

DEEP LEARNING: NEW APPROACH FOR DETECTING SCHOLAR EXAMS FRAUD

Sara EL KOHLI, Youssef JANNAJ, Mehdi MAANAN¹, Hassan RHINANE¹

¹ Earth Sciences Department, Faculty of Sciences Ain Chock, University Hassan II, Casablanca, Morocco

Commission 4, WG 7

KEY WORDS: Deep learning, cheating on exams, CNN, 3DCNN, image classification, Object detection, OpenCV.

ABSTRACT

Cheating in exams is a worldwide phenomenon that hinders efforts to assess the skills and growth of students. With scientific and technological progress, it has become possible to develop detection systems in particular a system to monitor the movements and gestures of the candidates during the exam. Individually or collectively. Deep learning (DL) concepts are widely used to investigate image processing and machine learning applications. Our system is based on the advances in artificial intelligence, particularly 3D Convolutional Neural Network (3D CNN), object detector methods, OpenCV and especially Google Tensor Flow, to provides a real-time optimized Computer Vision. The proposal approach, we provide a detection system able to predict fraud during exams. Using the 3D CNN to generate a model from 7,638 selected images and objects detector to identify prohibited things. These experimental studies provide a detection performance with 95% accuracy of correlation between the training and validation data set.

1. INTRODUCTION

An exam is a test at which the candidate proves the individual qualifications and skills, cheating during an exam is a dishonest behaviour, it has become a significant problem in the field of education, and one of the social evils that harm society, distort its values and lead to the robbery of rights, it is one of the biggest problems that education sector and teachers face. According to researchers and psychologists, the real reasons of cheat vary into several reasons (fear of failure, insufficient preparation, absence or irresponsibility of some supervisors during the exam) monitoring during exams is essential to prevent fraud. In the event of a proven cheating or attempted cheating, the supervisor, the head of the examination centre or any authorized person intervenes immediately. If the candidate has used a cheat sheet, a mobile phone or any other cheat object, they can be entered immediately. Usually, if supervisors are certain of the facts, they will take a report summarizing the situation. The report is then sent to the head of the examination centre. We might think that in order to supervise more students, no automation is necessary and that it would suffice to employ more people to monitor students remotely. While this is theoretically achievable, it is not possible to scale it for a course with a larger number of students taking their distance or face-to-face exam. This would require complex logistics. Nowadays, with the scientific and technological progress especially with several models of deep learning machine, it has become possible to develop a real-time cheating detection system as an alternative for human energies. The automated system will be able to detect the various moves that occur during the exam. It will be also capable to detect various kind of prohibited objects during the exams and record time of the sequence for later review by the supervisor. The current study is based on machine learning (ML) methods that include different types of neural network-based deep learning. These deep learning models have been successfully used in image-to-text conversion applications, as discussed in [1]. Our proposal system of detecting exams fraud is based on 3D convolutional neural network (3D CNN) which is a type of artificial neural network used in image recognition and processing, it is specifically designed to process pixel data. Object detection is the computer vision technique that allows us to identify and locate objects in an images and

videos. With this kind of identification and localization, object detection can be used in our suggestion system to count objects in a scene, determine, and track their precise locations, all while accurately labelling them. Image processing algorithms are also developed for detecting objects from videos, combining that with the image processing abilities of libraries like OpenCV, it is much easier today to build a real-time object detection system. In this research, the process of detecting scholar exams fraud is divide into two parts: using object detection system to identify prohibited items and notify the moderators, secondly the gesture recognition model to predicted abnormal gestures. The current training model of the proposal system achieve 95% accuracy and 0.13% as loss by using the 3D CNN.

