ISSN No:-2456-2165

Online Proctoring System: A Client Side Approach Using Deep Learning

Devesh Bedmutha
Information Technology
Pune Institute of Computer Technology
Pune, India

Purva Bapecha
Information Technology
Pune Institute of Computer Technology
Pune, India

Digambar Chaure
Information Technology
Pune Institute of Computer Technology
Pune, India

Piyush Bora
Information Technology
Pune Institute of Computer Technology
Pune, India

Rachna Karnavat
Project Guide
Pune Institute of Computer Technology,
Pune, India

Abstract:- An AI based Online Proctoring System isn't a new concept and many such capable exam portals do already exist. However, all of them have an unsolved design flaw which is server side processing. To detect any suspicious activity, the sites either take a snapshot of the examinee in regular intervals which is doable but is very weak, or they continuously send the video feed over to the server for processing which being comparatively more effective, is highly expensive. Sending video feeds of tens of thousands of students and processing them in real time can be very heavy on the server as well as costly for the client. To counter all these flaws, proposing an AI based proctoring system that securely works on the client side. Overall goal is to allow the face detection system and suspicious activity detection system to run on the client side which will significantly reduce the server load and dependency on the network. In this review paper we explored various algorithms for face verification, object detection, also reviewed pre-existing OPS systems and learned about their architecture.

Keywords:- *CNN (Convolution Neural Network), OPS (on-line proctoring system), TFOD (Tensorflow Object Detection).*

I. INTRODUCTION

In the wake of the digital age and the transformation of traditional educational paradigms, remote learning, or online education, has emerged as a dynamic force that is reshaping the landscape of knowledge dissemination. The advent of remote learning platforms has not only widened the horizons of accessibility to education but has also redefined the very essence of learning itself. This shift towards online education has brought a myriad of opportunities for learners of all ages, transcending geographical boundaries, socioeconomic constraints, and time limitations. Yet, as the digital learning revolution surges forward, it brings with it a unique set of challenges that necessitate innovative solutions. As the number of learners engaged in remote education continues to increase, so too does the need for a robust, reliable, and user-friendly online proctoring system

While the concept of AI-based proctored exam sites isn't new, there's a prevailing issue that plagues existing platforms: an unsolvable design flaw that hampers continuous monitoring of examinees. These platforms typically suffer from a glaring limitation - an inadequate proctor-to-examinee ratio. Consequently, they struggle to effectively monitor every student, and their internal suspicious activity detection systems come with inherent limitations. Existing solutions employ two primary methods to detect suspicious behavior. The first involves taking periodic snapshots of examinees, which, while feasible, offers weak surveillance capabilities. The alternative method entails constantly streaming video feeds to a central server for real-time processing. While more effective, this approach is exceedingly costly, burdening both servers and clients. Transmitting video feeds from tens of thousands of students and processing them in real-time places an immense strain on server resources, resulting in significant expenses. Moreover, the quality of the video feed depends on network stability, introducing additional vulnerabilities into the detection system. To address these critical flaws, introduced an innovative AI- based proctoring system that operates securely on the client side. The primary objective is to shift the burden of face detection and suspicious activity monitoring from the server to the client, thereby alleviating server load and reducing dependency on network infrastructure. This transition not only enhances the detection system's capabilities but also ensures a consistent video feed quality, unaffected by network fluctuations. Furthermore, system boasts exceptional scalability since all processing occurs on the client side, rendering it immune to fluctuations in the number of students.

II. LITERATURE SURVEY

The research paper [1] focuses on an in-depth examination of several online proctoring systems (OPS), including well- known platforms such as ProctorU, Kryterion, and Xproctor. These OPS rely on the utilization of webcams and micro- phones to closely monitor students' activities during online examinations, ensuring the integrity of the assessment process.

