Computer Vision-Based Unobtrusive Physical Activity Monitoring in School by Room-Level Physical Activity Estimation: A Method Proposition

Hans Hõrak [1,2]

1   Institute of Social Studies, University of Tartu, Lossi 36, 51003 Tartu, Estonia; hans.horak@ut.ee
2   iCV Lab, Institute of Technology, University of Tartu, Nooruse 1, 50411 Tartu, Estonia

check for
updates

Abstract: As sedentary lifestyles and childhood obesity are becoming more prevalent, research in the field of physical activity (PA) has gained much momentum. Monitoring the PA of children and adolescents is crucial for ascertaining and understanding the phenomena that facilitate and hinder PA in order to develop effective interventions for promoting physically active habits. Popular individual-level measures are sensitive to social desirability bias and subject reactivity. Intrusiveness of these methods, especially when studying children, also limits the possible duration of monitoring and assumes strict submission to human research ethics requirements and vigilance in personal data protection. Meanwhile, growth in computational capacity has enabled computer vision researchers to successfully use deep learning algorithms for real-time behaviour analysis such as action recognition. This work analyzes the weaknesses of existing methods used in PA research; gives an overview of relevant advances in video-based action recognition methods; and proposes the outline of a novel action intensity classifier utilizing sensor-supervised learning for estimating ambient PA. The proposed method, if applied as a distributed privacy-preserving sensor system, is argued to be useful for monitoring the spatio-temporal distribution of PA in schools over long periods and assessing the efficiency of school-based PA interventions.

1. Introduction

In the recent four decades, a 10-fold increase in the number of obese children and adolescents has been observed and it is estimated that almost one in every five children globally are overweight [1]. Meanwhile, physical inactivity (PI), which has been associated with various health risks [2], and which is also one of the main contributors to overweight, has been described as a global pandemic [3]. Concurrently, smartphones have become more accessible even to lower-income families and this is enabling screen time to increasingly compete with healthier activities in the temporal budgets of the youth even outside of their homes. Children and adolescents spend a large part of their time in school where their health behaviour can be researched and possibly influenced. So far, school-based physical activity (PA) interventions have mostly shown modest [4–7] and only temporary [8] effects on PA, if at all [9]. There are still many ambiguities in this field [10–12] due to limited evidence. To maximize impact on public health, evidence-based best practices of PA interventions should be determined before making large investments into scaling up the intervention programs [13,14].

2. Methods of Assessing Physical Activity of Children

Physical activity is defined as bodily movement via skeletal muscles that results in energy expenditure (EE) [15]. Measurement of PA in the context of the PI epidemic is mostly concerned with

assessing habitual PA and determining whether some populations of youth are meeting the established guidelines [16] of 60 min or more daily moderate to vigorous PA (MVPA) with moderate PA defined by the World Health Organization as 3–6 Metabolic Equivalent of Task units (METs) and vigorous PA above 6 [17]. MET is the PA intensity unit defined by the ratio of a person's working metabolic rate relative to their resting metabolic rate with individual metabolic differences normalized based on body weight [18]. The PA (proxy-) measures described below are often converted to this metric.

This work is concerned with brief expressions of PA (bodily movement lasting no more than a few seconds—PA microexpressions) that might be wholly observable in an indoor video camera's field of view (FoV). Therefore, the descriptions of methods do not go deeply into concerns of population-level PA inference, but rather the relation of the measurement techniques to age- and context-specific PA patterns observable in school.

Methods with varying levels of objectivity have been used in research on children's PA ranging from indirect approaches like survey questionnaires, interviews, and activity diaries to direct methods such as observation, and physical measurements like accelerometry, heart-rate monitoring, and doubly labelled water (for an overview see [19]). Assessing children's PA with indirect measures has shown to be unreliable, often overestimating PA [20–22]. Self-report or parent-assisted measures, while relatively cheap, suffer from reliability and validity issues concerning inaccuracy of assessment, recall, and social desirability bias [19,23].

Direct systematic observation can provide rich insight into the PA dynamics of a group of children in a specific context—one can observe the subjects' interactions with each other and their immediate environment while taking notes on the intensity and duration of PA these interactions entail. Results based on such observation, however, are not strictly reliable as the observer's senses are limited and interpretations subjective. Thorough training and "recalibration" of observers can somewhat mitigate these problems and increase comparability, but it also increases the cost of applying the method [24].

Direct physical measures are often used for estimating EE in epidemiological and kinesiological research. To this end, doubly labelled water (DLW) provides accurate measures of overall EE [25], but the method only allows EE assessment averaged over long periods of time (sampling rates counted in days), it is very intrusive, expensive, and does not directly measure the construct of PA. DLW's accuracy, however, has made it a useful tool for validation of the methods described here.

Heart rate monitoring can provide high sampling rates and is well correlated with EE, but the relation varies widely between and within individuals [26]. Consequently, thorough calibration for factors like age, sex, body weight, and physical fitness is required to assess EE via heart rate monitoring [27]. Further inference of PA from heart rate monitors benefits from the additional modality of movement measured with an accelerometer [28]. Combined heart rate and acceleration sensors have been deemed valid for assessing PA of children [29]. Before an overview of accelerometry, pedometers should be mentioned as a relatively cheap and reasonably valid option for assessing the PA levels of children [30–32]. However, pedometers are essentially single-axis inertial sensors that are individually calibrated for each subject to detect their stepping patterns, so these devices are not designed to register horizontal motion and cannot quantify the intensity of PA at a given moment.

Triaxial accelerometers can provide more information by quantifying the inertial forces on each of its three axes at high sampling rates (up to 100 Hz in practice). This allows modelling of acceleration vector magnitude (AVM) in 3D space which can be corrected for gravitation (Euclidean Norm Minus One g or ENMO) to obtain a measure of the force applied to the sensor by the subject. However, due to the restricted functionalities of popular wearable accelerometers, "activity counts" (arbitrary quantities reflecting PA intensity over fixed epochs that are calculated on board during measurement) are often used in practice [19]. Accelerometers have seen wide and methodologically varied application in PA research [33–37]. Although accelerometers provide rather good indication of PA intensity and sedentary behavior, especially when combined with additional sensors such as inclinometers and gyroscopes, decisions related to sensor data management and analysis remain somewhat subjective [37–40]. Specifically, devices of different manufacturers calculate activity counts

using different formulae (which are not always published), and there is no consensus on parameters of recording and methods of aggregating acceleration data to reflect comparably [41] the concepts of moderate and vigorous PA. This has led Migueles et al. to conclude "that it is not possible (and probably will never be) to know the prevalence of meeting the PA guidelines based on accelerometer data" [42].

Since researchers mostly cannot know whether the forces reflected in acceleration signals are truly applied by the subject or whether the device is worn as instructed (sensor jitter, vehicular transport, and non-compliant uses), machine learning approaches have gained popularity for classifying the type [43] and the intensity [44] of PA from wearable accelerometers. Fergus et al. [45] explored thoroughly several machine learning approaches and feature combinations to classify children's PA type and intensity based on wearable accelerometers achieving best performance on test data with a multilayer perceptron artificial neural network. Deep neural networks have achieved state of the art performance for the prediction of PAEE in pre-school children [46]. Machine learning becomes even more relevant when considering reduced study control of wrist-worn accelerometers compared to hip-wear [47] and especially for PA monitoring via smartphone sensors, where the researcher has even less control over the positioning of the sensors in relation to the body.

Smartphones contain various sensors that can provide relevant information about the intensity and type of PA while the subject is carrying the device. Accelerometer, magnetometer, gyroscope, and GPS have a clear association with PA, but additionally light, proximity and WiFi sensors, barometers, microphones, and cameras can provide extra modalities for PA analysis (for overview see [48]). The interactive nature of the smartphone also allows for attempts at influencing the users' PA [49], which itself is an important field of inquiry for promoting PA behavior change among the youth [50]. The growing popularity of smartphone and wearable fitness apps is leading to huge amounts of data potentially useful for large-scale PA analysis. However, the differences between devices and software, privacy concerns, and data ownership issues lead to a situation where unification and comparison of data collected by different companies is very difficult [51].

An overview of the advantages and disadvantages of PA assessment methods is presented in Table 1.

Table 1. Methods used to assess physical activity in children and adolescents.

