클라이언트 중심의 음악 장르 분류 프레임워크

굴람무즈타바1), 박은수2), 김승환2), 류은석2)

가천대학교1), 성균관대학교2)
mujtaba@gc.gachon.ac.kr1), espark804@skku.edu2), whitekomani@skku.edu2),
esryu@skku.edu2)

Client-driven Music Genre Classification Framework

Ghulam Mujtaba1), Eun-Soo Park2), Seunghwan Kim2), Eun-Seok Ryu2)

Gachon University1), Sungkyunkwan University2)

요약

We propose a unique client-driven music genre classification solution, that can identify the music genre using a deep convolutional neural network operating on the time-domain signal. The proposed method uses the client device (Jetson TX2) computational resources to identify the music genre. We use the industry famous GTZAN genre collection dataset to get reliable benchmarking performance. HTTP live streaming (HLS) client and server sides are designed locally to validate the effectiveness of the proposed method. HTTP persistent broadcast connection is adapted to reduce corresponding responses and network bandwidth. The proposed model can identify the genre of music files with 97% accuracy. Due to simplicity and it can support a wide range of client hardware.

1. INTRODUCTION

Music genre classification is one of the popular problems in machine learning. The problem has not only commercial business value, but also has many practical applications such as music recommendation [1]. Another application can be to automatically organize and tag by genre a huge music corpus. The convolutional neural network (CNN) can learn the features of music and most likely identify the music genre. Once the network identifies the genre (or sub-genres), that information can be used for music recommendation and discovery.

In past decades, CNNs have made a lot of progress in the computer vision area. Some of the networks even reached better accuracy than humans to classify images. Instead of classifying images, here we use CNN for music genre classification with mid-level time-frequency image representations (spectrograms) as inputs. We use the industry famous GTZAN dataset to get reliable performance benchmarking [2]. The network can classify music genres such as pop, jazz, rock, etc. We propose a unique client-driven music genre classification solution, that can identify its genre using a deep CNN operating on the time-domain signal. The proposed consist of two major sides i.e., The HLS server and client sides. To validate the performance of the proposed approach, the HLS client is configured on Nvidia Jetson TX2 an embedded AI computing device to demonstrate the performance. For the server-side perspective, the HLS server locally configured on Microsoft Windows internet information services (IIS) [3].

The main contributions are: (1) proposed a client-driven music genre classification framework. (2) Trained CNN model to identify the music genre. (3). Finally, HLS client and server sides configured locally to validate the effectiveness of the proposed method.

2. RELATED RESEARCH

Music genre classification has been a popular topic in music information retrieval since the seminal study [4]. Despite the extensive research, the field still presents open challenges nowadays, such as the ill-defined concept of genre, which is vague, fuzzy, and subject to human perceptions [5]. The CNNs have been actively used to find solutions for the tasks, such as music tagging [6], and genre classification [7]. To classify and organize large music corpus, enormous computational resources will be required at the server-side. The server-side computational resources demand can be reduced if we process it on the client-side. Now several devices are available in the market which has high computational resources (i.e., Nvidia Jetson TX2). This paper proposes a client-driven method that uses client-device computational resources to classify the music genre.

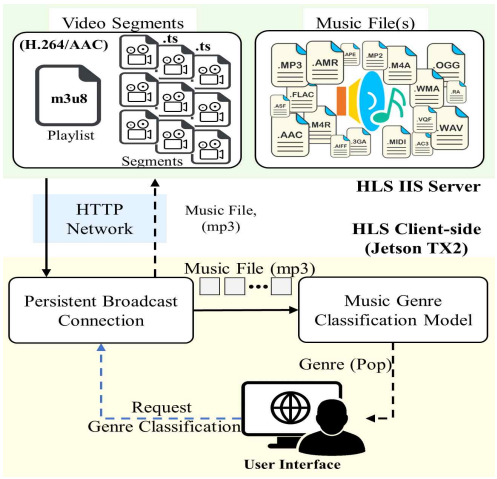

Figure 1: Proposed client-driven framework for music genre classification.