However, a significant portion of the paper is dedicated to shedding light on the pressing issues of security and privacy that accompany the adoption of these systems. Specifically, the paper delves into concerns surrounding user authentication methods, as well as the collection and handling of sensitive biometric data. Through this comprehensive exploration, the paper aims to provide a critical evaluation of OPS in the context of the broader education technology landscape, underlining the essential need for robust security measures and safeguards to protect students' privacy in an increasingly digital academic environment.

In the context of e-learning, a thorough and methodical review of 53 studies(published from Jan 2016 to July 2020) pertaining to online exam solutions is conducted in the paper [2]. It encompasses an extensive examination of both proposed tools which include FlexAuth, ViLLE, Secure Exam Environment(SEE),etc. and existing tools including ProVerif, Weka, OpenNLP, etc. in selected studies, while also delving into the global factors that influence their adoption and the interplay between these elements. The key findings of this review reveal the existence of 21 proposed tools and 25 established tools, as well as the identification of four pivotal global adoption factors i.e. Network Infrastructure, Hardware Requirements, Implementation Complexity and Training Re- quirements. Looking ahead, future research aims to build a predictive system that can forecast the likelihood of adopting online exam solutions.

The research in paper[3] highlights the considerable variability observed among commercial Online Proctoring (OLP) systems, presenting a dilemma for educators in their quest for the most fitting solution. While automated OLP systems bring undeniable advantages, the study acknowledges the accuracy challenges they encounter when identifying misconduct. As a solution, the research suggests a combined approach, urging institutions to invest in organizational strategies that integrate both automated and manual proctoring methods. Moreover, the study highlights the impracticality of video conferencing for large groups, underscoring the necessity for further research aimed at refining and enhancing automated OLP systems to address this limitation effectively.

In [4], the authors present an innovative monitoring system that utilizes a webcam to track candidate actions, including face movements, system usage, and audio data. These multi- modal features are integrated into a rule-based inference system, enabling the detection of potential malicious activities. This system plays a pivotal role in making decisions regarding user actions. Notably, the paper implements features like tab switching and USB detection. To identify suspicious actions, a threshold of 15 degrees is applied to the yaw angle. It's important to mention that audio and video processing does not occur on the client side, and the paper does not delve into extension detection. Furthermore, the study conducts a comparative analysis with existing systems, shedding light on the uniqueness and effectiveness of their proposed approach within the domain of user activity monitoring and security.

[5] is primarily focused on addressing the challenge of one-shot image recognition, a task that necessitates determining whether two images belong to the same class or not, even when the available data for each class is severely limited. To tackle this problem, the paper adopts a Siamese network architecture, characterized by the presence of two identical subnetworks with shared weights. These subnetworks receive input images and transform them into feature vectors. The paper employs a Contrastive Loss Function, described by the formula L(p, y), where y can take values of 1 or 0, signifying same or different class pairs. This loss function relies on the Euclidean distance (D) between feature vectors and a margin hyperparameter, which plays a crucial role in determining how far apart feature vectors of dissimilar pairs should be in the embedding space. In their experiments, the paper uses datasets like MNIST, consisting of black and white digits from 0 to 9, and OMNIGLOT, encompassing various symbols and signs[5]. The FaceNet paper's primary objective is to develop an efficient face recognition system that generates compact and highly discriminative face representations. This unified model is designed to excel in both face verification, determining if two faces belong to the same person, and face clustering, grouping faces of the same individual. Similar to the Siamese network, it employs two networks with shared weights. The Triplet Loss Formula, denoted as L(A, P, N), serves as a pivotal component, calculated as the maximum of the squared Euclidean distances between the anchor, positive, and negative face embeddings, with an added margin term. This margin is a crucial hyperparameter controlling the separation between positive and negative pairs. The authors tested their approach on the LFW dataset, achieving an impressive accuracy of 99.63%, and also evaluated its performance on the YouTube Faces dataset. The selection of appropriate triplets for training was a significant challenge addressed by the authors[6].

