Cognitive Assessment in Children through Motion Capture and Computer Vision: The Cross-your-Body task

Saif Iftekar Sayed[1], **Konstantinos Tsiakas**[2], **Morris Bell**[2], **Vassilis Athitsos**[1] **and Fillia Makedon**[1]
[1]University of Texas at Arlington, [2]Yale University
saififtekar.sayed@mavs.uta.edu, konstantinos.tsiakas@yale.edu, morris.bell@yale.edu, athitsos@uta.edu, makedon@uta.edu

ABSTRACT
This paper focuses on creating video-based human activity recognition methods towards an automated cognitive assessment system for children. We present the Activate Test for Embodied Cognition (ATEC), which assesses executive functioning in children through physical/cognitive tasks. Detecting activities for children is challenging due to high amount of random motion and variability. This paper focuses on creating a ubiquitous and non-intrusive activity recognition system for upper-body movements. Our proposed methods are evaluated on real-world data from children performing the Cross-your-Body task. The dataset includes 15 children performing 8 types of activities, resulting to 1900 annotated video samples.

CCS Concepts
•**Computing methodologies** → **Supervised learning by classification;**

Author Keywords
Human Activity Recognition, Body Pose Features, Cognitive Assessment

INTRODUCTION
Self-regulation, which generally refers to a complex of acquired, intentional skills involved in controlling, directing, and planning one's cognition, emotions and behaviors [20], is an important mechanism associated with variety of outcomes, including school readiness and performance [16]. Executive function refers to the mental processes that enable humans to plan, organize, problem-solve as well as manage their impulses, including cognitive flexibility, working memory, and inhibitory control [3]. Children who face deficits in executive functions are highly likely to present attention disorders [2]. ADHD or attention deficit hyperactivity disorder is a psychiatric neurodevelopmental disorder found in children and young adolescents and it can start as early as age 6 [6, 8]. Cognitive impairments in executive functions can not only cause bad performance in school settings, but can also show repercussions in family, employment and community settings which can result to several socioeconomic problems, resulting to low self-esteem and self-acceptance [7]. In order to quantify executive function in children, traditional assessments include either paper or computer-based activities, e.g., the NIH toolbox. However, recent studies suggest assessments which include physical activities, for example the Head-Toes-Knees-Shoulders (HTKS) task, which has been extensively tested on 208 children and elicits psychometric measures through physical performance [15].

Our research includes the development of ATEC; the ACTIVATE Test for Embodied Cognition, which includes a set of physical tasks with cognitive demands to assess executive function in motion. A core ATEC task is *Cross-your-Body*, which follows and extends the basic HTKS rules, and is designed to assess working memory and attention, bilateral coordination, rhythm and self-regulation. The HTKS rules include four behavioral activities: "touch your {*head, toes, knees, shoulders*}". The subject is initially instructed to touch the announced body part. Then, the task introduces task switching and requires the child to touch the body part in an "opposite" fashion (e.g. touch knees when told to touch shoulders).

Cross-your-Body requires the subject to touch the correct body part with the hand from the opposite side. Crossing the midline is an integral skill related to bilateral coordination that children learn from infancy. Poor midline crossing can affect reading (tracking with the eye from left to right) and writing (using their dominant hand across the writing page) skills. Moreover, Cross-your-Body is designed to assess rhythm; the child is asked to repeat each movement three times, alternating sides in a timely manner. Task performance is determined both in terms of accuracy (touch the correct part) and rhythm (perform movements in a rhythmic manner). Manual scoring requires a human rater to watch the videos and score the child based on the task rules (accuracy, rhythm) and can be time-expensive and often ambiguous.

The main purpose of our research is to build an automated scoring system for Cross-your-body, which detects and analyzes the performed activities to assess accuracy and rhythm. Current systems like Cognilearn [9] utilize state-of-the-art computer vision algorithms by capturing color frames from the Kinect V2 camera and provide an interface for motion capture and analysis. Deep Learning architectures were proposed as the backbone model [12] and tested on synthetic data with adults performing the task. For this paper, our dataset includes

iWOAR'19, September 16–17, 2019, Rostock, Germany.

