Journal of information and
communication convergence engineering
http://jicce.org

J. Inf. Commun. Converg. Eng. 21(4): 287-293, Dec. 2023 Regular paper

Automatic Poster Generation System Using Protagonist Face Analysis

Yeonhwi You, Sungjung Yong, Hyogyeong Park, Seoyoung Lee, and Il-Young Moon*, *Member, KIICE*

Department of Computer Science and Engineering, Korea University of Technology and Education, Cheonan 31253, Republic of Korea

Abstract

With the rapid development of domestic and international over-the-top markets, a large amount of video content is being created. As the volume of video content increases, consumers tend to increasingly check data concerning the videos before watching them. To address this demand, video summaries in the form of plot descriptions, thumbnails, posters, and other formats are provided to consumers. This study proposes an approach that automatically generates posters to effectively convey video content while reducing the cost of video summarization. In the automatic generation of posters, face recognition and clustering are used to gather and classify character data, and keyframes from the video are extracted to learn the overall atmosphere of the video. This study used the facial data of the characters and keyframes as training data and employed technologies such as DreamBooth, a text-to-image generation model, to automatically generate video posters. This process significantly reduces the time and cost of video-poster production.

Index Terms: AI, DreamBooth, Face Recognition, Keyframe, Poster

I. INTRODUCTION

Recent developments in the video industry have led to a surge in video content. YouTube, Netflix, Disney+, and Apple TV are major platforms that host and stream video content, and the number of platforms that provide videos is increasing.

This increase in video content requires techniques in organizing and managing large amounts of video data as well as methods for providing appropriate video summaries. These summaries are used to concisely convey the video content. Movie summaries are available to consumers in various forms, including plot synopses, character bios, thumbnails, and posters. These forms aim to provide consumers with a sense of what a movie concerns without watching the movie. Video summaries can be an important selection criterion for users consuming video content and promotional tool for video creators and distributors [1]. Modern film theory states that “all films are about nothing-nothing but character” [2]. This means that characters are the most important element in a movie. From the audience’s perspective, what makes a movie compelling and engaging is the audience wanting to understand the story of the movie’s characters, and interactions between the characters make the movie structure and content meaningful.

Because everything in movies and video revolves around a character, consumers are drawn to the characters, and we assume that the main character influences the video selection of consumers. Hence, the main character in both movies and videos is considered a selection criteria and an appropriate factor for promotion. Moreover, promotional services use recommendation systems to provide convenient choices to

Received 29 June 2023, **Revised** 10 October 2023, **Accepted** 23 October 2023

***Corresponding Author** Il-Young Moon (E-mail: iymoon@koreatech.ac.kr)
Department of Computer Science and Engineering, Korea University of Technology and Education, Cheonan 31253, Republic of Korea

Open Access https://doi.org/10.56977/jicce.2023.21.4.287 print ISSN: 2234-8255 online ISSN: 2234-8883

consumers. Recommendation systems have become an important research area in computer science, and numerous studies have proposed various approaches.

This study proposes a method that determines main movie characters based on trailers and then automatically generates posters based on the main characters.

The remainder of this paper is organized as follows. Section 2 describes related research on extracting people from videos and generating artificial images. In Section 3, we present a method for determining protagonists in movie trailers and an automatic poster-generation method based on the protagonists. Section 4 presents the experiments we conducted with both methods and their results. Finally, Section 5 concludes the paper and provides directions for future research.

II. RELATED RESEARCH

The development of artificial intelligence (AI) is characterized by improvements in facial recognition and automatic image generation technologies. Facial recognition is used in various fields and is the subject of ongoing research.

Since the introduction of feature-learning approaches, the evolution of facial recognition technology progressed from a holistic approach in the 1990s to a deep learning-based approach by the late 2000s [3].

A. Face Recognition

In the early 1990s, facial-recognition research became popular with the introduction of the histological eigenface approach [4]. Over time, low-dimensional expressions, such as linear subspaces, manifolds, and sparse presentations, were studied, and local feature-based face recognition methods gradually developed. Eventually, a deep learning-based face recognition method was studied [5].

