CONVOLUTIONAL NEURAL NETWORKS FOR LICENSE PLATE DETECTION IN IMAGES

Francisco Delmar Kurpiel, Rodrigo Minetto, Bogdan Tomoyuki Nassu
Federal University of Technology — Parana, Brazil
francisco@kurpiel.eng.br, rminetto@utfpr.edu.br, btnassu@utfpr.edu.br

Objectives

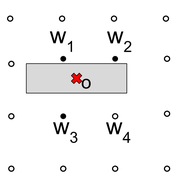
According to the R-CNN paper, precise object localization within the sliding-window paradigm with large strides is an open technical challenge.

- Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). **Rich feature hierarchies for accurate object detection and semantic segmentation**. IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2014), pages 580–587.

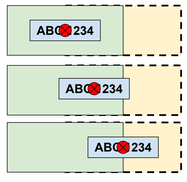
We propose a method for addressing this problem. The method estimates the location (center coordinates) of a single class of objects even when applied with a stride as high as 90px.

Proposed Method

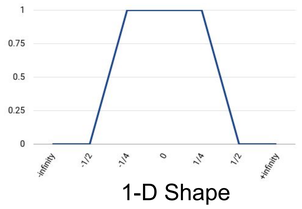
We propose a method for detecting the center of an object by processing the outputs from a convolutional neural network that is applied at multiple neighboring windows in an image, using large strides, with 50% overlap in each direction.



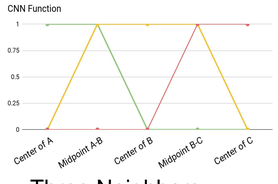
$$o_x = w_{1,x} + f(w_1, w_2)$$
$$o_y = w_{1,y} + f(w_1, w_3)$$
$$v_{w_1} = g\left(\frac{o_x - w_{1,x}}{w_{2,x} - w_{1,x}}, \frac{o_y - w_{1,y}}{w_{3,y} - w_{1,y}}\right)$$



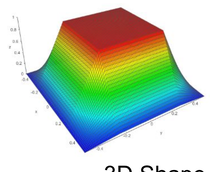
1-D View



1-D Shape



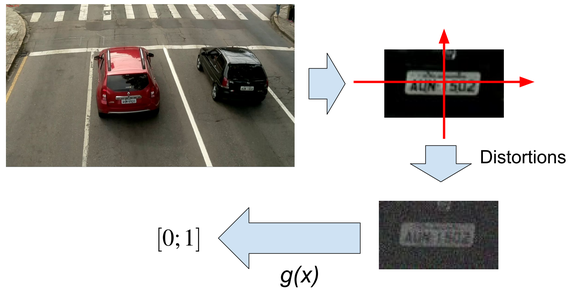
Three Neighbors



3D Shape

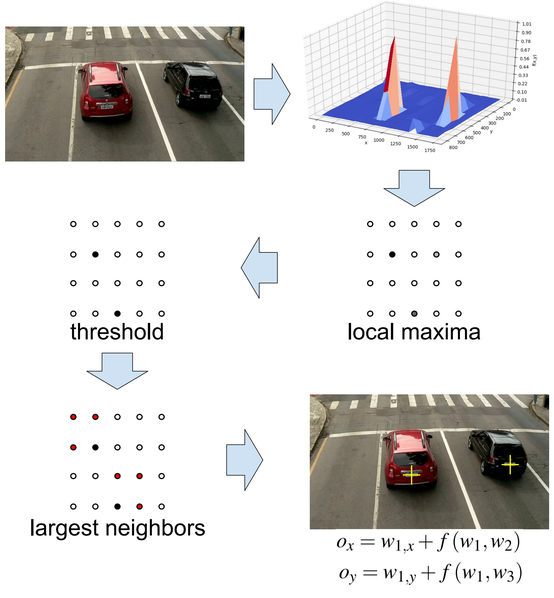
Training

A CNN is trained to perform regression of a function that maps an image to a scalar. This mapping is dependent on the offset of the center of a license plate, from the center of the window. Distortions are added to the images during training to increase the robustness of the model.



Inference

During inference the CNN is applied with stride equal to half the window size (60 X 90 on our tests), producing a very low resolution map. After this, the local maxima of all possible 2x2 regions are identified and a threshold (0.85 on our tests) is applied to them, identifying the approximate positions of all objects. The largest neighbors from each maximum are used in conjunction with the maxima to estimate the actual plate position.



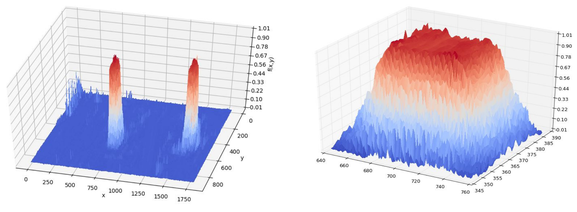
$$o_x = w_{1,x} + f(w_1, w_2)$$
$$o_y = w_{1,y} + f(w_1, w_3)$$

Results

We tested the method on a publicly available dataset for detecting license plates containing 1.829 full-HD images with 4.070 license plates. Our neural network was a simple 9-layer convolutional neural network:

ID	Input	Operation	DOF
M1	180 × 120 × 3	2 × CV 3 × 3, stride 3 × 3	56
M2	60 × 40 × 2	12 × CV 3 × 3, stride 3 × 3	228
M3	20 × 14 × 12	16 × CV 3 × 3	1,744
M4	20 × 14 × 16	16 × CV 3 × 3	2,320
M5	20 × 14 × 16	24 × CV 3 × 3, MP 2 × 2	3,480
M6	10 × 7 × 24	16 × CV 3 × 3	3,472
M7	10 × 7 × 16	2 × CV 3 × 3, MP 2 × 2	290
-	5 × 4 × 2	flatten	-
FC1	40	32 × FC	1,312
FC2	32	1 × FC LIN	33
Out	1		-
TOT			**12,935**

The trained neural network correctly models the desired function.



Our window size was 120 pixels wide by 180 pixels high, so the stride was 90 x 60. The CNN is applied 1.536 times on each 1080p image.

On a low-cost notebook with a Core i5 CPU and a NVidia GT-740M GPU the full frame processing takes 230ms.

230ms / Full HD Frame

The precision was evaluated on isolated frames against a method based on a descriptor+SVM approach [Luvizon et al, 2015]. Bounding boxes were inferred from the center by using the license plate's proportions and scaling according to the *y* position in the frame. We used the metrics from the PASCAL Visual Object Detection Challenge (intersection over union > 0.5). The values of precision, recall, F-measure and time are substantially better.

	Proposed	Luvizon et al
Precision	0.87	0.76
Recall	0.83	0.60
F-Measure	0.85	0.67
Time (s)	0.23	2.3

The precision is expected to increase substantially if we explore the correlation between consecutive frames (tracking).

Contributions

- A robust CNN-based license plate detector.
- A method using large strides for object detection tasks.
- A challenging dataset, freely available for research purposes.

Conclusions

- Our method correctly finds the center of object, even with large strides.
- It runs substantially faster and is more accurate than the reference method.
- The method is only applicable when there is only one kind of object and the entities have some distance.

Further Research

- Apply the same method to other kinds of objects besides license plates.
- Add the ability to operate on multiple objects.
- Remove restrictions regarding fixed window sizes and minimum distance between objects.
- Study the effect of tracking on the classification performance.