

Research Article

Collaborative Big Data Management and Analytics in Complex Systems with Edge 2021 eaCamera: A Case Study on AI-Based Complex Attention Analysis with Edge System

Chaopeng Guo , Peimeng Zhu , Feng Li , and Jie Song 

Northeastern University, No. 195, Chuangxin Road, Hunnan District, Shenyang 110169, Liaoning, China

Correspondence should be addressed to Jie Song; songjie@mail.neu.edu.cn

Received 23 September 2021; Revised 24 October 2021; Accepted 2 November 2021; Published 27 November 2021

Academic Editor: Xuyun Zhang

Copyright © 2021 Chaopeng Guo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

As an extension of cloud computing, edge computing makes up for the deficiency of cloud computing to a certain extent. Edge computing reduces unnecessary data transmission and makes a significant contribution to the real-time and security of the system due to its characteristics that are closer to the terminal equipment. In this paper, we study the problem of attention detection. Attentional concentration during some specific tasks plays a vital role, which indicates the effectiveness and performance of human beings. Evaluation of attentional concentration status is essential in many fields. However, it is hard to define the behavior features related to the variety of tasks and behaviors. To solve this problem, we propose an intelligent edge system for attention concentration analysis, eaCamera, to recognize attentional concentration behaviors of students at the edge. To make objective measurements and save the label cost, eaCamera utilizes AI approaches to find the concentration behaviors based on a behavior analysis model with two perspectives, namely, individual perspective and group perspective. Individual perspective indicates personal behavior changes in time dimension while group perspective indicates the changes of the behavior within a group behavior manner. To evaluate the proposed system, a case study is done within a primary school to evaluate student's performance in the classroom and offer teaching advice for teachers.

1. Introduction

At present, an attention concentration analysis system is used to track an individual's attention state to obtain its attention duration. Cognitive research shows that it is vital for success in any field of skilled performance [1]. Improving the period of attentional concentration is essential in lots of fields and scenarios.

However, the concentration status cannot be observed directly [2]. Researchers, especially in cognitive computing and computer vision, try to capture the concentration status-related features like head poses, eye movements, emotions, behaviors, and visual attention to estimate concentration status. In an aspect of different elements, two kinds of attention definitions have been adopted [3]:

(i) Object-driven: attention is the process of attending to objects. Attention is object-based, so attention can also be defined as how much you focus on an object.

(ii) Task-driven: the human brain is always focused on the task at hand, which is related to the high-level mental information in the human mind. When a human is doing a task, it is a kind of high-level information in the mind, guiding human attention [5].

Visual attention is a common approach based on object-driven attention that highlights the essential regions of an image where human observers would allocate their attention at first glance. The comprehensive method to represent visual attention is by eye fixation saliency object [6–8] or saliency map [9, 10]. In these methods, the saliency object or map is usually obtained by the eye-tracking equipment that records the eye fixations of the observer looking at the image.

To drive the research of attention analysis, we implement an edge intelligence system for attention analysis. The edging

machine collects data, processes the data through the countermeasure instance detection method [11], analyzes the attentional concentration, and generates a report. The data are regularly transmitted to the cloud server to avoid data loss.

Attention detection cannot avoid the processing of the massive video. If the video is transmitted to the cloud for processing each time, it will consume a lot of time and energy in data transmission and processing. Meanwhile, it introduces nonnegligible delay. In addition, our study uses cutting-edge image processing methods for attention detection. Recognizing attention through images includes multiple tasks, such as face recognition, attention recognition, and so on. Moreover, these tasks are carried out at the edge devices. The edge devices themselves are limited by computing resources, so it must be implemented by using lightweight techniques with certain accuracy. In the intelligent edge system implemented in this research, the edge device with limited computing resources acts as the agent of the cloud to process the video. Therefore, this study promotes the research of edge-supported complex systems to a certain extent.

In the attention concentration detection part of the intelligent edge system, the attentional concentration is defined based on task-driven attention presented by task-level features. The task-level features are related to the task type and its environment. Since there a variety of tasks and different environments, it is hard to describe or find all the task-related attentional behaviors. According to the research, facial expression can be recognized according to the detection of facial feature points [12].

In recent years, information technology has been more and more widely used in the education industry. With the thriving of AI technology, its applications in education have been increasing, with a promising potential to provide customized learning, offer dynamic assessments, and facilitate meaningful interactions in online, mobile, or blended learning experiences [13]. Online education services such as Google Classroom, Zoom, and Microsoft Team also emerge one after another [14]. In turn, the research on the education industry is conducive to the design of extensible neuromorphic complete hardware primitives and the corresponding chips [15]. Therefore, we focus on the enclosed multiple people space: educational environment. In such environments, we have the following observations:

- (i) Individual behavior will change over time, which indicates that the concentration status changes as well. For example, when the teacher asks students to read their textbook, there is a student who read the textbook at the beginning but looked at the window later. By comparing the behavior changes from reading the textbook to looking at the window, the concentration changes of the individual can be captured.
- (ii) Group behavior indicates the attentional concentration status in general. In the enclosed space,

especially like a classroom in the educational environment or a factor in the industry, most people follow the instrument and concentrate on finishing their tasks, which indicates that most people would have the same behavior according to the instructions. Therefore, the outlier behavior can be captured, and the people are in abstracted status.

According to the above observations, a study case, eaCamera (A Case Study on AI-based Complex Attention Analysis with Edge System), is introduced in this paper. EaCamera focuses on the educational scenario and enclosed multiple people space. It is deployed in a primary school classroom to obtain the concentration duration in the lecture time for students. At first, eaCamera accepts the raw video of the lecture. Later, the video goes through the compute vision pipeline and behavior analysis pipeline to analyze the students' concentration status in the classroom on edge machines. In the end, eaCamera provides static reports to teachers, which can be used to downstream teaching tasks, for example, analyzing teaching performance and improving teaching approaches. In eaCamera, we propose a novel attentional concentration analysis model, which captures task-related attentional behaviors from two perspectives, namely, individual and group perspectives. The analyzed model is an unsupervised learning process, which detects attentional behaviors automatically.

The rest of the paper is organized as follows: Section-related work summarizes related works that inspire this work; Section system architecture describes the system architecture of eaCamera and explains its mechanism; Section attentional concentration analysis proposes an attention control analysis module, which is the critical component within eaCamera; Section case result describes the study case in which we deploy eaCamera within a local primary school to evaluate the proposed approach; The Conclusion section summarizes the work.

2. Related Work

Many attentional concentration types of research within the computer science area focus on the object-driven attention model, and a few focus on the task-driven model. The attention concentration analysis system is undoubtedly a waste of time and resources to transmit data to the cloud for analysis, and the intelligent edge system can solve this problem. Therefore, we introduce the basis of the edge intelligence system: edge computing firstly. Then, we review visual-based attention works. Later, we review joint attention approaches, which inspire our group behavior model. In the end, some related deep learning models, considered fundamental approaches of this study, are introduced.

2.1. Edge Computing. Edge computing refers to an open platform integrating network, computing, storage, and application core capabilities on the side close to the object or data source to provide the nearest end services. Although the security design of Internet of things application based on edge computing is still in its infancy, there are many security

solutions of edge layer Internet of things recently, which makes edge computing have great application value [16]. Compared with the system architecture of cloud computing, edge computing architecture is closer to the edge side of users and terminal devices. Compared with the distance from the user to the cloud side, the distance from the user to the edge machine is negligible [17]. Therefore, users can get a faster response. At the same time, edge computing also has a cooperation mechanism [18], which solves the privacy and trust problems in large data-driven complex systems to a certain extent. Parallel optimization algorithms and joint optimization algorithm based on reinforcement learning can also be used in edge machines to reduce task execution delay and control additional resource consumption [19, 20].