Chaochao and Tang [6] proposed a multi-task learning method called Gaussian Face and solved the face authentication problem by increasing the efficiency of calculations and improving them based on a discrete Gaussian process latent variable model. We proposed a method for accelerating the inference and prediction of the model using Gaussian process approximation and anchor graphs, and we leveraged additional data to improve the generalization performance to automatically adapt to complex face changes and achieve human-level facial accuracy on the LFW dataset.

Sun et al. [7] presented a method that uses deep learning technology to develop effective feature representations for face recognition and designed a deep convolution network that uses both signals for face recognition and verification. Their method achieved a higher accuracy than that of previous studies and a higher accuracy on the LFW dataset by applying features to traditional face verification pipelines.

Zhang et al. [5] proposed the multi-task cascaded convolutional networks (MTCNN) method. This method comprises three neural networks based on a convolutional neural network (CNN): P-Net, R-Net, and O-Net. This approach employs joint learning, in which face classification, bounding box regression, and face landmark localization processes are conducted and learned simultaneously within each network. The results obtained from P-Net are used in R-Net, and those from both networks are used in O-Net. This approach enhanced the extraction speed and face detection accuracy by up to 95%. Subsequently, new methods emerged that divide facial images into regions using multiple CNNs for localized determinations and that then aggregate the results [8]. Acquiring substantial amounts of facial data is challenging. Therefore, researchers are exploring approaches that implement CNN models as Siamese networks, enabling high recognition rates with limited facial data [9].

Florian et al. [10] introduced FaceNet. FaceNet learns face images directly to efficiently recognize, validate, and cluster more efficiently. This study was the first to propose triplet loss, which is the most important concept in facial recognition models. FaceNet maps face images to Euclidean space such that the distance matches measurements of the face similarity, thereby facilitating the performance of face recognition, validation, and clustering tasks when using them as feature vectors when the mapping is generated. Online triple mining automatically generates learning data by automatically generating the triple learning data used in FaceNet and can be used efficiently.

In this paper, we propose learning using the triplets of roughly aligned matching/mismatching face patches using an online triple mining method in FaceNet, thereby improving the model performance while efficiently using the learning data. This method exhibited an excellent face recognition performance, and a large amount of processing was possible.

In this study, we utilized an algorithm developed by Florian et al. to recognize movie characters, classify characters through clustering, and determine the character with the most appearances as the main character.

This study contributes to the automatic generation of poster summaries by identifying main characters based on movie trailers.

B. Image Generation AI

Poster creation involves designing an overall layout and appropriately placing elements within it. In general, existing systems are limited to automatically generating layout options for poster production [11]. With the advancement of deep learning technology, both layout generation and image creation has become possible; even backgrounds can be created. Text-to-image generation technology can be used to generate images relevant to the textual information input.

Since DALL-E, most current image-generation technologies have been based on text-to-image services [12]. Ruiz et al. [13] used photo booths that utilize AI technology to generate various images of a given subject using only text prompts and some reference images. This technique enables the personalization of text-to-image diffusion models, thereby facilitating the synthesis of new realistic images contextualized in various scenes. In this study, we used text prompts and reference images as input, generated various images for a given topic, and used a text-image diffusion model to learn how to combine text and images for image generation.

DreamBooth has been validated through various experiments and applications, demonstrating that it can generate various images for a given subject and that personalizing text-to-image diffusion models is possible. In addition, DreamBooth can generate images while maintaining the core visual features of a given subject, thereby achieving realistic and unique features in the generated images [13]. Currently, to generate a desired character image, users must directly collect and provide the system with information about the character. This process is time-consuming. Furthermore, it requires considerable effort to select the design of an element and place it[14]. We aimed to minimize the human effort required for image generation by focusing on the process of creating training data for the image generation technology. In this study, we used the person recognition method of the Face Recognition module to obtain the face data of the main character in the video and used Google's Dreambooth, a text-to-image generation platform, to generate a video poster based on the main character.