Method	Positive Features	Negative Features	Participant Burden *	Cost **
Indirect measures				
PA diary, log Interviews, questionnaires	Inexpensive	Sensitive to cognitive development; inaccuracy; social desirability bias; recall bias.	++ +	– –
Direct measures				
Observation	Potential to capture a wide variety of PA expressions and related contextual factors	Subjective (limits of perception and individual interpretation); potentially reactive	–	-/+ depending on scale
Doubly labelled water	Accurate measure of EE	Does not directly reflect PA or activity types; very low sampling rate	++	+++
Heart-rate monitor	Reflects well aerobic activity	Only captures PA from aerobic activity; requires thorough calibration for each subject		+
Pedometer	Relatively inexpensive for a wearable sensor	Cannot accurately detect intensity of PA or capture PA microexpressions.	-	-
Wearable accelerometer	Widely field-tested and validated, machine learning enables PA type and specific activity recognition	Differences between devices; no consensus on acceleration signal processing and aggregation to standard PA indicators		+
Smartphone sensors	Rich sensor data; possibility to ask questions after detecting bouts of PA	Limited battery life; often not attached to body; differences between devices	-/+ depending on use	
Proposed computer vision approach	Unobtrusive; context specific; long measurement period	High initial investment; not yet validated	—	++ increasing returns

* "+" indicates relatively high participant burden and/or intrusiveness.** "+" indicates relatively high monetary and/or labor cost. Compiled by author based on [19,20,23,24,42,48].

Compared to adults, children's PA is intermittent in nature [52,53], so methods analyzing their PA patterns should consider higher sampling rates and shorter PA intensity estimation epochs than is required for measuring adults. It is also important to consider that it is more difficult to achieve high accelerometer wear protocol compliance in children, and especially early teens [54–56]. Individual-level measurement methods described above often assume recording and processing of personal information by researchers and are generally intrusive, requiring human research ethics reviews, the subjects' and their parents' informed consent, and the general bothering of subjects. Intrusiveness also potentially compromises the results by observer effects. Below, ways of overcoming these limitations in school-based PA research are explored.

3. Spatio-Temporal Distribution of Physical Activity in School

Schools are very specific semi-closed environments where children are required by law to spend lots of time. In OECD countries children spend on average 14% of their waking hours in compulsory classes during primary education (calculated assuming 8 h of sleep based on [57]). If one only considers the school year (September to May) and counts in recess between the classes, then it adds up to a large proportion of time spent in this specific environment. Parts of these spaces with differing attributes can facilitate more or less PA. Playground size has been shown to correlate with PA [58], but the evidence on the relation of other aspects of school architecture to PA is insufficient [59]. There is some evidence indicating playground redesigns', markings', and physical structures' positive effects on PA [60], but others have reached conflicting results [61]. Specifically, there is a lack evidence on which kind of playground equipment and their specific features have the strongest and longest-lasting effects on PA [58]. While one should strive to design the perfect playground for increasing all students' PA, one size might not fit all. Boys and girls of different ages have significantly different play preferences [62] and might require different stimuli for increasing PA [63]. Ethnographic evidence also suggests specific approaches to playground and classroom designs might be necessary for motivating the high-risk group of least physically active students [64]. In addition, Nicaise et al. [65] reported that the effects of playground redesign on PA might not reflect in wearable sensor data, while observations imply a positive effect.

Exergaming, as a branch of the emerging health behavior intervention paradigm of gamification [66–68], has received much attention [69–72] concerning the PI epidemic. The idea of taking advantage of the neurochemical reward mechanisms utilized in the gaming industry to achieve positive health outcomes is becoming increasingly relevant in the context of pervasive computing. Prevalence of smartphones and wearables in combination with increasing feasibility of integrating gaming hardware into the school environment provide a valuable opportunity for the gamification of PA. Baranowski et al. [73] defined the identification of optimal game designs for attaining PA change as an important research priority while emphasizing specific game context (e.g., recess on playground, in hallway or classroom before or after lunch) and context-specific game design elements (cooperation or competition with self or others while using various reward systems).

All of this infers a need for informed and efficient zoning of schools to facilitate increased PA for all students throughout the school year—ideally a custom design for each school, season, and day of week within a season. Location-based PA information can be useful for determining the areas that facilitate or hinder PA and for assessing the utilization rate of a playground, its sections or specific stationary PA equipment.

So far, the spatial distribution of physical activity in school has been studied in schoolyards using GPS combined with heart rate monitoring [74] and accelerometry [75,76]. However, GPS signals are sensitive to environmental factors such as tall buildings [77] and cannot reveal the altitude of the sensor, making it inapplicable in multi-level buildings.

4. Method Proposition

To better understand the spatial aspects of PA in school while also minimizing participant burden in the research process, a hypothetical method for assessing the spatio-temporal distribution of PA in school is proposed: ambient sensors capable of detecting the number of children at a location and classifying the intensity of their PA in real-time without recording any personal information. A computer vision application for accurate estimation of ambient PA based solely on video frames temporarily stored in random-access memory (RAM) could be a viable solution. Adopting a smart sensor with such capacity would allow researchers to delegate the processing of personal information to artificial intelligence thus obtaining PA estimations at a location without violating the subjects' privacy or bothering them at all. The proposed method can be considered as automated direct observation, except while the type and severity of human error during observation can be variable, algorithmic errors should be consistent and therefore easier to account for. Just one or a few of such sensors could suffice for assessing the effectiveness of stationary PA equipment and stimuli aimed at increasing PA at a specific location. Covering a school building with a distributed sensor system could provide a flow of location-based PA data at high temporal resolutions over any length of time that the system is maintained. Internet Protocol (IP) cameras with relatively wide FoV could be placed at strategic locations throughout the building, or alternatively with a uniform distribution to capture the PA in the building. As a semi-closed environment with students arriving and leaving based on a known time schedule, even a rather sparse distribution of the sensors could potentially reveal hallway-, floor-, and school-level PA patterns. Ability to detect long-term building-level changes in PA patterns enabled by continuous monitoring of ambient PA could open a new field of intervention research designs. Proposed sensors could also be useful for other settings where one is interested in assessing PA at a location in a privacy-preserving manner.

The output of a single sensor, or the basic measurement unit of the method, is currently envisioned as PA intensity of detected child during the length of the prediction epoch. One sample from a single sensor would be the PA intensity levels for each detection during a prediction period/frame range (varying number of values, depending on the number of children visible). In other words, the sensors would measure the intensities of brief displays of PA in FoV (ambient PA) as opposed to measuring PA of individuals. For example, one student can step out of the perceptive field of the sensor during a second and another in during the next second. Then, if both students were moving at the same PA intensity during the corresponding successive predictions, the measure of ambient PA would remain the same during the 2 s, even though originating from separate individuals.

The raw output of the proposed distributed sensor system could be aggregated and visualized on a 2D graph of a single floor plan or a 3D model of the whole school building where the size of a circle/sphere could represent the average number of students detected by a particular sensor during some period and a color scale could be used to represent the average intensity of the PA of the detections during the period. One can imagine a graph of a school floor plan where at the locations of the sensors a small blue dot would signify a single student standing still; a large purple circle indicating lots of detected students in the scene, but a medium average PA; a large red circle could indicate lots of students performing a group activity entailing vigorous PA. Similar visualizations could be done for various aggregations, computing the average number of detected children and the average PA intensity of the detections during a longer period at the sensor locations. These could then be further aggregated to reveal seasonality (average first recess PA distribution, average Monday within a semester PA distribution, etc.) or for pre-post intervention testing (average PA during the weeks before, during, and after an intervention).

For observing changes in whole school PA levels, additional measures can be taken to increase reliability. In schools that record the number of students in the building at the beginning of each lesson, the sensor system data could potentially be normalized by considering the number of students present and measures such as accessible floor area of the building and the floor area monitored by the sensors. Such "student-density" measures should increase over-time and between-school comparability of the

estimated ambient PA levels, especially during winter in colder climates when students remain indoors during recess.

Combining ambient PA measures with direct observations, interviews, and/or questionnaires would enable thorough analysis of PA in school. The value of such data could be further increased by simultaneously collecting rich contextual data such as lunch menu and its estimated sugar content, weather conditions, concurrent events, group vaccination, student sick leave rates, etc. Even though such a method would not have the capacity to reveal whether the students achieve their recommended hour of daily MVPA, it could be a valuable tool for assessing the capacity of an intervention to activate children on location and whether the PA reactions to intervention remain similar over time.

