SMPTE

MI

MOTION IMAGING JOURNAL

COVERING EMERGING TECHNOLOGIES FOR THE GLOBAL MEDIA COMMUNITY

TECHNICAL PAPER

KEYWORDS VIDEO SUMMARIZATION // THUMBNAIL EXTRACTION // NEURAL NETWORK // AI

O *paper* desta edição definitivamente nos mostra que a IA está mudando tudo o que conhecemos. Em um mundo onde cada segundo conta, a fusão entre inteligência artificial e criatividade está revolucionando a maneira como consumimos histórias. Imagine transformar horas de conteúdo em minutos de impacto, capturando a essência de cada cena com a precisão de um editor experiente. A NHK (*Japan Broadcasting Corporation*) não apenas imaginou — tornou realidade. Suas redes neurais desvendam a magia por trás das imagens, traduzindo movimentos de câmera e emoções em resumos vibrantes. E com *thumbnails* que pulsam como cartazes de cinema, cada clique se torna uma porta para novas experiências. Este é o futuro da narrativa: ágil, envolvente e infinitamente conectado. Podemos dizer que: Na era do instantâneo, a IA não apenas acelera o tempo — ela o transforma em arte!

Boa leitura!

Tom Jones Moreira

AI Image Analysis in Era of Short-Time Viewing

By Momoko Maezawa, Rei Endo, and Takahiro Mochizuki

As AI technologies in an era where short-time viewing is preferred, we presented automatic video summarization technologies and a thumbnail extraction technology to distribute summary videos quickly to social media. Many NHK broadcasting stations utilize these systems as tools to activate internet deployments of broadcast content.

Abstract

In the era when short videos are preferred, broadcasting stations have been enhancing momentums to distribute summary videos of broadcast content on social media. Therefore, we have developed automatic generation systems for news and program summary videos. Using a video summarization artificial intelligence (AI) that has learned the image composition and camerawork typical of important scenes, it is possible to automatically generate summary videos with a high quality close to videos edited by actual program production staff. Our systems include functions enabling users to easily modify the automatically generated summary videos. These systems have been on trial/practical use in many Japan Broadcasting Corp. (NHK) broadcasting stations. The generated summary videos are posted daily on social media. Furthermore, considering a program website is also important content that could boost viewer contact rates, we developed a support system for program website creation using an AI to extract thumbnails automatically. AI-generated thumbnail images can be used to develop program websites wit minimal effort. These technologies can streamline the production of internet content such as summary videos and program websites. Moreover, they will significantly boost internet deployments of various broadcasting programs.

With the increase in consumption of short video clips on the internet, broadcasters are attempting to boost user contact with their content by producing summary videos and distributing them on the internet. Editorial operations to create summary videos require certain specialties and high work costs. Automating the summary video production process for content that is updated daily, such as news programs, is especially desirable. Therefore, we are developing automatic video summarization technologies.

In this paper, we introduce a system that automates the production of news summary videos. Our automatic video summarization technology has the following features. In news footage, the importance of shots is strongly correlated with image composition and camerawork, such as zooming in on important people, panning to show items of interest in detail, and special shooting angles for buildings involved in incidents. We trained an artificial intelligence (AI) system to learn picture-making, such as the image composition and camerawork typical of summary videos produced by skilled editors. Moreover, our technology used the similarity of keywords with an anchor's introduction speech to evaluate the importance of each video segment.

This system summarizes 15–30-min news programs into about 1–2 min with about 20 summary videos produced and distributed daily. Summary videos that used to take a skilled editor more than one hour to produce can now be generated about 10–20 min after the end of the broadcast program, making it possible to produce and distribute summary videos without losing the immediacy of the news.

Regarding common programs not including an anchor's introduction, we developed a practical system to automatically generate summary videos. As with the news video summarization system, the AI on this system learned picture-making typical of important scenes using summary videos created by professional video production staff. This system has also been on trial use in several NHK regional broadcast stations to post summary videos on social media, such as X.

Along with summary videos, attractive thumbnails on program websites can boost the number of accesses to broadcast content. An attractive thumbnail contains an eye-catching image that is representative of the video. We developed a support system for program website creation using AI to select attractive thumbnails from videos. This AI was generated by learning how to compose effective thumbnails that represent the associated video.

The program's website and summary videos can link users to broadcast content, making these technologies a bridge between the internet and broadcasting.

Related Work

Video Summarization

This section describes practical examples of automatic summarization technology in video production sites.

The important scenes to be used in summary videos, such as successful attempts and scores, are clear in sports videos. Therefore, the use summarization technology is being promoted for practical use in various sports events.