In [7] conducts a comprehensive re-evaluation of face verification and recognition by introducing a model with over 120 million parameters. Their groundbreaking approach is trained on a massive dataset of 4 million facial images from Facebook, offering impressive accuracy of 97.35% on the LFW dataset. What sets this study apart is its reliance on raw RGB pixel values for feature extraction, bypassing the complexity of hand-crafted features. The strategic use of fiducial points and alignment techniques enhances performance. Even without frontalization, their 2D alignment achieves 94.3% accuracy, emphasizing the model's robustness. This research showcases the power of data-driven approaches and finely tuned alignment in the realm of facial recognition.

In [8], the authors explore post-video processing in the context of Zoom platform videos, employing a client-side approach to determine facial presence. Notably, this study does not delve into object detection or extension detection, offering a distinctive perspective on video analysis. This research highlights the potential for privacy-focused video processing and underscores the importance of client-side execution in simplifying the analysis of facial presence within Zoom videos.

Traditional object detection methodologies have paved the way for the evolution of computer vision techniques. One seminal contribution came in 2001 when Paul Viola and Michael Jones introduced a groundbreaking Facial Detection Algorithm, which played a foundational role in early face detection systems. In 2005, Navneet Dalal and Bill Triggs further advanced the field by publishing a paper on a feature descriptor known as Histograms Of Oriented Gradient (HOG). HOG provided an effective means of representing object features, particularly useful for pedestrian detection. Another significant milestone was reached in 2009 when Felzenszwalbet and his colleagues developed the Deformable Part Model (DPM), another pivotal feature-based model. DPM, with its capacity to handle complex object structures and deformations, emerged as a highly successful approach in the realm of object detection applications, significantly shaping the landscape of computer vision research and practical implementations[9].

In [10], the authors concentrate on computer vision applications for monitoring and proctoring while also focusing on object and activity detection. They use deep learning models like YOLO, Faster RCNN, and SSD to improve the effectiveness and precision of recognition. They create their models using real-time computer vision frameworks like OpenCV. Notably, the paper provides useful advice, advocating YOLO in situations where real-time performance is crucial and Faster RCNN and SSD in situations where accuracy is more important. This study offers a useful overview of the approaches and factors to be taken into account when solving the dynamic problems of computer vision in practical settings.

In [11], the authors extensively explored the critical advancements in object detection techniques. They begin by discussing the revolutionary "Integral Image" representation, which has significantly improved feature computation in computer vision. This representation, introduced in the early 2000s, laid the foundation for subsequent developments in the field. The paper then delves into the implementation of AdaBoost, an efficient learning algorithm that plays a pivotal role in selecting critical visual features. AdaBoost's accuracy and versatility have made it a central component in numerous object detection systems. To further enhance computational efficiency, the re- view also highlights the development of a cascade method that rapidly filters out background regions, focusing computational resources exclusively on potential object-like regions. These pioneering contributions collectively underscore the significance of the integrated approach of Integral Image, AdaBoost, and cascade methods in the evolution of object detection systems.

In [12] they discussed about increasing accuracy using three types of scaling depth, width and resolution. Depth here is the number of layers involved in neural network Width refers to addition of feature maps too many feature map may capture redundant information and too few feature maps may fail to capture essential information, Therefore a proper balance must be maintained. Resolution refers to number of dots per inches increasing image quality indeed helps to increase the accuracy. For finding the proper balance between all three factors they formulated a equation : $\alpha^{\phi} \cdot \beta^{\phi} \cdot \gamma^{\phi}$

Which represents depth, width and resolution respectively. By carrying out experiments and using GridSearchCv they have proposed values for these variables.