5. The Promise of Computer Vision

Motion detection and object (such as a human) tracking tasks have received much attention in the computer vision field [78–83]. When using stationary cameras in a school building where the background should be mostly static at a given length of time, background-subtraction methods [84,85] could potentially be used for obtaining a proxy for PA intensity as the ratio of black and white pixels in the subtracted image. However, such a simple approach would likely be sensitive to variance of scale, changes in illumination, and would not differentiate between the PA of children and any other motion. A more advanced approach to ambient PA estimation stems from human action recognition (HAR) (for an overview see [86]). HAR algorithms are usually developed and tested on datasets containing up to 101 actions such as brushing teeth, bowling, frisbee catch, playing guitar, baby crawling, band marching, etc. [87]. Since actions are defined through time, HAR research emphasizes temporal features (difference between consecutive video frames), while object recognition algorithms are mostly concerned with just the spatial information (the image). Thanks to advances in hardware and machine learning methods, HAR has seen rapid development in recent years [88]. Simultaneously applying two convolutional neural networks (CNNs), one for the spatial, and other for the temporal domains, has shown to be an effective approach for learning features of many abstract actions from video [89]. These two-stream methods have been shown to benefit from fusing together the spatial and temporal features learned by the separate networks to increase recognition accuracy. This fusion can be applied in the convolutional layers (early fusion) [90] or the fully connected layers (late fusion) [91,92], either approach can provide task-specific feature learning benefits.

Significant advances have recently been made in processing efficiency in action recognition. Several approaches have managed to reduce the computational cost of the task to enable real-time action recognition on established benchmarking datasets [93–95]. Singh et al. [93] translated the Single Shot Detector [96] network architecture, designed for rapid detection of multiple objects in images, to the action recognition task in the temporal domain resulting in capacity for online independent construction of multiple "action tubes" containing the humans whose actions are to be classified. By applying a novel greedy classification algorithm to the tubes, they achieved performance superior to state-of-the-art algorithms that are not capable of online action localization and did it all at real-time speeds. Such online capacity is especially important for the proposed method whereby PA intensity predictions are to be made at a constant frequency based on live video input.

Zhang et al. [94] combined several methods of knowledge transfer, enabling a CNN operating on low resolution motion vector images to utilize the knowledge of another CNN learned from high-resolution optical flow allowing reasonable action recognition performance at more than real time-speeds. Another approach [95] applied the efficient object detection architecture of YOLO^{v2} [97] to the output of FlowNet2 [98] (an optical flow estimation CNN) as the temporal stream and the same architecture to the spatial stream. Task-specific fine tuning and integration of FlowNet2 into the two-stream architecture in combination with early fusion of the spatial and temporal features enables end-to-end trainability and real-time speeds [95].

Another recent HAR innovation proposed by Li et al. [99] introduced convolutions exploiting the spatial correlations in images for efficient motion-based action localization by using a Long Short-Term

Memory cell (LSTM—a neural network mechanism that enables "remembering" previous states [100]) between convolutional instead of fully connected layers. Spatial correlations between the previous hidden state provided by the VideoLSTM and the current input reveal the likely location of an action based on motion, thereby making the whole process more efficient. Such motion-based attention can be especially useful for the intended setting utilizing stationary cameras. Building on this technique of motion-based attention, Zhao and Snoek developed an algorithm for detecting the spatiotemporal extent of actions by embedding the RGB spatial and optical flow temporal streams into a single two-in-one stream network [101]. Aside from simplifying the computation of action recognition, their approach also assigns motion direction to the actor as an extra feature distinctive to many actions (e.g., the difference between sitting down and standing up, or PA-entailing motion towards a direction relevant to the research questions studied with proposed smart sensors).

The processing speed and energy efficiency of proposed smart sensors could potentially also benefit from Deep Compression [102]. This technique, developed by Han et al., minimizes redundancies in deep neural networks by pruning ineffective connections, quantizing the weights and Huffman coding the resulting weights' distribution. Such compression, when applied to convolutional neural networks, was accompanied by a 3–4-fold increase in processing speeds and 3–7-fold increase in energy efficiency without significant loss of classification performance. The size reduction accompanied by Deep Compression allows to fit large neural networks in on-chip SRAM, thus removing the need for accessing DRAM during processing, which consumes the most power during neural network operation. Han et al. [103] propose a specific hardware design that would take full advantage of Deep Compression and power efficiency of SRAM-based computation (120-fold energy saving compared to DRAM-based implementations). Furthermore, novel hardware architectures utilizing the emerging Resistive RAM technology are being developed precisely with the goal of efficient neural network computation on very small chips [104,105]. While currently the proposed distributed sensor system is planned as a centralized computing implementation, the developments in specialized low-power artificial intelligence chips infer the possibility of a potential distributed computing implementation in the future.

Considering the machine learning task described below and potentially a low number of classes to be distinguished, using single-channel greyscale input might also be a viable option for further reducing network size and computational cost. Similarly, lower resolutions of input might be considered as the indoor environment forces relatively small distances between the subjects and the camera. Whether such approaches would be accompanied by severe loss of performance is to be determined with experimentation.

6. Action Intensity Classification by Acceleration Vector Magnitude Estimation

Supervised learning in video analysis is usually implemented by assigning semantically subjective labels to frames of video and learning the "typical" features from instances of visual data with such labels. One approach to action intensity classification would be to create a training dataset where the frames of moving children are annotated by visually assessing the intensity of PA displayed in a given range of video frames. Instead, this work proposes to annotate the data based on real accelerations measured from subjects in the training video. Such an approach could essentially fuse the sensors and the research fields of accelerometer-based PA monitoring and video-based HAR. Synchronizing accelerometers recording at 30 Hz with a video camera recording at 30 fps can create raw training data where each frame corresponds to three acceleration scores on the accelerometer's axes for each visible subject. AVM can be then calculated for each subject in each frame, which can then be preprocessed and aggregated (e.g., cumulative sum or average) according to the acceleration prediction frequency, forming the ground truth.

To explore the potential of such a dataset, a sample of proposed training data was collected using a Logitec C922 webcam (Logitech International S.A., Newark, CA, USA) and four Actigraph wGT3x-BT accelerometers (ActiGraph LLC, Pensacola, FL, USA) worn on the hips of three children

and one adult as actors. The camera's FoV covered a ~5 × ~4 m floor area with a fixed camera at ~2 m height. Table 2 presents the analysis of a 10.4-min synchronized clip where all four subjects or at least their torsos are mostly visible. For the most part, the clip contains structured play where the adult is acting as the game leader/instructor. It is important to note that due to the nature of the games played, the clip contains significant amounts of synchronized motion (game leader says "Go!" and children react by jumping/moving) which can increase the correlation coefficients. The bottom row of Table 2 corresponding to the section with least synchronized motion should represent better a real-world situation in a school hallway. Inversely, the webcam's auto-focus created some noise in the motion features which can somewhat reduce the correlations compared to using stable focus. Columns of the table present different proposed acceleration prediction frequencies as cumulative aggregates of both total motion information in the video (represented as total H.264-encoded motion vector magnitude per frame) and acceleration domains (represented as sum of all four subject's ENMO).

Table 2. Correlations between total visible physical activity* and motion information in video** cumulatively summed over various frequencies in a 10.4-min sample of proposed training data.

	Raw (30 Hz)	15 Hz	10 Hz	6 Hz	5 Hz	3 Hz	2 Hz	1 Hz
First 2 min	0.331	0.449	0.469	0.559	0.564	0.656	0.673	0.669
min 3–4	0.138	0.210	0.234	0.309	0.343	0.476	0.541	0.574
min 5–6	0.248	0.374	0.400	0.464	0.492	0.575	0.610	0.641
min 7–8	0.338	0.450	0.481	0.529	0.552	0.607	0.641	0.630
min 9–10	0.316	0.422	0.453	0.529	0.556	0.657	0.696	0.695
Whole clip (10.4 min)	0.279	0.387	0.416	0.486	0.511	0.602	0.640	0.646
Independent play (min 1.9–4.7)	0.140	0.217	0.240	0.312	0.350	0.477	0.532	0.574

All correlations statistically significant at $p < 0.001$. * Total visible PA represented by the sum of four subjects' ENMO measured with hip-worn Actigraph wGTX3-BT accelerometers. ** Motion information in video represented by total H.264-encoded motion vector magnitude per frame (sum of macroblock displacement distances for I (0), P, and B frames at 1920 × 1080 resolution extracted with modified version of MV-Tractus [106]) recorded with a Logitech C922 webcam.