III. SYSTEM MODEL AND METHODS

Although image generation technology has experienced significant advancements, its application in the creation of movie posters is rare. This study aimed to automatically generate protagonist-based movie posters utilizing face recognition and classification to identify main characters and generate poster images containing the protagonists. Fig. 1 overviews the system architecture. To create posters that capture the essence of a movie, keyframes from the film were extracted and used as training data for the AI poster-generation model. Additionally, facial recognition and clustering were performed on the entire video to calculate the frequency of character appearances throughout the video. The character with the highest frequency of appearance was identified as the protagonist. Subsequently, the protagonists' facial data were used for the training, and the AI model was used to generate movie posters that capture both the atmosphere of the film and main character. Only movie trailers were used as input data to reduce the amount of data. Table 1 lists the videos that this study used.

Table 1. Movie list

No	Title	Genre	Time
1	Coweb	Horror, Thriller	02:09
2	Gran Turismo	Drama, Sports	02:31
3	It Lives Inside	Drama, Horror, Thriller	02:23
4	Mad Heidi	Action, Comedy	01:47
5	Mending Line	Drama	02:32
6	Outpost	Horror	02:08
7	Strange Way of Life	Drama, Western	01:07
8	The Island	Action, Thriller	01:35
9	Two Tickets to Greece	Comedy	01:35
10	The Last Broadcast	Horror, Mystery	02:24

A. Keyframe Extraction

Keyframes were used to generate movie posters and understand the overall background and atmosphere of the films. We employed the YCbCr color model to extract keyframes from the video. YCbCr is a color-space representation method that captures a deep representation of the luminance component of an image while representing the chrominance components at a lower resolution, thereby reducing the amount of data required for the color representation. In this process, all video frames are converted into YCbCr histograms. When analyzing the variations in the histograms, the scene transitions in the video were determined

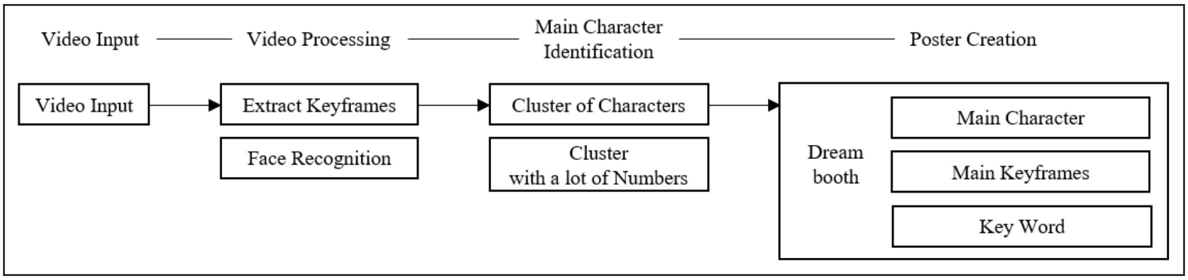


Fig. 1. System workflow.

to determine whether a keyframe should be extracted. If multiple similar frames are extracted from a specific frame range, that frame may be overfitted. To address this, sections in which scene transitions occurred rapidly were extracted and removed, thereby refining frames with similar backgrounds. This approach helps prevent overfitting and maintains a diverse set of keyframes.

B. Face Recognition and Clustering

We used the “face_recognition” Python package to perform facial recognition on the videos. After the video frames were extracted, we used the package to analyze the coordinates of 68 facial landmarks, including the eyes, nose, and mouth, to identify the faces. The analyzed coordinates were then converted into numerical representations (encodings) to enable facial recognition. When a new face was recognized, we used the package to calculate the Euclidean distance between the new face and all previously recognized faces. We used this measurement to determine the similarity between two faces. If the similarity with another face was below a predefined threshold, the faces were considered to be the same. If the similarity exceeded a threshold, a new face was added to a database of recognized faces. Through experimentation, a threshold value of approximately 0.5 was determined to be suitable for an effective classification.

C. Poster Generation

We employed Google’s DreamBooth algorithm to generate movie posters. DreamBooth is a pretrained text-to-image model that addresses the limitations of large-scale text-to-image models; these models often lack the ability to mimic target subjects accurately. DreamBooth finetunes a text-to-image model using a small set of images to generate images relevant to the user’s target topic.

This study aimed to embed the subject into the output domain of the model. Consequently, users can use a unique identifier associated with the subject to synthesize the images. Therefore, we employed DreamBooth to create movie posters that express the atmosphere of a given film. We used the facial images of the main characters and keyframes extracted from the video as training data.