In [13] they addresses the real-time tracking and positioning of 3D objects, offering improved object detection and tracking accuracy using neural networks like the Single-Shot MultiBox Detector (SSD). The study also presents benchmarks for Object Tracking with Motion Parameters (OTMP), which can be beneficial for tracking and monitoring users during online exams. The proposed Fast Depth-Assisted Single-Shot MultiBox Detector (FDA-SSD) model, which integrates spatial depth information, can be valuable for tracking users' movements and ensuring exam integrity in virtual proctoring systems. This research combines 3D tracking and object classification, making it potentially useful for detecting unauthorized actions during online examinations.

In [14] they emphasizes the growing importance of model efficiency in computer vision, particularly object detection. The authors explore various neural network design choices and propose key optimizations to enhance efficiency. Their innovations include the introduction of a weighted bi-directional feature pyramid network (BiFPN) for rapid multiscale feature fusion and a compound scaling method that uniformly scales resolution, depth, and width across networks. These optimizations result in the creation of "EfficientDet," a family of object detectors. Notably, EfficientDet-D7 achieves a remarkable 55.1 AP on COCO test-dev with only 77M parameters and 410B FLOPs, making it significantly smaller and less computation- ally demanding than prior detectors. This research underscores the increasing significance of model efficiency, particularly in resource-constrained real-world applications like robotics and self-driving vehicles.

In [15] they proposed the use of Haar-like features, which are simple rectangular patterns, for capturing object characteristics efficiently. These features could be adapted to represent a wide range of object patterns. The integral image idea was introduced to accelerate the computation of Haar-like features. Integral images allowed for the rapid calculation of sums within rectangular regions, which was crucial for the frame- work's efficiency. The cascade of classifiers was proposed to reject non-object regions early in the detection process. Weak classifiers were trained in stages, and objects that did not pass a stage were quickly rejected. This concept significantly improved the algorithm's speed, Highlighted the use of the AdaBoost algorithm for feature selection and classifier combination. AdaBoost focused on the most discriminative features, reducing the number of features required for detection.

In [16] they presents a pioneering approach in the field of computer vision and facial analysis. This work introduces the Multi-task Cascaded Convolutional Networks (MTCNN) framework, which addresses the dual challenges of face detection and facial landmark alignment within a unified, cascaded architecture. The core objectives are to achieve accurate face detection and precise localization of facial landmarks. MTCNN achieves these goals by utilizing a cascaded structure of three interconnected neural networks—Proposal Network (P-Net), Refinement Network (R-Net), and Output Network (O-Net). These networks employ multi-task learning to simultaneously predict bounding boxes for face detection, facial landmarks, and regression adjustments. The authors demon- strate the effectiveness of MTCNN through comprehensive experiments on diverse datasets, such as FDDB and WIDER Face, establishing MTCNN as a state-of-the-art solution for face detection and alignment tasks. This work has had a substantial impact in the field of computer vision and remains a cornerstone for researchers and practitioners aiming to address facial analysis challenges.

[19] introduces MobileNet, a family of efficient convolutional neural networks tailored for mobile and resource- constrained devices. MobileNet utilizes depthwise separable convolutions to reduce model size and computational cost while maintaining competitive accuracy. It has had a significant impact on real-world mobile and embedded vision applications by providing a balance between efficiency and performance.

In [20] the research is focused on using YOLOv3 for object detection under image degradation scenarios. It employed popular datasets like ImageNet, COCO, and VOC. Key degradation processes included blurring, rotation, noise, and cropping. Results showed that training with degraded images improved feature learning, leading to better performance in complex scenes. The model exhibited enhanced average precision and demonstrated robustness through training on degraded data, highlighting the practical applicability of YOLO-based models in challenging real-world conditions.