Analysis of the sample of raw training data shows moderate correlations between the temporal features in video and the accelerations of subjects in the scene. If one considers the task of the algorithm as regression of video to AVM of objects in FoV, then a 0.5 correlation with target in the temporal stream infers that even a relatively simple spatial stream architecture might provide the additional features for accurate estimation of AVM intervals. In a simple form, the temporal stream could quantify the total visible motion and the spatial stream could count the number of children to get average PA intensity per subject per prediction epoch. A more accurate model would separate the motion of children from other motion, noise, and changes in illumination while the RGB stream would not only count the children but based on features such as body position, and its variance along the action tube, also gain information about how much of the motion is associated with which subject. This can provide additional information on the nature of ambient PA displayed in a specific scene with several subjects. Deep learning models could become even more precise to reflect PA EE if training data labels were calculated accounting for the weight and/or the body mass index (BMI) of the actors. The formula for calculating the PA EE proxy labels for the whole dataset could likely be derived from thorough analysis of only a few clips where actors of various body types perform activities entailing the full range of PA intensities. Sliding window averaging of the accelerations prior to label aggregation can potentially enhance feature learning and the method's construct validity by reducing the effects of device jitter and sensor noise.

As the analysis presented in Table 2 concerns a clip where subjects do not step into or out of the cameras FoV, it does not reflect the eventual PA measurement setting very well. Since the proposed measurement units are defined by the prediction frequency (30 frames or 1 s for in Figure 1), some criteria should be developed for what should constitute a valid detection. For example, when a

child or their head/torso appears in the corner of the scene creating an action tube of length only 15 frames within a 30-frame epoch, then it likely should not be classified. Action tubes of length 25 within the epoch, on the other hand could already carry enough features to accurately assess the level of PA performed by the detected child during that second (perhaps they just stepped into the scene five frames after the beginning of the current prediction epoch or stepped out before the end of the prediction epoch). Such detection-validity thresholds should be determined by thorough experimentation and expert assessment of the algorithm's performance on the test set.
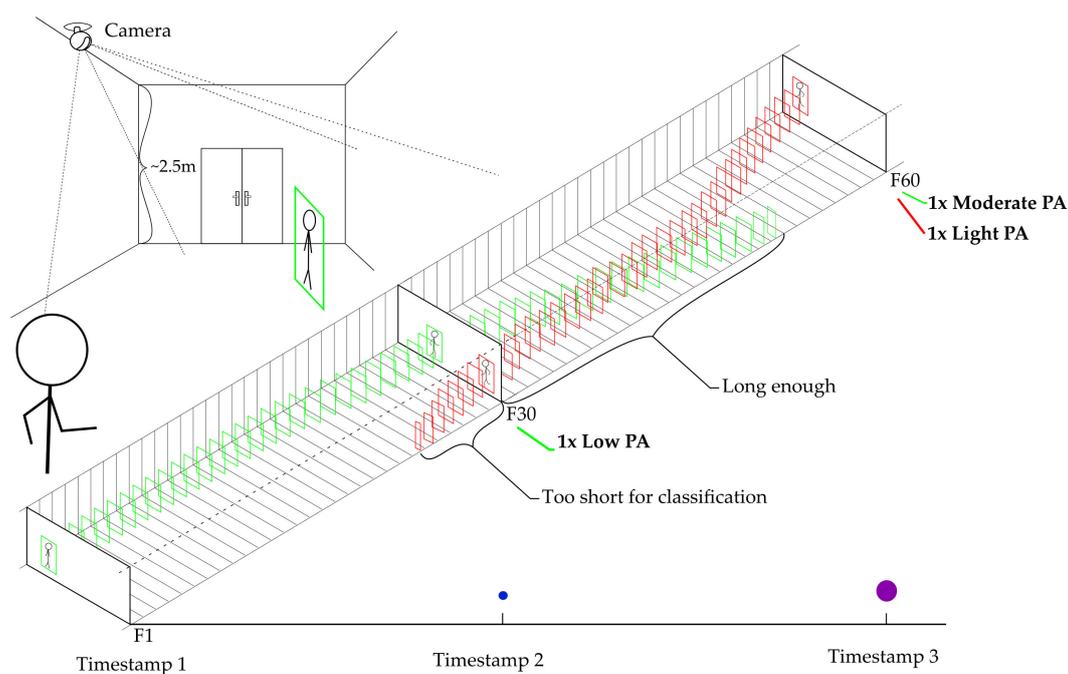
Figure 1. Classifying the intensity of ambient physical activity at a constant frequency (30 frames/~1 Hz).

The optimal period of aggregating acceleration scores to ground truth labels depends on the temporal resolution requirements of real-time implementation on one hand and the need for a good representation of age- (intermittency) and environment-specific (bouts wholly observable within the FoV of the camera in a school hallway) PA patterns on the other. In the early stages of such research a 5 Hz "data unification frequency" could be convenient so that data filmed with cameras using NTSC (29.97 or 59.94 fps) and PAL (25 or 50 fps) standards could be easily united in the dataset. However, this would restrict the selection of the final prediction epoch to lengths divisible by 200 ms and would not allow cumulative aggregation of accelerations if both types of cameras would be combined with a 30 Hz accelerometer sampling rate. On the other hand, using different acceleration sampling rates could compromise construct validity by distorting the ground truth. Additionally, to maintain some comparability with other accelerometer-based PA research, using NTSC cameras and 30 Hz acceleration sampling might be preferable, as it is the most common frequency used in PA accelerometry [40].

Aside from sampling frequency, the optimal FoV should also be determined for the cameras to be used in the collection of training data. Very large FoV or even 360° cameras could provide the best floor-area coverage per sensor when applied in school, but the distortions ("fisheye" effect) in such video could make the machine learning task much more difficult and therefore compromise measurement accuracy of the proposed sensors. Hence, there is likely a trade-off between the sensor's precision and the size of its perceptive field.

A great benefit of such continuous training data is that by selecting different starting moments (t + 15|10|6|5 frames and accelerations) as data augmentation prior to label aggregation, the size of the

dataset could be multiplied with little effort and this can be helpful for learning more general features of PA intensity displays.

This kind of action intensity classification can benefit from the advances in deep learning methods applied in HAR, but the task of the algorithm is fundamentally different. HAR is mostly concerned with recognizing actions with a specific relevant function (e.g., detecting getting up from bed in a smart home to start the coffeemaker), but the proposed method would attempt to classify the intensity of any human action and inaction regardless of the goal of the behavior. In other words, very different sequences of human body positions can fall under the same action intensity category, even though they may not represent the same actions. For example, the estimated AVM can be the same for two students moving at the same intensity, but one of whom is jumping while the other sprinting. Therefore, the variance in the appearance of features in each PA intensity cluster would be much higher than for classes representing specific actions. Nevertheless, the ground truth is based on directly measured accelerations, and as initial tests show (Table 2), acceleration as a feature of PA expression reflects well in video. As such, the proposed deep learning approach does not really belong to the domain of action recognition but is better described as a form of multimodal or sensor-supervised learning where the neural network learns the features in two-dimensional spatial data representing movement and based on this knowledge makes predictions in the form of AVM intervals. This action intensity classification task is essentially an ordinal regression problem—more variance in body positions per detected child (action tube entropy) and/or bigger displacement distance in the sequence of frames (action tube shape) should indicate a bigger AVM and a higher-order PA intensity class. Due to the somewhat linear nature of the task, the CNN architecture should benefit from class correlations for class distinction—features of vigorous PA can be somewhat similar to features of moderate PA, but should differ more from the features of light PA. Class correlations have been shown to improve performance [107] even when classifying abstract actions.

In general, as a classifier, the algorithm would be working with large overlapping feature spaces of ordinal classes. The optimal neural network architecture for this type of computer vision system is yet to be determined and partly depends on the quality and amount of training data available and necessary to learn PA intensity features well enough for application as a measurement technique.

7. Discussion

This work poses the following two hypotheses: (i) room-level measurement of PA is useful for determining best practices of school-based physical activity interventions; and (ii) modern computer vision technology is capable of privacy-preserving room-level physical activity estimation. A course of action is also proposed to test these hypotheses: (ii) deep learning on a dataset of synchronized video and accelerometry; and (i) location-specific or whole-school pre-post intervention analysis of data provided by proposed smart sensors.