Due to the geographical dispersion of cloud data centers, the storage and processing needs of billions of geographically distributed sensors are often not met. The result is network congestion and high-latency delivery in the service, which may cause a reduction in quality of service (QoS). Typically, edge computing is made up of traditional network elements such as routers, switches, proxy servers, and base stations (BSs). It can be placed closer to IoT sensors. These components are provided with a variety of computing, storage, networking, and other functions that can support the execution of service applications [21].

At present, there are many attention detection models. Still, due to the lack of advanced hardware resources, these prediction models cannot be applied to the analysis of daily life tasks. Edge intelligent system has high application value in this problem. EaCamera captures and processes the video at the edge side, analyzes the attention concentration status, and sends the generated data to the cloud side for display. This avoids a large amount of video data transmission and saves a lot of resources and time consumption.

2.2. Visual Attention. The typical workflow of detecting saliency-based visual attention is predicting a saliency object/map, then minimizing the loss, which stands for the difference between the prediction and the ground truth. To predict a saliency map or object, single-stream networks are used to extract the feature map in early works. However, a single-stream network is unable to extract multiple-scale cues. Thus, Xun Huang proposes to use multiple stream networks [22]. Due to the importance of feature map extraction in visual problems, many works have been done to study how to extract features. Bo Du and Wei Xiong propose models to extract features in an unsupervised framework and achieve excellent performance [23, 24]. Recently, a study proposed that the first layers of the network capture macroinformation and the latter layers capture detailed information [25]. A novel architecture is proposed to extract features by combining different layers [26, 27]. A detailed survey for saliency object detection can be found in work [28, 29].

While saliency-based visual attention study about the attention of an outside human image, in many actual circumstances, we need to infer the attention of a human from the third-person view of inside images. In a typical scenario,

we need to figure out the attentional concentration object inside an image or video.

Hyeonggyu Park proposes to infer human attention by using their eye movement patterns [30]. Participants were asked to view pictures while operating under different intentions, and a classic support vector machine algorithm was used to infer participants' attention. However, this method revealed low classification accuracy due to significant inter-individual variance and psychological factors underlying intentions. Thus, using eye movement alone is not sufficient to infer human attention.

Ping Wei proposes a probabilistic method to infer the third-person view human attention based on the latent intention by jointly modeling attention, intentions, and interactions [31]. It models the attention inference as a joint optimization with latent intentions. In this paper, an EA-based approach is adopted to learn latent intentions and model parameters. Given a video with human skeletons, a joint-state dynamic programming algorithm is utilized to infer the attention direction.

The above methods define attention as the process of attending to objects. Zhixiong Nan proposes a model to infer task-driven, inside-image human attention [3]. It defines human attention as attentional objects that coincide with the task the human is doing and suggests that a human finishes the task by doing several sub-tasks in a certain temporal order with a task in mind [32]. This literal uses a model that integrates the low-level visible human pose cue with the high-level invisible task encoding information to infer human attention inside a VR video.

In this paper, eaCamera takes advantage of visual attention approaches to implement the attentional concentration analysis system. We mainly follow the task-driven branch to evaluate the students' concentration status based on high-level behavior features [33]. However, in the implementation, we do not label the concentrated behavior and abstract behavior. A concentration analysis model with two perspectives is proposed to recognize the attentional concentration status automatically.

2.3. Joint Attention. Joint attention is a behavior in which two people focus on an object or event to interact with each other. It is a form of early social and communicative behavior [34]. Joint attention involves sharing a shared focus on something (such as other people, objects, concepts, or events) with someone else. It requires the ability to gain, maintain, and shift attention. For example, a parent and child may both look at a toy they are playing with or observe a train passing by. Joint attention (also known as "shared attention") may be gained by using eye contact, gestures (e.g., pointing using the index finger), and vocalizations, including spoken words (e.g. "look over there").

In cognitive computing and artificial intelligence, joint-attention-related approaches are used in building human-robot interaction systems. Wallace Lawson introduced joint attention mechanism in robotic systems based on the assumption that when robots have joint attention with a human collaborator [35], they are perceived as more

competent and more socially interactive [36]. They proposed a joint attention estimator that creates many possible candidates for joint attention and chooses the most likely object based on the human teammate's hand cues. The human attentional objects are found based on visual attention approaches.

Some kinds of literature use joint attention mechanisms to implement multi-perspective video analysis systems. Zhaohui Yang challenged cross-view video co-analysis and delivered a novel learning-based method, in which "joint attention" is used as core notion, indicating the shared attention regions that link the corresponding views [37].

EaCamera takes advantage of joint attention to evaluate students' concentration status, namely, the group perspective within the attentional concentration model. When we cannot determine the concentration status based on individual behavior changes, the status is determined according to the group manner, based on joint attention theory.

2.4. Deep Learning Models. With the development of computer vision and deep learning, more and more deep learning models are proposed to interpret images and videos. The primary tasks of these deep learning models include classification, object detection, segmentation, action recognition, etc. Object detection and facial landmark detection are two fundamental tasks within eaCamera to capture the behavior features of the student.

Object Detection, one of the most fundamental and challenging problems in computer vision, seeks to locate object instances from many predefined categories in natural images [38]. In recent years, plenty of deep learning models have been proposed to cope with the task, Fast-RCNN [39], Yolo [40], RetinaNet [41], CornerNet [42], and so on. In eaCamera, we choose yolov3 to implement the object detection module since its simplicity and high performance.

Facial Landmark Detection predicts the locations of the fiducial facial landmark points around facial components and facial contour to capture the rigid and nonrigid facial deformations due to head movements and facial expressions [43]. Lots of works have been done on this problem to detect facial key points automatically. DAN [44] is used to implement the facial landmark detection module. EaCamera uses facial key points to estimate the head pose and emotions, which are keys to determine the students' behavior.

3. System Architecture

EaCamera is an attentional concentration analysis system deployed in a primary school to automatically obtain the concentration duration of students in the lecture time and generate analysis reports for each student and provide advice to teachers for following education improvement. This section illustrates the system architecture and data processing pipeline to present eaCamera in the general view.

The architecture of eaCamera follows the systematic architecture method of cloud edge combination (see Figure 1), in which the masked parts are the main modules of eaCamera.

The basic hardware of the edge side includes the surveillance camera responsible for recording video and the host hardware for video processing: CPU, memory, and network. The system at the edge side is a Linux operating system, which is configured with a video processor module for processing video into images, a CV module for processing images, and a behavior analysis module for analyzing students' attention concentration. At the same time, the edge side has a data sending module to communicate with the cloud, which is responsible for sending the time series data of students' attention concentration state to the cloud side.

The basic hardware of the cloud side is host hardware: CPU, network, and database module responsible for long-term storage and query of data. The system at the cloud side is also a Linux operating system, which is configured with a data receiving module to receive data of the edge side. At the same time, the module is also responsible for generating an attention analysis report from the data transmitted from the edge side or the data in the database for display by the front-end module. Users can flexibly operate the front-end module according to the working mode of eaCamera to view the attention analysis report they want to see.

There are two working modes in eaCamera, namely, active mode and passive mode. In the active mode, the edge side will scan the local storage. If an unprocessed video is found, the video will be sent to the video processor module, CV module, and behavior analysis module to generate concentration status data and sent to the cloud side to create a report in real time. The video will be preferentially saved on the local storage at the edge side in the passive mode. When the user calls the process command on the front-end page of the cloud, the required videos will be processed to generate concentration status data, which will be sent to the cloud to create a report. Active mode provides fast result access because the video is processed in real time and sent to the cloud side. But, it consumes more resources. Passive mode saves resources and money, but it cannot provide an online query service, and users have to wait until the report is generated.

