Journal of Industrial Information Technology and Application *ISSN(Online): 2586-0852*
Vol.7 No. 4

Design of Face Analysis System for Online Interactive Remote Class Management Service

Myoung-Kyu Sohn[1,*], Sang-Heon Lee[2], and Hyunduk Kim[3]
[1,2,3] Division of Automotive Technology, DGIST, Daegu, Republic of Korea

Abstract. Recently, various remote conferencing systems are coming out in the market due to situations such as Covid-19. As this situation continues for a long time, the need for online learning is emerging in the educational environment as well. The online class management system is similar in overall concept to the online conference system in terms of video conversation but requires several different functions. Functions such as face authentication to check attendance and anti-spoofing to distinguish fake faces through photo or video playback are essential. In addition, it can provide services such as feedback through analysis of student concentration during the class. In this paper, we define the major functions of the online class management system and design the entire system. In particular, we propose a design and implementation method for a face analysis system.

Keywords; face recognition, anti-spoofing, online attendance system

1. Introduction

As the class pattern rapidly changes to remote non-face-to-face classes, the efficiency of each school's existing electronic attendance system is significantly reduced, students' concentration is reduced, and the learning effect is reduced as they search the internet or social network during the class. To address these issues in non-face-to-face classes, a system that can enhance the quality of the class and students' concentration is required.

In this paper, we design a non-face-to-face interactive remote class management system based on artificial intelligence and computer vision. Students can receive face

* Corresponding author: smk@dgist.ac.kr
Received: Aug 31, 2023; Accepted: Oct 12, 2023; Published: Dec 31, 2023

recognition-based electronic attendance services through cameras in desktop, tablet, or mobile environments. Through this, the management of attendance can be more convenient. In addition, interactive remote class management services can provide concentration analysis and face anti-spoofing from face analysis. Attendance has the characteristic of having to be executed at a specific time simultaneously. This paper focuses specifically on handling this traffic for deep learning inference.

2. System Design

Figure 1. shows the overall architecture of the online class management system. The system is divided into two parts: a server and a client, and the user's terminal becomes a client. The main functions of the user's terminal include a user interface for video conferences, a module for video transmission, and an encryption module for security. You can also get feedback on your degree of concentration during the class. The main functions of the server can be divided into a class management system and a face analysis system. The class management system has an online video conference system and also includes attendance management and statistics required for class management. The last is the face analysis system. It includes face detection[1], anti-spoofing[2,3,4], face recognition[5,6], and concentration analysis[7,8] modules.

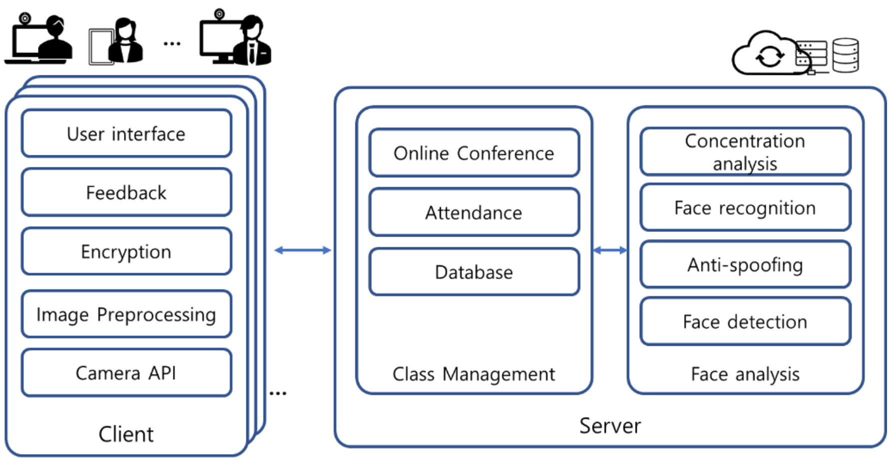

Figure 1. Overall architecture of online class management system

3. Implementation and Experiment

Large traffic rush on the server at a specific time since the user's attendance is generally made as soon as class starts. Since the server has to process the received user information simultaneously, it may take a lot of loads. Most of the load on the server is on deep learning inferences such as face detection, face recognition, and face anti-spoofing of the users. GPUs are used for inference in this system and the system can take advantage of the GPU characteristics to increase inference performance in terms of speed. The process of this method is shown in Figure 2.