DOI string from ACM form confirmation

collected data during the ATEC assessments with children between 5-10 years old in classroom environments.

The main contribution of our paper is a video-based activity recognition system for the Cross-your-Body task, which recognizes the *active hand* that performs the movement, estimates specific spatial hand positions for efficient feature extraction, while including low-confidence prediction class. Our experiments on real-world data indicate the efficiency of our method for reliable and user-independent activity prediction effective on scaling number of users. The structure of the paper is as follows: Firstly, we present related work on similar applications, highlighting the motivation of our work. Then, we present the system architecture and our experimental approach using machine learning techniques. We discuss our experimental protocol and results, describing the data collection and annotation process. Finally, we conclude with some final remarks and our future work.

RELATED WORK

Emerging technologies have influenced many medical related processes such as diagnosis, rehabilitation and treatment. Computer and data science have opened up another realm of capturing and analyzing data in an automated fashion. These implementations not only demand higher prediction accuracy but also focus on user engagement. Active video game play using consoles like Microsoft Kinect can help rehabilitation of children suffering from Cerebral Palsy [11]. Systems can also monitor the attention state of the child using eye-trackers towards user-friendly and personalized interfaces [4]. Moreover, virtual reality games have been developed for assessment and rehabilitation of children with attention deficit [19].

Inattention and/or hyperactivity or impulsivity symptoms can cause alterations in a person's human movements and reactions [1]. This is the main reason for exploring several sensor-based human activity recognition systems. Such sensors can be employed on the human body or can be placed in the surrounding environment. Hypothesis testing by studying the readings given by wearables showed significant differences for ADHD patients compared to non-ADHD controls [10, 13]. Recent advancements in deep learning have led to the use of convolution neural networks (CNN) to extract embedded acceleration patterns and provide objective measures to help diagnose ADHD [17], but such approaches can be obtrusive since the subject has to wear different types of wearable sensors.

Camera-based settings can provide an unobtrusive environment for data collection and computer vision and deep learning methods can be used to extract important spatio-temporal features and recognize patterns of interest. In a previous work, a camera-based system was proposed for the HTKS task [9] and evaluated on adults, which used deep learning techniques to extract body pose information for human activity recognition following a frame-based approach. In this work, we follow a segment-based approach, since the nature of activities involved in Cross-your-Body (CYB) is more complex compared to HTKS, i.e., crossing the midline and performing the task in rhythm. Moreover, our proposed methods are evaluated on real-world data from children performing the task.

CYB SYSTEM

The primary goal of the system is to reliably recognize the type of performed activity given a video segment. The system initially detects the subject and then tracks its hands over time to recognize the performed activity, as well as when the activity was performed. The overall system is illustrated in Figure [1]. The system receives body-motion data from Kinect and then it produces a set of spatio-temporal features used to predict the activity performed by the child. The system include two modules: the *Acquire* and the *Track* module. The Acquire module takes care of capturing and analyzing each frame to create an accurate skeleton vector for the entire video, which after preprocessing it is passed to the Track module for gesture recognition.

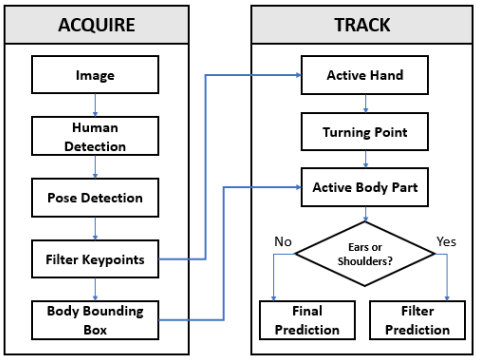


Figure 1: System Architecture

Acquire Module

The Acquire module initially fetches the RGB frames from the Kinect and detects the subject of interest, in order to detect and filter its 2d pose, divide the body into regions based on height in order to produce a filtered skeleton joint vector for activity recognition.

Human Detection

The first step of the process is to reliably detect the subject of interest. Due to the naturalistic environment, there are often multiple people in the background during the ATEC assessment was essential to first isolate the subject of interest in order to reduce computation complexity. YOLO v3 [18] was used to detect the humans in the scene because of his fast and accurate inference and then based on an empirically decided spatial threshold, the bounding box which fell in that criteria was chosen as the subject of interest which was then was passed to the pose estimation.

