Florida International University - University of Miami TRECVID 2017

Yilin Yan[1], Samira Pouyanfar[2], Yudong Tao[1], Haiman Tian[2], Maria Presa Reyes[2], Mei-Ling Shyu[1], Shu-Ching Chen[2], Winnie Chen[3], Tiffany Chen[3] and Jonathan Chen[4]

[1]Department of Electrical and Computer Engineering
University of Miami, Coral Gables, FL 33146, USA
[2]School of Computing and Information Sciences
Florida International University, Miami, FL 33199, USA
[3]School of Electrical and Computer Engineering
Purdue University, West Lafayette, IN 47907, USA
[4]Miami Palmetto Senior High School
Miami, FL 33156, USA
y.yan4@umiami.edu, spouy001@cs.fiu.edu, yxt128@miami.edu, htian005@cs.fiu.edu, mpres029@cs.fiu.edu, shyu@miami.edu, chens@cs.fiu.edu, chen1219@purdue.edu, chen1791@purdue.edu

Abstract

This paper demonstrates the framework and results from the team "Florida International University - University of Miami (FIU-UM)" in the TRECVID 2017 [1] Ad-hoc Video Search (AVS) task [2]. The following four runs were submitted:

- M_D_FIU_UM.17_1*: CNN features + Linear SVM*
- M_D_FIU_UM.17_2*: CNN features + Linear SVM + Scores from other groups*
- M_D_FIU_UM.17_3*: CNN features + Linear SVM + Rectified Linear Score Normalization*
- M_D_FIU_UM.17_4*: CNN features + Linear SVM + Scores from other groups + Rectified Linear Score Normalization*

In the first step, the features are extracted by the convolutional neural network (CNN) structure of GoogLeNet [3] for all runs. Then, the scores of each concept for the key frames are generated using the linear support vector machine (SVM) classifiers. For Run 2 *and Run* 4*, the scores generated by the ITI-CERTH team last year are integrated for score fusion and enhancement. Meanwhile, for Run* 3 *and Run* 4*, a new rectified linear score normalization algorithm is used. As a result, from the submission results, Run* 2 *outperforms the other three runs. The submission details are listed as follows.*

- Class*: M (Manually-assisted runs)*

- Training type*: D (IACC & non-IACC non-TRECVID data)*
- Team ID*: FIU-UM (Florida International University - University of Miami)*
- Year*: 2017*

1 Introduction

In year 2015 and the prior years, the TRECVID project [4] includes a semantic indexing (SIN) task which aims to recognize the semantic concept contained within a video shot. It has been well-acknowledged that there are several challenges in the SIN task, such as data imbalance, scalability, and semantic gap [5, 6, 7, 8, 9, 10, 11, 12]. In year 2016, the task was changed to the Ad-hoc video search (AVS) task which is to model the end user search use-cases. Comparing to the SIN task, the AVS task looks for not only the video segments that contain persons, objects, activities, locations, etc., but also the video segments of their combinations. The AVS task remains the same for this year.

The automatic annotation of semantic concepts in video shots can be essential for retrieval, categorization, and other video exploitations [13, 14, 15, 16, 17, 18, 19, 20]. The semantic concept retrieval research directions include (1) developing robust learning approaches that adjust to the increasing size and the diversity of the videos; (2) fusing the information from other sources such as audio and text; and (3) detecting the low-level and mid-level features that have a high discriminating capability [21, 22, 23, 24, 25, 26, 27, 28].

The size of the high-level semantic concepts provided by IACC remains the same as the AVS task of the previous year, which has 346 concepts in total. For each of the 346 semantic concepts, a list of ground truth labels is provided for training. Given the same test collection (IACC.3) and master shot reference, 30 Ad-hoc queries were released by NIST for testing. Each query can be a combination of some of the 346 concepts and/or some other concepts not included in the training set. Each participated group is allowed to submit a maximum of $1,000$ possible shots from the test collection for each query, which are ranked according to their likelihood of containing the target query. The submission result is rated by using the mean inferred average precision (mean xinfAP) [29] based on the assessment of a 2-tiered random sampling (1-200@100% and 201-1000@11.1%).

This paper is organized as follows. Section 2 describes our proposed framework and the specific approaches utilized for each run. Section 3 shows the submission results in details. Section 4 summarizes the whole paper and proposes some future directions to pursue and plan for next year.

2 The Proposed Framework

Our framework of the TRECVID 2017 AVS task is shown in Figure 1. In this year, key frames are already extracted from the videos in the IACC.3 collection and are provided to the participants. Thus, the key frame extraction step is skipped; while the key frames extracted in the previous years are used for training.