As a practical example for tennis, a summary video generation system developed by IBM was used at the 2018 Wimbledon Championships.[1] Using AI that has learned many point-scoring scenes collected from past game footage, it analyzes the cheers of the audience, movement of the players, scores of the players, etc., and automatically extracts the important scenes in the game. As for golf, at the 2019 Masters Tournament, IBM developed a technology to generate a summary video in a short time.[2] The importance of each shot is determined based on various factors, such as the presence or absence of specific actions (such as fist pumps), facial expressions, changes in the volume and tone of the cheers, and the position of the ball. Each player's importance is assigned automatically, and a summary video of each player is generated automatically. Practical examples for soccer include uniform number recognition, player movement analysis, and automatic summarization technology using techniques such as image trimming in accordance with the results. A summary video distribution service using this technology has started in Italy's Serie A.[3] As for basketball, to popularize and improve the value of the three-player basketball league, "3 x 3.EXE PREMIER," a summary video generation service is in operation that uses technology to detect objects such as scoreboards and recognize displayed characters.[4]

CONTENT THAT IS **UPDATED FREQUENTLY** IS WELL-SUITED TO THE INTERNET COMMUNITY, WHICH IS SENSITIVE TO TOPIC CHANGES.

In fields other than sports, it is challenging to define important scenes due to differences in editorial staff's points of view and viewers' tastes, and few technologies have been put to practical use.

Hakuhodo DY Media Partners Inc., Tokyo University of Science, and M Data Co., Ltd., have developed a trial version of a system that uses AI to automatically generate summary videos of dramas. They tested it for serial dramas broadcasted from July 2019.[5] However, this approach requires manually assigning metadata to each scene of a TV drama video, based on the performers' utterances and telop contents. Such data is attached to only some of the program videos owned by broadcasting stations. Therefore, we developed a practical technique that can automatically generate a summary video only from a program video.

Thumbnail Extraction

Various techniques have been proposed to select representative images from videos. These techniques use basic image information, such as color histograms, color layouts, texture features related to luminance co-occurrence, motion vectors between frames, and face sizes.[6,7,8] However, the various factors to be considered in selecting images are not always accurately reflected. In addition, a method using a convolutional neural network (CNN), which is the mainstream for tasks such as image classification, has also been proposed.[9] However, this technology is targeted toward YouTube, and because the number of video playbacks is used as training data for CNN, it is difficult to apply it to broadcast video.

The Aesthetic Visual Analysis (AVA) database[10] is a dataset in which images are divided into two classes, high quality, and low quality, based on aesthetic visual analysis by photographers. Jin et al. proposed a neural network (NN) that classifies images into high/low quality using the AVA database as training data.[11] First, the feature extraction network computes the image features, and then the classification network computes the two-class probability distribution from the image features. The selection of thumbnails by program production staff takes into consideration elements related to aesthetic preferences, such as image color and composition. For this reason, we apply this method in our program thumbnail extraction technology. Our technology can improve the efficiency of thumbnail selection work and drive the internet deployments of broadcast content.

News Video Summarization

Content that is updated frequently is well-suited to the internet community, which is sensitive to topic changes. At NHK, expectations are rising for the internet delivery of summarized news to increase the opportunities for viewers to experience broadcast contents. Therefore, we developed an automatic summarization system for news broadcast videos.[12] In this section, we describe the automatic summarization technology for news videos and the functions of the system.

Automatic Split into News Subjects

A typical news program video consists of a plurality of news items. Each news item consists of a part (lead video) in which the studio announcer reads the news, followed by a main story video. Since the content of each news item is completely different, to generate a summary video of the entire news program, it is necessary to automatically divide the news program video into news item units and summarize the main story video of each news item individually.

This section describes the flow of automatic segmentation into news items. First, a news program video is automatically divided into shots, and frame images are sampled from each shot. Then, each frame image is input to the "lead image determination AI" created in advance, and the prob-