III. NEED AND MOTIVATION

As a student have seen a lot of malpractices carried during an online assessment. Be it for placement activity carried by colleges/universities or some online certification courses everyone is shifting towards online exam. Now it is indeed beneficial for students who live in remote areas and still wants to upskill themselves, but a fair means to achieve these certifications will increase it's integrity. The rise of online assessment has brought undeniable benefits, particularly for students in remote areas seeking to enhance their skills and education. However, it's crucial to address the issue of malpractices that have become all too common in this environment. Whether it's during university placement activities or online certification courses, maintaining the integrity of these exams is paramount. While online assessments offer convenience and accessibility, ensuring a level playing field for all participants is equally essential. Efforts to tackle malpractices must go hand in hand with the expansion of online education, thus safeguarding the credibility and fairness of the certification process for all students, regardless of their geographical location or circumstances. There are often multiple or even several hundred students taking the same online exam for a course. Therefore, an effective approach to facilitate the proctoring of online exams should help teachers or other proctors easily and quickly identify the students who have possibly cheated in the online exam

IV. METHODOLOGY

By conducting a comprehensive literature survey to under- stand the prevailing methods in online proctoring systems, The findings revealed that most systems follow a client- server architecture, where the model processing predominantly occurs on the server-side.

A. Existing Approach

In existing systems, various models have been employed for face verification. Notable models include FaceNet, which employs a loss function designed to maximize the distance between an anchor-positive image pair and an anchor-negative image pair. Siamese Models, trained for one-shot image classification, have proven effective, particularly in scenarios with limited face datasets. Facebook's DeepFace introduced techniques for face alignment, significantly enhancing accu- racy on datasets like LFW (Labeled Faces in the Wild).

ISSN No:-2456-2165

EfficientNet, on the other hand, proposed a strategy for balancing network depth, image resolution, and feature map size, ultimately improving model performance. The Single-Shot MultiBox Detector (SSD) stands out in real-time object detection, especially for face detection in proctoring systems. Its low processing requirements make it perfect for client-side processing in online exam proctoring. YOLO, or You Only Look Once, is a pioneering real-time object detection algorithm that has revolutionized computer vision. Its core strengths lie in its ability to detect multiple objects in a single pass through a neural network, ensuring exceptional speed and efficiency. YOLO employs a grid system, dividing the input image into cells, with each cell responsible for predicting object presence, bounding boxes (including their coordinates and dimensions), and class probabilities. The algorithm also utilizes Non-Maximum Suppression to eliminate redundant bounding boxes, retaining the most confident predictions.

B. Proposed Approach

- *Client-Side Processing:* In this research, the aim is to develop an online proctoring system with a unique approach. Unlike traditional systems, model will execute on the client- side. This approach offers multiple advantages, such as reduced bandwidth usage and minimized mobile data consumption during online exams. model will be stored in cloud object storage and loaded into a React app, ensuring efficient and responsive execution.
- *Face Verification:* To verify the user's identity, the Siamese technique is used to create embeddings for each user. These embeddings will be securely stored in a structured database. During the exam, the real-time identity verification is performed by comparing the user's face with the stored embeddings.
- *Object Detection:* For object detection, particularly identifying items like mobile phones, fine-tuning is done by the MobileNet model. MobileNet, a family of efficient convolutional neural network architectures, has made significant contributions to object detection in the field of computer vision. Recognized for its efficiency and speed, MobileNet models are particularly well-suited for real-time object detection tasks on resource-constrained devices, such as smart- phones and IoT devices. This efficiency is achieved through the use of depthwise separable convolutions, which reduce the computational load and model size. MobileNet comes in various versions, including MobileNetV1, MobileNetV2, and MobileNetV3, each offering improvements in terms of accuracy, speed, and compactness. These models find applications in real-time pedestrian detection, face detection, and general object detection. While they may not achieve the same level of accuracy as larger, more complex models, their balance between speed and accuracy makes them an excellent choice for scenarios where real-time processing is critical. MobileNet can be integrated with popular object detection frameworks, making it a versatile and efficient solution for a wide range of object detection applications.
- *Extension Detection:* The proposed system will also tackle issues related to browser extensions that may interfere with the exam. The plan to develop a custom browser ex- tension that will detect and disable flagged extensions during the exam. Additionally, the exam is held in full-screen mode, limiting the user's ability to interact with other browser tabs, along with it disabling JavaScript code inspection to prevent users from altering client-side behavior.
- *Support for Multiple Question Types:* Online proctoring system will provide flexibility to test setters. They can choose from a variety of question types, including multiple-choice questions (MCQs), coding questions, and subjective questions. For coding questions, a system is implemented using Docker containers to execute and evaluate user's code against predefined test cases. This containerized approach enables scalability as it allows multiple instances to run in parallel, accommodating varying workloads.
- By incorporating these elements into proposed methodology, the aim to create a robust and efficient online proctoring system that enhances the integrity of online exams while offering a user-friendly experience.