2. RELATED WORK

Cheating on exams can result in serious consequences for the student who cheats, or breaches the exam rules and guidelines [2], this may lead students to face multiple risks like failing, being put on academic probation, or getting kicked out of school. This is why the analysis of the impact of student behaviours has been a focus of interest of research in recent years, it has also been described those perspective and demographic factors affect cheating also [3]. Most previous cheating research was based on retrospectively self-reported survey data, thus there is a need of more objective data-based evidence regarding what features are more dominant. Even though most cheating studies are based on user features, there are some studies that show the addiction of cheating on contents, instructional goals [4]. To this end, early work done by Roger [5], suggested installing the proctoring security program into computers, which enables continuous monitoring of each student's computer screen on an instructor control view. Similarly, Cluskey et al. [6], suggested an eight-step model for reducing potential cheating by students, which amalgamated several criteria that were influential to the exam, such as the duration of the exam, the time allocated for answering each question, etc. Nevertheless, this model has limitations given that it required the students to use a different browser in order to access the examination applications, and the consequent necessity for the teacher to change the questions every semester by using the randomised approach. Newly, owing to the

uprising of deep learning, gesture recognition approaches have been presented with inspiring performances compared to the traditional methods. Deep learning-based methods turn out to be popular by reason of its capability to extract spatiotemporal features from the raw video input automatically. A convolutional neural network (CNN) was firstly used for removing spatial features for static images, but it has been extended to deal with different input types or to extract different types of features. Various approaches [7], [8], [9], [10] have applied CNN to treat sequential frames of a video as multi-channel inputs for the purpose of video classification. In this section, we briefly described the previous methods relevant to our work, while significant enhancement on fraud detection has been reported, new challenges appear with different input modalities and restrictions.

3. METHODOLOGY AND ANALYSIS

3.1 Methodology

The methodology applied in the present system [Figure1] structured in two main axes TRAINING and INFERENCE including images as input dataset to guarantee the best performance. These images are divided in three phases [Test, Validation, Training] each phase contain three tags of categories [cheating, try cheating or not cheating]. The phase of TRAINING contains as an input multiple images about 7,638 approximatively; this labeled data is collected one by one from different angles, in order to create the perfect model. The CNN will generate as an output a trained model to recognize the certain types of gestures. The phase of INFERENCE, is the major phase, the trained model takes the N-dimensional array as input for predicting the correct category from the list [Cheating, trying to cheat, not cheating]. Other algorithms based on the same input also detect the forbidden objects. The result is a prediction of the correct category [Cheating, trying to cheat, not cheating], according to the student behavior.

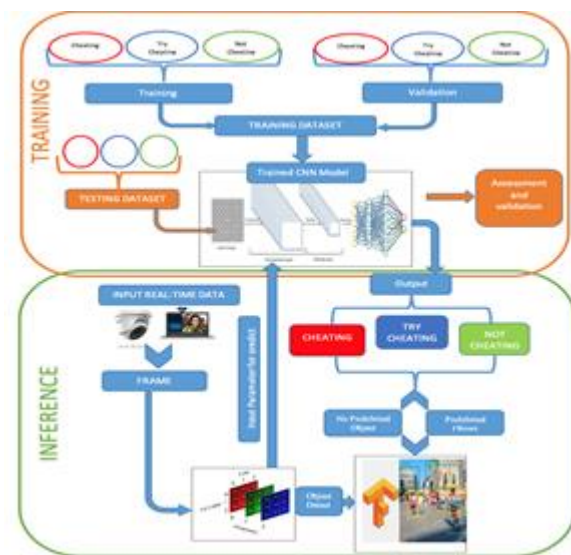


Figure 1. Methodology used

3.2 Data preprocessing

Data preprocessing is an important part of the before training and testing phase. Using a configuration with a total of 7600+ images of the dataset and shape of (400 x 400), each image was

duplicated for reading and training. This step is important before moving towards post-processing, because we have to look at what type of data we have collected and which part of it will be useful for the purpose of training, testing, and for obtaining better accuracy.

We trained a traditional deep neural network to classify user body movements. After several iterations of hyperparameters tuning and training data optimization/augmentation, a Cheating on exams can result in serious consequences for the student who cheats, or breaches the exam rules and guidelines [2], this may lead students to face multiple risks like failing, being put on academic probation, or getting kicked out of school. This is why the analysis of the impact of student behaviors has been a focus of interest of research in recent years, it has also been described those perspective and demographic factors affect cheating also [3]. Most previous cheating research was based on retrospectively self-reported survey data, thus there is a need of more objective data-based evidence regarding what features are more dominant. Even though most cheating studies are based on user features, there are some studies that show the addiction of cheating on contents, instructional goals [4].

