

## LABORATORY WORK 4 – PREDICTION OF VISUAL SALIENCY IN VIDEO SEQUENCES BY DEEP CNN

The visual saliency concept refers to the visual property that some objects around us are more perceived than others, they "capture our attention."

It is very important for a complex biological system (humans, animals) to be able to quickly detect potential prey, predators, or others in a world of visual information. Primates or other animals that filter visual information in order to focus their unconscious process of recognition of objects on a limited number of objects called "salient".

Many computer vision works are inspired by this visual property to focus their calculations on the salient areas. In the following figure, it is easy to see that for a saliency prediction in natural video an object in motion, here the two people running, attract human gaze.



The left figure shows the original frame. The right figure shows a "heat map" of the saliency values, called "saliency map" automatically calculated using an algorithm exploiting the properties of human visual system.

In this example, it is visually easy to conclude that the automatic extraction algorithm of salience shows very high performance. Nowadays, prediction of saliency in images and video by machine learning methods is an intensive research subject. Deep Convolutional Neural Networks (Deep CNNs) have proven to be a powerful mean for prediction of salient areas in stills. By adding residual motion information to the input of Deep CNNs, the architecture of the network ChaboNet developed in the LABRI to predict saliency in natural video is summarized in next figure.

The learning of Deep CNN parameters is frequently done with the technique of Stochastic Gradient Descent (SGD). Generally, for a simple supervised learning the SGD algorithm is as follows:

repeat until convergence

$$\begin{cases} V_{i+1} = \epsilon \cdot V_i - \gamma \cdot \alpha \cdot W_i - \alpha \cdot 1/N \cdot \sum \frac{\partial}{\partial W_i} f(W_i) \\ W_{i+1} = W_i + V_{i+1} \end{cases}$$

where  $W_i$  are network parameters of each convolutional layer at the iteration  $i$ ,  $W_i$  is randomly initialized from a gaussian distribution,  $\alpha$  is the learning rate,  $\epsilon$  is a momentum coefficient,  $\gamma$  is the weight decay the sum is taken over a batch of size  $N$  and  $f$  is the loss function to minimise.  $V_i$  presents the previous weight update that is initialized at zero. The main SGD problem is that it converges to a local optimum even if  $\alpha$  is constant. However, it is still the best learning algorithm when the training set is large.

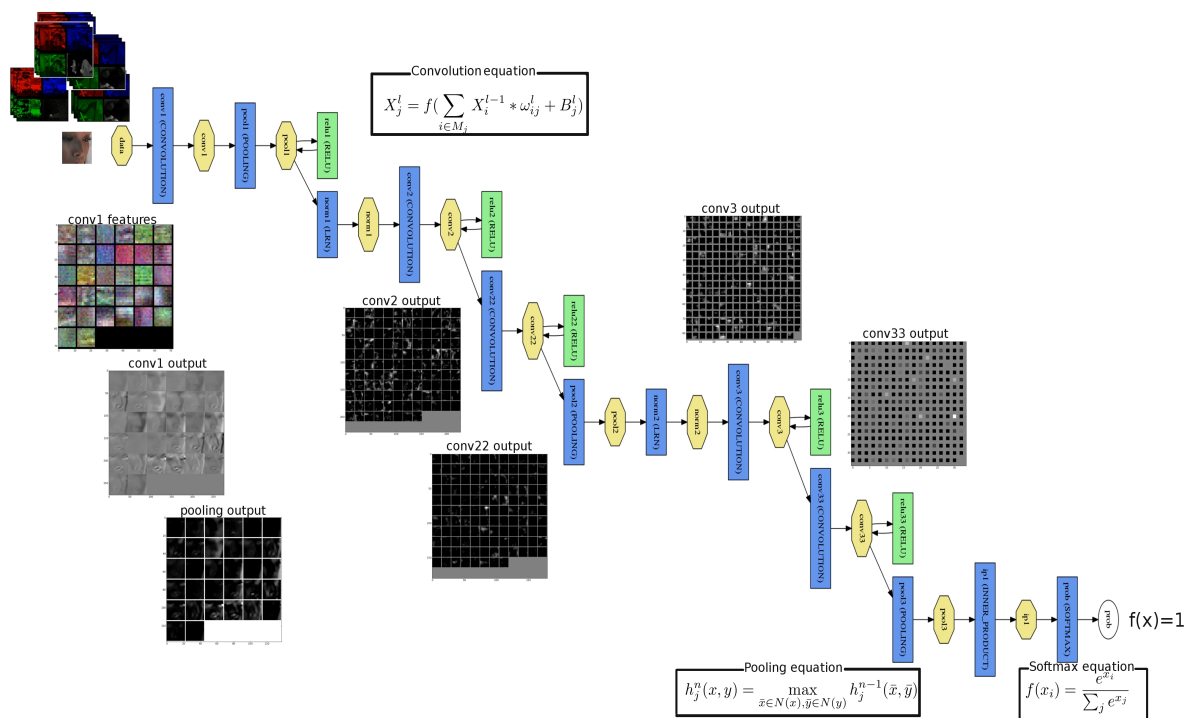


Figure : "ChaboNet" deep CNN to saliency prediction in natural video.

The model of "ChaboNet" network is already defined and trained. The training step needs an interest computational time. To simplify, you will just use it to start the prediction of saliency in available videos. The first point of interest: **how to use a trained model to create a dense saliency map**. It is visually easy to conclude that the automatic prediction of salience shows very high performance. However this is perhaps not necessarily the case for all the frames of the video.

The second interest point in this tutorial: **how to create a "heat map" of the saliency values**.

You will write the C++ code in order to visualize the obtained saliency.

## 1- RECOVERY OF THE DATA BASE AND THE CODE

Download the dataset and code to the following link : (here the predicted saliency map are already computed) and extract the archive in your working folder.

[http://dept-info.labri.fr/~schaabou/TD\\_2017\\_2018/lab4\\_etudiant.tar.gz](http://dept-info.labri.fr/~schaabou/TD_2017_2018/lab4_etudiant.tar.gz)

### Description of the archive Lab4:

```
├── Lab4
│   ├── Data
│   │   ├── video
│   │   ├── model
│   │   └── ChaboNet_Saliency
│   ├── lab4.cpp
│   ├── Classifier.hpp
│   ├── Classifier.cpp
│   └── classification.cpp
```

The folder "Data" is composed of three folder . First , the folder video contains two videos "test\_rgbframe" and "test\_tmpframe", and folder of patches that contains different testing patches. Second the folder "model", contains the pretrained model and the label. And finally, the ChaboNet\_Saliency will contain the results of your Laboratory work.

## 2- FIRST TESTS

During this project you will test the classification.cpp by insert the needed parameters. Here, we will just classify one patch.  
So, compile the code in a created build folder.  
Visualize the obtained results.

## 3- PREDICTION OF CHABONET SALIENCY MAP:

Classification.cpp give just a class for a input patch. Now, we want to generate a saliency map by browsing all patches of video' frame.

Open the lab4.cpp and finish the missed codes.

## 4- HEAT MAP

We have predicted the saliency. Now we want to visualize the saliency on the frame's video. As output, we will obtain a heat map for each frame.

Move to the code Lab4bis folder and create the heat map for the video.

## 5- REPORT

You are asked to write a report containing the following parts:

- Introduction / Presentation of results (some exemple of heat map) and Conclusion

Send your report, named MYNAME\_Lab4.tar.bz2 , to [souad.chaabouni@u-bordeaux.fr](mailto:souad.chaabouni@u-bordeaux.fr) in an archive that also contains the C++ code before midnight of Saturday, October 29, 2017.