2.1 CNN Feature Extraction

A pre-trained deep learning model called AlexNet [30] was utilized in our framework last year. AlexNet is a CNN structure trained on the ImageNet database for object feature extraction and contains five convolutional layers and three fully-connected layers. In this year, the GoogLeNet [3] which is a 22 layers deep network as shown in Figure 1 is used. The CNN features are extracted for all the training and testing key frames from the pool5 layer, i.e., the last layer before the softmax layer with $1,000$ dimensions. The GoogLeNet structure is well trained and proven to achieve great performance.

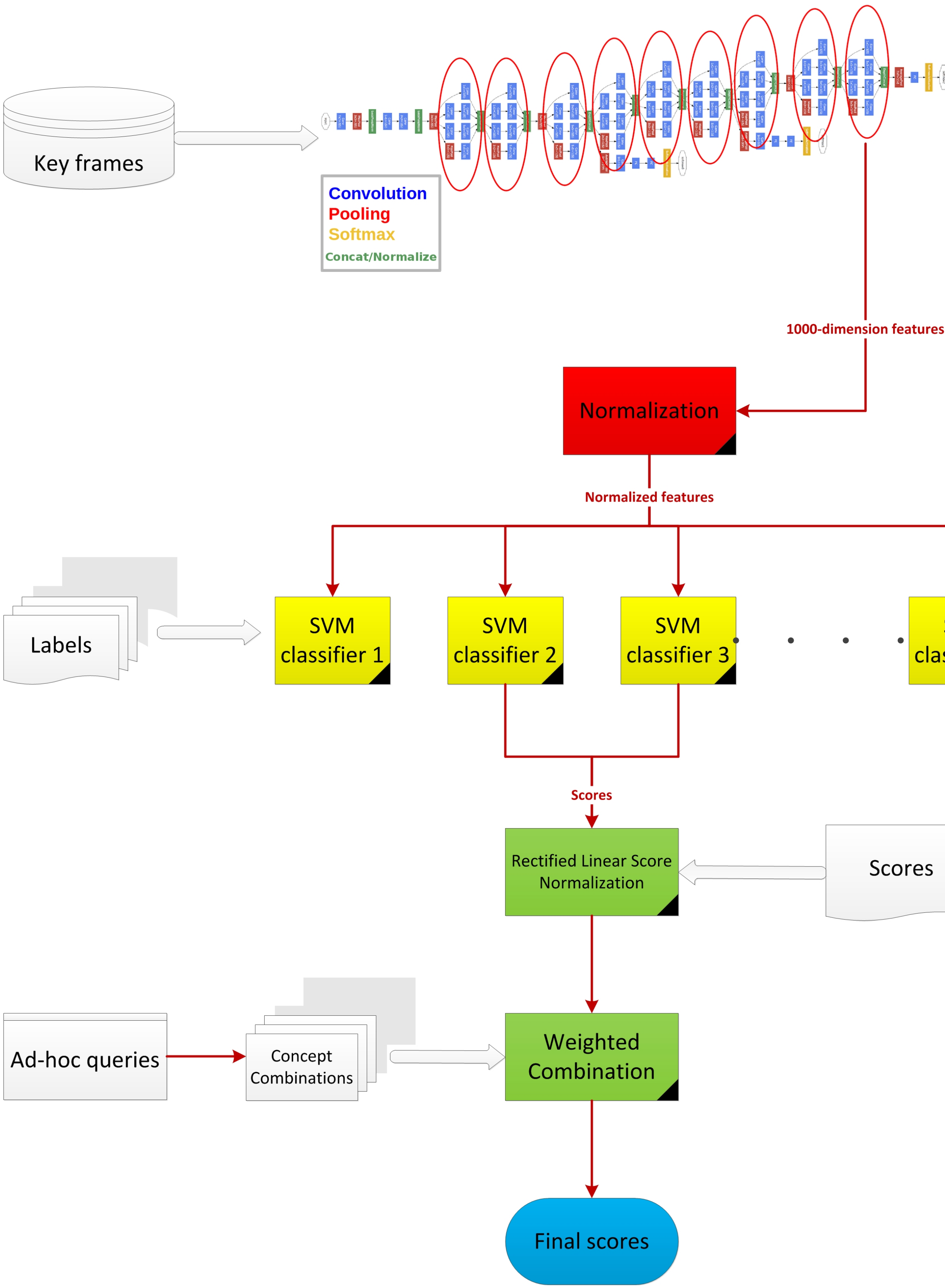


Figure 1: The designed framework for the TRECVID 2017 AVS task

2.2 Classification

After feature extraction, the feature vectors are normalized for each key frame. Support Vector Machine (SVM), one of the state-of-the-art algorithms in the data mining area [31, 32, 33, 34, 35, 36, 37, 38] including multimedia classification, is used for classification and concept score generation. The general idea of SVM is to build a separating hyperplane to classify the data instances so that the geometric margin is maximized. For efficiency, linear kernels are adopted in our proposed framework.