To this end, early work done by Roger [5], suggested installing the proctoring security program into computers, which enables continuous monitoring of each student's computer screen on an instructor control view. Similarly, Cluskey et al. [6], suggested an eight-step model for reducing potential cheating by students, which amalgamated several criteria that were influential to the exam, such as the duration of the exam, the time allocated for answering each question, etc.

Nevertheless, this model has limitations given that it required the students to use a different browser in order to access the examination applications, and the consequent necessity for the teacher to change the questions every semester by using the randomized approach.

Newly, owing to the uprising of deep learning, gesture recognition approaches have been presented with inspiring performances compared to the traditional methods.

Deep learning-based methods turn out to be popular due to its capability to extract spatiotemporal features from the raw video input automatically. A convolutional neural network (CNN) was firstly used for removing spatial features for static images, but it has been extended to deal with different input types or to extract different types of features. Various approaches [7], [8], [9], [10] have applied CNN to treat sequential frames of a video as multi-channel inputs for the purpose of video classification. In this section, we briefly described the previous methods relevant to our work, while significant enhancement on fraud detection has been reported, new challenges appear with different input modalities and restrictions after training for 80 epochs, the model achieved 95% of accuracy (a) and 0.13% as a loss (b), our final model had good accuracy and was able to satisfy our approach requirements, as shown in [Figure 2].

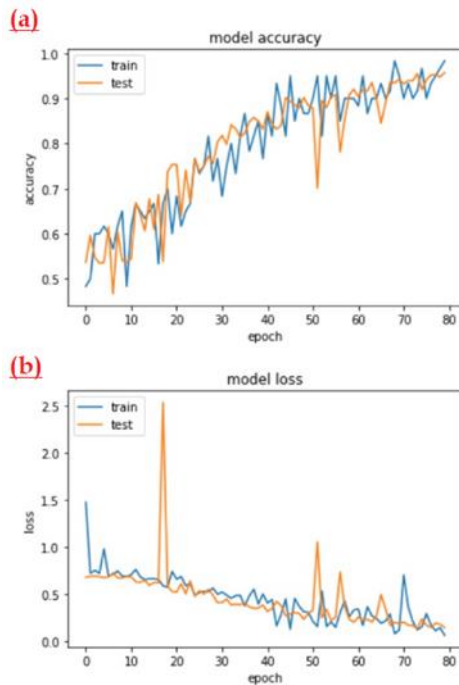


Figure 2. Model accuracy (a) - Model Loss (b)

3.3 Analysis

The idea is to make this new system learn normal behavior, then anything that does not fit into that normal is classified as abnormal and requires human supervision. The most common methods of doing this are 3D convolutions or recurrent networks which allow the temporal aspect of the sequence to be considered. Machine learning models are able to learn from patterns of normal behavior

They are very fast to adapt to changes in that normal behavior and can quickly identify patterns of fraud transactions. This means that the model can identify Cheating Students even when they only have the attention to cheat. It is absolutely necessary to keep the temporal information which is crucial to analyze a behavior. Without this, a learning model would only see a series of inconsistent information and have no notion of the time that connects them. To remedy this, two methods exist, convolutions or recurring networks. These two methods are for example used to analyze abnormal videos, extract information from videos, or recognize gestures (optimization of 3D convolutions). The goal of the first method, 3D convolutions, is to convolve on the two spatial dimensions of an image, plus a "temporal" a series of images. This is achieved through the use of three-dimensional kernels as shown in Figure 2 (unlike 2D convolutions, see Figure 3). Here, all the weights correspond to the two spatial dimensions for the same image. As for the other weights, they correspond to a single point in space, but for different images in time. Several 3D convolutions can be applied in a row or a single one is enough to get rid of the temporal data, and return to more usual models processing two-dimensional data.

The other approach can be to apply a recurring network [12], [13]. These networks, unlike convolutions which have a fixed time window that you move, work by iterating over the elements of the sequence while keeping the past in internal memory. However, since these cannot process a series of two-

dimensional data, a convolutional network (2D) is applied to each image to reduce them to vectors of one dimension. They will then be exploitable by a recurring network. Also, we can pass the images one by one to this network or give it the difference between two consecutive images. With these methods, temporal information is preserved. We can therefore obtain spatial information to detect abnormal behavior (e.g. the student is no longer in the field of the camera), but above all temporal information linking several images which individually could be considered as compliant, but considered abnormal when analyzed in sequence.