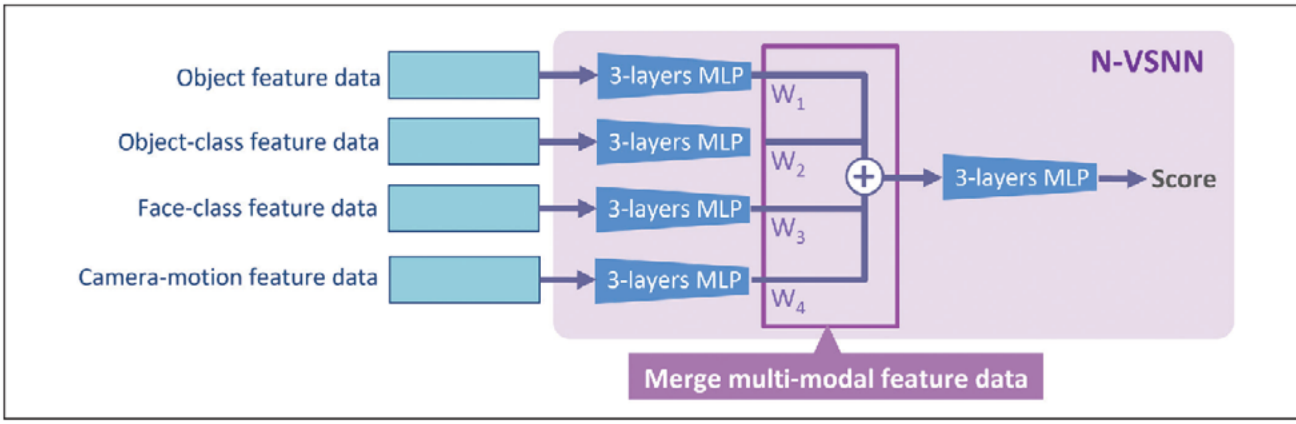

FIGURE 1. News video summarization neural network.

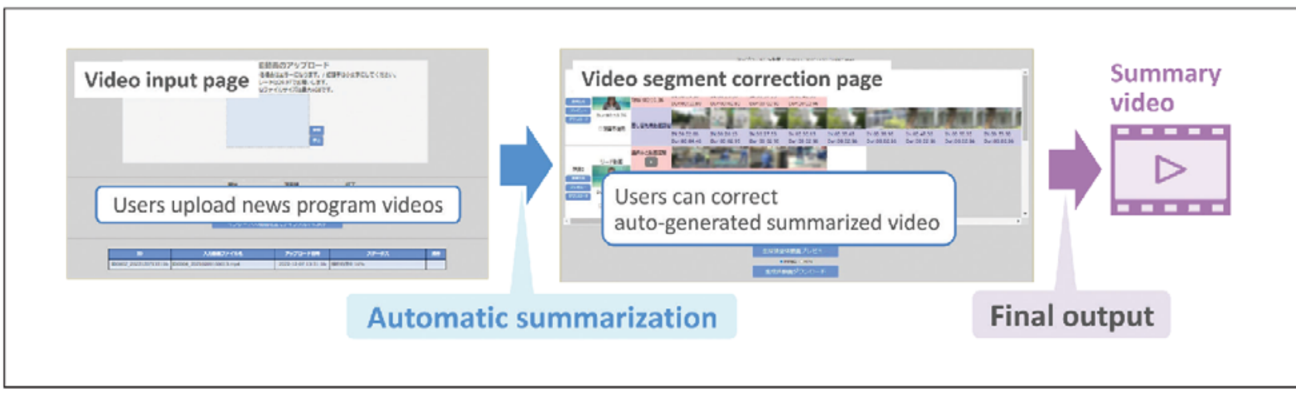

FIGURE 2. Overview of news video summarization system.

ability (score) of each image being included in the lead video is calculated. This process uses a support vector machine that has learned the image features of lead images in various news programs. Next, in each shot, the average score of the images belonging to it is determined. Finally, shots with average scores equal to or higher than a threshold are defined as lead videos. Shot sequences between the lead videos are defined as main story videos and intervals of each news item are determined.

News Video Summarization NN

A video summary of each news item is generated by extracting important video segments from the main story video. We describe the News Video Summarization Neural Network (N-VSNN) for estimating the importance score of each video segment. **Figure 1** shows the structure of N-VSNN. In video production at a broadcasting station, various elements unique to television, such as the type and size of the subject, composition, and camera movement, are considered. Therefore, we made it possible to input multiple modal features to N-VSNN and adopted the following four types of image features (hereafter referred to as "feature datasets") as feature data.

- *Subject feature:* Each image sampled at equal intervals from the video segment is input to the existing trained image classification CNN model,[13] and the feature vector is obtained from the intermediate layer. The average feature vector (2,048 dimensions) for all images is taken as the subject feature.
- *Object class feature:* In the object feature calculation process, the vector (1,000 dimensions) obtained in the same way using the final layer instead of the intermediate layer is used as the object class feature.
- *Face class feature:* From each image sampled in the same way as the subject feature, the face region image detected by Kawai et al.'s method[14] is input to the existing trained face classification model[15] to obtain a feature vector. Then, for each feature vector, a similarity vector is generated using 1,000 face classes created by clustering a large number of face feature vectors. This similarity vector consists of the inner product of the feature vector and the center vector of each class. A face-class feature (1,000 dimensions) is obtained by multiplying and summing the similarity vectors for all face regions by a weighting factor in accordance with the face size.
- *Camera movement feature:* The video segment is divided equally into three sub-segments, the first half, the middle part, and the second half. A vector connecting motion histograms (6 regions x 8 directions) in each sub-segment is taken as a camera motion feature (144 dimensions).

