Master ST Computer Science

Deep Learning in Computer Vision

University of Bordeaux
Faculty S&T
Academic year 2018 – 2019

Lab 5
(2 hours, evaluated)

Neural Networks
Implementing a Convolutional Neural Network with Keras for action recognition in videos

Introduction

In this lab, we will see how to build a convolutional neural network with the Keras framework in Python aimed at recognizing actions in videos. We will use the UCF101 dataset, that gathers Youtube videos in 101 categories. Some actions to recognize are: "Apply eye makeup", "Apply lipstick", "Playing guitar", "Playing violin", "Skate boarding", "Skiing", ...

As usual, we give an incomplete python program. Some instructions will be used as they are and you do not need to change them. We will comment some of them but we do not need to understand them all.

So first, you need to download the file lab5_skeleton.py available at:
http://dept-info.labri.fr/~mansenca/DLCV2018/lab5_skeleton.tar.bz2

You will need to activate the work environment with:
source /net/ens/DeepLearning/tensorflow/bin/activate

In order to classify actions, we will use a **siamese** convolutional neural network. This network will take both a slice of the video and the corresponding optical flow, and produce the probabilities to belong to the 101 action classes.
We will build this network progressively, one branch at a time, and study the contribution of each architectural change. Each branch has the same architecture:
Conv.→ MaxPooling → Conv. → MaxPooling → Conv. → MaxPooling → F.C. → F.C.

The Conv. Layers all have filters of size 3, zero padding and Relu as activation function. The number of filters is respectively 30, 60 and 80. The MaxPooling layers all have a size of 2. For the first Fully Connected layer we will use 500 neurons.

First extract frames and compute optical flow locally by running lab5_0.py

1. Convolutional Neural Network for action recognition: color branch

Program a first convolutional neural network to recognize the 101 actions using color information from videos.

1. Complete the file to define the model with the architecture defined above.
2. Load the provided pre-trained weights and apply your model on selected videos.
3. Train your network and measure its accuracy.

2. Convolutional Neural Network for action recognition: optical flow branch

Program a second convolutional neural network to recognize the 101 actions using optical flow from videos.

4. Complete the file to define the model with the architecture defined above.
5. Load the provided pre-trained weights and apply your model on selected videos.
6. Train your network and measure its accuracy.

3. Convolutional Neural Network for action recognition: siamese network

Program a first convolutional neural network to recognize the 101 actions using both color information and optical flow from videos. This siamese network will merge the output of the two branches [before their last F.C. layer] with a concatenation layer.

7. Complete the file to define the model with the final siamese architecture.
8. Load the provided pre-trained weights and apply your model on selected videos.
9. Train your network and measure its accuracy.

You must send an archive with your code and report.
The report should contain a the accuracy that you got with each network.
This archive must be named LASTNAME_firstname_Lab5_DLCV.tar.bz2 and send to : boris.mansencal@labri.fr with DLCV in the e-mail subject.