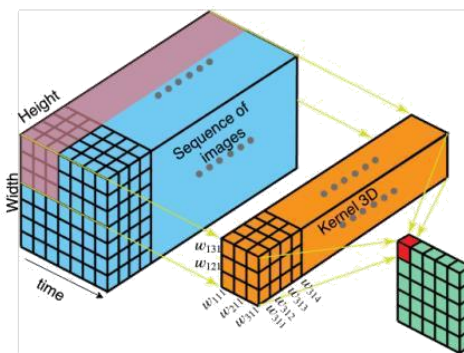


Figure 3. Example of a 3D convolution where the convolution kernel is a cube instead of a square as for a 2D convolution

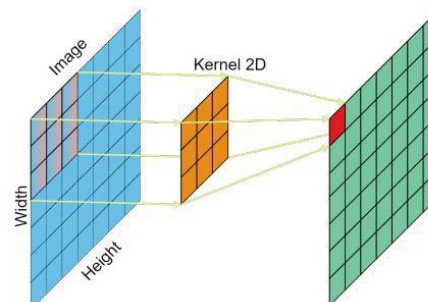


Figure 4. Example of a 2D convolution. The goal is for the network to learn the weights of the convolution kernel (in orange)

3.4 Gesture Recognition Model

Gesture recognition is an important topic in computer vision because of its wide range of applications to identify gestures. Actions from images and video sequences are challenging due to issues such as background image and lighting ratio. Gesture recognition is the alternative user interface for providing real-time data to the computer. Instead of typing with keys or tapping on a touch screen, a motion sensor perceives and interprets movements as the primary source of the input data. This is what happens between the time when a gesture of fraud is made and the computer reacts. A camera feeds image data into the trained model, then, the fraud system identifies meaningful gestures from the predetermined gestures where each gesture is matched to one of the categories [Cheating, trying to Cheat, not Cheating]. The system then correlates each registered real-time gesture, interprets the gesture and identify meaningful gestures that match the good category. Most of the

gesture recognition methods exploit the human pose estimation, however often the pose estimation in the video can often be inconsistent because of independent errors in the sequence of frames. Additionally, the extraction of skeleton joints may not be successful when the background contains multiple objects or human beings. In this paper, we try to solve these problems through the use of segmentation of the target person. In this part of our development, we considered existing body movement recognition algorithms. Presently, mainstream algorithms are frequently based on analyzing sequential video frames. Although these algorithms could meet our necessities.

3.5 Object detection

Object detection is a computer vision technique that deals with detecting instances of semantic objects of a certain class, this computer technology allows us to identify whether the students are taking any illegal objects. Object detection is also a supervised problem, which means that training the model on labeled data is an important step [14]. The object detection network is trained on the annotated data until it can find regions in images that correspond to each kind of object [Figure 5].

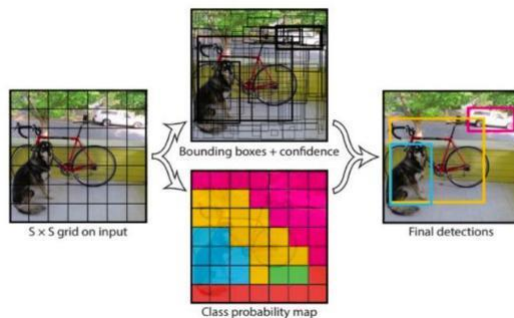


Figure 5. Process of object detection

Our dataset contains five cheating classes, which are giving cheat sheet, exchanging papers, using earphones, mobile phone or any illegal object [Figure 6].

