Human Activity Recognition Project Proposal

1. Problem to be solved

In this project, we aim to solve the problem of recognizing human activities from 2D images. We do this by dividing the problem into two consecutive parts. Given new unseen activity images, first we estimate the location of body joints using a trained regression neural network, then we use the network's predicted estimates as inputs for a second neural network that has been previously trained to recognize the main activity in the image. We will evaluate the performance of both network models on a publicly available dataset. Deliverables for this project will be a text report detailing the methods and data analysis techniques used, the code and tools used to analyze, model and test the data. This will be in addition to a slide deck to be used for brief project presentations in the future.

2. Clients interested in the solution

Activity Recognition is of interest to video surveillance applications while pose estimation is important for motion capture systems.

3. Dataset

We will use the MPII Human Pose Dataset, Version 1 available at: http://human-pose.mpi-inf.mpg.de/#download. The dataset consists of 25,000 images of about 40,000 people with annotated body joints. Each image in the dataset contains activities performed by one person or more. Activities are grouped into 20 categories with 410 classes or types. Each image is annotated with a single activity label that shows the main activity contained in the image.

In addition to the images and annotations, videos are provided for each image with one preceding and one following unannotated frame that can be used for additional model testing. Sizes of all images and annotations are 12.9 GBytes and 12.5 MBytes respectively.

4. Methods

A general overview of the project workflow is shown in figure 1. Following the traditional data science process, we start with acquiring and processing the data into an appropriate format for the main project tasks. We will use Python 3 packages to implement all project steps. The provided dataset annotations are stored in Matlab objects and structs which can be converted into Python dictionaries easily using the scipy library. Since our main data are 2D images, convolutional network architectures are appropriate to implement our learning models.

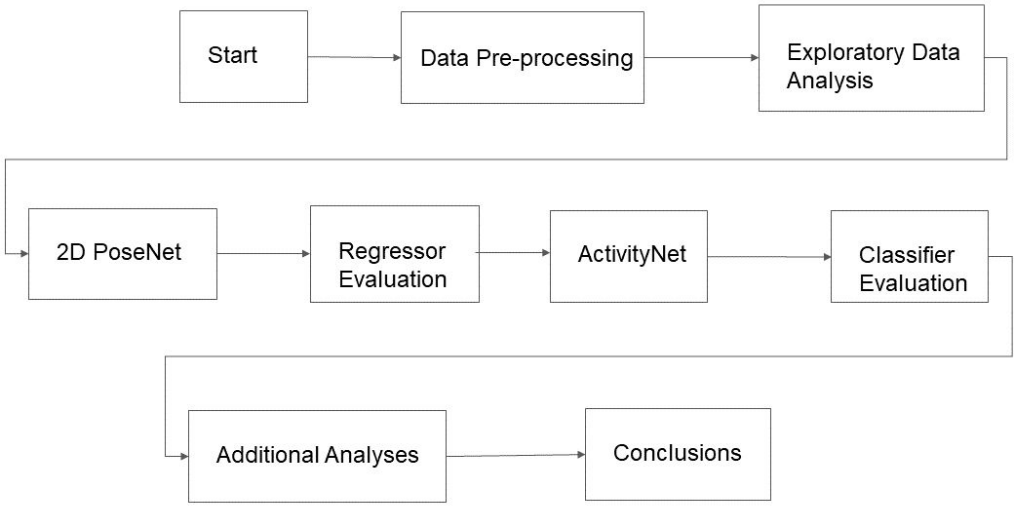

Figure 1: Project Steps and Milestones

References:

Mykhaylo Andriluka, Leonid Pishchulin, Peter Gehler, Schiele, Bernt, *2D Human Pose Estimation: New Benchmark and State of the Art Analysis*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014