German University in Cairo
Faculty of Media Engineering and Technology
DMET1067 - Deep Learning in Computer Vision
Spring Semester 2022

Course Project

Milestone 3

Announced 25.05.2022. Deadline 02.06.2022. Total Points 60

Project Topic:

In this project we will address the problem of action recognition from Egyptian movie scenes. It is an image classification task, where video frames are classified based on people actions. For example: Hand shaking, two-people talking, group of people talking, walking, walking in groups, running, driving, dining individually, dining with a group, dancing, etc. We shall consider old and recent movies, colored and gray level videos. The classification is on the frame level.

The project will go through few milestones. For each milestone, a specific list of deliverables on a specific deadline will be required. The students will be guided to form workgroups, to prepare training and testing datasets, to construct a deep convolutional neural network for image classification, to apply pre- and post-processing steps, and to display, evaluate, and interpret the results.

Driving

Group Dining -

Prof. Dr Mohammed A.-M. Salem
Mohammed.salem@guc.edu.eg

Milestone 3: Building a CNN for Action Recognition

Announced 25.05.2022. Deadline 02.06.2022. Total Points 60

Milestone 3 Objectives:

The objective of this milestone is to build, train and test a convolutional neural network for image classification. The model should be used with the help of the created dataset in Milestone 2 to classify an Egyptian movie frame as one of 10 pre-determined action classes.

1- Select 10 classes from the shared dataset in the link: ??????????????????????
2- Use the data in these classes for training/validation and testing your model. Use the rule 70/20/10 for the split ratios.
3- Check the data before the training and do any necessary cleaning and standardization.
4- Chose the model hyper-parameters carefully and build the model using any preferable libraries in Python or MATLAB.
5- Train the model until it reaches an acceptable classification accuracy. Then test the model with the testing data. No need to repeat the experiment several times in this milestone.
6- Report the model architecture and the initial testing results.

Model Architecture:

1- Build a CNN with 3 convolutional layers, each of which uses ReLU as activation function, and followed by a 2x2 maxpooling filters. Use the default stride for the maxpooling filters.
2- The filter sizes are 3x3, 5x5, and 7x7 for the three successive convolution layers, respectively. Use planner convolution only.
3- The convolution blocks are followed by a flattening operation.
4- Use the flat vector as an input to a fully connected layers with only one hidden layer. Use the Sigmoid as the activation function for this layer.
5- Use the Softmax function for the output layer.
6- All other hyper-parameters than what have been mentioned here are left to be determined by the developing team.

Delivery:

All deliverables must be compressed together in one (*.zip) file. Name the file as ***T***_M3.zip, where ***T*** is the team 4-characters short name. Submit your work to dlcv.guc@gmail.com.

1- Submit a Jupiter text file or .m for your well-commented code. Specify in clear text the selected hyper-parameters and any used libraries.
2- A powerpoint slides for the visualization of the created model architecture annotated with the dimensions of the layers and in/out data. Present the selected classes with some samples of the inputs and the obtained results in the same file. Add one slide for the model hyper-parameters, number of training iterations and any additional notes.
3- A textfile (*.txt) with the team information including the names and contacts of the team members.

Grading:

Any delayed deliver will be subject to 50% deduction.

Milestone 2: Building Training and Testing Dataset

Announced 17.05.2022. Deadline 24.05.2022. Total Points 10

Milestone 2 Objectives:

The main objective of this milestone is to create training and testing dataset. This objective will be reached by the collaboration of all groups. Each team will be responsible to create part of the dataset and together we will have good large and challenging dataset.

1- Each team will address as many classes as the number of members.
2- The team should collect 1000 frame per class. That is for a team with 4 members, they should collect 4000 frames for 4 classes. For a team with 3 members, they should collect 3000 frames for 3 classes.
3- Each class should be addressed by two teams at least.
4- Each class should be addressed by five teams at most.
5- To select classes, use the form on the following link:
https://docs.google.com/spreadsheets/d/1y_GfxbdUlEMfP8PFvIPLIbKY9LLzTVJaL4XUuGbjrMk/edit?usp=sharing
The addressed classes: { Action with animals, Applying Makeup, Baby crawling, Blood scenes, Blowing Candles, Cooking, Crying, Dancing, Driving, Farming, Group dining, Group fighting, Group talking, Gun man, Horse Riding, Hula Hoop, Jumping Rope, Knitting, Laughing, Pair talking, Playing Football, Playing Guitar, Playing Piano, Running, Shaking Hands, Single dining, Smoking, Swimming, Two-people fighting, Walking}
6- To ensure diversity, an Egyptian movie should be processed by one team only. Therefore, teams are asked to register the movies that they are working on, on a global excel sheet at this link:
https://docs.google.com/spreadsheets/d/1_5WqJ4Ezgg8p3qVd2nbbEXyMTjRavOBGAoKv7_4OtVs/edit?usp=sharing
In this form, please specify the classes extracted from each movie.
On the other hand a team can process multiple movies. Thus, it is possible to have multiple rows in this form for each team. Please specify the classes extracted from each movie, use the short names of the classes.

Technical Notes:

To have consistent dataset that we can all use, let's fix some technical features.

1- File names and type: Frames selected to be added to the dataset should be saved as PNG images and should be named based on the following template:

C_T_N.png

C is the class name as appeared in the column "class short name". Use capital letters only.

T is 2 digits represent the team. Please use be sure that it fixed in length and unique.

N is 4 digits serial number. It starts by 0 and is fixed to be 4 digits.

2- Sampling rate: To avoid redundant frames, you are asked to save key-frames only in the dataset. There are serval ways to extract key frames from a video, use any of these. Any algorithm, code, or a tool could be shared freely across teams.

An alternative solution, is to reduce the frame rate to 10 frames-per-second at most, then to apply a manual inspection step to delete blurred and very-similar frames.

3- Image dimensions: The frame dimension should be fixed to 640x480 (landscape).

4- Image colors: It is better to consider colored movies, but gray-level movies could be included too.

Delivery:

Submit the created data using the following form. Put the files of each class in a (*.zip) file. Thus, if a team of 4 members has completed the data for 4 classes, then a team member should use this form to upload 4 (*.zip) files.

Grading:

The total point for this milestone is 10. The points will be calculated as *min(number of uploaded frames/(100*number of team members), 10)*. Any delayed deliver will be subject to 50% deduction.

Regulations:

1- You may work in teams up to **4** students. Small group to ease the communication among group members.

2- The deadline of the first task is on Tuesday 24/05/2022 at 11:59 pm.

3- Submit your work to dlcv.guc@gmail.com.

Project Plan:

Task	Announcement	Deadline
Create & register the team	07.05.2022	12.05.2022
Create training and testing dataset	17.05.2022	24.05.2022
Build and train a simple CNN to classify actions in an Egyptian movie frames.	30.05.2022	02.06.2022
Apply model modification and optimization. Apply pre-processing and pot processing steps.	02.06.2022	11.06.2022