2D Body Pose Estimation

Microsoft Kinect V2 has an RGB capture resolution of 1920x1080 pixels with a Time-of-Flight depth data as an 512x424 resolution image [21]. The field of view for depth is 70 degrees horizontally and 60 degrees vertically [14]. In this paper, a Kinect V2 is used for acquisition since it tracks more joints and has a higher motion tracking accuracy, with greater stability. Kinect's SDK provides it's own stock SDK that can be used to get the 3D body pose of the skeleton, but the problem is that Kinect's skeletal tracking doesn't perform

well under occlusions [22]. In our work we are still using kinect, since it gives us the color and the depth channels of the environment. Currently we are considering only the color modality of the acquisition for our analysis as it is much more consistent and less noisy than kinect's skeletal tracker.

For locating the joints in the RGB images, we leverage the recent advancements in deep learning where data has been trained on millions of images encompassing scenarios like self-occlusions and networks like OpenPose [5] can be very useful to provide accurate estimation of body pose. We have employed the skeleton map result based on the 2016 COCO keypoints dataset challenge and the skeleton structure provided by openpose is as shown in the figure 2. Each joint is represented by a 2D vector in the cartesian co-ordinate space. The extracted tensor for a video can be expressed as follows:

$$P_i = [B_1, B_2, ..., B_{18}], i = [1, 2, ..., n] \tag{1}$$

Where P_i is a set of 18 2D keypoints location representing respective body joints for a given frame i in a video sample.

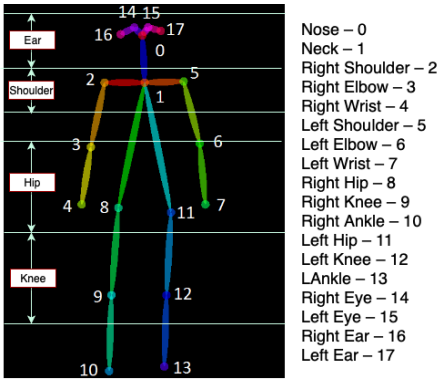


Figure 2: Openpose Skeleton Map

Filter Keypoints

These points in a video sample are further filtered, where intermediate body points are interpolated in case of misclassifications.

Body Bounding Box

After filtering the keypoints, the first frame of the video sample is used to divide the body into 4 areas. This area is based on the required class labels of ears, shoulders, hip and knees. The height of the person is computed using the distance between ear and ankle with a padding of 50 pixels. Using a fixed width consistent with all the subjects the body is spatially divided into 4 parts using a fixed percentage of height per body part. The divided region helped the classifier understand which part the subject was trying to touch and the use of these regions will be explained in the active body part prediction section.

Track

Tracking modules involves classifying the activity by using the 2d keypoints. This involves finding which hand was active in other words which hand was performing the gesture/activity and which body part it was touching/interacting with. This involves finding the relevant frames in the video which gives the maximum information for correct classification as well as extracting those spatio-temporal features. The tracking module first finds which hand was active, then tracking the spatial positions of the palm decides where the touching of body happened based on the velocity and curvature of the palm trajectory and eventually classifies the active body part. These steps will be elaborated in the following sections.

Detect Active Hand

Before tracking the hand it was important to compute the palm position and not the wrist position and since the body pose estimator of OpenPose didn't give the palm position, an approximate estimation of palm was done by extending the vector passing from the elbow through wrist by a magnitude of 1.25 times the magnitude of the vector between elbow and palm, where the elbow vector is added by the scalar element-wise (Eq. 2).

$$\vec{B}_{Palm} = \vec{B}_{elbow} + \|\vec{B}_{wrist} - \vec{B}_{elbow}\| * 1.25 \tag{2}$$

The experimental protocol dictated that a valid touch of body part is supposed to be done by the opposite hand-body pair (midline crossing). So to check whether a palm is on the other side of the body a reliable anchor point was needed to decide the horizontal center of the body. Based on the data visualizations and the experimental protocol, subjects were instructed to stand at a fixed location in the scene, hence their feet position is fixed in the whole video and can act as anchor points. The vector passing from midpoint of the 2 ankle and parallel to the y-axis was considered as a border dividing the body into left and right side.