2.3 Rectified Linear Score Normalization

A new rectified linear score normalization algorithm is proposed this year. Different from traditional score normalization algorithms, a rectified function is applied as shown in Algorithm 1. The idea here is to eliminate the effect of "bad" scores of a concept in an Ad-hoc query before the next step. If the score of a concept in a frame is greater than a certain threshold, the concept is considered definitely to appear in that frame and thus its score is set to 1. Meanwhile, if the score of a concept in a frame is lower than a certain threshold, it is very unlikely to appear in that frame. Thus, the score would be set to 0 in this case to eliminate its effect in the fusion stage as the scores of those selected concepts will be multiplied in the next stage.

The thresholds $threshold_{high}$ and $threshold_{low}$ are generated based on our empirical studies. Generally speaking, they are calculated based on the positive to negative ratio (PN_{ratio}) in the train data set. If the scores are good enough, strict thresholds will be used, i.e., setting more high scores to 1 and setting more low scores to 0. In this study, $threshold_{high}$ is set to the score ranked half PN_{ratio} on average, while $threshold_{low}$ is set to the score ranked 5 times PN_{ratio} on average. The sample score distribution before rectified linear normalization is depicted in Figure 2(a) and the distribution after applying rectified linear normalization is presented in Figure 2(b).

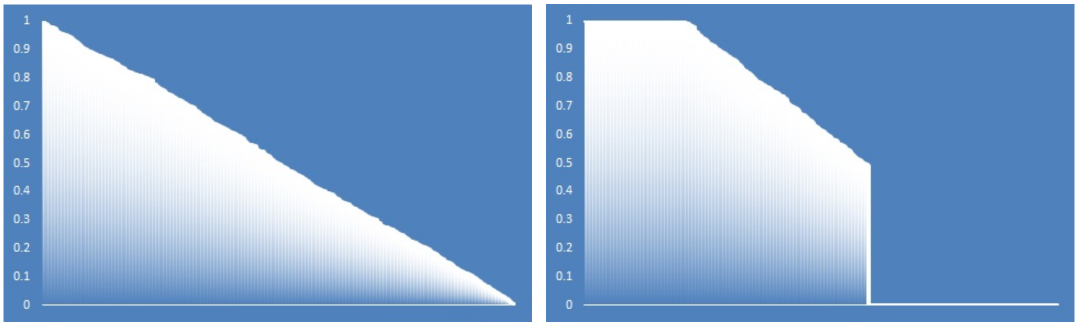
Algorithm 1 The proposed rectified linear score normalization algorithm.

for all scores **do**
  if $score \geq threshold_{high}$ **then**
    $score_{normed} = 1$;
  else if $score \leq threshold_{low}$ **then**
    $score_{normed} = 0$;
  else
    $score_{normed} = \frac{score}{threshold_{high}}$.
  end if
end for

2.4 Query Formulation and Score Combination

All the four runs are manually-assisted runs. For an initial ad-hoc query, a group member formulates it into a combination of concepts based on its topic and query interface without the knowledge of the collection or the search results. For queries containing the concepts not included in the 346 concepts from the training set (e.g., the concept "couch" in "Find shots of a man sitting down on a couch in a room"), a similar concept like "sofa" is selected.

Meanwhile, in order to increase the number of object categories and have more options in score fusion, the pre-trained ImageNet models are also applied to generate the scores for ImageNet concepts. The GoogLeNet model for ImageNet1000, provided by the Caffe development team, is applied to the frames in the test data set. Take the same query as an example. The scores of the concept "studio couch, day bed" in ImageNet are utilized for score combination in the next step. The scores are generated from the pool5 layers as well.



(a) (b)

Figure 2: Sample score distribution (a) before & (b) after rectified linear normalization.

In order to compare our scores with others, we integrate the scores from the ITI-CERTH team which performed very well last year. Since the performance of the weighted geometric mean of the scores is often better than the weighted arithmetic mean, which has been reported by several groups, we fuse the scores of the selected concepts by calculating their weighted geometric mean. The same weights are used for score combination based on our empirical studies in all runs as follows, where N is the total number of concepts integrated.

$$score_{query} = \prod_{i}^{N} (score_{concept_i})^{weight_i} \tag{1}$$

3 Experimental Results

3.1 Data

Given the information (including the IACC.3 test collection, master shot reference, and 30 Ad-hoc queries) released by NIST and the concept definitions, a list of at most 1,000 shot IDs from the test collection for each ad-hoc query was returned and ranked according to their likelihood of containing the target query. TRECVID 2017 test data set (IACC.3) contains $4,593$ Internet Archive videos with the durations between 6.5 and 9.5 minutes (144GB, 600 hours in total). The training data set combines the development and testing data sets from the 2010 to 2015 issues of the SIN task, namely the IACC.1.tv10.training, IACC.1.A-C, and IACC.2.A-C data sets. Each contains about 200 hours of videos drawn from the IACC.1 and IACC.2 collections using videos with durations ranging from 10 seconds to (3.5 to 6.4) minutes, respectively. The TRECVID data set has been widely used in research [39, 40, 41, 42].