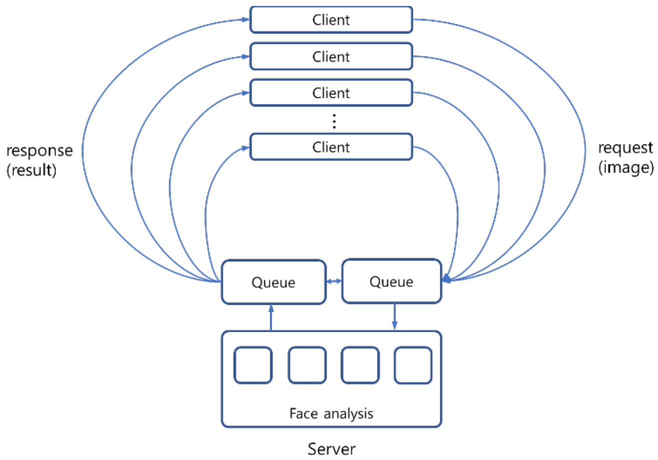

Figure 2. Process method for deep learning inference

Users transmit their face images or encrypted face images to the server for class attendance. The transmitted image is sent to the queuing module instead of directly to the face analysis module. In this queuing module, images are transformed into tensors to be input to the deep learning module, and tensors are also collected and made into batches. By making batches of the image tensors in this way, the bottleneck of input and output to the GPU memory can be reduced and we can take advantage of the multi-core characteristic of the GPU. [9,10]. An experiment was conducted to verify the inference performance for such batch processing.

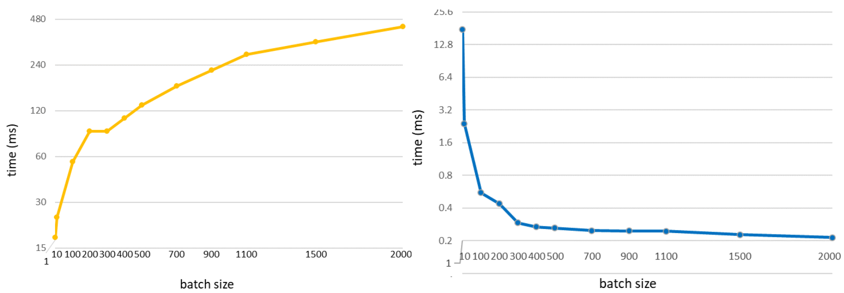

Figure 3. Inference performance according to the batch size. Inference time of a batch(upper) and inference time of an image(lower)

To verify the effect of the tensor batch, experiments were conducted in the anti-spoofing module. Inference performance was measured according to the batch size. Here, the batch size is defined as the number of image tensors in a batch. Figure 3 shows the inference performance for different batch sizes and the performance for one image on a logarithmic scale. The average processing time for one batch increases as the batch size increases as expected. When calculating the processing time based on an image, it can be seen that the processing time is significantly improved as the batch size increases. However, the image processing time does not decrease after the batch size increases to some extent. In addition, there is a limit to increasing the batch size due to the GPU memory capacity in terms of hardware. In real-world scenarios, a large batch size may result in a large delay in response to certain users, as the response may be given to the user after one batch has been completely processed. Therefore, the overall system must be designed in consideration of the trade-off between batch size and response time for overall users.

4. Conclusion

We have discussed issues that may arise when deep learning systems are deployed in real service. Experiments were conducted to solve these problems in designing the online attendance system. To respond to the heavy traffic of user requests, tensor batch, which can maximize the multi-core performance of the GPU through the queuing system, was applied and the experiment result was shown.

References

[1] Deng, Jiankang, et al. "Retinaface: Single-shot multi-level face localisation in the wild." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.

[2] Yu, Zitong, et al. "Deep learning for face anti-spoofing: A survey." arXiv preprint arXiv:2106.14948 (2021).

[3] Liu, Yaojie, Amin Jourabloo, and Xiaoming Liu. "Learning deep models for face anti-spoofing: Binary or auxiliary supervision." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[4] Li, Xiaobai, et al. "Generalized face anti-spoofing by detecting pulse from face videos." 2016 23rd International Conference on Pattern Recognition (ICPR). IEEE, 2016.

[5] Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

[6] Fuad, Md Tahmid Hasan, et al. "Recent advances in deep learning techniques for face recognition." IEEE Access 9 (2021): 99112-99142.Cha, Seunghui, and Wookhyun Kim. "Concentration analysis by detecting face features of learners." 2015 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM). IEEE, 2015.

[7] Li, Penghua, et al. "Driver distraction detection using octave-like convolutional neural network." IEEE Transactions on Intelligent Transportation Systems (2021).

[8] Oh, Seungtaek, Jaewon Moon, and Seungwoo Kum. "Application of Deep Learning Model Inference with Batch Size Adjustment." 2022 13th International Conference on Information and Communication Technology Convergence (ICTC). IEEE, 2022.

[9] Choudhury, Anamitra Roy, et al. "Variable batch size across layers for efficient prediction on CNNs." 2020 IEEE 13th International Conference on Cloud Computing (CLOUD). IEEE, 2020