In both modes, the pipeline of processing a video is the same (see Figure 2). The pipeline mainly includes a video processor, computer vision module, behavior analysis module, and front-end. The functionality and process procedure for each module is shown below.

- (1) *Video Processor.* A long-term service at the edge side, responsible for streaming video into images. The video reader uses a parameter, sampler frequency, to control the fps of the output images stream to improve the processing speed. We take 12 fps as the default setting.
- (2) *Computer Vision Module.* There are three functions within the CV module: face detector, id assigner, and landmark detector. The face detector draws a bound box for each student in all images generated by the video reader. Id assigner gives an id to each student in all images and maintains the assignments. In the

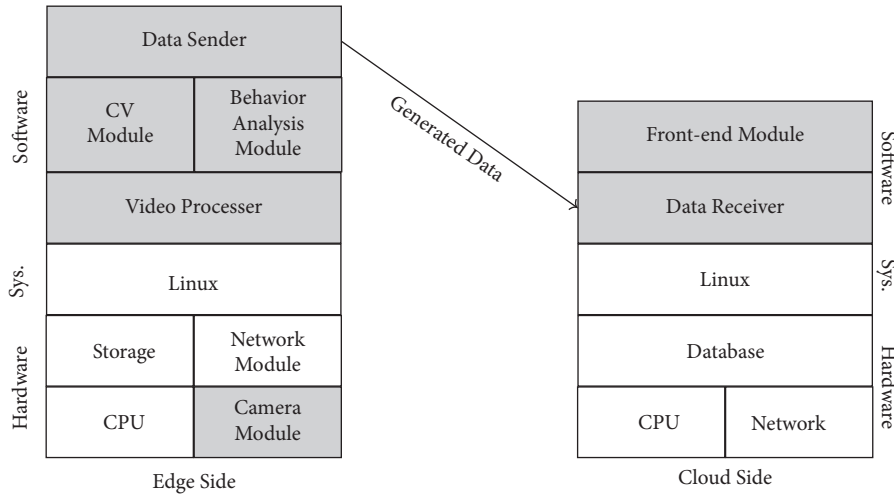


FIGURE 1: System architecture (Masked parts are main module of eaCamera).

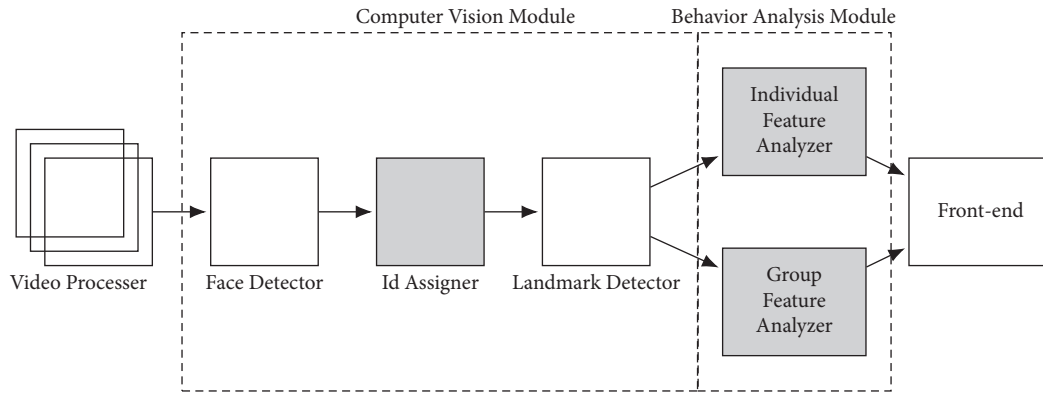


FIGURE 2: Data processing pipeline of eaCamera.

end, the landmark detector generates facial key points according to the bounded areas.

- (3) *Behavior Analysis Module.* The responsibility of the behavior analysis module recognizes the attentional concentration behaviors. The module includes two components, namely, individual feature analysis and group feature analyzer, which extract attentional features from two perspectives. All features are combined later to detect the attentional concentration status of each student.
- (4) *Front-End.* A web-based user application where the user gives a query about the classroom or student name to obtain the attentional concentration analysis reports. Besides, suppose the passive working mode is set. In that case, users can schedule analysis tasks according to their query conditions in the front-end, and the system notifies users when the reports are generated.

In Figure 2, the marked components Id Assigner, Individual Feature Analyzer, and Group Feature Analyzer are key modules of eaCamera. The novel models and algorithms are proposed in this paper to analyze student's attentional concentration status effectively.

4. Attentional Concentration Analysis

Attentional concentration analysis describes human concentration status at a certain time. Let us consider the concentration status sequence along with the time. We can explain the student concentration status during lecture time, evaluate their learning performance, and propose advice for improving teaching methods to teachers. This section focuses on the id assignment algorithm, behavior capture model, and attentional concentration analysis model used in eaCamera to build up the attentional concentration analysis system.

4.1. Id Assignment Algorithm. If we combine the functions of Face Detector and Id Assigner, it is an object tracking system [45], in which we try to track all the people in a video and assign a unique id to each of them. To be noticed, when we process a video, we process a sequence of images, and it is hard to keep the same id for the same person in all images, which is defined as an id switching problem. To conquer the id assignment task, two strategies can be used, face recognition and image similarity approach. In face recognition, the features of each person need to be collected in advance.

When the images are processed, we recognize who they are, and the id is related to their personal information, like name and gender. In the image similarity approaches, the chopped images from different video frames are compared. If the similarity of two images is higher than a user-defined threshold, we say two chopped images are assigned the same id. Deep neural networks can obtain the similarity and the features of the image. However, two approaches are unsuitable for eaCamera for the following reasons: (1) Face recognition needs to collect personal facial information, which causes privacy issues and needs a lot of labor; (2) Image similarity is time-consuming since the feature extraction neural network and the similarity comparison neural network need to be trained in advance and chopped images need to be labeled as well.

We apply a location-based id assignment strategy based on the assumption that students' seats are relatively fixed in the classroom. Figure 3 shows an example of 9 id assignments, and Algorithm 1 shows the id assignment algorithm. The id assignment process is shown as follows:

- (1). For all the images from a video, the photo with a maximized number of bounding boxes is found (Line 1 in Algorithm 1).
- (2) The bounding boxes are reshaped to the maximum size of the bounded box within the image (Line 2 in Algorithm 1), and the reshaped bounded boxes are denoted as standard boxes with unique ids. In Figure 3, the dashed rectangles are bounded boxes generated from the face detector, and the solid lines are standard boxes. The id of a standard box is set to (x, y) , which is a pair of coordinates within the image.
- (3) For the rest of the images, each bounded box is assigned with the id of the closest standard bounded box. The similarity is measured according to the intersection-over-union (IoU) function shown as equation (1), in which A_{box} and A_{standard} indicate the area of the bounded box and standard box, respectively.

$$\text{IoU}(B_{\text{Box}}, B_{\text{standard}}) = \frac{(A_{\text{box}} \cap A_{\text{standard}})}{(A_{\text{box}} \cup A_{\text{standard}})}. \quad (1)$$

4.2. Behavior Feature Model. To evaluate students' attentional concentration status, eaCamera extracts the attentional features based on student's behavior. To be noticed, students sit down in the classroom most of the lecture time. Therefore, general behaviors like sitting, writing, and reading are unsuitable for attentional concentration analysis. Besides, general behaviors are continuous action, and the start and end times are estimated within a video by action localization techniques [46]. However, the abstracted behaviors might be transient, and action localization techniques might fail in this scenario.