The overall framework of the TRECVID 2017 AVS task contains three stages:

1. Model training: using TRECVID 2010-2012 training videos as the training data.
2. Model evaluation: using TRECVID 2013-2015 training videos as the testing data to evaluate the framework and tune the parameters of the models.
3. Model testing: using TRECVID 2010-2015 training videos as the TRECVID 2017 training data, and TRECVID 2017 testing videos as the testing data to generate the ranking results for the submission.

3.2 Evaluation

All (30) queries are evaluated by assessors at NIST after pooling and sampling. Measures (indexing) are shown as follows [43].

1. Mean extended inferred average precision (mean xinfAP) [29] which allows the sampling density to vary so that it can be 100% in the top strata. This is the most important one for average precision.

2. As in the past years, other detailed measures based on recall and precision are generated and given by the sample_eval software provided by the TRECVID team.

3.3 Performance

All of the measures below were based on the assessment of a 2-tiered random sampling (1-200@100% and 201-1000@11.1%) of the full submission pools and the sample_eval software was used to infer the measures. Figures 3, 4, 5, and 6 present the performance of our ad-hoc query video search results. The x-axis are the concept numbers; while the y-axis present the inferred average precision values. More clearly, Table 1 shows the inferred mean average precision (MAP) values of the first 5, 10, 15, 20, 30, 100, 200, 500 and 1000 shots. The numbers of the returned inferred true shots and the mean xinfAP values are shown in Table 2.

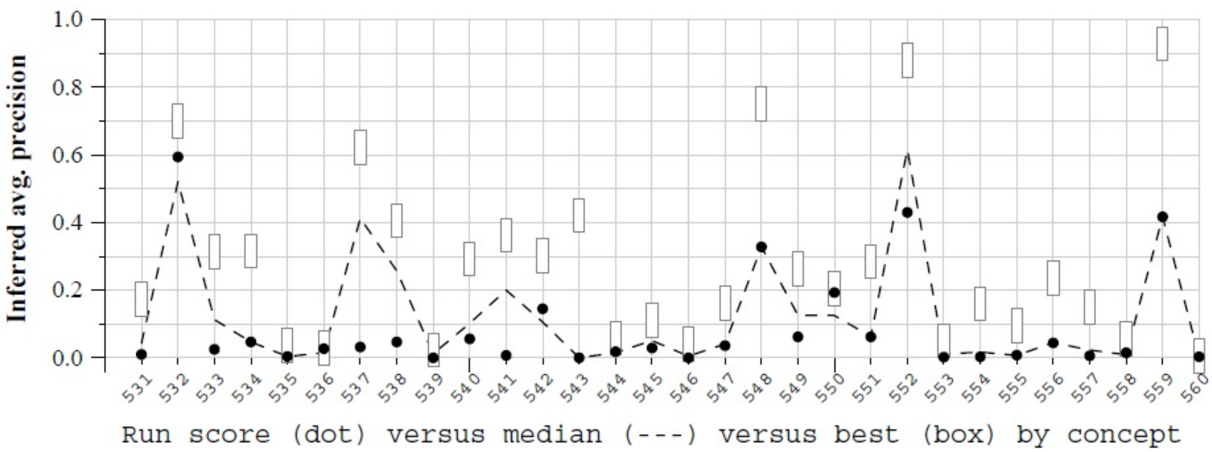


Figure 3: Run scores (dot) versus median (—) versus best (box) for *M_D_FIU_UM.17_1*

Framework	5	10	15	20	30	100	200	500	1000
M_D_FIU_UM.17_1	0.340	0.333	0.329	0.320	0.310	0.230	0.196	0.144	0.120
M_D_FIU_UM.17_2	0.447	0.420	0.404	0.398	0.383	0.334	0.287	0.202	0.158
M_D_FIU_UM.17_3	0.353	0.353	0.340	0.322	0.311	0.252	0.235	0.160	0.138
M_D_FIU_UM.17_4	0.420	0.387	0.384	0.383	0.384	0.309	0.270	0.191	0.154

Table 1: The MAP values at first *n* shots for all 4 runs

The evaluation results demonstrate that the performance of our framework is above the average of all the groups, while we perform the best for queries 537, 538, 539, and 544. By applying the rectified linear score normalization algorithm before the score fusion, the better performance in Run 3 than Run 1 is achieved. For most queries, the proposed normalization algorithm improves the xinfAP values, especially for queries 537, 538, and 552. While the result of Run 4 is similar to that of Run 2, the rectified linear score normalization step helps improve the xinfAP value for queries 538 and 552 as well, and makes our team performs the best for query 538.