V. SYSTEM ARCHITECTURE AND WORKFLOW

A. Existing System Architectural Components

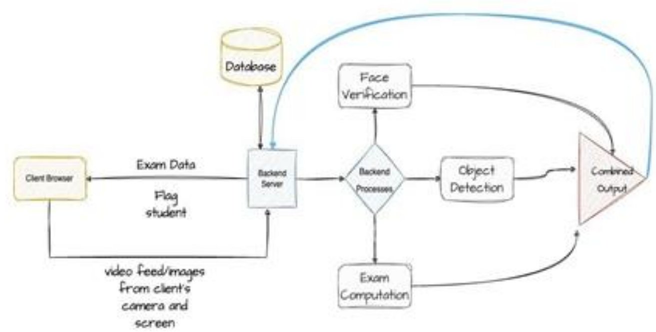

Fig. 1. Architecture of Exisiting Systems

In all the existing system they had the basis architecture in common. They used backend server for face verification, object detection, score calculations. This results were stored in database which adds up cost of storage. Talking about the flow all the required exam data was sent to client by backend server and the required video/images are taken from client to server for further processing. This systems were not filtering out the required images were suspicious activity is detected, they used to store entire feed or images at a interval. Some of the system also used audio for proctoring user but it didn't proved much effective. It took high processing power and didn't produced significant findings.

VI. MARKET SURVEY AND STATISTICS

To analyze the participation of countries in online exams research [2], the selected studies were thoroughly investigated to identify the contributing institutes and corresponding countries. They identified 25 countries that contributed to online exams research. It was analyzed that china and India are the leading contributors in online exams research as 8 studies belong to each of them. Furthermore, Indonesia is the third leading contributor on the list with 7 studies where Teknologi Bandung, Indonesia is a leading institute with 3 studies.

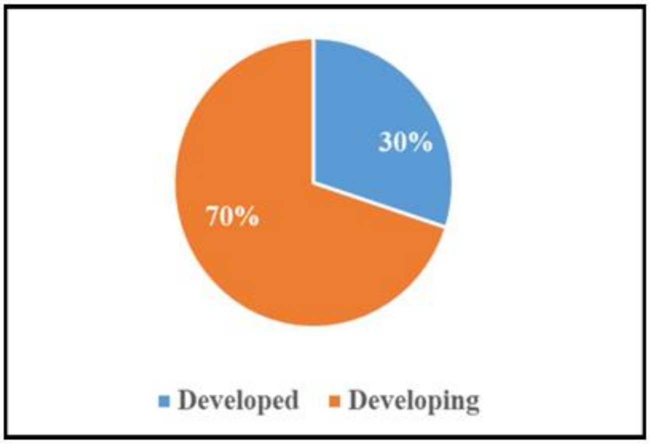

Fig. 2. Countries Participation in past five years. [2].

