1 Team details

- Team name: UPC-STP
- Team name leader: Xavier Giro-i-Nieto
- Team leader address: Campus Nord UPC (Modul D5), Jordi Girona 1-3, 08034 Barcelona Catalonia, +34 934 015 769, xavier.giro@upc.edu
- Rest of team members: Andrea Calafell, Matthias Zeppelzauer, Amaia Salvador
- Team website URL: `https://imatge.upc.edu`, `http://mc.fhstp.ac.at/`
- Affiliations: Universitat Politecnica de Catalunya, St. Poelten University of Applied Sciences, Austria

2 Contribution details

- Title of the contribution: Fine-tunning a CNN trained with objects and locations for cultural event detection
- Final score: 0.670 (on the validation set)
- General method description: We used the AlexNet CNN architecture pre-trained with ImageNet, Places and both datasets. The fully connected layers of the three CNNs were fine-tunned first with the training ChaLearn dataset. Then we applied a method called fracking, which shows multiple times to our networks those images that cause most confusion to them, with the goal that it will learn better decision boundaries between the classes. Figure 1 exemplifies this technique. Then, the CNNs were fine-tunned with the validation dataset. We extracted features from layers 6, 7, and 8 from all three networks and trained separate 1-vs-all linear SVM classifiers from them (9 separate classifiers). The probabilistic outputs of these 9 different SVMs were fused with a top-level SVM which provided the a class score for each image. The submitted rankings were based on these scores. The overall processing pipeline of our approach is shown in Figure 2.
- References
  - Calafell-Oros A. Fine-tuning a Convolutional Network for Cultural Event Recognition. Bachelor thesis report. ETSETB-UPC. 2015
  - B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva. "Learning Deep Features for Scene Recognition using Places Database." Advances in Neural Information Processing Systems 27 (NIPS), 2014.

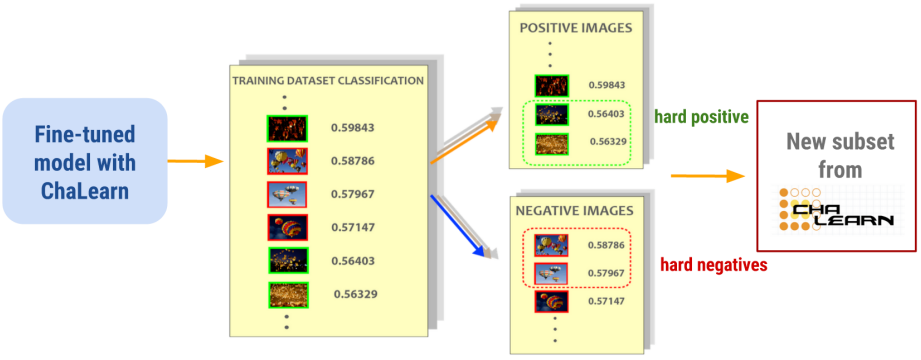

Figure 1: Fracking the training dataset for hard samples.

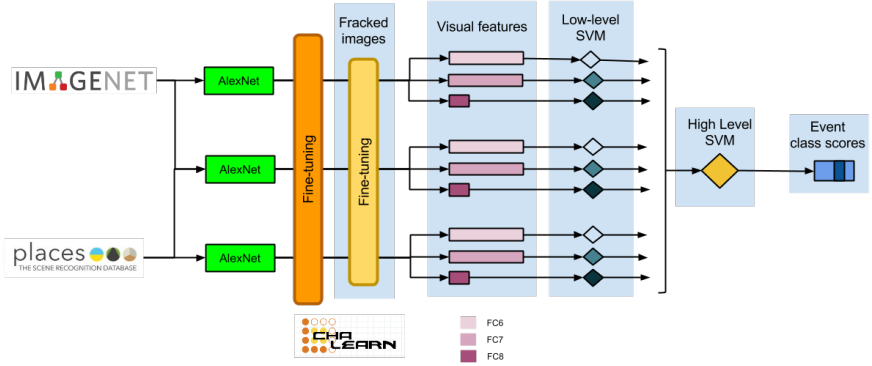

Figure 2: Diagram of the method.

– Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., and Darrell, T. (2014, November). Caffe: Convolutional architecture for fast feature embedding. In Proceedings of the ACM International Conference on Multimedia (pp. 675-678). ACM.

– Wang, L., Wang, Z., Du, W., and Qiao, Y. (2015). Object-Scene Convolutional Neural Networks for Event Recognition in Images. arXiv preprint arXiv:1505.00296.

– Simo-Serra, E., Trulls, E., Ferraz, L., Kokkinos, I., and Moreno-Noguer, F. (2014). Fracking deep convolutional image descriptors. arXiv preprint arXiv:1412.6537.

3 Data Preprocessing

- Describe features used or data representation model (if any): Features were extracted from fully connected layers 6, 7 and 8 from fine-tunned versions of AlexNet (CaffeNet implementation).
- Dimensionality reduction technique applied (if any): None
- Segmentation strategy used (if any): None
- Other techniques/strategy used not included in previous items FOR DATA PREPROCESSING (if any):

4 Classification details

- Classifier or method used to train and validate your results: We tested both the soft-max provided in CaffeNet as well as linear SVMs on each layer and all linear SVMs fused by another SVM. The last option (fusion of SVM outputs) resulted in the best results.
- Large scale strategy (if any): None
- Compositional model used (scene context representation), i.e. pictorial structure (if any): None
- Other technique/strategy used not included in previous items FOR CLASSIFICATION (if any): -

5 Global Method Description

- Total method complexity analysis: The method required an important amount of computation in terms of GPU for fine-tunning the CNNs and extracting their features. Of course this step become more crucial when different approaches and strategies are explored. Training of non-linear SVM models took also a considerable amount of time. Thus, we restricted to linear SVMs for the challenge.
- Which pre-trained or external methods have been used (for any stage, if any): CaffeNet trained with ImageNet images (providede by Berkeley), and AlexNet pre-trained with ImageNet and Places (provided by MIT).
- Qualitative advantages of the proposed solution: It is very flexible and applicable to any image classification problem. The combination of a CNN trained for objects and a second ones for locations allows capturing the two scales in the images.
- Results of the comparison to other approaches (if any): The results using SVMs performed much better than the ones using the SoftMax Layer from the DNNs.

- Novelty degree of the solution and if is has been previously published: The solution is basically an extension of our previous submission for the CVPR workshop by using more CNNs trying to capture both the objects and the locations of the events. Preleminary results were published in the bachelor thesis of Andrea Calafell (Calafell-Oros A. 2015) but most of them were not applicable to this challenge because they aimed at using noisy external data, which was not allowed in this competition.

6 Other details

- Language and implementation details (including platform, memory, parallelization requirements): The fine-tunning of CNNs was performed in Python, while the training of SVMs was run in Matlab. For SVM we used the libsvm and liblinear packages.

- Human effort required for implementation, training and validation?: A bachelor student become familiar with the work during the whole Spring semester, as worked full time on the task during 15 days. She was supported by an associated professor working part-time during one week. Another associated professor and Phd student provided logistic support in an average basis of one hour a day during two weeks.

- Training/testing expended time? Extract the features from the CNN took 90 hours.

- General comments and impressions of the challenge? It is a nice task because it is well defined and fun. However the communication and usability may be improved, as the Codalab platform is a bit messy to use. Awards are very impressive in this workshop, as well as the announcement of a Journal issue coming soon.