EE4213
Human Computer Interaction
Individual Project

Interactive Musical Instruments Learning App (using Hand Tracking)

Medium Article:
https://medium.com/@yuvrajpatra/interactive-musical-instruments-learning-app-using-hand-tracking-and-generative-ai-da7ec79c3084

PATRA Yuvraj
55907774

1. Objective and Vision

This project aims at creating a human-computer interface application that helps bring alive the experience of learning music in a way like never before. Learning in classrooms often brings along with it a certain amount of boredom unless there is some application that can make things exciting and rejuvenate young minds. To make this more interesting, exciting new technology with the right guided approach can make learning not only more effective but also spike a general inclination towards asking questions, and critical thinking and creativity. The idea of this project is to enable young primary school kids to learn more about different instruments through an interactive web app and make it more interesting using hand tracking so they can learn about any instrument of their choice, from a set of options, learn more about it and also have the opportunity to play that instrument simply with their hand gestures, not requiring dexterity of any kind at their tender age to enjoy the experience.

2. Methodology

2.1. Hand Tracking

Hand tracking technology is the cornerstone of this project. Hand tracking has significantly changed and enhanced user experiences across technological and industrial platforms. It has been the foundation for sign language comprehension, menu controls with hand gestures as well as all applications involved in the next most cutting-edge area of Human Computer Interaction: *Augmented Reality*.

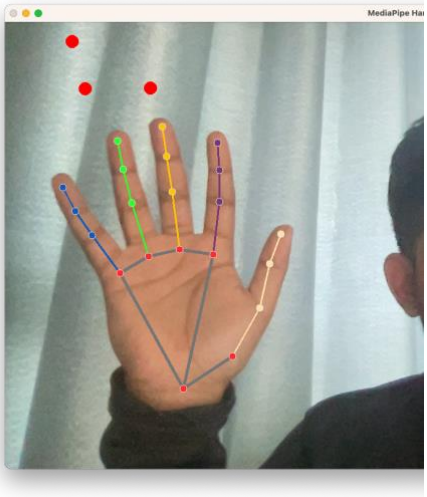

Figure 1 Hand Tracking Demo

Accurate measurement of hand tracking has been a difficult task in the field of computer vision. However, Google's MediaPipe Hands is an excellent high-fidelity finger and hand tracking solution which is driven by a ML (machine learning) pipeline that infer 3 dimensional landmarks from a single captured image frame. The MediaPipe Hands ML algorithm can also extend to both hands and this

technology begs for thorough research in this area with a scope of a huge variety of creative and new applications, one of which has been developed and pursued in this project. The pipeline consists of two machine learning models namely the Palm Detection Model and the Hand Landmark Model. This project uses hand gestures like hand gestures used in American Sign Language. A model was newly trained on synthesized data with the help of cv2 and cvzone python packages to create this project's training dataset for hand sign detection.

2.2. Deep Learning

A convolutional neural network was trained with synthesized training data by capturing numerous cropped images of the hand gestures using cv2. The different labels to this training data set were defined and the Keras model was trained as an image classification model which could classify a hand gesture into one of the 8 classes each used to denote a unique note in one octave of musical notes from C4 to C5 (piano as reference).