IV. RESULTS

A. Data Extraction

To enable the automatic generation of poster images, we extracted keyframes from movie trailers. The keyframe extraction rate, which is the percentage of keyframes compared with the total number of frames, exhibited an average compression rate of 2.93% for the 10 movies. This demonstrates a significant efficiency in obtaining and refining the training data. Table II compares keyframe extraction rates based on the total number of frames.

Table 2. Overall frame and keyframe capacity comparison (average compression rate of 2.93%)

No.	Title	Frame	Keyframe	Capacity Reduce Rate
1	Coweb	3,108	81	2.61%
2	Gran Turismo	3,633	151	4.16%
3	It Lives Inside	3,439	71	2.06%
4	Mad Heidi	2,568	82	3.19%
5	Mending Line	3,665	122	3.33%
6	The Outpost	3,077	109	3.54%
7	Strange Way of Life	1,624	45	2.77%
8	The Island	2,292	77	3.36%
9	Two Tickets to Greece	2,290	65	2.84%
10	The Last Broadcast	4,324	62	1.43%

In this study, we obtained facial characters by conducting face recognition on movie trailers, and we applied clustering to groups of similar characters. Figs. 2 and 3 show the face recognition and clustering results, respectively.

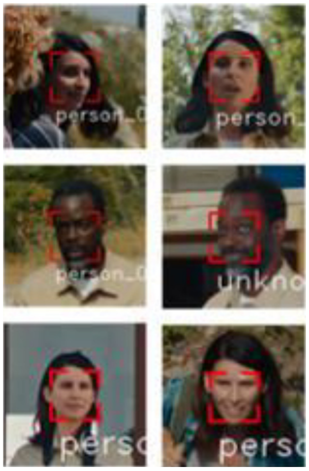


Fig. 2. Face recognition processing example.

We considered characters that appear in posters as representative of their respective movie. Therefore, we treated the character that appears most frequently in a movie trailer as the main character. We verified whether the 10 candidate characters from the 10 movie trailers we analyzed matched the actual main character. In particular, we calculated the importance of each character based on their appearance frequency and used this information as training data for the movie poster generation.

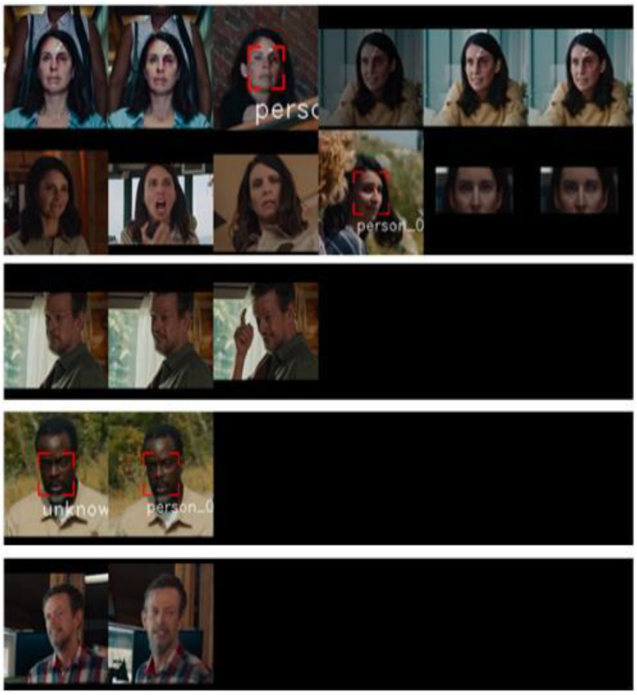


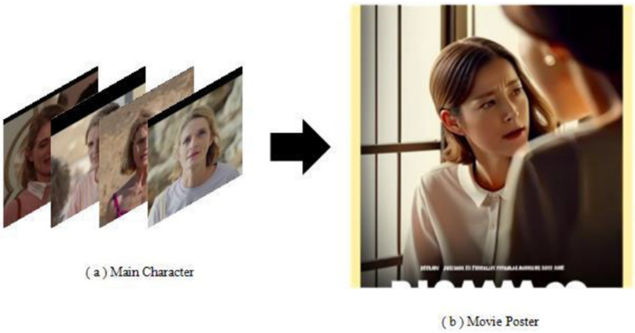
Fig. 3. Clustering processing example.