In the previous network, for each modal, 256-dimensional intermediate data is generated through a three-layer multi-layer perceptron (MLP), and the four intermediate data are multiplied by weighting factors W_i (i = 1, ..., 4) and added. When N-VSNN is trained, W_i is updated along with the MLP parameters. Finally, the importance of video segments is output through a post-network consisting of three-layer MLPs.

As training data, we used about 200 summary videos distributed in the past on NHK's news website[16] and news main story videos as sources for each. N-VSNN was generated by training so that a high score would be output for sections of the program video included in the summary video, and a low score would be output for other sections. By using N-VSNN, is it possible to extract video segments with picture-making unique to important scenes in news programs, such as zooming in on important people, panning to show the subject in detail, and special shooting angles for buildings involved in the incident.

Speech Content Score

In the lead video, the announcer gives an overview of the news. Therefore, we consider that the interval where the lead video and the utterance content are similar is important and we introduced the "utterance content score" that expresses the similarity to the summarization process.

In this section, we describe the calculation method of the utterance content score. Keywords are extracted by recognizing the voice of the lead video, and each keyword is vectorized by Word2Vec (hereinafter referred to as a keyword vector). Keyword groups are extracted from each utterance period of the main story video by utterance period detection processing and speech recognition processing, and keyword vector groups are obtained in the same manner as the lead video. For each utterance segment, the degree of similarity between keyword vector groups with the lead video is used as the utterance content score.

News Video Auto-Summarization System

This section describes the News Video Auto-Summarization System developed using the aforementioned processes and N-VSNN. **Figure 2** shows the flow of generating a summary video using this system. The user uploads a news program video to be summarized on the video input page. After uploading, a summary video is automatically generated by the following process.

1. Using the method of Automatic Split into News Subject sections, the news video is divided into news items, and steps 2–7 are performed for each news item.
2. A main story video is divided into shot units, and each shot is divided into fixed-length video segments.
3. For each video segment n, a feature dataset is computed and input to N-VSNN to compute the importance $S_{NN}(n)$.
4. The utterance content score $S_{SP}(m)$ is calculated for each utterance section m using the method in the Speech Content Score section.
5. A total score S(n) for each video segment is calculated using Equation (1).

   $$S(n) = S_{NN}(n) + S_{SP}(m^*)\, R(n, m^*) \qquad (1)$$

   Here, m* is the index of the utterance segment with the highest overlapping rate with the video segment n, and R(n, m*) represents the overlapping rate.
6. The moving image segments are shot out in descending order of S(n). When the total duration of the clipped sections exceeds the length of the lead video, the process ends, and the video segments are connected to generate a summary video.
7. The audio of the summary video is replaced with that of the lead video. Since the content of the lead video is the outline explanation of the news by the announcer, this processing can be expected to improve the explainability of the summary video.

To put the system into practical use, in addition to the function of automatically generating summary videos, it is necessary to perform confirmation and correction work to deploy it to media other than broadcasting. For example, consideration must be given to personal information and the right to use video materials outside of broadcasting. Therefore, we implemented a "video segment correction page" to easily correct the automatically generated summary video. On this screen, it is possible to replace the video segment that constitutes the summary video with another candidate and adjust the IN/OUT points of each video segment. The video summary of each news item that has undergone correction work is connected to a dedicated logo video in between to output the final video summary.

With this system, it is possible to generate summary videos of 15–30-min news programs by automatic processing in about 10–20 min and correction work in about 15 min . Full-scale practical use of this system began in the fall of 2022 and NHK headquarters and many regional broadcasting stations distribute news summary videos generated by this system on social media daily. Furthermore, this system was successfully introduced across all NHK broadcasting stations in the fall of 2023.

Program Video Summarization

For general programs other than news, there is a growing interest in internet development, and there is a demand for a mechanism that automatically generates a summary video. This section describes an NN for program video summarization that we devised[17] and an automatic program video summarization system that we developed using this NN.

Program Video Summarization NN

Figure 3 shows the structure of the Program Video Summarization NN (P-VSNN) for estimating the importance score of each video segment in a program video. Many program videos are long, from tens of minutes to hours. In the process of summarizing long videos, it is common to evaluate the importance of each video segment by considering the video content of the section, nearby sections, and the video content of the entire video. Therefore, P-VSNN is input with feature datasets calculated on multiple timescales, such as nearby shots and the entire video, instead of the feature dataset for the video segment for which the score is to be estimated.