3. PROPOSED FRAMEWORK

This section explains the proposed client-driven music genre classification framework. The proposed system architecture is shown in Figure 1. The proposed approach mainly focusses on the client-side implementation. The HLS client-side configured on Nvidia Jetson TX2 an embedded AI computing device. It is a GPU based board with Nvidia 256 core pascal architecture along with 64-bit hex-core ARMv8 CPU, stacked with a memory of 8 gigabytes and 59.7 GB/s 128-bit the interface of memory data transfer capacity [8]. Jetpack 4.3 SDK is used to automate the basic installations on Jetson TX2, which includes the Board Support Packages, libraries especially for deep learning and computer vision. Jetson TX2 supports several energy profiles; among those profiles, Max-N used in experiments.

The client-side consists of two major components: HTTP persistent connection to download the music file, music genre classification model. Due to faster for frequent data exchanges, and better performance [9], HTTP persistent connection is used to download the music file(s). The music genre classification model is trained on the GTZAN dataset. Even though some drawbacks and limits are indicated [10], it is still one of the widely used datasets as a benchmark for music genre classification. The dataset consists of 1,000 songs and 10 genres. Each clip is 30-second long with the sampling rate of 22,050 HZ, 16 bits.

4. IMPLEMENTATION DETAILS AND RESULTS

To train the dataset each audio clip of 30-seconds split into 3 seconds with the size 19000×129×128×1 (samples × time × frequency × channels). Initially, all the audio files transformed into as mel-spectrograms and chromagram, both of which are a 2D array in terms of time and feature value (see Figure 2). Computing the spectrograms makes extensive use of the librosa library for audio processing. This is a standard approach to processing music and speech because the mel-scale corresponds well with human sound perception.

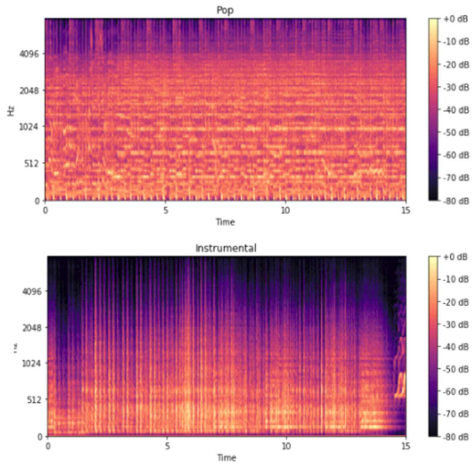

Figure 2: Spectogram for Pop (top), Instrumental (below) clips

The features extracted using the VGG16 convolutional neural network with a small modification. We have used Keras toolbox for features extraction and train the network on GeForce RTX 2080 Ti GPU. SGD optimizer is used to train the network with learning rate of 0.001 and decay of 0.01 and 0.9 momentum. The model reaches an accuracy of 97.02% in the validation set. Figure 3 shows the confusion matrix of genre predictions. Figure 4 shows the training and validation accuracy of the model during the training process. The trained model is used in Jetson TX2 to classify the music genre. The proposed approach process only music file in the client-side without extracting from a video, thus it requires minimum computational resources. The computational demand on the server-side can be reduced further if we process more operations (i.e., trailer generation [11]) in the client-device. In addition, it can support the privacy-preserving solutions using efficient encryption techniques [12].