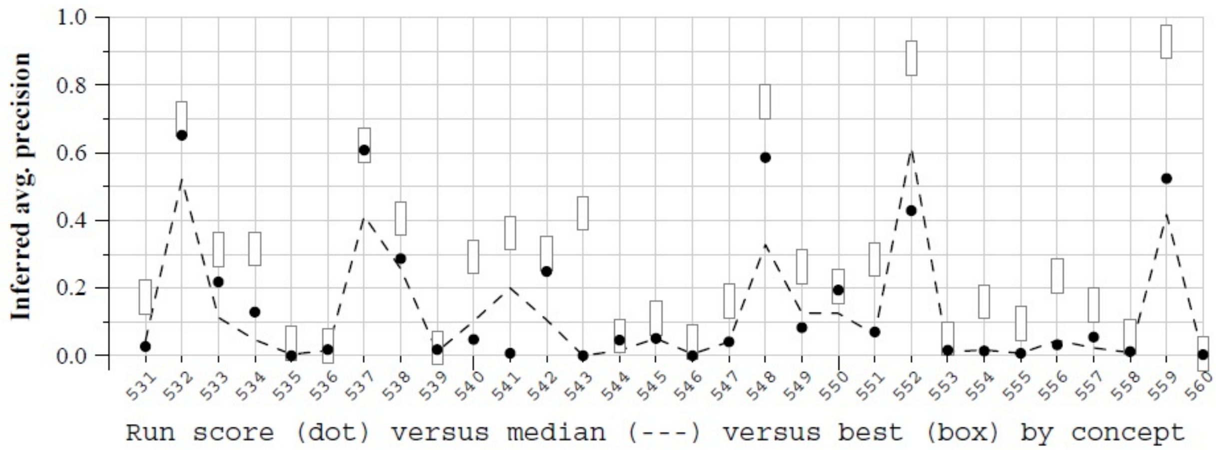


Figure 4: Run scores (dot) versus median (—) versus best (box) for *M_D_FIU_UM.17_2*

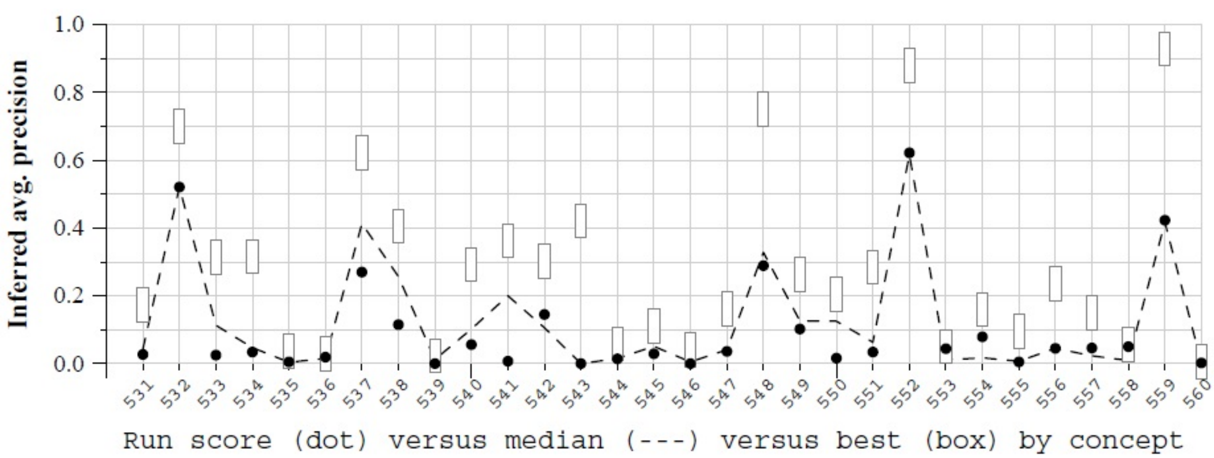


Figure 5: Run scores (dot) versus median (—) versus best (box) for *M_D_FIU_UM.17_3*