EaCamera defines the student behavior features based on their emotions and head poses. Figure 4 shows an output example of the landmark detector. To be noticed, the example is a chopped facial image according to Id Assignment

Algorithm. The landmark detector gives 68 key points for each facial image. We only consider six key points to describe behavior features. Six key points present the left eye, the right eye, the nose, the left cheek, the right cheek, and the chin, respectively, and the points are denoted as a set $\mathbf{P} = \{(x_{el}, y_{el}), (x_{er}, y_{er}), (x_n, y_n), (x_{cl}, y_{cl}), (x_{cr}, y_{cr}), (x_c, y_c)\}$. The behavior features are extracted according to Algorithm 2.

In Algorithm 2, the features are given by a distance vector $\mathbf{V} = \{d_{el}, d_{er}, d_{cl}, d_{cr}, d_c\}$, where d_{el} presents the distance between the left eye point and the nose point. The distance is computed by Euclid distance shown as equation (2). We use distance-based features rather than vision-based features for the following reasons:

- (1). It is efficient to obtain features for all facial images. There are more than 50 students in the classroom, and we need to collect behavior features for each of them. Therefore, the behavior feature model needs to be efficient and straightforward.
- (2) The distance-based behavior features describe the head poses and emotions. The head pose is essential in the classroom to describe the concentration status since most of the time, students focus on the teacher and blackboard. Our eaCamera is installed in the front of the classroom. Therefore, the distances are symmetrical if the student is looking at the front. However, if the distances are unsymmetrical, we say the student is looking at the left or right sides.
- (3) Since we consider the distances between the eyes and the nose and the distances between the cheeks and the nose, the behaviors like lowering one's head and raising one's head can be found.

$$d_t = \sqrt{(x_t - x_n)^2 + (y_t - y_n)^2}, \quad t \in \{el, er, cl, cr, c\}. \quad (2)$$

4.3. Attentional Concentration Analysis Model. The behavior feature model can find the head pose changes and the emotion changes. However, it is not sufficient to automatically analyze the attentional concentration status because we need to point out the concentration and abstracted behavior manually. Namely, we have to label the data. Therefore, we propose an attentional concentration analysis model to recognize the concentration-related behaviors. In eaCamera, the attentional concentration is described in two perspectives, namely, individual perspective and group perspective based on the following assumption:

4.3.1. Individual Perspective. Students' attentional concentration status changes during lecture time. However, primary students pay attention to the lecture content most of the time. Therefore, we try to observe the changes in students' behavior features, and the attentional concentration status can be captured from an individual perspective. Besides, primary students are always concentrated at the beginning of the lecture or the teacher's instruction. Figure 5 shows an example of individual attentional

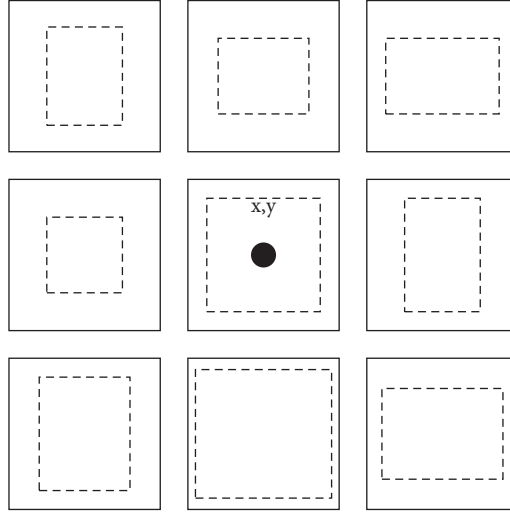


FIGURE 3: Id assignment strategy.

Input: L_{Box} : The list of bounded box lists. Each bounded box list includes a list of four coordinates ($L_{\text{Coordinate}}$) for a video frame. The bounded box is a list of coordinates.

Output: The list of id for all bounded box L_{ID}

- (1) $\text{frameIndex}, \text{maxNumberOfBox} \leftarrow \text{FindFrameWithMaxNumberBoxes}(L_{\text{Box}})$
- (2) $L_{\text{StandardBoxes}} \leftarrow \text{GenerateStandardBoxes}(L_{\text{Box}}, \text{frameIndex}, \text{maxNumberOfBox})$
- (3) $L_{\text{ID}} \leftarrow \{\}$
- (4) **for** $L_{\text{Coordinate}}$ in L_{Box} **do**
- (5) $\text{idAssignment} \leftarrow \text{MatchStandardBox}(\text{StandardBox}, L_{\text{Coordinate}})$
- (6) Append ($L_{\text{ID}}, \text{idAssignment}$)
- (7) **end for**
- (8) **return** E_n ;

ALGORITHM 1: Id Assignment Algorithm.



FIGURE 4: An output example of the landmark detector.

Input: P: The list of key points. Each key point includes a list of two coordinates
Output: V: A feature vector of the facial image

- (1) $\{(x_{el}, y_{el}), (x_{er}, y_{er}), (x_n, y_n), (x_{cl}, y_{cl}), (x_{cr}, y_{cr}), (x_c, y_c)\} \leftarrow \text{ExtractKeyPoints}(P)$
- (2) $V \leftarrow \{\}$
- (3) **for** x, y in $\{(x_{el}, y_{el}), (x_{er}, y_{er}), (x_{cl}, y_{cl}), (x_{cr}, y_{cr}), (x_c, y_c)\}$ **do**
- (4) $d \leftarrow \text{DistanceBetween}((x, y), (x_n, y_n))$
- (5) Append (V, d)
- (6) **end for**
- (7) **return** V;

ALGORITHM 2: Behavior feature generation algorithm.

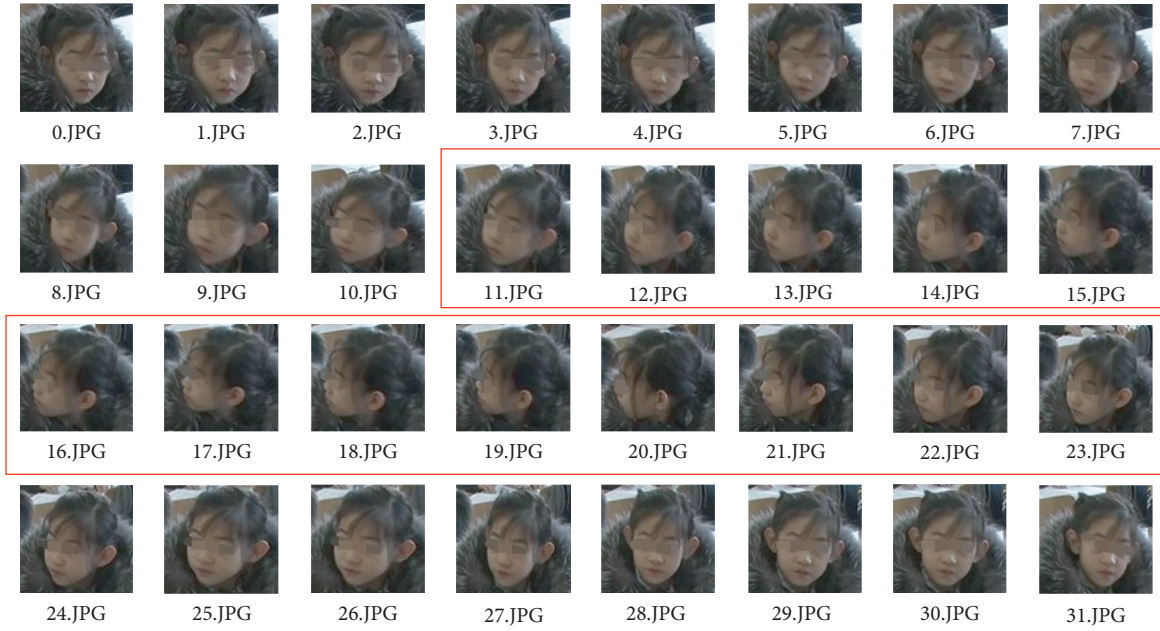


FIGURE 5: Example of individual attentional concentration changes.

concentration changes, in which the abstracted behavior is recognized by finding the outliers within behavior features in all frames.