In the network of the first stage, for each modal, feature data of multiple time scales are integrated by a 1-dimensional (1D) CNN and Max pooling, and 256-dimensional interme-

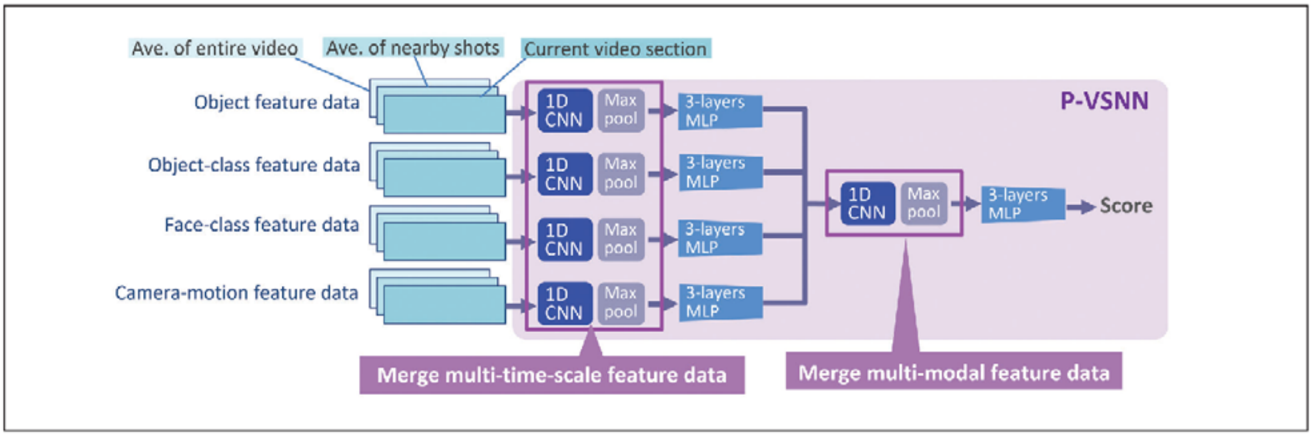

FIGURE 3. Program video summarization neural network.

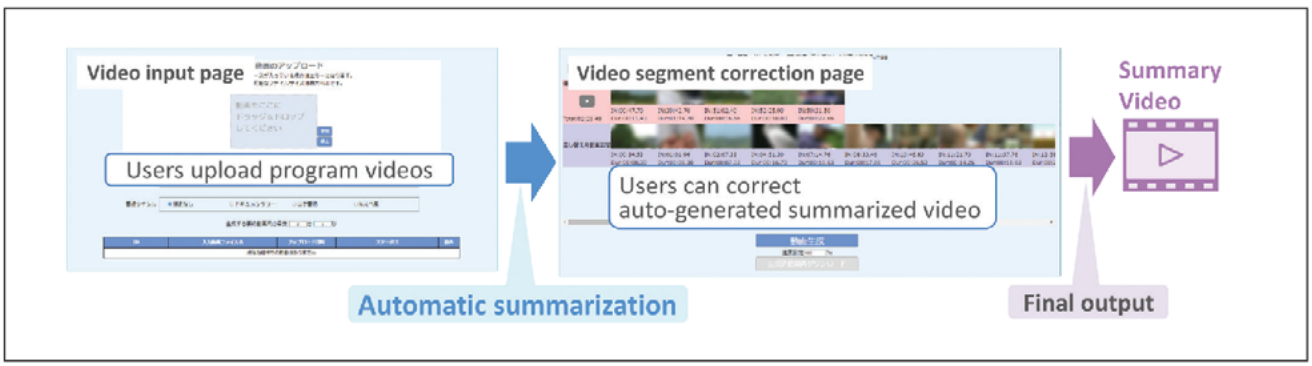

FIGURE 4. Overview of program video summarization system.

diate data are output through a three-layer MLP. In the latter network, the intermediate data of each modal is integrated by a 1D CNN and Max pooling, and the score of the video segment is output through a three-layer MLP.

P-VSNN was trained using dozens of program summary videos created by professional video editing staff, and each program video as learning data. By using P-VSNN, it is possible to extract video segments with picture-making (how to shoot the subject, composition, camera work, etc.) unique to the important scenes of the program video.

Program Video Auto-Summarization System

Using the aforementioned P-VSNN, we have developed an automatic program video summarization system. **Figure 4** shows the flow of program summary video generation by this system. The user uploads a program video to be summarized on the video input page. At that time, it is possible to specify the approximate length of the generated summary video (=T_{SUM}). After uploading the program video, a summary video is automatically generated by the following process.