Framework	# of inferred true shots returned	Mean xinfAP values
M_D_FIU_UM.17_1	3608	0.088
M_D_FIU_UM.17_2	4731	0.147
M_D_FIU_UM.17_3	4125	0.102
M_D_FIU_UM.17_4	4623	0.145

Table 2: The numbers of inferred true shots returned and the mean xinfAP values

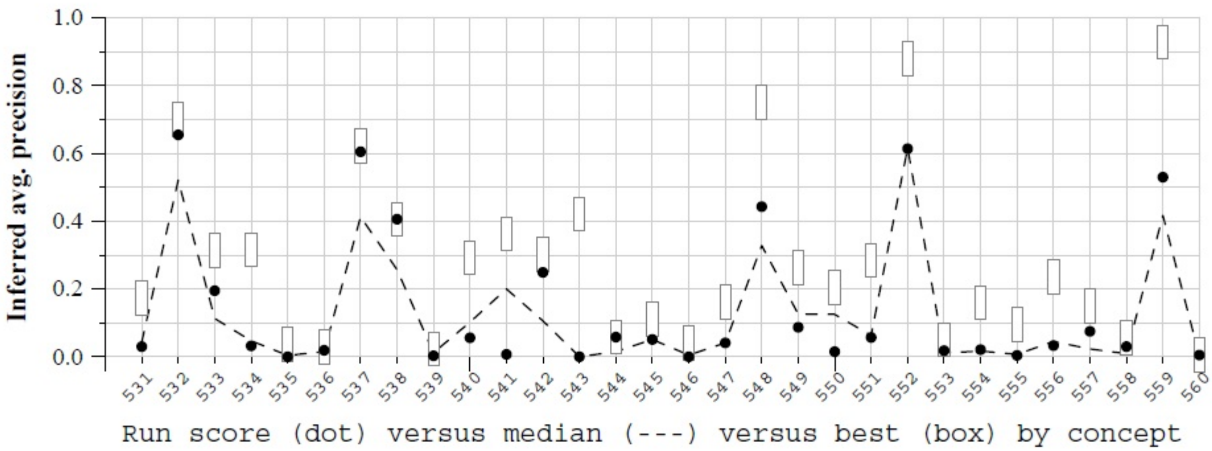


Figure 6: Run scores (dot) versus median (—) versus best (box) for *M_D_FIU_UM.17_4*

4 Conclusion and Future Work

In this notebook paper, the framework and results of team FIU-UM in the TRECVID 2017 AVS task are summarized. It can be seen that a lot of improvements still need to be done based on the results. The following directions will be investigated.

- In our framework, only global features are utilized. The object-level features can also be explored by R-CNN (Regional CNN).
- Non-linear SVM classifiers need to be adopted to address the data imbalance issue.
- More advanced CNN structures can be integrated and scores from them can be fused.
- Temporal correlations can be considered to reach a better performance.
- More training data should be collected by a general purpose search engine like Google using the query definition to further improve the retrieval accuracy.

In this year, there are still several very "hard" queries which challenged all groups. For instance, the query "Find shots of children playing in a playground" was not easy to interpret since the concept "playground" varies a lot. Furthermore, it is considered hard to find images related to the query "Find shots of a person holding opening closing or handing over a box" by Google. Therefore, it will be helpful to exchange ideas and thoughts with other groups so that novel approaches can be developed for further performance improvements.

References

[1] George Awad, Asad Butt, Jonathan Fiscus, David Joy, Andrew Delgado, Martial Michel, Alan F. Smeaton, Yvette Graham, Wessel Kraaij, Georges Qunot, Maria Eskevich, Roeland Ordelman, Gareth J. F. Jones, and Benoit Huet. Trecvid 2017: Evaluating ad-hoc and instance video search, events detection, video captioning and hyperlinking. In *Proceedings of TRECVID 2017*. NIST, USA, 2017.

[2] Alan F. Smeaton, Paul Over, and Wessel Kraaij. High-Level Feature Detection from Video in TRECVid: a 5-Year Retrospective of Achievements. In Ajay Divakaran, editor, *Multimedia Content Analysis, Theory and Applications*, pages 151–174. Springer Verlag, Berlin, 2009.

[3] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[4] George Awad, Cees G. M. Snoek, Alan F. Smeaton, and Georges Qunot. Trecvid semantic indexing of video: A 6-year retrospective. *ITE Transactions on Media Technology and Applications*, 4(3):187–208, 2016. Invited paper.

[5] Shu-Ching Chen, Arif Ghafoor, and R. L. Kashyap. *Semantic Models for Multimedia Database Searching and Browsing*. Kluwer Academic Publishers, Norwell, MA, USA, 2000.