4.3.2. Group Perspective. Teaching is a group activity. When most students pay attention to the lecture content, they have the same action pattern, which indicates students' attentional concentration status. Since most of the students focus on the lecture content in primary school, we use this character to recognize the outliers who are nonconcentrated. Figure 6 shows an example of group attentional concentration status comparison. In the figure, the teacher was demonstrating the lecture content. A few students did not look at the teacher but focused on the desk or the textbook, which can be recognized as abstracted status since they are the outliers within all students.

According to the lecture content, the object, which students pay attention to, changes during the lecture time.

The objective could be the teacher, the blackboard, the textbook, and so on. The individual perspective model only captures the behavior changes over time for a student, and it cannot capture the behavior changes caused by lecture content changes. Therefore, the group perspective model determines the attentional concentration status if the behavior feature change happens for a single student.

In the individual perspective model, the first step is to obtain the attentional concentration feature baseline. We assume that the primary student is concentrated at the beginning of the lecture or at the beginning of the teacher's instruction. Equation (3) shows a baseline feature matrix for a student, which consists of 1440 behavior feature vectors. The baseline feature matrix is extracted from the first 2 minutes of the class, which includes 120 seconds in total and 12 frames for each second. According to Chinese classroom habits, teachers will say something important in the first two minutes. We can think that the students' attention state is



FIGURE 6: Example of group attentional concentration status comparison.

the most concentrated in the first 2 minutes of class, and it is the most accurate to compare the students' state in the first 2 minutes with the follow-up. At the same time, according to our calculation method, 1440 eigenvectors will be generated from the data in the first 2 minutes, and this amount of data has no calculation pressure on the edge machine.

$$M = \begin{bmatrix} d_{el}^1 & d_{er}^1 & d_{cl}^1 & d_{cr}^1 & d_c^1 \\ d_{el}^2 & d_{er}^2 & d_{cl}^2 & d_{cr}^2 & d_c^2 \\ \dots & \dots & \dots & \dots & \dots \\ d_{el}^{1440} & d_{er}^{1440} & d_{cl}^{1440} & d_{cr}^{1440} & d_c^{1440} \end{bmatrix}. \quad (3)$$

A baseline vector is obtained in the following steps: (1) The frequency diagram is done for each row within the feature matrix. (2) Choose the group of data with maximum frequency. (3) Compute the mean value of the group data as a part of the baseline. Figure 7 shows an example of frequency diagram for $[d_{el}^1, d_{er}^1, \dots, d_{el}^{1440}]$. The class interval is set to 0.5. The majority distances are between 1.0 cm and 1.5 cm. Let us say the mean value is 1.2 cm. Then, we set $d_{el}^* = 1.2$ cm. Using the same strategy, we can obtain a standard baseline vector to present attentional concentration status of a student, denoted as $V^* = \{d_{el}^*, d_{er}^*, d_{cl}^*, d_{cr}^*, d_c^*\}$.

In a short period, the teaching object remains the same. To evaluate the attentional concentration status, we only compare the current student behavior feature V with the attentional concentration baseline V^* . A distance, Offset, between two features is obtained according to equation (4). When $\text{Offset}(V, V^*) < \varepsilon$, we say the students are concentrated, in which ε is the pre-defined parameter present in concentration tolerance.

$$\text{Offset}(V, V^*) = \sqrt{\sum_{t \in \{el, er, cl, cr, c\}} (d_t - d_t^*)^2}. \quad (4)$$

The object that students pay attention to changes according to the teaching content. It causes attentional concentration baseline changes, which makes the individual perspective model fail. The group perspective model is introduced to solve this problem. In the individual perspective model, if the offset between the current behavior feature and the standard feature is higher than ε , the student might be in the abstracted status or still in concentration status with

another attentional object. Therefore, we observe the others to decide whether the student is still concentrated.

In the individual perspective model, when an abstracted status is detected, namely, $\text{Offset}(V, V^*) \geq \varepsilon$, then we use the group perspective model to detect the attentional object change. In the group perspective model, the evaluation of a student's attentional concentration status is done in the following steps:

- (1) For all students, we compute the offset matrix G shown as equation (5), in which each row presents offsets for a student within 2 seconds (12 frames for each second).
- (2) Use ε as a threshold to compute the binarization matrix of G , denoted as G_b .
- (3) Summarize each column of G_b , denoted as group indicator, which indicates the number of students whose attentional concentration is changed.
- (4) Since most of the primary students are concentrated in the class, if more than half of the students' attentional concentration status changes, we say the attentional object changes currently. In this case, the concentration standard baselines for all students are recomputed.
- (5) Otherwise, we say the student is in abstracted status.

$$G = \begin{bmatrix} \text{Offset}(V_1^*, V_1^i) & \text{Offset}(V_1^*, V_1^{i+1}) & \dots & \text{Offset}(V_1^*, V_1^{i+24}) \\ \text{Offset}(V_2^*, V_2^i) & \text{Offset}(V_2^*, V_2^{i+1}) & \dots & \text{Offset}(V_2^*, V_2^{i+24}) \\ \dots & \dots & \dots & \dots \\ \text{Offset}(V_n^*, V_n^i) & \text{Offset}(V_n^*, V_n^{i+1}) & \dots & \text{Offset}(V_n^*, V_n^{i+24}) \end{bmatrix}. \quad (5)$$

5. Case Result

EaCamera is deployed in a primary school in Liaoning province, China. In the school, each classroom contains 40 students, and the lecture time is usually 40 minutes. In eaCamera, we choose state-of-art deep learning models to implement face detectors and landmark detectors. The face detector is implemented based on YoloV3 [47] and ArcFace [48], and the landmark detector is implemented based on DAN [44]. In this section, we first briefly introduce the

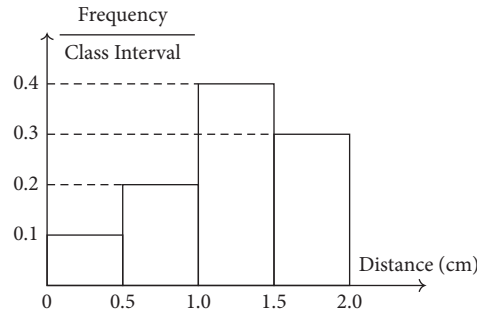


FIGURE 7: Example of frequency diagram.

structure of the network we use, and then we show the attentional concentration analysis result for each component.

As the calculation of eaCamera is mainly carried out at the edge side, we have selected several lightweight network implementations on the premise of ensuring accuracy.

- (i) YoloV3: Yolo3 adopts a network structure called darknet-53 (including 53 convolution layers). It draws lessons from the practice of residual network, sets up fast links between some layers, forms a deeper network level and multi-scale detection, which improves the detection effect of mean Average Precision(mAP) and small objects [47].
- (ii) ArcFace: ArcFace is a new loss function for face recognition based on additive angular margin loss. Its focus is to directly maximize the classification boundary in the angle space [48].
- (iii) DAN: DAN is based on AlexNet network, which explores the adaptation relationship between source and target. As a representative method of deep transfer learning, it makes full use of the transferable characteristics of deep network, and then introduces the maximum mean discrepancy (MMD) in statistical learning, which has achieved good accuracy [44].