Table 3. Comparison of main character predictions

No.	Title	Main Character	Predicted Main Character	Matched or Not
1	Coweb	Lizzy Caplan, Antony Starr, Cleopatra Coleman	Lizzy Caplan	Matched
2	Gran Turismo	David Harbour, Orlando Bloom, Archie Madekwe	David Harbour	Matched
3	It Lives Inside	Megan Suri, Neeru Bajwa, Mohana Krishnan	Mohana Krishnan	Matched
4	Mad Heidi	Alice Lucy, Max Rudlinger, Casper Van Dien	Alice Lucy	Matched
5	Mending Line	Brian Cox, Perry Mattfeld, Sinqua Walls	Sinqua Walls	Matched
6	Outpost	Ato Essandoh, Dallas Roberts, Beth Dover	Beth Dover	Matched
7	Strange Way of Life	Pedro Pascal, Ethan Hawke, Jose Condessa	Ethan Hawke	Matched
8	The Island	Jackson Rathbone, Michael Jai White, Gillian White	Michael Jai White	Matched
9	Two Tickets to Greece	Laure Calamy, Olivia Cote, Kristin Scott Thomas	Olivia Cote	Matched
10	The Last Broadcast	David Beard, Lance Weiler, Stefan Avalos	Jim Seward	Not Matched

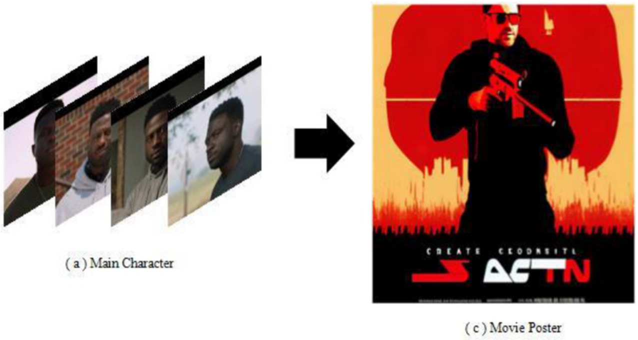
B. Automatic Poster Generation

We used the facial data of the main characters extracted via face recognition and clustering as training data for the AI model and then generated movie posters featuring the main characters. Note that the backgrounds in the posters sometimes included unrelated elements that were irrelevant to the movie.


(a) Main Character

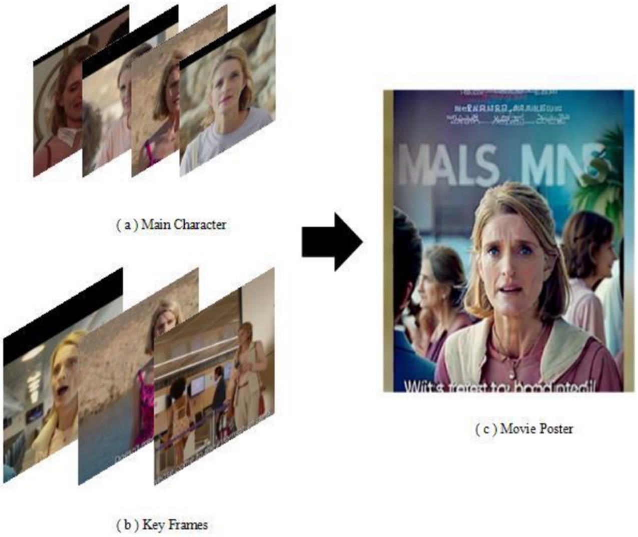
(b) Movie Poster

Fig. 4. Automatically generated movie poster using main characters.