Construct validity of the proposed sensor to measure ambient PA is currently difficult to estimate as, to the knowledge of the author, ambient PA has not been researched in this manner. However, deep neural networks continue to perform tasks previously thought to be impossible for machines, and due to the nature of the training data, construct validity for a single sensor can be thoroughly assessed. Besides visually analyzing the PA displays and corresponding predictions and ground truth labels in the test set, the sensors could also be validated in the field, potentially using additional sensors (heart-rate monitor, thermometer and/or thermal camera) in combination with subject-specific attributes such as body weight, BMI, the weight of their clothes and back-pact, hardness and density of their shoe-soles (potential effects of footwear on hip-worn accelerometer signals and inferred PA microexpression intensity). Assessing construct validity of proposed distributed sensor system to measure school-level PA would be much more difficult. The sensors could either be strategically placed (ends of hallways and larger open areas) or alternatively by maintaining uniform distances between sensors and establishing a standard sensor-FoV-to-floor-area ratio. The former, cheaper approach could be viable to compare school-level PA over time, but comparison between schools of

different architecture would be rather limited. The latter case would allow increased between-school comparability, but the distributed sensor systems would be more complex and expensive. Ideally, such sensors could also be deployed with total coverage by maintaining some FoV overlap between the sensors. This would enable stitching together a whole-school perceptive field that would allow seamless tracking of individuals and their PA levels. Currently, such an approach seems excessive and a more sparse deployment of wide-FoV sensors should suffice to capture PA distribution patterns well enough to test hypotheses.

Occlusion in crowded scenes and the presence of adults in the building threaten the reliability of such a method. These issues should be addressed early on while creating the machine learning dataset. For the purposes of reducing occlusion, the sensors should be placed relatively high, so the spatial stream could more reliably count the subjects. For this, the training data should be collected with cameras fixed at heights varying from 2 to 3 m at viewing angles that maximize the perceptive field at specific heights—this should also ensure applicability of the sensors in various architectural settings. The data should also contain crowded scenes of various PA distributions among the crowd. To avoid false positives due to the school personnel, grownups could be included in the training dataset, but either not annotating them at all or adding a label "non-detection". The latter case would provide researchers with additional information regarding their research questions (e.g., teachers actively implementing a PA intervention), however, this would also change the regressional nature of the machine learning task and therefore increase computational complexity. Coming back to the idea of school as a semi-closed system, assumptions could be made that occlusion and grownup false positives follow a somewhat constant distribution at least within a semester of a school year. For good measure, events that bring more adults or larger crowds into the building or out of it, should be recorded as contextual data. Considering these notions, the inference of student PA distribution from such a sensor system might yet be valid even at considerable occlusion and adult presence rates. Since the proposed method would be gathering high-resolution data throughout the school year ideally several years in a row, the sheer amount of data would likely enable detection of relevant school-level PA patterns.

Aside from technical and statistical issues, human factors could threaten such a method as well. Even when certifying such a sensor system as truly privacy-preserving with no possibility to retrieve video frames from RAM, the teachers, parents of students, and the wider public might not trust such activity. A camera in a school, even when called a "smart sensor", might cause concerns regarding potential surveillance and security of the data. Therefore, informing the public with an adequate science communication strategy could have an important role when attempting to apply such a method.

Since ambient PA has not been studied before, testing specific interventions might not be the only scientific value of such measurement, there is also a large explorative component to this research. This new form of data could potentially—lead to discovery and new hypotheses and not necessarily only concerning students' health behavior. New insights into crowd and pedestrian behavior dynamics and communal building architecture could be gained in addition to currently unforeseeable phenomena. For this, it would be important to collect dense contextual data alongside the ambient PA distribution and supporting indirect measures.

8. Conclusions and Future Work

Proliferation of physical inactivity is increasing the demand for PA research and for effective large-scale manipulation of health behavior, best practices of school-based PA interventions need to be developed. Verification of the effectiveness of interventions could benefit from unobtrusive privacy-preserving monitoring of ambient PA in schools. To this end, recent advances that have enabled real-time recognition of many rather complex actions seem promising. This work proposes a novel method for unobtrusive ambient PA monitoring whereby the processing of personal data is delegated to deep learning neural networks while maintaining enough validity and reliability to draw meaningful inferences on the PA at the location.

Future work priorities should entail development of the synchronized dataset of video and accelerometry, preferably in a way such that it could be shared between researchers and used for benchmarking—if deemed ethical, this could be achieved by contracting child actors. Developing a segmentation-based annotation tool can likely simplify data annotation by indicating the start and termination of labelling when subjects step into and out of the frame. Once raw accelerations are annotated to the video frames, different options for constructing labels should be explored—different weights on the accelerometers' vertical and horizontal axes, cutting acceleration peaks to potentially ease learning and normalization of accelerations by actor BMI and other individual attributes to better reflect the measurement construct. This should eventually be followed by testing different machine learning frameworks to come to a real-time capable model for a privacy-preserving sensor.

Conflicts of Interest: The author declares no conflict of interest.

References

1. World Health Organization. *World Health Statistics 2018: Monitoring Health for the SDGs*; World Health Organization: Geneva, Switzerland, 2018; ISBN 978-92-4-156558-5.
2. Lee, I.-M.; Shiroma, E.J.; Lobelo, F.; Puska, P.; Blair, S.N.; Katzmarzyk, P.T. Effect of physical inactivity on major non-communicable diseases worldwide: An analysis of burden of disease and life expectancy. *Lancet* **2012**, *380*, 219–229. [CrossRef]
3. Kohl, H.W.; Craig, C.L.; Lambert, E.V.; Inoue, S.; Alkandari, J.R.; Leetongin, G.; Kahlmeier, S. The pandemic of physical inactivity: Global action for public health. *Lancet* **2012**, *380*, 294–305. [CrossRef]
4. Burns, R.D.; Fu, Y.; Podlog, L.W. School-based physical activity interventions and physical activity enjoyment: A meta-analysis. *Prev. Med.* **2017**, *103*, 84–90. [CrossRef] [PubMed]
5. Metcalf, B.; Henley, W.; Wilkin, T. Effectiveness of intervention on physical activity of children: Systematic review and meta-analysis of controlled trials with objectively measured outcomes (EarlyBird 54). *BMJ* **2012**, *345*, e5888. [CrossRef] [PubMed]
6. Johnstone, A.; Hughes, A.R.; Bonnar, L.; Booth, J.N.; Reilly, J.J. An active play intervention to improve physical activity and fundamental movement skills in children of low socio-economic status: Feasibility cluster randomised controlled trial. *Pilot Feasibility Stud.* **2019**, *5*, 45. [CrossRef] [PubMed]
7. Lonsdale, C.; Lester, A.; Owen, K.B.; White, R.L.; Peralta, L.; Kirwan, M.; Diallo, T.M.O.; Maeder, A.J.; Bennie, A.; MacMillan, F.; et al. An internet-supported school physical activity intervention in low socioeconomic status communities: Results from the Activity and Motivation in Physical Education (AMPED) cluster randomised controlled trial. *Br. J. Sports Med.* **2019**, *53*, 341–347. [CrossRef] [PubMed]
8. González-Cutre, D.; Sierra, A.C.; Beltrán-Carrillo, V.J.; Peláez-Pérez, M.; Cervelló, E. A school-based motivational intervention to promote physical activity from a self-determination theory perspective. *J. Educ. Res.* **2018**, *111*, 320–330. [CrossRef]
9. Love, R.; Adams, J.; Sluijs, E.M.F. van Are school-based physical activity interventions effective and equitable? A meta-analysis of cluster randomized controlled trials with accelerometer-assessed activity. *Obes. Rev.* **2019**, *20*, 859–870. [CrossRef]
10. Van Sluijs, E.M.F.; McMinn, A.M.; Griffin, S.J. Effectiveness of interventions to promote physical activity in children and adolescents: Systematic review of controlled trials. *BMJ* **2007**, *335*, 703. [CrossRef]
11. Dobbins, M.; Husson, H.; DeCorby, K.; LaRocca, R.L. School-based physical activity programs for promoting physical activity and fitness in children and adolescents aged 6 to 18. *Cochrane Database Syst. Rev.* **2013**. [CrossRef]
12. Mura, G.; Rocha, N.B.F.; Helmich, I.; Budde, H.; Machado, S.; Wegner, M.; Nardi, A.E.; Arias-Carrión, O.; Vellante, M.; Baum, A.; et al. Physical Activity Interventions in Schools for Improving Lifestyle in European Countries. *Clin. Pract. Epidemiol. Ment. Health* **2015**, *11*, 77–101. [CrossRef] [PubMed]
13. Reis, R.S.; Salvo, D.; Ogilvie, D.; Lambert, E.V.; Goenka, S.; Brownson, R.C. Scaling up physical activity interventions worldwide: Stepping up to larger and smarter approaches to get people moving. *Lancet* **2016**, *388*, 1337–1348. [CrossRef]

14. Naylor, P.-J.; Nettlefold, L.; Race, D.; Hoy, C.; Ashe, M.C.; Wharf Higgins, J.; McKay, H.A. Implementation of school based physical activity interventions: A systematic review. *Prev. Med.* **2015**, *72*, 95–115. [CrossRef] [PubMed]

15. Caspersen, C.J.; Powell, K.E.; Christenson, G.M. Physical activity, exercise, and physical fitness: Definitions and distinctions for health-related research. *Public Health Rep.* **1985**, *100*, 126–131. [PubMed]

16. Janssen, I.; LeBlanc, A.G. Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. *Int. J. Behav. Nutr. Phys. Act.* **2010**, *7*, 40. [CrossRef] [PubMed]

17. WHO. What Is Moderate-Intensity and Vigorous-Intensity Physical Activity? Available online: https://www.who.int/dietphysicalactivity/physical_activity_intensity/en/ (accessed on 11 June 2019).