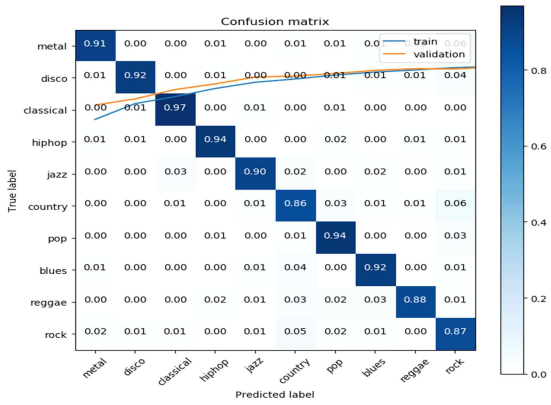

Figure 3: Confusion matrix of genre predictions

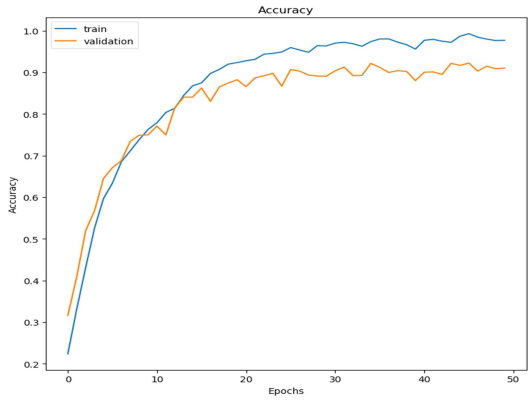

Figure 4: Training and validation accuracy during training on GTZAN dataset

5. CONCLUSION

This paper presents a client-driven music genre classification framework. Jetson TX2 configure locally to analyze the music genre classification using its resources. HTTP persistent connection adapted to download the music file(s) to reduce CPU usage and roundtrips. We use the industry famous GTZAN genre collection dataset to get reliable performance benchmarking. The proposed model can identify the genre of music files with 97% accuracy. As several devices are available in the market having a high computational resource such as Nvidia Jetson TX2, their computational resources can be used to reduce demand on the server-side.

ACKNOWLEDGMENT

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터육성지원사업의 연구결과로 수행되었음" (IITP-2020-2017-0-01630)

참 고 문 헌

〔1〕 Chen, Xu, Yongfeng Zhang, Qingyao Ai, Hongteng Xu, Junchi Yan, and Zheng Qin. "Personalized key frame recommendation." In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 315-324. ACM, 2017.

〔2〕 Sturm, Bob L. "An analysis of the GTZAN music genre dataset." In Proceedings of the second international ACM workshop on Music information retrieval with user-centered and multimodal strategies, pp. 7-12. ACM, 2012.

〔3〕 O'Leary, Mike. "IIS IIS IIS and ModSecurity." In Cyber Operations, pp. 789-819. Apress, Berkeley, CA, 2019.

〔4〕 Tzanetakis, George, and Perry Cook. "Musical genre classification of audio signals." IEEE Transactions on speech and audio processing 10.5 (2002): 293-302.

〔5〕 Aucouturier, Jean-Julien, and Francois Pachet. "Representing musical genre: A state of the art." Journal of New Music Research 32.1 (2003): 83-93.

〔6〕 Dieleman, Sander, and Benjamin Schrauwen. "End-to-end learning for music audio." In 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6964-6968. IEEE, 2014.

〔7〕 Chiliguano, Paulo, and Gyorgy Fazekas. "Hybrid music recommender using content-based and social information." In 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2618-2622. IEEE, 2016.

〔8〕 Jetson TX2 Module. (2020, June). Retrieved from https://developer.nvidia.com/embedded/jetson-tx2.

〔9〕 Zurawski, Richard. "The Hypertext Transfer Protocol and Uniform Resource Identifier." In The Industrial Information Technology Handbook, pp. 456-478. CRC Press, 2004.

〔10〕 Sturm, Bob L. "The state of the art ten years after a state of the art: Future research in music information retrieval." Journal of New Music Research 43, no. 2 (2014): 147-172.

〔11〕 Mujtaba, Ghulam, and Eun-Seok Ryu. "Client-Driven Personalized Trailer Framework Using Thumbnail Containers." IEEE Access 8 (2020): 60417-60427.

〔12〕 Mujtaba, Ghulam, Muhammad Tahir, and Muhammad Hanif Soomro. "Energy Efficient Data Encryption Techniques in Smartphones." Wireless Personal Communications 106.4 (2019): 2023-2035.