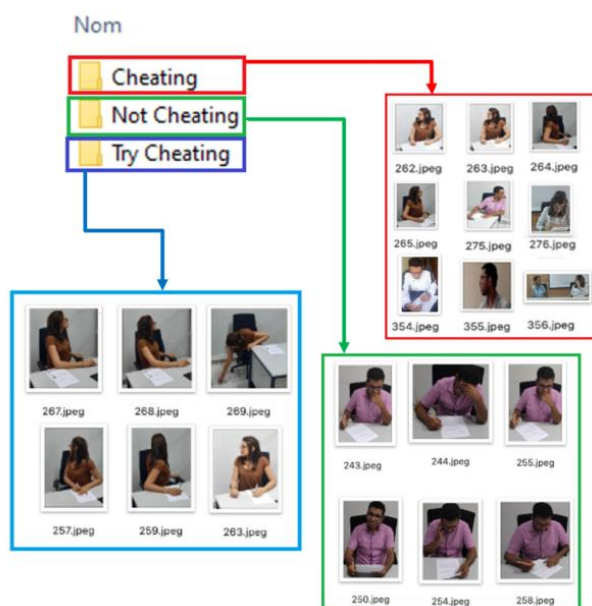


Figure 6. System classes

4. RESULTS AND DISCUSSION

According to the Moroccan Minister of National Education, Vocational Training, Higher Education and Scientific Research, the phenomenon of cheating in the baccalaureate exams for the year 2021 has been addressed; 4235 cases of fraud were detected, an increase of 116 percent compared to 2020, this increase is due to several factors, In this way, the educational sector can continue to keep beating those cheats Cheaters will continue to use new technology to dodge current rules and restrictions, and educational institutions must continue to fight this on an equal footing. Our connected world is transforming rapidly, and we must be able to use the benefits of this change and especially the deep and machine learning rather than ignoring the problem and hoping that it goes away. This New approach for detecting scholar exams fraud has so many advantages like reducing human resources, energy, minimizing the logistics and especially guarantee no cheating. To create this new system, the major step is to generate the best model, the training, validation and testing phases are obligatory in the input data set, each phase contain three tag of categories (try cheating, cheating, or not cheating), both components are combined into a single app that generate descriptions based on the sequence to predict cheating gestures. More formally, it is necessary to encode the data in a vector space of greater dimension, which is used for the training. The system will then learn a decision boundary and text description among this data, in order to obtain a better representation for the classification of the student's behavior. In general, there are several deep learning algorithms like Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) for image classification, in this proposal approach we used the three algorithms in order to make a comparison and to select the best and the most efficient. Therefore, the choice was to the Convolutional Neural Network (CNN) the efficient recognition algorithm for the image classification due to the high accuracy and the assessment of the performance testing in real-time images. The table below [figure 7] illustrate the different accuracies.

Algorithm Result	3D CNN MODEL	RNN MODEL	LSTM MODEL
ACCURACY	0.95	0.72	0.76
Loss	0.13	0.66	0.52
Real-Test (600 images)	89% (534)	58% (348)	65% (390)
Treatment Time	8h43m34s87 ms	9h24m00s01 ms	7h58m28s98 ms

Figure 7. Classification algorithm comparison

Additionally, the system is able to discriminate any element belonging to this space. It is possible to classify behaviors considered normal on one side and all the abnormal elements on the other. For example, if the student raises the right arm, the system has never seen this behavior before; it will then classify this as abnormal and write a short description according to the category cheating or try cheating [Figure 8]. It is then up to the supervisor to decide whether this behavior deserves more attention or not. Accordingly, it was necessary to provide enough normal behaviors, so that the system can learn and generalize about the behavior of students during the exam period. As a result, this approach may reduce the number of supervisors and avoid the irresponsibility of some moderators

and achieving equal opportunities among the examinees. Therefore, this type of method cannot be successful without human intervention, and a high quality of material, but this can drastically reduce the attention required.