1. A program moving image is divided into shot units, and each shot is divided into fixed-length moving image sections.
2. For each video segment n, feature datasets at three-time scales (this video segment, average of nearby shots, average of entire video) are computed and input into P-VSNN to calculate importance S(n).
3. The video segments are cut out in descending order of S(n). At that time, to reduce the sense of the incongruity of the sound at the connection point of the moving image section, if the cut-out point is in the middle of the speech section, it can be changed to the start or end point of the speech automatically. An existing library[18] was used to detect speech segments from videos. When the total length of the clipped sections exceeds T_{SUM}, the process ends, and the video segments are connected to generate a summary video.

In this system, we implemented a "video segment correction page" in the same way as the news summarization system, with functions by easy mouse operations such as replacing the video segment that constitutes the summary video with another candidate, adjusting the IN/OUT points of each video segment, and so on. After performing correction work as necessary, the final summary video is output. The user can use this system to generate a summary video with approximately half the program's length and only about 5–10 min of correction work. Currently, several NHK regional broadcasting stations are using this system on a trial basis, and video summaries of cameraman-led programs and other programs generated on this system are being distributed to social media.

Program Thumbnail Extraction

Similar to video summaries, thumbnail images of programs

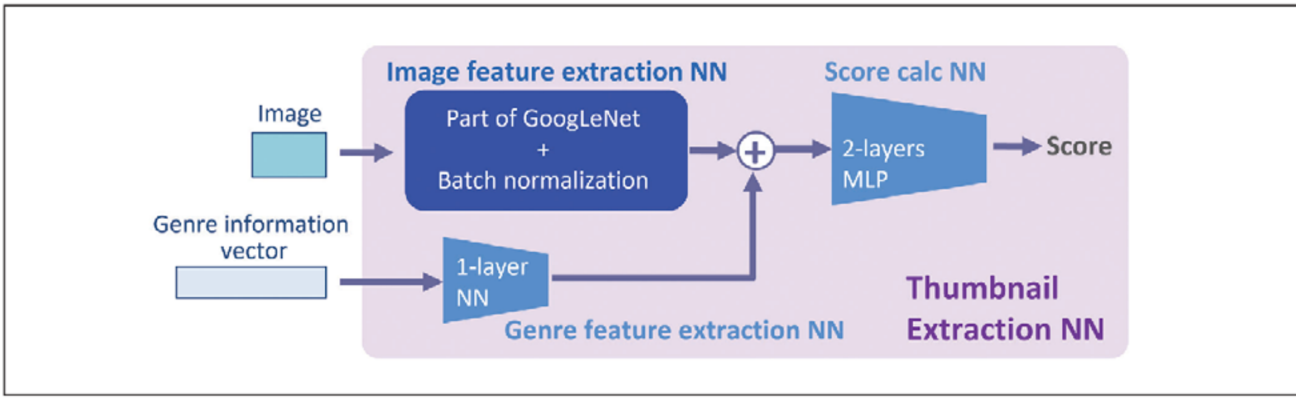

FIGURE 5. Thumbnail extraction neural network.

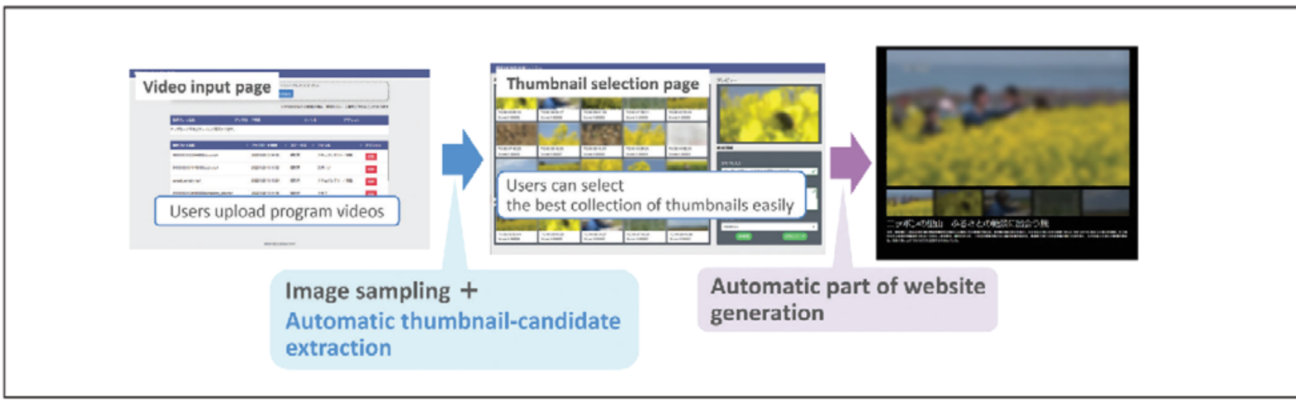

FIGURE 6. Overview of support system for program website creation.