$$\vec{B}_{midankle} = \frac{\vec{B}_{left_{ankle}} + \vec{B}_{right_{ankle}}}{2} \tag{3}$$

For a video sample, let C_{left} and C_{right} be the set of frame indices in a video sample where the system predicted that the hands are in opposite sides of the body. These 2 sets are then passed to a filter where:

$$C_{left}[n] = \begin{cases} n, & \text{if } \vec{B}_{midankle_x} - \vec{B}_{leftpalm_x} > 0 \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

$$C_{right}[n] = \begin{cases} n, & \text{if } \vec{B}_{midankle_x} - \vec{B}_{rightpalm_x} < 0 \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

After getting the frame indices where the left and right palms cross the midline, these indices were further analyzed to get the active hand which is computed by analyzing the velocities of the palms. The protocol for a correct body movement indicates that subjects moves their hands from a rest position, cross the midline, touch the body part of the opposite side and then bring the hand back to rest by crossing the midline again. More formally, this means that the subject's hand will cross the midline twice (once while approaching and once while leaving) and the hand's velocity vector in the x direction will have opposite direction as it cross the midline. Note that there might be cases when the subject's hand may be in the cross state and still they did touch the body part. If both hands were in cross state, then the hand which crosses the midline later

Task ID	Task Nature	Video Instructed - "Cross your body touch your.."	Actual Movement Intended
1	Cross Body - Trial 1	E,S,H,K	E,S,H,K
2	Cross Body - Trial 2	E,S,H,K	E,S,H,K
3	Cross Body Ears - Knees	E,K,K,E,K,E,E,K	K,E,E,K,E,K,K,E
4	Cross Body Hips - Shoulders	S,H,H,S,H,S,S,H	H,S,S,H,S,H,H,S
5	Cross Body Hips - Combined	E,H,K,S,K,H,E,K,H,S,E,S	K,S,E,H,H,E,H,K,E,S,H,K,H

Table 1: Cross-your-Body versions and rules. Trials 1, 2 do not have cognitive demands; the rest of them introduce task switching

was assumed as the active hand. If none of the hands were in a state of cross, then the classifier is not confident of the prediction and not undergo further steps.

Get Active Body Part

Once the system detects the active hand, the next step of the algorithm is to identify which body part is touched. This is the crux of the system as based on the data analysis for the kids, there is a very high intra-class variability on the style of how a subject performs an activity in terms of distance from the palm and the intended body part touch and velocities the palm approaches and leaves a body part after touching. The trajectory of an active hand's palm can be considered as a curve defined in a parametric form by equations $x = x(t)$ and $y = y(t)$, where t is time and x and y are the co-ordinates of the palm. So, a curvature at any point on the trajectory can be given as:

$$K = \frac{|x'y'' - y'x''|}{[(x')^2 + (y')^2]^2} \tag{6}$$

Here x' and x'' are the first and second derivative of the x co-ordinates and similarly for y-co-ordinates. Before getting the points of curvature, the trajectory is smoothed by using a 1 dimensional smoothing filter. Using the spatial positions of the trajectory, further they were filtered based on the velocity of the hand. An empirical threshold of 2 was chosen to filter the positions. Once the spatial positions of the hands are known where the palm trajectory showed highest curvature and the palm was moving slowly, then the mean of these spatial positions is taken and using the bounding boxes of the relevant body parts as shown in figure 2, the final prediction is done. If the prediction is ear or shoulder it goes through further processing of ear and shoulder classification module which computes more spatio-temporal features and produces a final prediction (ear/shoulder) by passing the features to a decision tree algorithm.