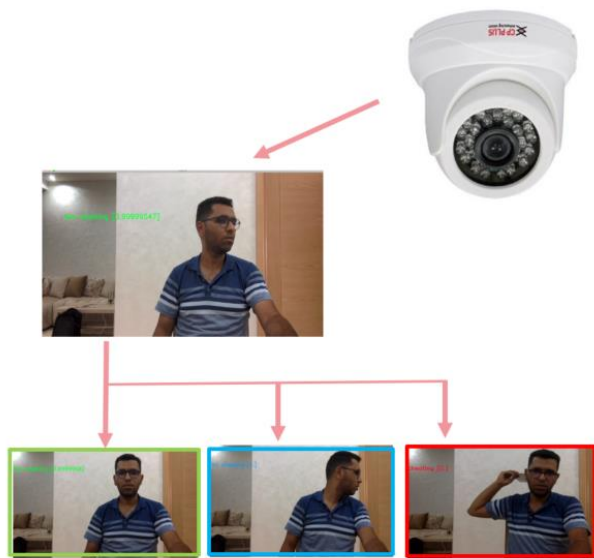


Figure 8. Real-time detection of scholar exams fraud

5. CONCLUSIONS

Many scholars have sought to map out how and why students cheat, but few have addressed the larger significance of cheating as a social phenomenon. If the high school exit exam is indeed “a social exam,” then cheating on that exam can be considered a social statement. We have argued that cheating is a way of surviving the “social exam” and claiming greater opportunities for students in futures. In this paper, we presented an original and new approach for detecting scholar exams fraud, with gesture recognition by using multiple RGB images. We have designed a cheating video description system that detects cheating activities between students and generates text description describing the cheating events. The gesture recognition model achieves recognition of the cheating gestures with 95% accuracy. With the help of deep learning algorithms and neural networks, machines can be taught to see and interpret images in the way required for this particular system. Progress in the implementation of AI-based image processing is impressive and opens a widespread range of opportunities in this field. In closing although this model may not work perfectly every single time and there is a lot of what we can do to improve its performance, it gives us a starting point to build on these fundamentals a lot of improvements.

REFERENCES

- [1] Y. LeCun, Y. Bengio, G. Hinton. 2015. "Review: Deep Learning." *Nature* Vol. 521 Halaman 436 - 444.
- [2] Examination information and rules: Examinations Office The University of Adelaide Adelaide, South Australia 5005 Australia- June 2021 <https://www.adelaide.edu.au/student/exams/examinations>
- [3] K. D. Bogle, “Effect of perspective, type of student, and gender on the attribution of cheating,” *Proceedings of*

Oklahoma Academic Science. Oklahoma City, vol. 80, pp. 91–97, 2000

- [4] Y. Kauffman and M. F. Young, “Digital plagiarism: An experimental study of the effect of instructional goals and copy-and-paste affordance,” *Computers & Education*, vol. 83, pp. 44–56, 2015.

- [5] C. F. Roger, “Faculty perceptions about e-cheating during online testing,” *Journal of Computing Sciences in Colleges*, vol. 22, no. 2, pp. 206–212, 2006

- [6] G. R. Cluskey, C. R. Ehlen, and M. H. Raiborn, “Thwarting online exam cheating without proctor supervision,” *Journal of Academic and Business Ethics*, vol. 4, no. 1, pp. 1–7, 2011.

- [7] Feichtenhofer, C.; Pinz, A.; Zisserman, A. Convolutional two-stream network fusion for video action recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1933–1941.

- [8] Karpathy, A.; Toderici, G.; Shetty, S.; Leung, T.; Sukthankar, R.; Fei-Fei, L. Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 23–28 June 2014; pp. 1725–1732.

- [9] Kopuklu, O.; Kose, N.; Rigoll, G. Motion fused frames: Data level fusion strategy for hand gesture recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Salt Lake City, UT, USA, 18–22 June 2018; pp. 2103–2111.

- [10] Molchanov, P.; Yang, X.; Gupta, S.; Kim, K.; Tyree, S.; Kautz, J. Online detection and classification of dynamic hand gestures with recurrent 3D convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 27–30 June 2016; pp. 4207–4215.

- [11] Lee, H.K.; Kim, J.H. An HMM-based threshold model approach for gesture recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 1999, 21, 961–973.

- [12] Hussein, M.E.; Torki, M.; Gowayed, M.A.; El-Saban, M. Human action recognition using a temporal hierarchy of covariance descriptors on 3d joint locations. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, Beijing, China, 3–19 August 2013; Volume 13, pp. 2466–2472.

- [13] Dardas, N.H.; Georganas, N.D. Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques. *IEEE Trans. Instrum. Meas.* 2011, 60, 3592–3607. [CrossRef]

- [14] Feichtenhofer, C.; Pinz, A.; Zisserman, A. Convolutional two-stream network fusion for video action recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1933–1941