18. Jetté, M.; Sidney, K.; Blümchen, G. Metabolic equivalents (METS) in exercise testing, exercise prescription, and evaluation of functional capacity. *Clin. Cardiol.* **1990**, *13*, 555–565. [CrossRef] [PubMed]

19. Warren, J.M.; Ekelund, U.; Besson, H.; Mezzani, A.; Geladas, N.; Vanhees, L. Assessment of physical activity–A review of methodologies with reference to epidemiological research: A report of the exercise physiology section of the European Association of Cardiovascular Prevention and Rehabilitation. *Eur. J. Cardiovasc. Prev. Rehabil.* **2010**, *17*, 127–139. [CrossRef] [PubMed]

20. Adamo, K.B.; Prince, S.A.; Tricco, A.C.; Connor-Gorber, S.; Tremblay, M. A comparison of indirect versus direct measures for assessing physical activity in the pediatric population: A systematic review. *Int. J. Pediatr. Obes.* **2009**, *4*, 2–27. [CrossRef] [PubMed]

21. Steene-Johannessen, J.; Anderssen, S.A.; van der Ploeg, H.P.; Hendriksen, I.J.M.; Donnelly, A.E.; Brage, S.; Ekelund, U. Are Self-report Measures Able to Define Individuals as Physically Active or Inactive? *Med. Sci. Sports Exerc.* **2016**, *48*, 235–244. [CrossRef] [PubMed]

22. Gorzelitz, J.; Peppard, P.E.; Malecki, K.; Gennuso, K.; Nieto, F.J.; Cadmus-Bertram, L. Predictors of discordance in self-report versus device-measured physical activity measurement. *Ann. Epidemiol.* **2018**, *28*, 427–431. [CrossRef] [PubMed]

23. Mindell, J.S.; Coombs, N.; Stamatakis, E. Measuring physical activity in children and adolescents for dietary surveys: Practicalities, problems and pitfalls. *Proc. Nutr. Soc.* **2014**, *73*, 218–225. [CrossRef] [PubMed]

24. McKenzie, T.L.; van der Mars, H. Top 10 Research Questions Related to Assessing Physical Activity and Its Contexts Using Systematic Observation. *Res. Q. Exerc. Sport* **2015**, *86*, 13–29. [CrossRef] [PubMed]

25. Schoeller, D.A. Measurement of Energy Expenditure in Free-Living Humans by Using Doubly Labeled Water. *J. Nutr.* **1988**, *118*, 1278–1289. [CrossRef] [PubMed]

26. Li, R.; Deurenberg, P.; Hautvast, J.G. A critical evaluation of heart rate monitoring to assess energy expenditure in individuals. *Am. J. Clin. Nutr.* **1993**, *58*, 602–607. [CrossRef] [PubMed]

27. Dugas, L.R.; Merwe, L.V.D.; Odendaal, H.; Noakes, T.D.; Lambert, E.V. A Novel Energy Expenditure Prediction Equation for Intermittent Physical Activity. *Med. Sci. Sports Exerc.* **2005**, *37*, 2154–2161. [CrossRef]

28. Brage, S.; Brage, N.; Franks, P.W.; Ekelund, U.; Wareham, N.J. Reliability and validity of the combined heart rate and movement sensor Actiheart. *Eur. J. Clin. Nutr.* **2005**, *59*, 561. [CrossRef] [PubMed]

29. Corder, K.; Brage, S.; Wareham, N.J.; Ekelund, U. Comparison of PAEE from combined and separate heart rate and movement models in children. *Med. Sci. Sports Exerc.* **2005**, *37*, 1761–1767. [CrossRef] [PubMed]

30. McNamara, E.; Hudson, Z.; Taylor, S.J.C. Measuring activity levels of young people: The validity of pedometers. *Br. Med. Bull.* **2010**, *95*, 121–137. [CrossRef] [PubMed]

31. Schneider, M.; Chau, L. Validation of the Fitbit Zip for monitoring physical activity among free-living adolescents. *BMC Res. Notes* **2016**, *9*, 448. [CrossRef]

32. Mooses, K.; Oja, M.; Reisberg, S.; Vilo, J.; Kull, M. Validating Fitbit Zip for monitoring physical activity of children in school: A cross-sectional study. *BMC Public Health* **2018**, *18*, 858. [CrossRef]

33. Godfrey, A.; Conway, R.; Meagher, D.; ÓLaighin, G. Direct measurement of human movement by accelerometry. *Med. Eng. Phys.* **2008**, *30*, 1364–1386. [CrossRef] [PubMed]

34. Plasqui, G.; Westerterp, K.R. Physical Activity Assessment With Accelerometers: An Evaluation Against Doubly Labeled Water. *Obesity* **2007**, *15*, 2371–2379. [CrossRef] [PubMed]

35. Troiano, R.P.; McClain, J.J.; Brychta, R.J.; Chen, K.Y. Evolution of accelerometer methods for physical activity research. *Br. J. Sports Med.* **2014**, *48*, 1019–1023. [CrossRef] [PubMed]

36. Ward, D.S.; Evenson, K.R.; Vaughn, A.; Rodgers, A.B.; Troiano, R.P. Accelerometer use in physical activity: Best practices and research recommendations. *Med. Sci. Sports Exerc.* **2005**, *37*, S582–S588. [CrossRef] [PubMed]

37. Cain, K.L.; Sallis, J.F.; Conway, T.L.; Van Dyck, D.; Calhoon, L. Using Accelerometers in Youth Physical Activity Studies: A Review of Methods. *J. Phys. Act. Health* **2013**, *10*, 437–450. [CrossRef] [PubMed]

38. Trost, S.; Loprinzi, P.; Moore, R.; Pfeiffer, K. Comparison of Accelerometer Cut Points for Predicting Activity Intensity in Youth. *Med. Sci. Sports Exerc.* **2011**, *43*, 1360–1368. [CrossRef]

39. Brug, J.; van der Ploeg, H.P.; Loyen, A.; Ahrens, W.; Allais, O.; Andersen, L.F.; Cardon, G.; Capranica, L.; Chastin, S.; De Bourdeaudhuij, I.; et al. Determinants of diet and physical activity (DEDIPAC): A summary of findings. *Int. J. Behav. Nutr. Phys. Act.* **2017**, *14*, 150. [CrossRef]

40. Migueles, J.H.; Cadenas-Sanchez, C.; Ekelund, U.; Delisle Nyström, C.; Mora-Gonzalez, J.; Löf, M.; Labayen, I.; Ruiz, J.R.; Ortega, F.B. Accelerometer Data Collection and Processing Criteria to Assess Physical Activity and Other Outcomes: A Systematic Review and Practical Considerations. *Sports Med.* **2017**, *47*, 1821–1845. [CrossRef]

41. Smith, M.P.; Standl, M.; Heinrich, J.; Schulz, H. Accelerometric estimates of physical activity vary unstably with data handling. *PLoS ONE* **2017**, *12*, e0187706. [CrossRef]

42. Migueles, J.H.; Cadenas-Sanchez, C.; Tudor-Locke, C.; Löf, M.; Esteban-Cornejo, I.; Molina-Garcia, P.; Mora-Gonzalez, J.; Rodriguez-Ayllon, M.; Garcia-Marmol, E.; Ekelund, U.; et al. Comparability of published cut-points for the assessment of physical activity: Implications for data harmonization. *Scand. J. Med. Sci. Sports* **2019**, *29*, 566–574. [CrossRef]