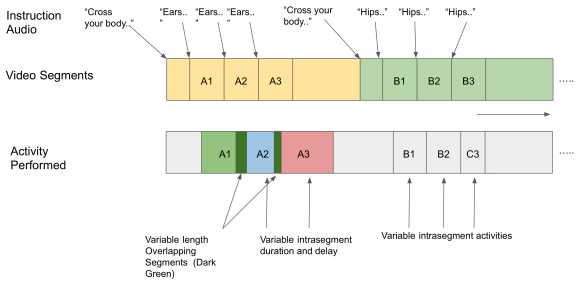


Figure 3: Temporal Analysis of activities and rules

The spatio-temporal features into consideration are:

- Hand Shoulder and Hand Ear Distance: To compare that the hand was much closer to ear or shoulder, euclidean distances between the active hand's palm and the opposite side's ear and shoulder were computed taking for the spatial positions where the curvature of the trajectory was maximum and velocity was low. Note, there can be a multiple points where this criteria of curvature and velocity may be true over time in a video so a mean of these euclidean distances was taken to compare if it was close to ear or shoulders.

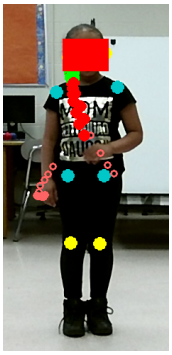
- Shoulder-Palm-Ear Angle: This is a very important feature that can be used to differentiate the touching of ear or shoulder. For example the angle made by the left shoulder, left elbow and left palm will be much closer to the angle made by the left shoulder, left elbow and right ear compared to left shoulder, left elbow and right shoulder when the actual activity performed was left hand touching right ear. Using the formula 7, we can compute an angle between 3 joints and the above logic will yield into 3 angles namely Θ_{palm}, $\Theta_{shoulder}$ and Θ_{ear} resulting into addition of information for better prediction between ears and shoulder.

$$\begin{cases} \overrightarrow{AB} = A - B \\ \overrightarrow{BC} = B - C \\ \Theta = \dfrac{\sum_{i=1}^{n} \cos^{-1}\left(\dfrac{\overrightarrow{AB_i} \cdot \overrightarrow{BC_i}}{|\overrightarrow{AB_i}| \cdot |\overrightarrow{BC_i}|}\right)}{n} \end{cases} \tag{7}$$

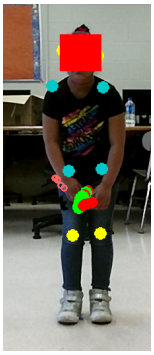
System Protocol

In the context of our research study, children were participated to perform the ATEC activities, including the Cross-your-Body task. For our experiments, we created our dataset including data from 15 participants performing five versions of the task in 2 sessions with a gap of 2 months. In order to ensure a high-fidelity assessment system, all instructions are pre-recorded and same for all children. A large screen is used to display a theme-based music video, where the on-screen host, Aliza, instructs the child to perform the task following her "Cross-your-Body" song. Two Kinects (front and side) are used to capture the movements. The distance between the subject and both of the Kinects is 2m. Table 1 illustrates the task versions. Before each task, the subject is shown a task instruction, as well as a demonstration video clip , explaining the task rules. For example, for Trials 1 and 2, the child is told to perform three touches, touching the announced body part, using the hand from the opposite side and alternating sides. For the rest of the tasks, the child is instructed to follow the "opposite" rules; task 3 switches ears and knees, task 4

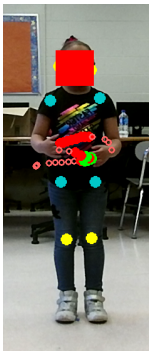
switches shoulders and hips, while task 5 includes both rules. Every subject undergoes through the same process and there is no prior instructions given other than the video instructions.



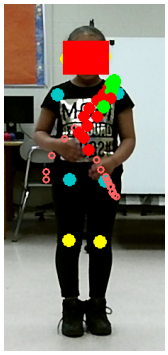
(a) Left Hand Right Ear



(b) Left Hand Right Knee



(c) Right Hand Left Hip



(d) Left Hand Left Shoulder

Figure 4: Sample predictions for a subject. Body joint positions are highlighted in yellow (ear), cyan (shoulders and hips) and yellow (knee). Active hand positions for a hand are in thick red, while unfilled red circles indicate an inactive hand. Points with high curvature and low speed are in green.