Figure 8 shows the result of the face detector. The red box is the bounded box for each student's face, and the number is the model's confidence for detecting facial objects. Figure 9 shows the result of the landmark detector and key point filter. To be noticed, after the face detector, the id assigner gives a unique id for each bounded box. However, the face detector might fail to generate a bounded box in some students. In this case, the id assigner ignores these ids, which indicates that the student is assigned with an abstraction label in this frame.

Figures 10 and 11 show examples of concentrated behaviors and abstracted behaviors. In Figure 10, student A follows the teacher's instruction. When the teacher required students to read the textbook, student A stared down her head and paid attention to it. When the teacher asked to look at the blackboard, student A was raising her head as well. In Figure 11, student B's behaviors are adverse. Therefore, we say she was abstracted.

In the front-end module, eaCamera generates statistic reports of the chosen lecture. To simplify the description, we only show the result of 18 students, whose id is assigned from numbers 1 to 18. To be noticed, a student with a lower id indicates that his position is more close to the blackboard and the teacher.

Figures 12 and 13 show the concentrated duration report and the corresponding statistic report. As can be seen from Figure 12, except for individual examples, the attention concentration time decreases as the ID increases. This shows that the closer the students are to the blackboard, the more focused they are in class, which is consistent with the reality. The closer the students are to the teacher, they may be more afraid to be distracted. In Figure 13, two-thirds of the students' attention concentration time is 30–40 minutes. The time of a class is 45 minutes, that is, most primary school students' attention is focused on the class time, which is consistent with our hypothesis.

EaCamera can give analysis reports according to user requirements or user-defined scripts. Users can generate corresponding reports for viewing according to their own needs. For example, users can view the analysis report in a certain day or even in a customized time period, the attention concentration analysis report of all people in a class, or the average attention duration of each class in different courses. Through different choices, users can carry out different target analysis. Figure 14 shows the compression of mean concentration duration between different lectures, namely, English, Chinese Literature, and Math. The result shows that students are likely pay more attention to Math considering the difficulty of the course and its importance. Using different analysis reports, teachers may change or improve their teaching approaches to get better performance.

6. Discussion

In this section, we compare eaCamera with other attention-related studies to analyze the advantages and disadvantages of eaCamera.

Traditionally, the attention concentration status of students is collected manually by human observers. With the development of science and technology, more and more attention detection methods have been proposed:



FIGURE 8: Result of face detector.

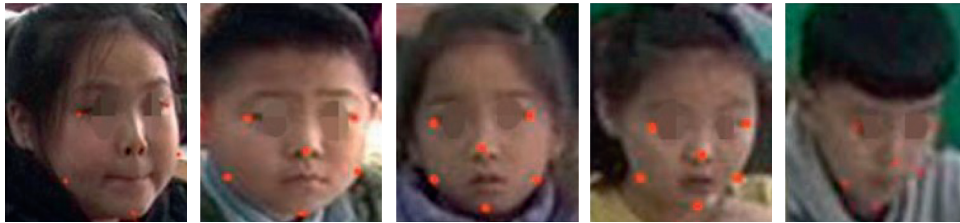


FIGURE 9: Result of key point filter.

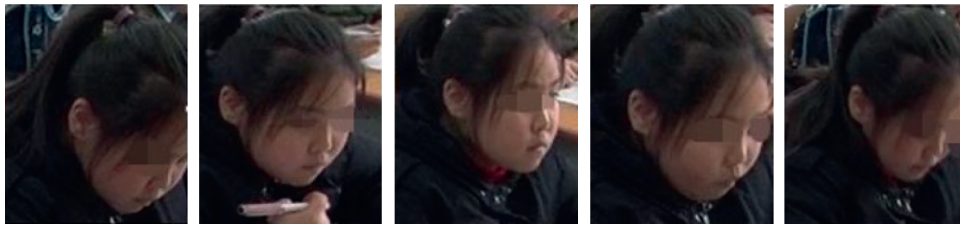


FIGURE 10: Example of student A concentrated status.

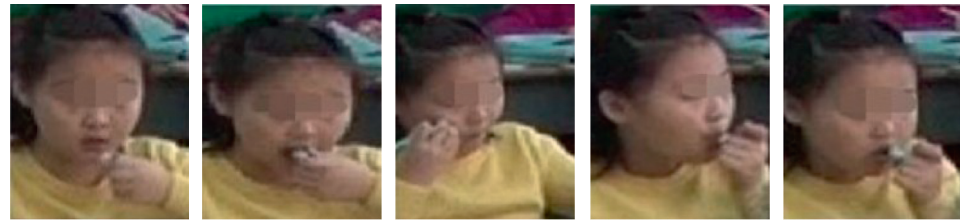


FIGURE 11: Example of student B abstracted status.

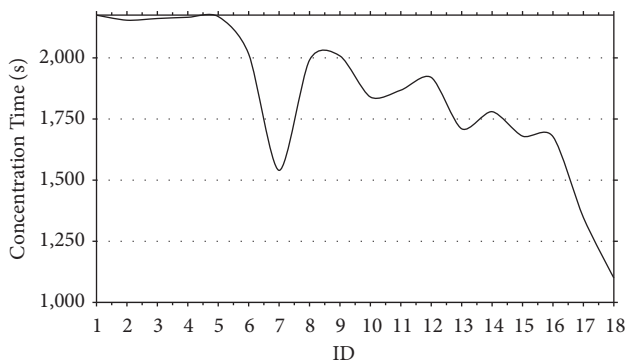


FIGURE 12: Concentrated duration report for all students.

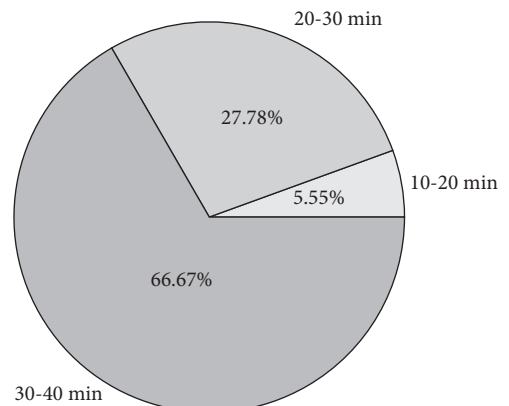


FIGURE 13: Concentration duration statistic report.

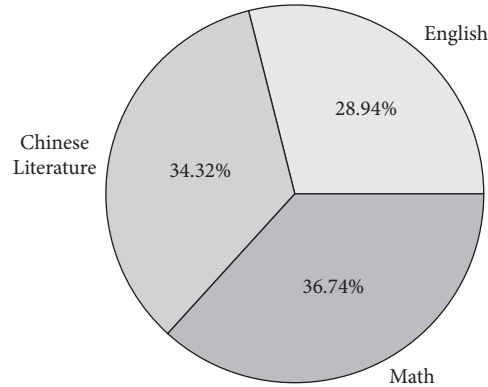


FIGURE 14: Mean concentration duration statistic report in aspects of different lectures.

TABLE 1: Comparison of eaCamera with other studies.