posted on program websites and social media are indispensable items for improving the rate of contact with broadcast content. The presentation of eye-catching thumbnails can be expected to increase the number of viewers accessing programs. However, selecting thumbnails is not an easy task because it is necessary to carefully consider the subject's color, composition, content and so on. NHK disseminates information using many program websites and social media, and improving the efficiency of thumbnail selection work is necessary. This section describes our developed technology for extracting thumbnails and supporting website creation programs.

Thumbnail Extraction NN

The program genre has a significant effect on the thumbnail selection criteria. For example, an image showing the performers is preferred for a drama program, and for a travel program, the beauty of the scenery is given priority. Therefore, we developed a method to select thumbnails from program videos by scoring images using a NN that has learned both images and program genre information.[19] This method makes it possible to select images considering the trend of thumbnails for each program genre.

Figure 5 shows an overview of our developed thumbnail extraction NN. The input is the images sampled from the program video and the genre information vectors representing the program genre (drama, travelogue, variety, etc.), and the output is the score representing the image's suitability as a thumbnail. A genre information vector is a binary vector representing whether a program belongs to each genre by 1/0, and the number of dimensions is 8, which is the number of genres. If the output score equals or exceeds a predetermined threshold value, it is selected as a thumbnail candidate.

Thumbnail extraction NN consists of three networks: an image feature extraction network that computes image features, a genre feature extraction network, and a score computation network.

The image feature extraction network uses the network structure of an existing method[11] that evaluates the visual artistry of images. A structure using GoogLeNet[20] and Batch Normalization[21] generates a 1024-dimensional feature vector. The genre feature extraction network is a one-layer fully connected NN and generates feature vectors (1,024 dimensions) with the genre information vector described later as input. The outputs of the image and genre feature extraction networks are added and input to the score calculation network. The score calculation network, a two-layer MLP, outputs a score between 0.0 and 1.0 using the sigmoid function.

As training data, we used about 6,500 images sampled from various program videos and three-level labels (great/good/bad) assigned to each image by program production directors and editorial assistants regarding suitability as program representative images. We set the correct scores for great, good, and bad images to be 1.0, 0.5, and 0.0, respectively, and trained the thumbnail extraction NN using the image and genre information as input. Using this NN makes it possible to extract thumbnails that reflect the image selection skills of professional program production staff.

Support System for Program Website Creation

As previously mentioned, NHK publishes information on many program websites, and there is a demand for a mechanism to streamline the selection of thumbnail images to be posted on the website along with the creation of the website. Therefore, we have developed a system that supports the creation of program websites using the thumbnail extraction NN previously described.

Figure 6 illustrates the process flow of this system. The user uploads a program video, selects the genre of the program, and then a representative image is extracted. Next, shot division of the uploaded video and image sampling processing are automatically performed. The thumbnail extraction NN calculates image scores for the sampled images. Images with high scores are presented as candidate images, and users select the image they want to use. At that time, if necessary, it is possible to input text such as a program outline to be posted on the webpage. Finally, a portion of the website is automatically generated using the selected thumbnail image. Users can preview the generated website and download the HTML files and thumbnail images.

The system will allow for easy creation of a program website and the possibility of further enhancements. In the future, we aim for practical use after trials at the program production site.

Conclusion

As AI technologies in an era where short-time viewing is preferred, we presented automatic video summarization technologies and a thumbnail extraction technology to distribute summary videos quickly to social media. Many NHK broadcasting stations utilize these systems as tools to activate internet deployments of broadcast content. In addition, the program website creation support system using the thumbnail extraction technology is expected to be a practical tool for reducing the cost of creating websites, which is essential to grasp the program easily. In the future, we will work to improve the performance of each technology by retraining AI using data accumulated during trials and practical use and refining the systems based on the needs of program production sites. A mechanism to automatically retrain AI while operating the system could also be necessary.

