, 729–744 (2009)
14. Ranjan, R., Patel, V.M., Chellappa, R.: Hyperface: a deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. PAMI **41**, 121–135 (2017)
15. Rao, Y.: Automatic vehicle recognition in multiple cameras for videosurveillance. Vis. Comput. **31**, 271–280 (2015)
16. Rizvi, S.T.H., Patti, D., Björklund, T., Cabodi, G., Francini, G.: Deep classifiers-based license plate detection, localization and recognition on gpu-powered mobile platform. Fut. Internet **9**, 66 (2017)
17. Shuang-Tong, T., Wen-Ju, L.: Number and letter character recognition of vehicle license plate based on edge Hausdorff distance. In: PDCAT (2005)
18. Silva, S.M., Jung, C.R.: Real-time Brazilian license plate detection and recognition using deep convolutional neural networks. In: SIBGRAPI. IEEE (2017)
19. Soumya, K.R., Babu, A., Therattil, L.: License plate detection and character recognition using contour analysis. IJATCA (2014)
20. Špaňhel, J., Sochor, J., Juránek, R., Herout, A., Maršík, L., Zemčík, P.: Holistic recognition of low quality license plates by CNN using track annotated data. In: AVSS (2017)

21. Walker, J., Doersch, C., Gupta, A., Hebert, M.: An uncertain future: forecasting from static images using variational autoencoders. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9911, pp. 835–851. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46478-7_51
22. Zhang, X., Zhao, J., LeCun, Y.: Character-level convolutional networks for text classification. In: NIPS (2015)
23. Zhang, Y., Yang, Q.: A survey on multi-task learning. arXiv preprint arXiv:1707.08114 (2017)
24. Zhang, Z., Luo, P., Loy, C.C., Tang, X.: Facial landmark detection by deep multi-task learning. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8694, pp. 94–108. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10599-4_7