They further found the geographical distribution of online exams contributions, it can be analyzed that Asia is the leading contributor with 62% (33 studies) of overall conducted research. The closest competitor is Europe that stands with 17% (9 studies) on the list. Similarly, North America and Africa stand fourth and fifth on the list with 13% (7 studies) and 6% (3 studies), respectively. Oceania (New Zealand) is last standing on the list with merely 2% of the selected publications. In short, most of the online exams research has been conducted in Asian countries during the past five years. Diverse online proctoring platforms have employed distinct methods for administering exams. The table that contrasts the proposed solution with the current solutions is shown below. The proposed strategy does all of the monitoring of student activities online, reducing the need for human intervention, in contrast to the majority of the other options, which employ physical proctor methods. In addition, the proposed system makes use of additional features for voice recognition and camera surveillance. Most crucially, the suggested solution keeps track of the active window to prevent users from switching tabs in their browsers.

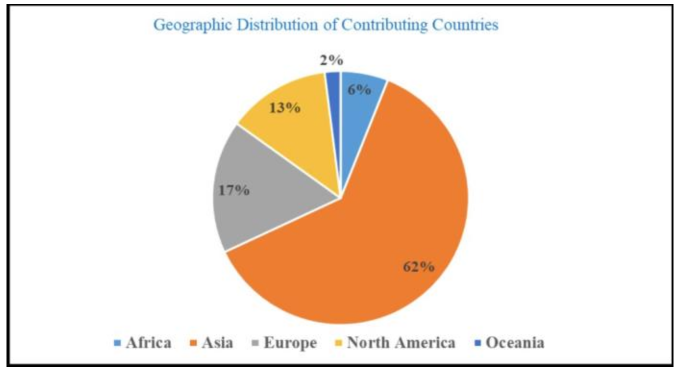

Fig. 3. Summary of Online Exam Participation based on five countries. [2].

ONLINE PROCTORING TOOLS FEATURE COMPARISON				
Features	Kryterion	Software Secure	ProctorU	Proposed System
Physical Proctor for monitoring	YES	YES	YES	NO
Voice recognition	NO	NO	NO	YES
Usage of webcam	YES	YES	YES	YES
Continuous Internet	YES	NO	YES	YES
Automatic active window capture	NO	NO	NO	YES

Fig. 4. Comparison of various Online Proctoring System Features. [4].

VII. CONCLUSION

By researching on existing OPS, face verification, and object detection. Existing methods that perform very well on datasets like LFW and YT are discussed, also proposed a new approach for the OPS platform which has a lot of benefits discussed above. Overall the proposed system overcomes the drawbacks of many existing systems and it will be the one-stop solution for any online exam be it MCQ, Coding Challenge, or subjective questions.

REFERENCES

[1]. Anjali Rajendra Yelkar,Reena Vishwas Pawar,Diksha Anil Khopkar,Pallavi Pravin Ratwadkar, Raghvendra Omprakash Singh, ”ONLINE EXAM PROCTORING SYSTEM”, International Journal of Creative Research Thoughts (IJCRT), ISSN:2320-2882, Volume.10, Issue 1, pp.e954-e959, January 2022, Available at:http://www.ijcrt.org/papers/IJCRT2201560.pdf

[2]. A.W. Muzaffar, M. Tahir, M.W. Anwar, Q. Chaudry, S.R. Mir, Y. Rasheed, ”A Systematic Review of Online Exams Solutions in E- Learning: Techniques, Tools, and Global Adoption,” IEEE Access, vol. 9, pp. 32709-32723, 2021. DOI: 10.1109/ACCESS.2021.3060192.

[3]. Arno`, S., Galassi, A., Tommasi, M., Saggino, A., & Vittorini, P. (2021). State-of-the-Art of Commercial Proctoring Systems and Their Use in Academic Online Exams. International Jour- nal of Distance Education Technologies (IJDET), 19(2), 55-76. http://doi.org/10.4018/IJDET.20210401.oa3

[4]. S. Prathish, A. N. S. and K. Bijlani, ”An intelligent system for on- line exam monitoring,” 2016 International Conference on Information Science (ICIS), Kochi, India, 2016, pp. 138-143, doi: 10.1109/IN-FOSCI.2016.7845315.

