**A Review on Violence Video Classification Using Convolutional Neural Networks**

In this work, the author evaluated previous works that were proposed as a solution to the ways video classification could be possible. At first, he reviews [1] where the author of that paper fuses frame features to come up with a global descriptor. The random forest[4] method was replaced by k-means whereas the bag of words method was replaced by the combination of Vector of Locally Aggregated Descriptor (VLAD)[5] along with Fisher Kernel. On the other hand, the author in [2] proposed to use a convolutional neural net using a subset from the ImageNet classes selected particularly for violence detection. Then the researcher came up with a two-stream CNN framework to extract detectors on frames that were still and also moving. Lastly, a third Long Short Term Memory (LSTM)[6] models are applied on top of the two-stream CNN network. Other than that other audio and motion features are also taken out as extra features. The author could achieve a mean average precision of 0.296.

Meanwhile, in [3], the researcher used computer vision techniques using a local descriptor approach[7] for violence detection from Hollywood movies using large-scale segmental feature extraction.

In conclusion, the author of the paper proposed a framework with which an effective classification model can be achieved. The steps are given below:

1.Data Collection

2. Pre-Processing the data, using data augmentation and wavelet Filtering.

3. Partitioning the data into Training and Testing with ratio of 8:2 respectively.

4. Passing it through Convolutional Neural Network with sliding window optimization.

5.Subsampling the data on the training phase

6. Passing it through a Fully Connected Layer.

7. Evaluating performance on the basis of Accuracy and Computational Time.

**References**

1. Mishna, F., Khoury-Kassabri, M., Gadalla, T., Daciuk, J.: Risk factors for involvement in cyber bullying: Victims, bullies and bully-victims. Child Youth Serv. Rev. 34(1), 63–70 (2012)
2. Dai, Q., Zhao, R.W., Wu, Z., Wang, X., Gu, Z., Wu, W., Jiang, Y.G.: Fudan-Huawei at MediaEval 2015: Detecting Violent Scenes and Affective Impact in Movies with Deep Learning (2015)
3. Eyben, F., Weninger, F., Lehment, N., Schuller, B., Rigoll, G.: Affective video retrieval: violence detection in Hollywood movies by large-scale segmental feature extraction. PLoS ONE 8(12), e78506 (2013)
4. M. Pal (2005) Random forest classifier for remote sensing classification, International Journal of Remote Sensing, 26:1, 217-222, DOI: [10.1080/01431160412331269698](https://doi.org/10.1080/01431160412331269698)
5. Giuseppe Amato,Paolo Bolettieri,Fabrizio Falchi,Claudio Gennaro.:Large Scale Image Retrieval Using Vector of Locally Aggregated Descriptors
6. Martin Sundermeyer, Ralf Schluter, and Hermann Ney.:LSTM Neural Networks for Language Modeling
7. Bermejo Nievas, E., Deniz Suarez, O., Bueno García, G., Sukthankar, R.: Violence detection in video using computer vision techniques. In: Real, P., Diaz-Pernil, D., Molina-Abril, H., Berciano, A., Kropatsch, W. (eds.) CAIP 2011. LNCS, vol. 6855, pp. 332–339. Springer, Berlin, Heidelberg (2011). doi:10.1007/978-3-642-23678-5\_39