

Competition Track Evaluation Setup

The First International Workshop on Action Recognition with a Large Number of Classes in conjunction with ICCV, Sydney, Australia, December 2013

The goal of the competition track is to encourage cross-site algorithm comparison, and to provide the participants supports in algorithm evaluation. For both recognition and detection experiments, we provide standard train/test partitions and sample result submission files in order to make sure the results are directly comparable across different methods. Participants of the competition track will be asked to submit the outputs of their methods in the specified format which we will use for performing the evaluations.

Task Definitions:

- 1. Recognition:** For each of the 101 action classes, predict the presence/absence of the action in each test video clip.
- 2. Detection:** For a predefined set of 24 action classes, identify the bounding box(es) of the predicted actions in each test video clip.

Data Resources:

Both tasks in the competition track are based on video data from the UCF101 dataset. To download the entire dataset, [click here](#).

The clips of one action class are divided into 25 groups, each having 4-7 clips. Clips from the same group normally share common background and actor(s). All videos were downloaded from YouTube with manual filtering of irrelevant ones. The videos are saved in avi files compressed using DivX codec. For more details of the dataset, [see the UCF101 technical report](#).

Various low-level features (STIP, DFT, SIFT), class-level attributes, and bounding box annotations of actions for 24 classes are provided on the [download page](#) of the workshop webpage. Participants are encouraged to leverage these resources or compute/annotate their own.

Evaluation Setups:

1. Recognition Task

Participants are required to adopt [the provided three train/test splits](#). In each split, clips from 7 of the 25 groups are used as test samples, and the rest for training.

For each split, test results should be organized according to the [sample submission file](#), which has the following format:

```
[video_name_1]      [class_label]
[video_name_2]      [class_label]
[video_name_3]      [class_label]
[video_name_4]      [class_label]
.
.
.
[video_name_n]      [class_label]
```

video_name_X and *action label* are the clip name and corresponding class label (ranging from 1 to 101). Each row shows the result of one video, and therefore, the file should include *n* lines, which *n* is the number of test clips in the split. Participants need to strictly follow the submission format, otherwise they have the risk of not having their results evaluated and posted on the official evaluation report and webpage.

Also, it is optional, yet encouraged, to provide the confidence scores of all 101 classes for each clip, using the following format:

```
[video_name_1]      [confidence_class_1] [confidence_class_2]...[confidence_class_101]
[video_name_2]      [confidence_class_1] [confidence_class_2]...[confidence_class_101]
[video_name_3]      [confidence_class_1] [confidence_class_2]...[confidence_class_101]
[video_name_4]      [confidence_class_1] [confidence_class_2]...[confidence_class_101]
.
.
.
[video_name_n]      [confidence_class_1] [confidence_class_2]...[confidence_class_101]
```

The confidence values will be used for plotting and comparing the precision-recall curves of the submissions. The official performance measure of the recognition task is the average classification accuracy over all classes. The performance of each method will be the mean of average accuracies over the three provided test splits.

2. Detection Task

Participants are given a set of training samples for each of the 24 classes, with local action bounding box annotations. For each frame of the videos, participants need to provide an action label and a predicted bounding box of the action. The train/test splits and the bounding boxes annotations are provided on the [download page](#) of the workshop website.

For each train/test split, results of the detection task should be organized according to the [sample submission file](#), which has the following format:

```
[video_name_1] [frame_number] [class_label] [upper_left_x] [upper_left_y] [width] [height] [confidence_value]
```

Therefore, each row will contain one bounding box in each frame of a test video. The performance is evaluated based on the correctness of the predicted action and its bounding box. The popular “intersection-over-union” criterion will be adopted; a detection is correct if the overlap equals or is larger

than 0.2. ROC curves will then be plotted for each submitted result and mean AUC will be used as the single-value measure of detection performance.

Result Submission:

Please click [here](#) to submit your results to the competition. Each participant may submit the results of up to five runs for each task based on different configurations of their systems. All the results for each task should be zipped into a single file named by [organizationName-taskName.zip]. Within the zipped folder, results from different runs should be placed in separate folders named by [Run-Number]. Further under the folder, there should be 3 files complying with the specified submission format, where each file is generated based on one of three train/test splits.

Schedule:

For both recognition and detection tasks, results should be submitted before **November 7th**. Evaluated performances will be returned in 3 weeks after the submission deadline.

Important Date:

Submission deadline:

Evaluation results due (competition track): November 7, 2013

Papers due (research track): November 7, 2013*

Competition results announcement: November 28, 2013

Review results and author's notice: October 7, 2013 (early deadline submissions)

Camera ready: October 11, 2013 (early deadline submissions)

Publication date: As per conference schedule

Workshop date: December 7, 2013 (Full-day)

The authors who wish to have their papers published in the ICCV proceedings should submit their papers by the early deadline of **September 7.*

Note that for first year of the competition, labels of the test samples for both tasks are made available which can be used for calculating the accuracy by the participants. However, the submitted results should not be optimized according to the labels to allow a fair and realistic comparison.