DRIVER DROWSINESS DETECTION
MINI REPORT 6

BASU VERMA
(142002007)

under the guidance of

Dr Satyajit Das

IIT PALAKKAD

INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD
PALAKKAD - 678557, KERALA

Contents

List of Figures

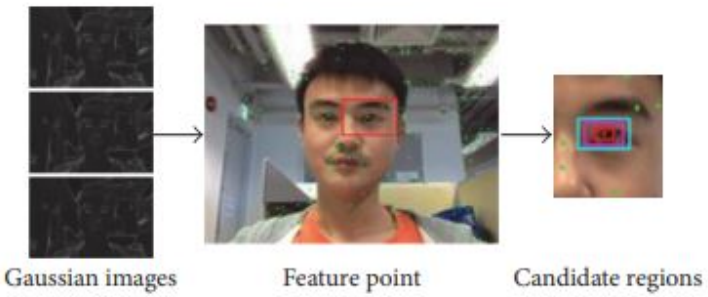
Part. 1

Literature review of Eye Detection using CNN

[1] In this paper, they found that the pupil and iris were darker than other parts of the eye. The locations of the local extreme points in the image are more likely to be the rough center positions of the eyes. To find these extreme points, they convolved the facial image with three Gaussian kernels of diferent variances to obtain Gaussian images G_n(x,y,σ n) = (1, 2, 3). Each pixel in G_2(x, y, σ2) was compared with its 5 × 5×3 neighborhood pixels in G_1(x, y, σ_1), G_2(x, y, σ_2)and G_3(x, y, σ_2) If the pixel (x,y) is the maximum or minimum in its neighborhood, its local gradient value Gr(x, y) was calculated as follows:

$$\begin{aligned} \mathrm{Gr}(x,y) &= \sqrt{(G_2(x+1,y,\sigma_2) - G_2(x-1,y,\sigma_2))^2 + (G_2(x,y+1,\sigma_2) - G_2(x,y-1,\sigma_2))^2}, \\ G_n(x,y,\sigma_n) &= \frac{1}{2\pi\sigma_n^2}\exp\left(-\frac{x^2+y^2}{2\sigma_n^2}\right) * I(x,y), \end{aligned} \tag{1}$$

where G_n(x,y,σ n) the convolution of the Gaussian kernel and the facial image. They selected the top N extreme points with the largest gradient value as the candidate feature points. They aim to ensure that the candidate regions can completely cover the eye region and make the number of candidate feature points as small as possible. Then, they generated three diferent sizes of candidate eye regions R_iC_m(i = 1, 2, 3) centered on each candidate feature point C_m(m = 1, 2, 3, . . .), which ensures that the generated candidate eye regions can completely cover the eye region as shown in figure.



1.0.1 CNN Architecture

In CNN Architecture, three sub-CNNs were built and each carries the same structure. In each sub-CNNs, the first layer was a convolutional layer with a kernel size of 5 × 5 pixels, two pixel strides, and one padding, and the convolution layer was followed by a maximum pooling layer with a window size of 3 × 3 and two pixel strides. The second layer was a convolutional layer with a kernel size of 3 × 3 pixels, one pixel stride, one padding, and no pooling layer. The third layer was similar to the first layer, except that the convolutional kernel size was 3 × 3 pixels. Trough three stages of convolution and pooling, in which the convolutional layer learned the edges, eye structure, and other basic features, the pooling layer helped the networks to be robust to details in the changes. Next, they used fully connected (FC) layers to combine a deeper knowledge and produce the fnal region label and confidence index of each candidate region. Finally, they choose the candidate region with the maximum index as the eye region and according the region label to classify the left or right eyes. They then used the coordinates of this region and restored it to the original facial image. All CNNs' weights were initialized based on ImageNet's weight values for fine tuning, which will help them to train the network with faster convergence and obtained good experimental results.

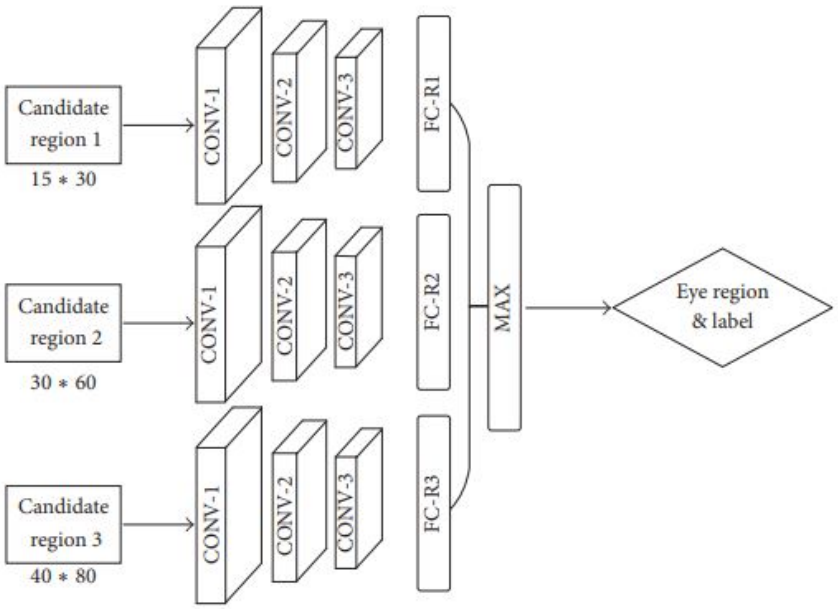


Fig. 1.1 CNN structure of above method.

References

[1] B. Li and H. Fu, "Real time eye detector with cascaded convolutional neural networks," *Applied Computational Intelligence and Soft Computing*, vol. 2018, p. 1439312, Apr 2018. [Online]. Available: https://doi.org/10.1155/2018/1439312