Team name	SBU_CS
Team leader name	Heeyoung Kwon
Team leader address, phone number and email	61 University Heights Dr. Stony Brook, NY 11790 631-480-7455, heekwon@cs.stonybrook.edu
Rest of team members	Kiwon Yun Dimitris Samaras Minh Hoai Nguyen
Team website URL (if any)	

Title of the contribution	Cultural event classification by Convolutional Neural Networks
General method description	With using Convolutional Neural Network (CNN) features on multiple sub-regions, Spatial Pyramid Matching(SPM) framework is exploited.
References	[1] Chatfield, Ken, et al. "Return of the devil in the details: Delving deep into convolutional nets." arXiv preprint arXiv:1405.3531 (2014). [2] Regularized Max Pooling for Image Categorization Hoai, M., Proceedings of British Machine Vision Conference. (2014) [3] Lazebnik, Svetlana, Cordelia Schmid, and Jean Ponce. "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories." Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. Vol. 2. IEEE, 2006. [4] J. A. K. Suykens and J. Vandewalle. Least squares support vector machine classifiers. Neural Processing Letters, 9(3):293–300, 1999.

Describe data preprocessing techniques applied (if any)	
Describe features used or data representation model (if any)	The feature extraction is generally followed as [2]. We first resize the image to 224 × 224 pixels (the desired input for our CNN network), then divide the image into 100 subwindows as described in [2]. We extract a 128-dimensional feature vector for each subwindow using the medium CNN architecture (CNN-M) described by Chatfield et al. [1].
Dimensionality reduction technique applied (if any)	

Segmentation strategy used (if any)	
Classifier or method used to train and validate your results (if any)	Spatial Pyramid Matching (SPM) [3] based on Least-Sqaures Support Vector Machines (LSSVM) [4] is used as classifiers. SPM works by partitioning the image into increasingly fine sub-regions and aggregating local features found inside each sub-region. LSSVM, also known as kernel Ridge regression, has been shown to perform equally well as SVM in many classification benchmarks. LSSVM has a closed-form solution, which is a computational advantage over SVM.
Large scale strategy (if any)	

Transfer learning strategy (if any)	
Compositional model used (scene context representation), i.e. pictorial structure (if any)	
Other technique/strategy used not included in previous items (if any)	
Method complexity analysis	N/A

Qualitative advantages of the proposed solution	
Results of the comparison to other approaches (if any)	We got mAP = 0.5239 on validation set when using a Pyramid Histogram of Visual Words (PHOW) descriptor and RG chromaticity with a Bag of Word (BoW) framework. We also got mAP = 0.4914 with the CNN features extracted from the entire image without subregions.
Novelty degree of the solution and if is has been previously published	We evaluated that the performance of the CNN-based approach for the cultural event classification. Although similar approaches have already shown impressive levels of performance on various image classification tasks, the importance of the spatial relationship was also considered. Our method is inspired by [2].

Language and implementation details (including platform, memory, parallelization requirements)	
Human effort required for implementation, training and validation?	No. All you need to do is just runnning CultureEvent.main() function on matlab. CPU info: Intel(R) Xeon(R) CPU E5-2667 @ 2.90GHz
Training/testing expended time?	Extracting deep learning features takes around 20s per image, and the training/testing time take around 350s in total.
General comments and impressions of the challenge	It was fun since I can apply several techniques on real world dataset, even though it was hard to analyze results due to the number of classes.