(a) Main Character

(c) Movie Poster

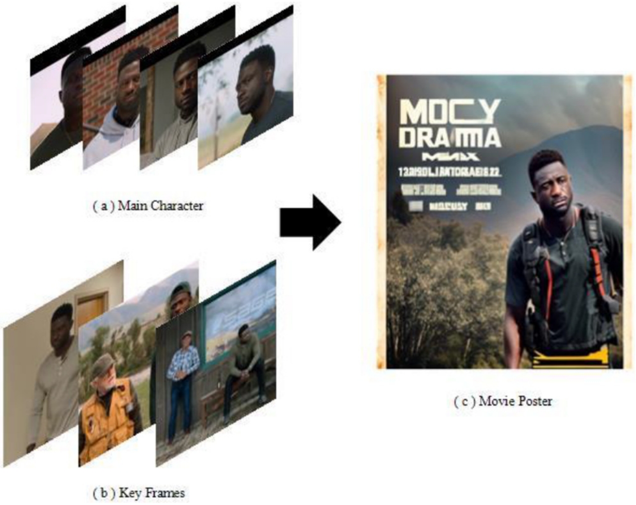
Fig. 5. Automatically generated movie poster using main characters.


(a) Main Character

(b) Key Frames

(c) Movie Poster

Fig. 6. Automatically generated movie poster using both main characters and keyframes.



(a) Main Character

(b) Key Frames

(c) Movie Poster

Fig. 7. Automatically generated movie poster using both main characters and keyframes.

Using the proposed method, we trained the model by incorporating keyframes in addition to the facial data of the main characters to generate movie posters. The posters depicted the overall atmosphere of the movie, and we observed the generation of posters featuring the main characters.

V. CONCLUSIONS AND FUTURE WORK

A. Conclusion

We proposed and implemented an approach that extracts the facial data of characters appearing in movies, clusters them, and uses keyframes to generate movie posters. We performed the keyframe extraction using YCbCr color values; we observed a keyframe extraction rate of less than 3% of the total number of frames in a video. This significantly reduced the training time compared with training on the entire video. When training the AI model only on the facial data of the main characters, we obtained posters that accurately included the main characters but featured backgrounds unrelated to the movie genre or atmosphere. With additional training on movie keyframes, we obtained movie posters that captured the overall atmosphere and character. Automatically generated movie posters that reflect the preferences of movie audiences help in movie selection and recommendation. Compared with the manual poster production method, the method proposed in this study demonstrates that automatically generating movie posters from movie trailers may significantly reduce the time and cost of production, thereby contributing to the development of the movie industry.

B. Future Research Works

In this study, we propose a method for automatically generating posters. We analyzed keyframes and protagonist faces in movie trailers. The results confirm the feasibility of automatically generating posters utilizing the keyframes of the main characters in the target videos. In the future, we plan to optimize the automatic generation of posters by considering individual preferences. In this study, the target data was limited to that obtained from movie trailers, because of copyright issues. The completeness of the posters was limited because of the lack of data required to learn the keyframes and the faces of the main characters. However, if the keyframes and faces of the main characters are extracted from the entire video content (the full movie), sufficient training data for automatic generation can be obtained. This is expected to improve the quality of the final poster. Furthermore, we plan to enhance the user experience and efficiently improve the movie poster generation process by providing ultra-personalized posters. We may accomplish this by considering the actor’s face or the mood that individuals want in addition to video keyframes and the protagonist’s face.

ACKNOWLEDGEMENTS

This research was supported by the Basic Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education (No. 2021 R1I1A3057800) and this results was supported by “Regional Innovation Strategy (RIS)” through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (MOE) (2021RIS-004).

REFERENCES

[1] J. Sang and C. Xu, “Character-based movie summarization,” in *Proceedings of the 18th ACM International Conference on Multimedia*, pp. 855-858, 2010. DOI: 10.1145/1873951.1874096.

[2] J. Monaco, “How to Read a Film: The Art, Technology, Language, History and Theory of Film and Media,” Oxford University Press, New York, 1982.

[3] M. Wang and W. Deng, “Deep face recognition: A survey,” *Neurocomputing*, vol. 429, pp. 215-244, Mar. 2021. DOI: 10.1016/j.neucom.2020.10.081.

[4] W. W. Bledsoe, “The Model Method in Facial Recognition,” Panoramic Research, Inc., Palo Alto: CA, Technical Report, PRI 15, 1964.

[5] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, “Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks,” *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499-1503, Oct. 2016. DOI: 10.1109/LSP.2016.2603342.