A temporal analysis of the activities performed vs instructed can be seen in fig 3 which indicates the progression of a task and is divided in 2 parts: video segments and activity performed. Referring to the row of video segments, the main task is made of several sub-tasks and is highlighted in yellow and green respectively. A sub-task begins when the instruction video starts saying "Cross your body.." while a task segment begins when the actual body part to touch was said. The instruction time gap between every task segment is 1 second and there are 3 segments in every sub-task. Each task segment is an activity of touching a body part and there were overlapping of activities, in other words if A1 and A2 are 2 task segments of touching ear, the subject might be partially touching the ear or moving hand away from that ear while lifting the other hand and approaching the other ear intended for A2. Also since there involves processing of working memory a subject would perform the activity with varying delay after the instruction and since there also involves switching of rules, the activity performed may or may not be correctly done as instructed.

EXPERIMENTS AND RESULTS

Based on the problem definition 8 activity classes were chosen as LHRE, RHLE, LHRS, RHLS, LHRH, RHLH, LHRK, RHLK, where LH stands for left hand and RH stands for right hand. Also there was a ninth class as nooo indicating system low-confidence. The recorded videos were segmented based on the timestamps of the presented stimuli and a frame level activity annotation was performed resulting into 1900 video samples, where the average frame length of the video was 28. Each annotated segment refers to one (out of three) movements. First 5 subjects were used to set the thresholds required by the algorithm (e.g., bounding box) and the next 10 subjects were used for testing.

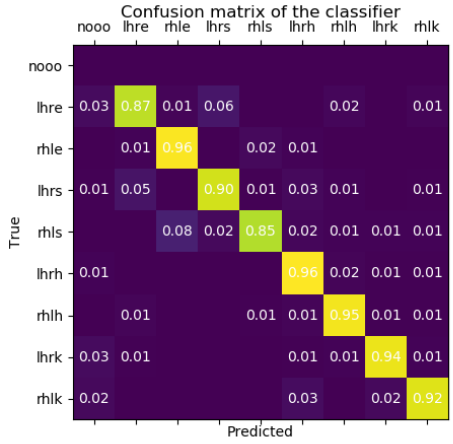


Figure 5: Confusion matrix for the test split

As illustrated in figure 4, these are some sample predictions of the system and it can be clear that system could focus more on the spatial locations in the trajectory of the hand where it could extract maximum information. The accuracy of the system was measured for these 10 test subjects by comparing with the ground truths and the system could achieve an overall 89.95%. The confusion matrix for the predictions can be seen as in figure 5. Based on the confusion matrix, the prediction of ear or shoulder needs further fine-tuning as the system still gets confused since the distance between ear and shoulder is small and the palm position prediction is not able to capture the fine motion of palm. One way to improve this is to use the depth modality or skin detection for better segmentation of hand and in-turn help to compute the distance and angles between palm, ear and shoulder much more accurately.

CONCLUSION AND FUTURE WORK

A video-based activity recognition system for cognitive assessment in children was presented. Data were collected from the Cross-your-Body task during the ATEC administration with children between ages 5-10. Overall 1900 video samples were segmented and annotated and the system gave an overall accuracy of 89.95%. The automated system was also tested with manual scoring and gave accurate results as comparatively. The system successfully applied temporal modelling dependencies to capture the aforementioned activities. Moving forward, the system will be extended to perform automated scoring given the task rules. Our ongoing work on temporal localization of the activity will provide us with insights on

how to automatically score both for accuracy (which part is touches) and rhythm (when the touch occurs). Intelligent interfaces will be used to provide the experts with intuitive data visualization to enhance their decision making.

ACKNOWLEDGEMENTS

This work was partially supported by National Science Foundation grants IIS 1565328 and IIP 1719031.