References

1. How AI picks the highlights from Wimbledon fairly and fast, 2019. [Online]. Available: https://www.ibm.com/blog/how-ai-picks-the-highlights-from-wimbledon-fairly-and-fast/
2. AI-generated Highlights Tell the Story of the Masters, 2019. [Online]. Available: https://blog.video.ibm.com/ai-video-technology/ai-generated-highlights-tell-story-of-the-masters-2019/
3. Serie A agrees AI highlights deal with WSC Sports, 2022. [Online]. Available: https://www.sportbusiness.com/news/serie-a-agrees-ai-highlights-deal-with-wsc-sports/
4. Providing automatic highlight creation service using NTT DoCoMo Co., Ltd., and AI - Easily extract only the highlight scenes you want and reduce the burden of video editing, 2021. [Online]. Available: https://3x3exe.com/premier/news20220601-3/
5. Hakuhodo DY Media Partners Develops an Automatic Drama Digest Video Generation System Together with the Tokyo University of Science: Trial Operation with a New Drama Program Broadcast on Tuesdays by TBS, 2019. [Online]. Available: https://www.inter-bee.com/2019/en/magazine/production/detail/?id=871
6. Y. Gao, T. Ahang, and H. Xiao, "Thematic video thumbnail selection," *Proc. of ICIP*, pp. 4333-4336, 2009.
7. H.-C. Lian, X.-Q. Li, and B. Song, "Automatic video thumbnail selection," *Proc. of Internat. Conf. on Multimedia Technol.*, pp. 242-245, 2011.
8. Y. Song, M. Redi, J. Vallmitjana, and A. Jaimes, "To click or not to click: Automatic selection of beautiful thumbnails from videos," *Proc. of ACM International Conference on Information and Knowledge Management*, pp. 659-668, 2016.
9. N. Arthurs, S. Birnbaum, and N. Gruver, "Selecting youtube video thumbnails via convolutional neural networks," *Technical Report*, Stanford, 2017.
10. N. Murray, L. Marchesotti, and F. Perronnin, "AVA: A large-scale database for aesthetic visual analysis," *Proc. of CVPR*, pp. 2408-2415, 2012.
11. X. Jin, et al., "ILGNet: Inception modules with connected local and global features for efficient image aesthetic quality classification using domain adaptation," *IET Computer Vision*, 13.2: 206-212, 2019.
12. T. Mochizuki, Y. Kawai, N. Fujimori, M. Maezawa, R. Endo, Y. Asami, "Prototype of support system for news summary video production (in Japanese)," *J. Instit. Image Inform. and Tel. Engin.*, 77(2): 262-271, 2023.
13. D. Tran, L. Bourdev, R. Fergus, L. Torresani, M. Paluri, "Learning spatiotemporal features with 3D convolutional networks," *Proc. of ICCV*, pp. 4489-4497, 2015.
14. Y. Kawai, R. Endo, N. Fujimori, T. Mochizuki, "Study of face recognition using deep neural network (in Japanese)," *Proc. of Forum of Information Technology (FIT)*, no. 3, H-003, pp.103-104, 2019.
15. Y. Kawai, R. Endo, N. Fujimori, T. Mochizuki, "Face detection using cascaded convolutional network (in Japanese)," *Proc. of ITE Annual Convention*, no. 3, 22B-1, 2018.
16. NHK NEWS WEB, [Online]. Available: https://www3.nhk.or.jp/news/movie.html
17. T. Mochizuki, Y. Kawai, "Program video summarization by fusion of multi-modal/timescale features using 1D-CNN (in Japanese)," *Proc. of IEICE General Conference*, D-12-26, 2022.
18. Real Python, The Ultimate Guide to Speech Recognition [Online]. Available: https://realpython.com/python-speech-recognition/
19. M. Maezawa, R. Endo, N. Fujimori, T. Mochizuki, "A study on the selection of thumbnail images considering a genre of TV programs (in Japanese)," *IEICE Technical Report*, vol. 121, no. 155, PRMU2021-15, pp. 48-51, 2021.
20. C. Szegedy, et al., "Going deeper with convolutions," *Proc. IEEE Conf. on Comput. Vis. and Pattern Recognotion.* (CVPR), 2015.
21. S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *Proc. 32nd International Conference on Machine Learning (ICML)*, 448-456, 2015.

About the Authors

Momoko Maezawa received BE and ME degrees from Keio University, Japan, in 2017 and 2019 respectively. She is a research engineer with the Science and Technology Research Laboratories, NHK (Japan Broadcasting Corporation).

Rei Endo received BE, ME, and PhD degrees in engineering from Keio University, Japan, in 2008, 2010, and 2013, respectively. He is a research engineer with the Science and Technology Research Laboratories, NHK.

Takahiro Mochizuki received BE, ME, and PhD degrees in engineering from the Tokyo Institute of Technology, Japan, in 1994, 1996, and 2010, respectively. He is a senior research engineer with the Science and Technology Research Laboratories, NHK.