43. Mannini, A.; Sabatini, A.M. Machine Learning Methods for Classifying Human Physical Activity from On-Body Accelerometers. *Sensors* **2010**, *10*, 1154–1175. [CrossRef] [PubMed]

44. Clark, C.C.T.; Barnes, C.M.; Stratton, G.; McNarry, M.A.; Mackintosh, K.A.; Summers, H.D. A Review of Emerging Analytical Techniques for Objective Physical Activity Measurement in Humans. *Sports Med.* **2017**, *47*, 439–447. [CrossRef] [PubMed]

45. Fergus, P.; Hussain, A.J.; Hearty, J.; Fairclough, S.; Boddy, L.; Mackintosh, K.; Stratton, G.; Ridgers, N.; Al-Jumeily, D.; Aljaaf, A.J.; et al. A machine learning approach to measure and monitor physical activity in children. *Neurocomputing* **2017**, *228*, 220–230. [CrossRef]

46. Chowdhury, A.K.; Tjondronegoro, D.; Zhang, J.; Hagenbuchner, M.; Cliff, D.; Trost, S.G. Deep learning for energy expenditure prediction in pre-school children. In Proceedings of the IEEE Conference on Biomedical and Health Informatics, Las Vegas, NA, USA, 4–7 March 2018.

47. Trost, S.G.; Zheng, Y.; Wong, W.-K. Machine learning for activity recognition: Hip versus wrist data. *Physiol. Meas.* **2014**, *35*, 2183–2189. [CrossRef] [PubMed]

48. Morales, J.; Akopian, D. Physical activity recognition by smartphones, a survey. *Biocybern. Biomed. Eng.* **2017**, *37*, 388–400. [CrossRef]

49. Bort-Roig, J.; Gilson, N.D.; Puig-Ribera, A.; Contreras, R.S.; Trost, S.G. Measuring and Influencing Physical Activity with Smartphone Technology: A Systematic Review. *Sports Med.* **2014**, *44*, 671–686. [CrossRef]

50. Lau, P.W.; Lau, E.Y.; Wong, D.P.; Ransdell, L. A Systematic Review of Information and Communication Technology–Based Interventions for Promoting Physical Activity Behavior Change in Children and Adolescents. *J. Med. Internet Res.* **2011**, *13*, e48. [CrossRef]

51. Hicks, J.L.; Althoff, T.; Sosic, R.; Kuhar, P.; Bostjancic, B.; King, A.C.; Leskovec, J.; Delp, S.L. Best practices for analyzing large-scale health data from wearables and smartphone apps. *Npj Digit. Med.* **2019**, *2*, 45. [CrossRef]

52. Welk, G.J.; Corbin, C.B.; Dale, D. Measurement Issues in the Assessment of Physical Activity in Children. *Res. Q. Exerc. Sport* **2000**, *71*, 59–73. [CrossRef]

53. Nilsson, A.; Ekelund, U.; Yngve, A.; Söström, M. Assessing Physical Activity among Children with Accelerometers Using Different Time Sampling Intervals and Placements. *Pediatr. Exerc. Sci.* **2002**, *14*, 87–96. [CrossRef]

54. Crocker, P.R.E.; Holowachuk, D.R.; Kowalski, K.C. Feasibility of Using the Tritrac Motion Sensor over a 7-Day Trial with Older Children. *Pediatr. Exerc. Sci.* **2001**, *13*, 70–81. [CrossRef]

55. Van, P.C.; Harnack, L.; Schmitz, K.; Fulton, J.E.; Galuska, D.A.; Gao, S. Feasibility of using accelerometers to measure physical activity in young adolescents. *Med. Sci. Sports Exerc.* **2005**, *37*, 867–871. [CrossRef]

56. Colley, R.; Gorber, S.C.; Tremblay, M.S. Quality control and data reduction procedures for accelerometry-derived measures of physical activity. *Health Rep.* **2010**, *21*, 63–64. [PubMed]

57. OECD. *Education at a Glance 2018: OECD Indicators*; OECD Publishing: Paris, France, 2018.

58. Delidou, E.; Matsouka, O.; Nikolaidis, C. Influence of school playground size and equipment on the physical activity of students during recess. *Eur. Phys. Educ. Rev.* **2016**, *22*, 215–224. [CrossRef]

59. Brittin, J.; Sorensen, D.; Trowbridge, M.; Lee, K.K.; Breithecker, D.; Frerichs, L.; Huang, T. Physical Activity Design Guidelines for School Architecture. *PLoS ONE* **2015**, *10*, e0132597. [CrossRef] [PubMed]

60. Ridgers, N.D.; Stratton, G.; Fairclough, S.J.; Twisk, J.W.R. Long-term effects of a playground markings and physical structures on children's recess physical activity levels. *Prev. Med.* **2007**, *44*, 393–397. [CrossRef] [PubMed]

61. Hamer, M.; Aggio, D.; Knock, G.; Kipps, C.; Shankar, A.; Smith, L. Effect of major school playground reconstruction on physical activity and sedentary behaviour: Camden active spaces. *BMC Public Health* **2017**, *17*, 552. [CrossRef] [PubMed]

62. Pawlowski, C.S.; Andersen, H.B.; Troelsen, J.; Schipperijn, J. Children's Physical Activity Behavior during School Recess: A Pilot Study Using GPS, Accelerometer, Participant Observation, and Go-Along Interview. *PLoS ONE* **2016**, *11*, e0148786. [CrossRef] [PubMed]

63. Murillo Pardo, B.; García Bengoechea, E.; Generelo Lanaspa, E.; Bush, P.L.; Zaragoza Casterad, J.; Julián Clemente, J.A.; García González, L. Promising school-based strategies and intervention guidelines to increase physical activity of adolescents. *Health Educ. Res.* **2013**, *28*, 523–538. [CrossRef] [PubMed]

64. Pawlowski, C.S.; Andersen, H.B.; Tjørnhøj-Thomsen, T.; Troelsen, J.; Schipperijn, J. Space, body, time and relationship experiences of recess physical activity: A qualitative case study among the least physical active schoolchildren. *BMC Public Health* **2016**, *16*, 16. [CrossRef] [PubMed]

65. Nicaise, V.; Kahan, D.; Reuben, K.; Sallis, J.F. Evaluation of a Redesigned Outdoor Space on Preschool Children's Physical Activity During Recess. *Pediatr. Exerc. Sci.* **2012**, *24*, 507–518. [CrossRef] [PubMed]

66. Deterding, S.; Sicart, M.; Nacke, L.; O'Hara, K.; Dixon, D. Gamification. Using Game-design Elements in Non-gaming Contexts. In Proceedings of the CHI '11 Extended Abstracts on Human Factors in Computing Systems, Vancouver, BC, Canada, 7–12 May 2011. [CrossRef]

67. King, D.; Greaves, F.; Exeter, C.; Darzi, A. 'Gamification': Influencing health behaviours with games. *J. R. Soc. Med.* **2013**, *106*, 76–78. [CrossRef] [PubMed]

68. Hamari, J.; Koivisto, J.; Sarsa, H. Does Gamification Work? A Literature Review of Empirical Studies on Gamification. In Proceedings of the 47th Hawaii International Conference on System Sciences, Hawaii, HI, USA, 6–9 January 2014. [CrossRef]

69. Bailey, B.W.; McInnis, K. Energy Cost of Exergaming: A Comparison of the Energy Cost of 6 Forms of Exergaming. *Arch. Pediatr. Adolesc. Med.* **2011**, *165*, 597–602. [CrossRef] [PubMed]

70. Perron, R.M.; Graham, C.A.; Feldman, J.R.; Moffett, R.A.; Hall, E.E. Do exergames allow children to achieve physical activity intensity commensurate with national guidelines? *Int. J. Exerc. Sci.* **2011**, *4*, 257–264. [PubMed]

71. Gao, Z.; Chen, S.; Stodden, D.F. A Comparison of Children's Physical Activity Levels in Physical Education, Recess, and Exergaming. *J. Phys. Act. Health* **2015**, *12*, 349–354. [CrossRef]