[6] L. Chaochao and X. Tang, “Surpassing human-level face verification

performance on LFW with GaussianFace," in *Proceedings of the AAAI Conference on Artificial Intelligence*, Austin: TX, pp. 3811-3819, 2015. DOI: 10.1609/aaai.v29i1.9797.

[7] Y. Sun, X. Wang, and X. Tang, "Deeply learned face representations are sparse, selective, and robust," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston: NY, pp. 2892-2900, 2015. DOI: 10.1109/CVPR.2015.7298907.

[8] K. Yoon and J. Choi, "Compressed Ensembleof Deep Convolutional Neural Networks withGlobal and Local Facial Features for ImprovedFace Recognition," *Journal of Korea Mul-timedia Society*, vol. 23, no. 8, pp. 1019-1029, Aug. 2020. DOI: 10.9717/kmms.2020.23.8.1019.

[9] Y. Ha, J. Park, and J. Shim, "Comparison of Face Recognition Performance Using CNN Models and Siamese Networks," *Journal of Korea Multimedia Society*, vol. 26, no. 2, pp. 413-419, 2023. DOI: 10.9717/kmms.2023.26.2.413.

[10] S. Florian, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston: NY, pp. 815-823, 2015. DOI: 10.1109/CVPR.2015.7298682.

[11] S. Tabata, H. Yoshihara, H. Maeda, and K. Yokoyama, "Automatic layout generation for graphical design magazines," in *SIGGRAPH '19: Special Interest Group on Computer Graphics and Interactive Techniques Conference, ACM SIGGRAPH 2019 Posters*, Los Angeles: CA, pp. 1-2, 2019. DOI: 10.1145/3306214.3338574.

[12] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, and I. Sutskever, "Zero-Shot Text-to-Image Generation," in *Proceedings of the 38th International Conference on Machine Learning*, vol. 139, pp.8821-8831, 2021.

[13] N. Ruiz, Y. Li, V. Jampani, Y. Pritch, M. Rubinstein, and K. Aberman, "Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation," *arXiv preprint arXiv: 2208.12242*, Aug. 2023. DOI: 10.48550/arXiv.2208.12242.

[14] S. Guo, Z. Jin, F. Sun, J. Li, Z. Li, Y. Shi, and N. Cao, "Vinci: An Intelligent Graphic Design System for Generating Advertising Posters," in *Proceedings of the 2021 CHI Conference on Human Factors inComputing Systems (CHI '21)*, Yokohama, Japan, pp. 1-17, 2021. DOI: 10.1145/3411764.3445117.

Yeonhwi You

received his B.S. degree in computer science and engineering in 2022 from Korea University of Technology and Education, Cheonan, Republic of Korea. He is currently pursuing a M.S. degree from the Department of Computer Science and Engineering at Korea University of Technology and Education. His current research interests include artificial intelligence, big data, and recommendation systems.

Sungjung Yong

received his M.S. degree in computer science and engineering in 2020 from Korea University of Technology and Education, Cheonan, Republic of Korea. He is currently pursuing a Ph.D. from the Department of Computer Science and Engineering at Korea University of Technology and Education. His current research interests include artificial intelligence, web services, and recommendation systems.

Hyogyeong Park

received her B.S. degree in computer science and engineering in 2021 from Korea University of Technology and Education, Cheonan, Republic of Korea. She is currently pursuing a M.S. degree from the Department of Computer Science and Engineering at Korea University of Technology and Education. Her current research interests include artificial intelligence, web services, big data, and recommendation systems.

Seoyoung Lee

received her B.S. degree in computer science and engineering in 2022 from Korea University of Technology and Education, Cheonan, Republic of Korea. She is currently pursuing a M.S. degree from the Department of Computer Science and Engineering at Korea University of Technology and Education. Her current research interests include artificial intelligence, web services, and computer vision.

Il-Young Moon

has been a professor at the Department of Computer Science and Engineering, Korea University of Technology and Education, Cheonan, Republic of Korea since 2005. He received his Ph.D. from the Department of Aeronautical Communication and Information Engineering, Korea Aerospace University in 2005. His current research interests include artificial intelligence, wireless internet and its applications, and mobile IP.