[6] Xiuqi Li, Shu-Ching Chen, Mei-Ling Shyu, and Borko Furht. An effective content-based visual image retrieval system. In *Proceedings of the IEEE International Computer Software and Applications Conference*, pages 914–919, August 2002.

[7] Lin Lin, Guy Ravitz, Mei-Ling Shyu, and Shu-Ching Chen. Effective feature space reduction with imbalanced data for semantic concept detection. In *Proceedings of the IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing*, pages 262–269, 2008.

[8] Qiusha Zhu, Lin Lin, Mei-Ling Shyu, and Shu-Ching Chen. Feature selection using correlation and reliability based scoring metric for video semantic detection. In *Proceedings of the IEEE Fourth International Conference on Semantic Computing*, pages 462–469, 2010.

[9] Mei-Ling Shyu, Thiago Quirino, Zongxing Xie, Shu-Ching Chen, and Liwu Chang. Network intrusion detection through adaptive sub-eigenspace modeling in multiagent systems. *ACM Trans. Auton. Adapt. Syst.*, 2(3), September 2007.

[10] Mei-Ling Shyu, Choochart Haruechaiyasak, and Shu-Ching Chen. Category cluster discovery from distributed www directories. *Information Sciences*, 155(3):181–197, 2003.

[11] Shu-Ching Chen and R.L. Kashyap. Temporal and spatial semantic models for multimedia presentations. In *Proceedings of the 1997 International Symposium on Multimedia Information Processing*, pages 441–446, 1997.

[12] Lin Lin, Mei-Ling Shyu, Guy Ravitz, and Shu-Ching Chen. Video semantic concept detection via associative classification. In *Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on*, pages 418–421. IEEE, 2009.

[13] Shu-Ching Chen, Srinivas Sista, Mei-Ling Shyu, and R. Kashyap. Augmented transition networks as video browsing models for multimedia databases and multimedia information systems. In *Proceedings of the 11th IEEE International Conference on Tools with Artificial Intelligence*, pages 175–182, November 1999.

[14] Xiuqi Li, Shu-Ching Chen, Mei-Ling Shyu, and Borko Furht. Image retrieval by color, texture, and spatial information. In *Proceedings of the 8th International Conference on Distributed Multimedia Systems*, pages 152–159, September 2002.

[15] Xin Chen, Chengcui Zhang, Shu-Ching Chen, and Min Chen. A latent semantic indexing based method for solving multiple instance learning problem in region-based image retrieval. In *Proceedings of the Seventh IEEE International Symposium on Multimedia*, pages 37–44, December 2005.

[16] Mei-Ling Shyu, Kanoksri Sarinnapakorn, Indika Kuruppu-Appuhamilage, Shu-Ching Chen, LiWu Chang, and Thomas Goldring. Handling nominal features in anomaly intrusion detection problems. In *Research Issues in Data Engineering: Stream Data Mining and Applications, 2005. RIDE-SDMA 2005. 15th International Workshop on*, pages 55–62. IEEE, 2005.

[17] Shu-Ching Chen, S.H. Rubin, Mei-Ling Shyu, and Chengcui Zhang. A dynamic user concept pattern learning framework for content-based image retrieval. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 36(6):772–783, November 2006.

[18] Mei-Ling Shyu, C. Haruechaiyasak, Shu-Ching Chen, and Na Zhao. Collaborative filtering by mining association rules from user access sequences. In *Proceedings of the International Workshop on Challenges in Web Information Retrieval and Integration*, pages 128–135, April 2005.

[19] Shu-Ching Chen, Mei-Ling Shyu, Chengcui Zhang, and R. L. Kashyap. Identifying overlapped objects for video indexing and modeling in multimedia database systems. *International Journal on Artificial Intelligence Tools*, 10(4):715–734, 2001.

[20] Xin Chen, Chengcui Zhang, Shu-Ching Chen, and Stuart Rubin. A human-centered multiple instance learning framework for semantic video retrieval. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 39(2):228–233, 2009.

[21] Shu-Ching Chen, Mei-Ling Shyu, and R.L. Kashyap. Augmented transition network as a semantic model for video data. *International Journal of Networking and Information Systems*, 3(1):9–25, 2000.

[22] Mei-Ling Shyu, Shu-Ching Chen, and Rangasami L Kashyap. Generalized affinity-based association rule mining for multimedia database queries. *Knowledge and Information Systems*, 3(3):319–337, 2001.

[23] Qiusha Zhu, Lin Lin, Mei-Ling Shyu, and Shu-Ching Chen. Effective supervised discretization for classification based on correlation maximization. In *Proceedings of the IEEE International Conference on Information Reuse and Integration*, pages 390–395, 2011.