72. Huang, H.-C.; Nguyen, H.V.; Cheng, T.C.E.; Wong, M.-K.; Chiu, H.-Y.; Yang, Y.-H.; Teng, C.-I. A Randomized Controlled Trial on the Role of Enthusiasm About Exergames: Players' Perceptions of Exercise. *Games Health J.* **2018**, *8*, 220–226. [CrossRef] [PubMed]

73. Baranowski, T.; Blumberg, F.; Buday, R.; DeSmet, A.; Fiellin, L.E.; Green, C.S.; Kato, P.M.; Lu, A.S.; Maloney, A.E.; Mellecker, R.; et al. Games for Health for Children—Current Status and Needed Research. *Games Health J.* **2015**, *5*, 1–12. [CrossRef]

74. Fjørtoft, I.; Kristoffersen, B.; Sageie, J. Children in schoolyards: Tracking movement patterns and physical activity in schoolyards using global positioning system and heart rate monitoring. *Landsc. Urban Plan.* **2009**, *93*, 210–217. [CrossRef]

75. Dessing, D.; Pierik, F.H.; Sterkenburg, R.P.; van Dommelen, P.; Maas, J.; de Vries, S.I. Schoolyard physical activity of 6–11 year old children assessed by GPS and accelerometry. *Int. J. Behav. Nutr. Phys. Act.* **2013**, *10*, 97. [CrossRef]

76. Andersen, H.B.; Klinker, C.D.; Toftager, M.; Pawlowski, C.S.; Schipperijn, J. Objectively measured differences in physical activity in five types of schoolyard area. *Landsc. Urban Plan.* **2015**, *134*, 83–92. [CrossRef]

77. Kerr, J.; Duncan, S.; Schipperjin, J. Using Global Positioning Systems in Health Research: A Practical Approach to Data Collection and Processing. *Am. J. Prev. Med.* **2011**, *41*, 532–540. [CrossRef] [PubMed]

78. Moeslund, T.B.; Granum, E. A Survey of Computer Vision-Based Human Motion Capture. *Comput. Vis. Image Underst.* **2001**, *81*, 231–268. [CrossRef]

79. Hu, W.; Tan, T.; Wang, L.; Maybank, S. A Survey on Visual Surveillance of Object Motion and Behaviors. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **2004**, *34*, 334–352. [CrossRef]

80. Luo, W.; Xing, J.; Milan, A.; Zhang, X.; Liu, W.; Zhao, X.; Kim, T.-K. Multiple Object Tracking: A Literature Review. *arXiv* **2014**, arXiv:14097618 Cs.

81. Thida, M.; Yong, Y.L.; Climent-Pérez, P.; Eng, H.; Remagnino, P. A Literature Review on Video Analytics of Crowded Scenes. In *Intelligent Multimedia Surveillance: Current Trends and Research*; Atrey, P.K., Kankanhalli, M.S., Cavallaro, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 17–36. [CrossRef]

82. Zhou, H.; Hu, H. Human motion tracking for rehabilitation—A survey. *Biomed. Signal Process. Control* **2008**, *3*, 1–18. [CrossRef]

83. Acampora, G.; Cook, D.J.; Rashidi, P.; Vasilakos, A.V. A Survey on Ambient Intelligence in Healthcare. *Proc. IEEE* **2013**, *101*, 2470–2494. [CrossRef] [PubMed]

84. Piccardi, M. Background subtraction techniques: A review. In Proceedings of the 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583), The Hague, The Netherlands, 10–13 October 2004; Volume 4, pp. 3099–3104. [CrossRef]

85. Barnich, O.; Droogenbroeck, M.V. ViBe: A Universal Background Subtraction Algorithm for Video Sequences. *IEEE Trans. Image Process.* **2011**, *20*, 1709–1724. [CrossRef]

86. Aggarwal, J.K.; Ryoo, M.S. Human Activity Analysis: A Review. *ACM Comput. Surv.* **2011**, *43*, 16:1–16:43. [CrossRef]

87. Soomro, K.; Zamir, A.R.; Shah, M. UCF101: A Dataset of 101 Human Actions Classes from Videos in the Wild. *arXiv* **2012**, arXiv:12120402 Cs.

88. Herath, S.; Harandi, M.; Porikli, F. Going deeper into action recognition: A survey. *Image Vis. Comput.* **2017**, *60*, 4–21. [CrossRef]

89. Simonyan, K.; Zisserman, A. Two-Stream Convolutional Networks for Action Recognition in Videos. In *Advances in Neural Information Processing Systems 27*; Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q., Eds.; Curran Associates Inc.: Red Hook, NY, USA, 2014; pp. 568–576.

90. Wang, X.; Gao, L.; Wang, P.; Sun, X.; Liu, X. Two-Stream 3-D convNet Fusion for Action Recognition in Videos with Arbitrary Size and Length. *IEEE Trans. Multimed.* **2018**, *20*, 634–644. [CrossRef]

91. Feichtenhofer, C.; Pinz, A.; Zisserman, A. Convolutional Two-Stream Network Fusion for Video Action Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016. [CrossRef]

92. Park, E.; Han, X.; Berg, T.L.; Berg, A.C. Combining multiple sources of knowledge in deep CNNs for action recognition. In Proceedings of the 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA, 7–10 March 2016. [CrossRef]

93. Singh, G.; Saha, S.; Sapienza, M.; Torr, P.; Cuzzolin, F. Online Real-Time Multiple Spatiotemporal Action Localisation and Prediction. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017. [CrossRef]

94. Zhang, B.; Wang, L.; Wang, Z.; Qiao, Y.; Wang, H. Real-Time Action Recognition with Deeply Transferred Motion Vector CNNs. *IEEE Trans. Image Process.* **2018**, *27*, 2326–2339. [CrossRef] [PubMed]

95. Ali, A.; Taylor, G.W. Real-Time End-to-End Action Detection with Two-Stream Networks. In Proceedings of the 2018 15th Conference on Computer and Robot Vision (CRV), Toronto, ON, Canada, 9–11 May 2018; pp. 31–38. [CrossRef]

96. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In Proceedings of the 2016 14th European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 8–16 October 2016; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer International Publishing: Berlin, Heidelberg, Germany, 2016; pp. 21–37.

97. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 22–25 July 2017; pp. 6517–6525. [CrossRef]

98. Ilg, E.; Mayer, N.; Saikia, T.; Keuper, M.; Dosovitskiy, A.; Brox, T. FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 22–25 July 2017; pp. 1647–1655. [CrossRef]

99. Li, Z.; Gavrilyuk, K.; Gavves, E.; Jain, M.; Snoek, C.G.M. VideoLSTM convolves, attends and flows for action recognition. *Comput. Vis. Image Underst.* **2018**, *166*, 41–50. [CrossRef]

100. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]

101. Zhao, J.; Snoek, C.G.M. Dance with Flow: Two-in-One Stream Action Detection. *arXiv* **2019**, arXiv:190400696 Cs.

102. Han, S.; Mao, H.; Dally, W.J. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding. *arXiv* **2015**, arXiv:151000149 Cs.

103. Han, S.; Liu, X.; Mao, H.; Pu, J.; Pedram, A.; Horowitz, M.A.; Dally, W.J. EIE: Efficient inference engine on compressed deep neural network. In Proceedings of the 2016 ACM/IEEE 43rd Annual International Symposium on Computer Architecture (ISCA), Seoul, Korea, 18–22 July 2016; pp. 243–254. [CrossRef]

104. Chi, P.; Li, S.; Xu, C.; Zhang, T.; Zhao, J.; Liu, Y.; Wang, Y.; Xie, Y. Prime: A novel processing-in-memory architecture for neural network computation in reram-based main memory. In Proceedings of the 2016 ACM/IEEE 43rd Annual International Symposium on Computer Architecture (ISCA), Seoul, Korea, 18–22 July 2016; pp. 27–39. [CrossRef]

105. Zhu, Z.; Sun, H.; Lin, Y.; Dai, G.; Xia, L.; Han, S.; Wang, Y.; Yang, H.A. Configurable Multi-Precision CNN Computing Framework Based on Single Bit RRAM. In Proceedings of the 2019 56th Annual Design Automation Conference (DAC), Las-Vegas, NV, USA, 2–6 June 2019; p. 56. [CrossRef]

106. Jishnu, P. MV-Tractus: A simple and fast tool to extract motion vectors from H264 encoded video streams. *Zenodo* **2018**. [CrossRef]

107. Yi, Y.; Wang, H.; Zhang, B. Learning correlations for human action recognition in videos. *Multimed. Tools Appl.* **2017**, *76*, 18891–18913. [CrossRef]