System	Function			Cost		Method		
	Attention analysis	Real-time warning	Generate report	Long-time storage	Expensive	Cheap	Tradition	New
Study 1	✓					✓	✓	
Study 2	✓	✓			✓			✓
Study 3	✓	✓			✓			✓
eaCamera	✓		✓	✓		✓		✓

- (i) Study 1: Sujan Poudyal uses image processing technology to extract features from the student data captured from the monitoring system, and uses three data mining methods (SVM, decision tree, and KNN) to classify students' attention patterns [49].
- (ii) Study 2: Xin Zhang proposed that wearable devices can be used to analyze students' attention state in class. The system integrates head movement module, pen movement module, and visual focus module to accurately analyze students' attention level in class [50].
- (iii) Study 3: Shimeng Peng has developed an attention perception system (AAS) based on electroencephalography (EEG) signals to accurately identify students' attention level, which has high application potential and can provide online teachers with timely early warning of low attention level feedback in an e-learning environment [51].

We compared eaCamera with the above three studies in terms of function, cost, and method of use (see Table 1). Study 1 uses traditional methods such as data mining to analyze the data captured from the camera. Its function is relatively simple. It can only analyze the attention state of students, but it is not real time, and does not support the generation of reports and long-term storage of data; Study 2 and Study 3, respectively, use wearable devices and brain wave analysis devices to monitor students' attention state in class in real time, which can provide real-time alarm. The method is novel, but compared with cameras, the cost of their devices is very high. EaCamera adopts the method of edge calculation, the main functions are set at the edge side, and the hardware only needs cameras. It uses novel methods

such as CV to realize the function of attention detection. At the same time, due to the cloud side, our research can realize the long-term storage of data, so as to facilitate the long-term comparative analysis and customize the analysis report. It can be seen that eaCamera uses novel methods to complete more comprehensive functions at a lower cost. However, eaCamera does not support the real-time warning function, which is also the further work of eaCamera.

7. Conclusion

We propose a novel edge intelligence system for attention analysis, eaCamera, in this case study, which is deployed in a primary school in China to evaluate students' attentional concentration status during lecture time. EaCamera gathers videos of lectures and utilizes deep learning models to extract behavior features of students. Later, the features are processed by an attentional concentration analysis model proposed to capture students' concentrated and abstracted behaviors automatically.

In eaCamera, we mainly focus on three modules. (1) Id assignment algorithm is used to assign a unique id to students within the same lecture. The assigned id keeps the same id for the same person within a video to indicate each student's identity. (2) Behavior feature model is adopted to generate the behavior feature vectors for each student in a video frame. A behavior feature is a vector that contains distances between facial key points. (3) Attentional concentration analysis model is proposed to capture students' attentional concentration status. Based on the above three modules, we implement an edge intelligent system for attention concentration analysis, which not only provides a certain reference value for the research of attention

concentration detection but also promotes the application of science and technology in the educational environment. At the same time, the system also makes a certain contribution to the research of edge intelligent complex system.

EaCamera can be used in enclosed multi-people spaces to analyze users' attentional states according to two perspectives, namely, individual perspective and group perspective. Individual perspective captures the behavior feature changes over time, and group perspective captures the behavior feature outlier within a group of people. EaCamera relies on the assumption that the majority of people within a group are concentrated. Therefore, eaCamera cannot work within some environments that do not have the above assumption. Besides, eaCamera relies on computer vision technique, namely, its anti-interference ability is limited. If the faces are blocked, then eaCamera would not be able to produce the correct analyses. In further work, we want to expand the usage scenario of eaCamera, for example, diver notification system, factory monitoring system, office monitoring system, driver monitoring system, and hospital operation monitoring system. In order to improve the existing shortcomings of eaCamera, we can expand the recognition scope of eaCamera in the future, recognize the state of attentional concentration according to human behavior, and make a certain contribution to pedestrian behavior recognition and other scenarios in automatic driving. In the current implementation, the facial-based behavior feature model is adopted considering lecture and primary school characters. A novel behavior feature model is needed to capture the behavior features in a different scenario in further work.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] A. Moran, "Concentration: attention and performance," *The Oxford Handbook of Sport and Performance Psychology*, Oxford University Press, Oxford, UK, pp. 117–130, 2012.
- [2] Y. Wang, Y. Shi, J. Du, Y. Lin, and Q. Wang, "A CNN-based personalized system for attention detection in wayfinding tasks," *Advanced Engineering Informatics*, vol. 46, Article ID 101180, 2020.
- [3] Z. Nan, T. Shu, R. Gong et al., "Learning to infer human attention in daily activities," *Pattern Recognition*, vol. 103, Article ID 107314, 2020.
- [4] B. J. Scholl, "Objects and attention: the state of the art," *Cognition*, vol. 80, no. 1-2, pp. 1–46, 2001.
- [5] A. Seemann, *New Developments in Psychology, Philosophy of Mind, and Social Neuroscience*, MIT Press, Cambridge, MA, USA, 2011.
- [6] T. Liu, Z. Yuan, J. Sun et al., "Learning to detect a salient object," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, pp. 353–67, 2011.
- [7] J. Wang, B. Cao, P. Yu, L. Sun, W. Bao, and X. Zhu, "Deep learning towards mobile applications," in *Proceedings of the 2018 IEEE 38th International Conference on Distributed Computing Systems*, pp. 1385–1393, IEEE, Vienna, Austria, July 2018.
- [8] W. Wang, J. Shen, R. Yang, and F. Porikli, "Saliency-Aware video object segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 1, pp. 20–33, 2018.
- [9] A. Borji, M. N. Ahmadabadi, B. N. Araabi, and M. Hamidi, "Online learning of task-driven object-based visual attention control," *Image and Vision Computing*, vol. 28, no. 7, pp. 1130–1145, 2010.
- [10] W. Wang, J. Shen, J. Xie, M. M. Cheng, H. Ling, and A. Borji, "Revisiting video saliency prediction in the deep learning era," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, pp. 220–237, 2021.
- [11] K. Han, Y. Li, and B. Xia, "A cascade model-aware generative adversarial example detection method," *Tsinghua Science and Technology*, vol. 26, no. 6, pp. 800–812, 2021.
- [12] T. Qian, "Emotion recognition system based on distributed edge computing," *Computer Science*, vol. 48, pp. 638–643, 2021.
- [13] K. Zhang and A. B. Aslan, "AI technologies for education: recent research & future directions," *Computers and Education: Artificial Intelligence*, vol. 2, Article ID 100025, 2021.
- [14] A. Agarwal, S. Sharma, V. Kumar, and M. Kaur, "Effect of E-learning on public health and environment during COVID-19 lockdown," *Big Data Mining and Analytics*, vol. 4, no. 2, pp. 104–115, 2021.
- [15] Y. Zhang, P. Qu, and W. Zheng, "Towards "general purpose" brain-inspired computing system," *Tsinghua Science and Technology*, vol. 26, no. 5, pp. 664–673, 2021.
- [16] K. Sha, T. A. Yang, W. Wei, and S. Davari, "A survey of edge computing-based designs for IoT security," *Digital Communications and Networks*, vol. 6, no. 2, pp. 195–202, 2020.
- [17] Z. Tong, F. Ye, M. Yan, H. Liu, and S. Basodi, "A survey on algorithms for intelligent computing and smart city applications," *Big Data Mining and Analytics*, vol. 4, pp. 155–172, 2021.
- [18] L. Yuan, Q. He, S. Tan et al., "CoopEdge: a decentralized blockchain-based platform for cooperative edge computing," in *Proceedings of the Web Conference 2021*, pp. 2245–2257, ACM, Ljubljana Slovenia, April 2021.
- [19] J. Luo, J. Li, L. Jiao, and J. Cai, "On the effective parallelization and near-optimal deployment of service function chains," *IEEE Transactions on Parallel and Distributed Systems*, vol. 32, no. 5, pp. 1238–1255, 2021.
- [20] J. Cai, Z. Huang, L. Liao, J. Luo, and W.-X. Liu, "APPM: adaptive parallel processing mechanism for service function chains," *IEEE Transactions on Network and Service Management*, vol. 18, no. 2, pp. 1540–1555, 2021.
- [21] X. Zheng, M. Li, and J. Guo, "Task scheduling using edge computing system in smart city," *International Journal of Communication Systems*, vol. 34, no. 6, 2021.
- [22] X. Huang, C. Shen, X. Boix, and Q. Zhao, "SALICON: reducing the semantic gap in saliency prediction by adapting deep neural networks," in *Proceedings of the International Conference on Computer Vision*, pp. 1–9, IEEE, Santiago, Chile, December 2015.
- [23] B. Du, Y. Wang, C. Wu, and L. Zhang, "Unsupervised scene change detection via latent dirichlet allocation and multivariate alteration detection," *IEEE Journal of Selected Topics in*