REFERENCES

1. American Psychiatric Association and others. 2013. *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.
2. Russell A Barkley. 1997. Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. *Psychological bulletin* 121, 1 (1997), 65.
3. John R Best and Patricia H Miller. 2010. A developmental perspective on executive function. *Child development* 81, 6 (2010), 1641–1660.
4. Aude Billard, Ben Robins, Jacqueline Nadel, and Kerstin Dautenhahn. 2007. Building Robota, a mini-humanoid robot for the rehabilitation of children with autism. *Assistive Technology* 19, 1 (2007), 37–49.
5. Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2018. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. In *arXiv preprint arXiv:1812.08008*.
6. Eileen Cormier. 2008. Attention deficit/hyperactivity disorder: a review and update. *Journal of pediatric nursing* 23, 5 (2008), 345–357.
7. CA Dendy. 2008. Executive FunctionâĂIJWhat is this anyway?âĂİ. *Retrieved September* 18 (2008), 2008.
8. David W Dunn and William G Kronenberger. 2003. Attention-deficit/hyperactivity disorder in children and adolescents. *Neurologic clinics* (2003).
9. Srujana Gattupalli, Dylan Ebert, Michalis Papakostas, Fillia Makedon, and Vassilis Athitsos. 2017. Cognilearn: A deep learning-based interface for cognitive behavior assessment. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*. ACM, 577–587.
10. Elizabeth Hotham, Miranda Haberfield, Susan Hillier, Jason M White, and Gabrielle Todd. 2018. Upper limb function in children with attention-deficit/hyperactivity disorder (ADHD). *Journal of Neural Transmission* 125, 4 (2018), 713–726.
11. Jennifer Howcroft, Sue Klejman, Darcy Fehlings, Virginia Wright, Karl Zabjek, Jan Andrysek, and Elaine Biddiss. 2012. Active video game play in children with cerebral palsy: potential for physical activity promotion and rehabilitation therapies. *Archives of physical medicine and rehabilitation* 93, 8 (2012), 1448–1456.
12. Eldar Insafutdinov, Leonid Pishchulin, Bjoern Andres, Mykhaylo Andriluka, and Bernt Schiele. 2016. Deepercut: A deeper, stronger, and faster multi-person pose estimation model. In *European Conference on Computer Vision*. Springer, 34–50.
13. Hye Jin Kam, Kiyoung Lee, Sun-Mi Cho, Yun-Mi Shin, and Rae Woong Park. 2011. High-resolution actigraphic analysis of ADHD: A wide range of movement variability observation in three school courses-a pilot study. *Healthcare informatics research* 17, 1 (2011), 29–37.
14. E Lachat, H Macher, MA Mittet, T Landes, and P Grussenmeyer. 2015. First experiences with Kinect v2 sensor for close range 3D modelling. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 40, 5 (2015), 93.
15. Megan M McClelland, Claire E Cameron, Robert Duncan, Ryan P Bowles, Alan C Acock, Alicia Miao, and Megan E Pratt. 2014. Predictors of early growth in academic achievement: The head-toes-knees-shoulders task. *Frontiers in psychology* 5 (2014), 599.
16. FREDERICK J Morrison, C Cameron Ponitz, and Megan M McClelland. 2010. Self-regulation and academic achievement in the transition to school. *Child development at the intersection of emotion and cognition* 1 (2010), 203–224.
17. Mario Muñoz-Organero, Lauren Powell, Ben Heller, Val Harpin, and Jack Parker. 2018. Automatic extraction and detection of characteristic movement patterns in children with ADHD based on a convolutional neural network (CNN) and acceleration images. *Sensors* 18, 11 (2018), 3924.
18. Joseph Redmon and Ali Farhadi. 2018. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767* (2018).
19. Albert A Rizzo, J Galen Buckwalter, Todd Bowerly, Cheryl Van Der Zaag, L Humphrey, Ulrich Neumann, Clint Chua, Chris Kyriakakis, Andre Van Rooyen, and D Sisemore. 2000. The virtual classroom: a virtual reality environment for the assessment and rehabilitation of attention deficits. *CyberPsychology & Behavior* 3, 3 (2000), 483–499.
20. Dale H Schunk and Barry J Zimmerman. 1997. Social origins of self-regulatory competence. *Educational psychologist* 32, 4 (1997), 195–208.
21. Jeremy Steward, Derek Lichti, Jacky Chow, Reed Ferber, and Sean Osis. 2015. Performance assessment and calibration of the Kinect 2.0 time-of-flight range camera for use in motion capture applications. *FIG Working Week 2015* (2015), 1–14.
22. Liuyang Zhou, Zhiguang Liu, Howard Leung, and Hubert PH Shum. 2014. Posture reconstruction using kinect with a probabilistic model. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology*. ACM, 117–125.