[24] Dianting Liu, Yilin Yan, Mei-Ling Shyu, Guiru Zhao, and Min Chen. Spatio-temporal analysis for human action detection and recognition in uncontrolled environments. *International Journal of Multimedia Data Engineering and Management*, 6(1):1–18, January 2015.

[25] Shu-Ching Chen, Mei-Ling Shyu, and Chengcui Zhang. Innovative shot boundary detection for video indexing. In Sagarmay Deb, editor, *Video Data Management and Information Retrieval*, pages 217–236. Idea Group Publishing, 2005.

[26] Lin Lin and Mei-Ling Shyu. Weighted association rule mining for video semantic detection. *Int. J. Multimed. Data Eng. Manag.*, 1(1):37–54, January 2010.

[27] Shu-Ching Chen, Mei-Ling Shyu, and Chengcui Zhang. An intelligent framework for spatio-temporal vehicle tracking. In *Proceedings of the 4th IEEE International Conference on Intelligent Transportation Systems*, pages 213–218, August 2001.

[28] Tao Meng, Yang Liu, Mei-Ling Shyu, Yilin Yan, and Chi-Min Shu. Enhancing multimedia semantic concept mining and retrieval by incorporating negative correlations. In *Proceedings of the IEEE International Conference on Semantic Computing*, pages 28–35, June 2014.

[29] Emine Yilmaz, Evangelos Kanoulas, and Javed A. Aslam. A simple and efficient sampling method for estimating ap and ndcg. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 603–610, New York, NY, USA, 2008. ACM.

[30] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[31] Chao Chen, Qiusha Zhu, Lin Lin, and Mei-Ling Shyu. Web media semantic concept retrieval via tag removal and model fusion. *ACM Transactions on Intelligent Systems and Technology*, 4(4):61:1–61:22, October 2013.

[32] Yilin Yan, Jun-Wei Hsieh, Hui-Fen Chiang, S.-C. Cheng, and Duan-Yu Chen. Plsa-based sparse representation for object classification. In *Proceedings of the 22nd International Conference on Pattern Recognition*, pages 1295–1300, August 2014.

[33] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, September 1995.

[34] Tao Meng and Mei-Ling Shyu. Leveraging concept association network for multimedia rare concept mining and retrieval. In *Proceedings of the IEEE International Conference on Multimedia & Expo*, pages 860–865, July 2012.

[35] Yilin Yan, Min Chen, Saad Sadiq, and Mei-Ling Shyu. Efficient imbalanced multimedia concept retrieval by deep learning on spark clusters. *International Journal of Multimedia Data Engineering and Management (IJMDEM)*, 8(1):1–20, 2017.

[36] Yilin Yan and Mei-Ling Shyu. Enhancing rare class mining in multimedia big data by concept correlation. In *Proceedings of the 2016 IEEE International Symposium on Multimedia (ISM)*, pages 281–286, December 2016.

[37] Yilin Yan, Qiusha Zhu, Mei-Ling Shyu, and Shu-Ching Chen. A classifier ensemble framework for multimedia big data classification. In *Proceedings of the 2016 IEEE 17th International Conference on Information Reuse and Integration (IRI)*, pages 615–622, July 2016.

[38] Saad Sadiq, Yilin Yan, Mei-Ling Shyu, Shu-Ching Chen, and Hemant Ishwaran. Enhancing multimedia imbalanced concept detection using vimp in random forests. In *2016 IEEE 17th International Conference on Information Reuse and Integration (IRI)*, pages 601–608, July 2016.

[39] Yilin Yan, Mei-Ling Shyu, and Qiusha Zhu. Supporting semantic concept retrieval with negative correlations in a multimedia big data mining system. *International Journal of Semantic Computing*, 10:247–268, 2016.

[40] Yilin Yan, Min Chen, Mei-Ling Shyu, and Shu-Ching Chen. Deep learning for imbalanced multimedia data classification. In *2015 IEEE International Symposium on Multimedia (ISM)*, pages 483–488, December 2015.

[41] Yilin Yan, Yang Liu, Mei-Ling Shyu, and Min Chen. Utilizing concept correlations for effective imbalanced data classification. In *Proceedings of the IEEE 15th International Conference on Information Reuse and Integration*, pages 561–568, August 2014.

[42] Yilin Yan, Mei-Ling Shyu, and Qiusha Zhu. Negative correlation discovery for big multimedia data semantic concept mining and retrieval. In *2016 IEEE Tenth International Conference on Semantic Computing (ICSC)*, pages 55–62, February 2016.

[43] Alan F. Smeaton, Paul Over, and Wessel Kraaij. Evaluation campaigns and trecvid. In *MIR '06: Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval*, pages 321–330, New York, NY, USA, 2006. ACM Press.