- Applied Earth Observations and Remote Sensing*, vol. 11, no. 12, pp. 4676–4689, 2018.
- [24] W. Xiong, L. Zhang, B. Du, and D. Tao, “Combining local and global_Rich and robust feature pooling for visual recognition,” *Pattern Recognition*, vol. 62, pp. 225–235, 2017.
- [25] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, “Return of the devil in the details: delving deep into convolutional nets,” in *British Machine Vision Conference*, M. F. Valstar, A. P. French, and T. P. Pridmore, Eds., BMVA Press, University of Nottingham, Nottingham, UK, pp. 1–11, 2014.
- [26] M. Kümmerer, L. Theis, and M. Bethge, “Deep gaze I: boosting saliency prediction with feature maps trained on ImageNet,” in *Proceedings of the International Conference on Learning Representations*, Y. Bengio and Y. LeCun, Eds., pp. 1–12, San Diego, CA, USA, May 2015.
- [27] W. Wang and J. Shen, “Deep visual attention prediction,” *IEEE Transactions on Image Processing*, vol. 27, no. 5, pp. 2368–2378, 2018.
- [28] W. Wang, Q. Lai, H. Fu, J. Shen, H. Ling, and R. Yang, “Salient object detection in the deep learning era: an in-depth survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, p. 1, 2021.
- [29] C. Baoyuan, L. Yitong, and S. Kun, “Research on object detection method based on FF-yolo for complex scenes,” *IEEE Access*, vol. 9, pp. 127950–127960, 2021.
- [30] H. Park, S. Lee, M. Lee, M.-S. Chang, and H.-W. Kwak, “Using eye movement data to infer human behavioral intentions,” *Computers in Human Behavior*, vol. 63, pp. 796–804, 2016.
- [31] P. Wei, D. Xie, N. Zheng, and S.-C. Zhu, “Inferring human attention by learning latent intentions,” in *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 1297–1303, Melbourne, Australia, August 2017.
- [32] X. Zhang, N. Mlynaryk, S. Japee, and L. G. Ungerleider, “Attentional selection of multiple objects in the human visual system,” *NeuroImage*, vol. 163, pp. 231–243, 2017.
- [33] Y. Xu, X. Hong, Q. He, G. Zhao, and M. Pietikäinen, “A task-driven eye tracking dataset for visual attention analysis,” in *Advanced Concepts for Intelligent Vision Systems*, S. Battiato, J. Blanc-Talon, G. Gallo, W. Philips, D. Popescu, and P. Scheunders, Eds., Springer International Publishing, Berlin, Germany, pp. 637–648, 2015.
- [34] P. Mundy, “A review of joint attention and social-cognitive brain systems in typical development and autism spectrum disorder,” *European Journal of Neuroscience*, vol. 47, no. 6, pp. 497–514, 2018.
- [35] W. Lawson, A. M. Harrison, E. S. Vorm, and J. G. Trafton, “Joint attention estimator,” in *Proceedings of the Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 331–333, ACM, Cambridge UK, March 2020.
- [36] C.-M. Huang and A. L. Thomaz, “Effects of responding to, initiating and ensuring joint attention in human-robot interaction,” in *Proceedings of the 2011 RO-MAN*, pp. 65–71, IEEE, Atlanta, GA, USA, July 2011.
- [37] Z. Yang, Y. Wang, X. Chen et al., “CARS: continuous evolution for efficient neural architecture search,” in *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pp. 1826–1835, IEEE, Seattle, WA, USA, June 2020.
- [38] L. Liu, W. Ouyang, X. Wang et al., “Deep learning for generic object detection: a survey,” *International Journal of Computer Vision*, vol. 128, no. 2, pp. 261–318, 2020.
- [39] R. Girshick, “Fast R-CNN,” in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1440–1448, IEEE Computer Society, Santiago, Chile, December 2015.
- [40] J. Redmon and A. Farhadi, *YOLOv3: An Incremental Improvement*, <https://arxiv.org/abs/1804.02767>, 2018.
- [41] X. Li, H. Zhao, and L. Zhang, *Recurrent RetinaNet: A Video Object Detection Model Based on Focal Loss*, L. Cheng, A. C. S. Leung, and S. Ozawa, Eds., Springer International Publishing, Berlin, Germany, 2018pp. 499–508, Neural Information Processing.
- [42] H. Law and J. Deng, “CornerNet: detecting objects as paired keypoints,” *International Journal of Computer Vision*, vol. 128, no. 3, pp. 642–656, 2020.
- [43] M. Bodini, “A review of facial landmark extraction in 2D images and videos using deep learning,” *Big Data and Cognitive Computing*, vol. 3, no. 1, p. 14, 2019.
- [44] M. Kowalski, J. Naruniec, and T. Trzcinski, “Deep alignment network: a convolutional neural network for robust face alignment,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 2034–2043, IEEE, Honolulu, HI, USA, July 2017.
- [45] W. Luo, J. Xing, A. Milan, X. Zhang, W. Liu, and T.-K. Kim, “Multiple object tracking: a literature review,” *Artificial Intelligence*, vol. 293, p. 103448, 2021.
- [46] S. Narayan, H. Cholakkal, F. S. Khan, and L. Shao, “3C-Net: Category count and center loss for weakly-supervised action localization,” in *Proceedings of the International Conference on Computer Vision*, pp. 8678–8686, OpenReview.net, Seoul, Republic of Korea, October 2019.
- [47] J. Qi, C. Wang, L. Cheng, S. Jiang, X. Zhang, and H. Jing, “YOLOFKP: dense face detection based on YOLOv3 key point network,” in *Proceedings of the 2020 9th International Conference on Computing and Pattern Recognition*, pp. 187–191, ACM, Xiamen China, October 2020.
- [48] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, “ArcFace: additive angular margin loss for deep face recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4690–4699, IEEE, Long Beach, CA, USA, June 2019.
- [49] S. Poudyal, M. J. Mohammadi-Aragh, and J. E. Ball, “Data mining approach for determining student attention pattern,” in *Proceedings of the 2020 IEEE Frontiers in Education Conference (FIE)*, pp. 1–8, IEEE, Uppsala, Sweden, October 2020.
- [50] X. Zhang, C.-W. Wu, P. Fournier-Viger, L.-D. Van, and Y.-C. Tseng, “Analyzing students’ attention in class using wearable devices,” in *Proceedings of the 2017 IEEE 18th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, pp. 1–9, IEEE, Macau, China, June 2017.
- [51] C.-M. Chen, J.-Y. Wang, and C.-M. Yu, “Assessing the attention levels of students by using a novel attention aware system based on brainwave signals: novel attention aware system based on brainwave signals,” *British Journal of Educational Technology*, vol. 48, no. 2, pp. 348–369, 2017.