[5]. G. R. Koch, ”Siamese Neural Networks for One-Shot Image Recogni- tion,” in Proceedings of [Conference Name], 2015

[6]. F. Schroff, D. Kalenichenko and J. Philbin, ”FaceNet: A unified em- bedding for face recognition and clustering,” 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 815-823, doi: 10.1109/CVPR.2015.7298682.

[7]. Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, ”DeepFace: Closing the Gap to Human-Level Performance in Face Verification,” in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2014, doi: 10.1109/CVPR.2014.220.

[8]. Yang, X., Wu, D., Yi, X., Lee, J. H. M., and Lee, T., “iExam: A Novel Online Exam Monitoring and Analysis System Based on Face Detection and Recognition”, ¡i¿arXiv e-prints¡/i¿, 2022. doi:10.48550/arXiv.2206.13356.

[9]. A. B. Amjoud and M. Amrouch, ”Object Detection Using Deep Learn- ing, CNNs and Vision Transformers: A Review,” in IEEE Access, vol. 11, pp. 35479-35516, 2023, doi: 10.1109/ACCESS.2023.3266093.

[10]. M. S. S. Reddy, P. R. Khatravath, N. K. Surineni and K. R. Mulinti, ”Ob- ject Detection and Action Recognition using Computer Vision,” 2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS), Coimbatore, India, 2023, pp. 874-879, doi: 10.1109/IC- SCSS57650.2023.10169620.

[11]. Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition 1:I-511- I-518 vol.1 DOI:10.1109/CVPR.2001.990517

[12]. M. Tan and Q. V. Le, ”EfficientNet: Rethinking Model Scaling for Con- volutional Neural Networks,” in Proceedings of the 36th International Conference on Machine Learning, 2019, pp. 6105-6114.

[13]. Wang, Zihao & Yang, Sen & Shi, Mengji & Qin, Kaiyu. (2022). FDA-SSD: Fast Depth-Assisted Single-Shot MultiBox Detector for 3D Tracking Based on Monocular Vision. Applied Sciences. 12. 1164. 10.3390/app12031164.

[14]. M. Tan, R. Pang and Q. V. Le, ”EfficientDet: Scalable and Efficient Object Detection,” 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 10778- 10787, doi: 10.1109/CVPR42600.2020.01079.

[15]. P. Viola and M. Jones, ”Rapid Object Detection using a Boosted Cascade of Simple Features,” in Proceedings of the 2001 IEEE Com- puter Society Conference on Computer Vision and Pattern Recog- nition, CVPR 2001, Kauai, HI, USA, 2001, pp. I-511-I-518, doi: 10.1109/CVPR.2001.990517.

[16]. K. Zhang, Z. Zhang, Z. Li and Y. Qiao, ”Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks,” in IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499-1503, Oct. 2016, doi: 10.1109/LSP.2016.2603342.

[17]. Abdul Razak, Siti Fatimah Soon, Liew Ibrahim, Siti Zainab Lee, Chin-Poo Lim, Kian. (2017). Facilitating examination process via exam monitoring system. 187-192. 10.1109/SCORED.2017.8305396.

[18]. Razan Bawarith, Dr. Abdullah Basuhail, Dr. Anas Fattouh and Prof. Dr. Shehab Gamalel-Din, “E-exam Cheating Detection System” International Journal of Advanced Computer Science and Applications(IJACSA), 8(4), 2017. http://dx.doi.org/10.14569/IJACSA.2017.080425

[19]. Howard, Andrew G., Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. ”MobileNets: Efficient Convolutional Neural Networks for Mobile Vi- sion Applications.” arXiv preprint arXiv:1704.04861 (2017).

[20]. C. Liu, Y. Tao, J. Liang, K. Li and Y. Chen, ”Object Detection Based on YOLO Network,” 2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China, 2018, pp. 799-803, doi: 10.1109/ITOEC.2018.8740604.