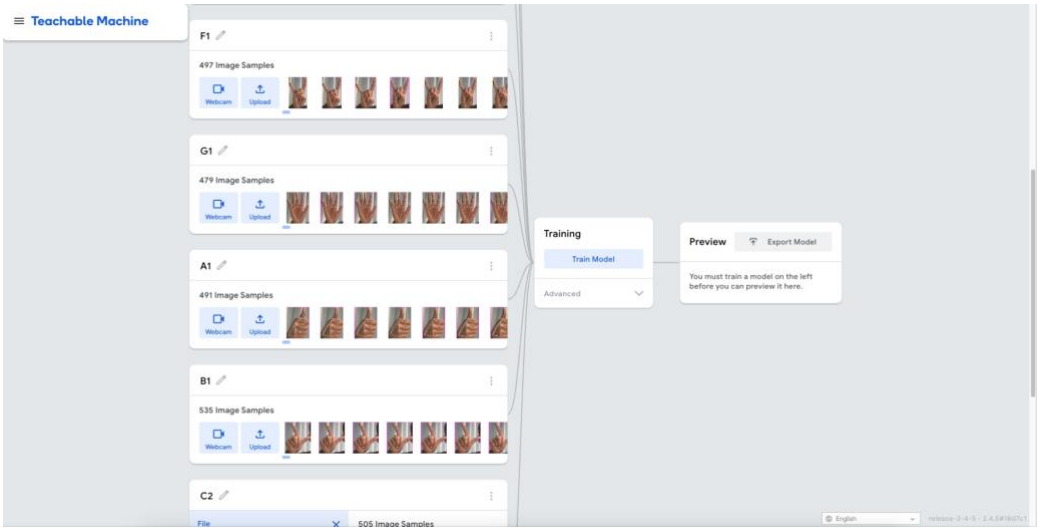

Figure 2 Google Teachable Machine training a CNN classification model for Hand Gesture Recognition. Retrieved from: https://teachablemachine.withgoogle.com/

The tool used for model training was the Google Teachable Machine which helps train simple classification models using CNN architecture by uploading input data files (images for each classification class) (approximately 350-400 images per classification type). Figure 2 shows a brief overview of the user-friendly interface of the Teachable Machine by Google where one can simply upload the image files as well as the label to which they are to be associated with (target variable) upon training it with the CNN classification model. Figure 3 shows a brief overview of the architecture of a CNN.

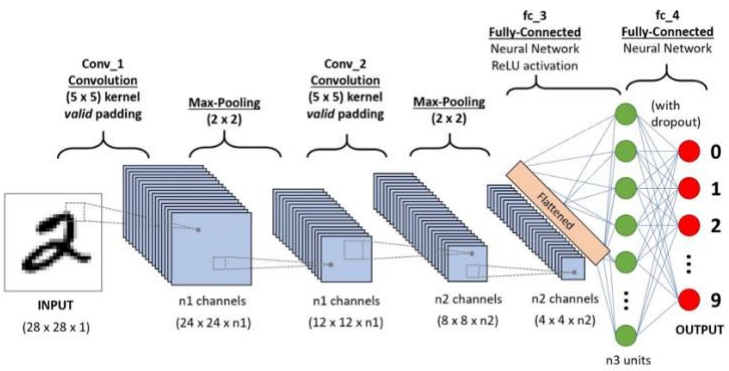

*Figure 3 Convolutional Neural Network containing Sequential, Conv2D, MaxPooling2D and Dense layers. Retrieved from: Source: https://miro.medium.com/max/1400/1*uAeANQIOQPqWZnnuH-VEyw.jpeg*

2.3. Music Instrument and Sound Generation

Music Instrument selection will be done with the scamp package for music provided by Python. With the help of this package an instrument can be selected and notes or chords on that instrument can be played for a certain duration with a certain volume. This procedure is used repeatedly for generating music by the system. For example, one can play the sound of a Piano at a pitch of 60 corresponding to 261 Hz for C4 on an actual piano. Similar this extends to 7 other notes that were used in this project for one octave to suffice basic music generation requirements.

2.4. Streamlit Web App

Streamlit is an open-source app framework for Machine Learning and Data Science work. We use Streamlit to create a simple Web App interface that helps the kids interact and learn more about musical instruments. We also use this interface to integrate it with one of the most famous technologies that took the world by surprise earlier this year: ChatGPT.

2.5. ChatGPT

The Streamlit App interface has ChatGPT being accessed at the backend with API calls. ChatGPT is the revolution in AI research and development that took the world by surprise in November 2022. Since this project requirement is not as complex, we resort to a simpler text-davinci-003 model to make API calls to retrieve information that is displayed onto the screen of the Streamlit Web App.

2.6. Synthesia.io

Synthesia.io is another interesting application of AI. “Synthesia is an AI video generation platform that enables you to quickly create videos with AI avatars, in over 120 languages. It includes templates, a screen recorder, a media library, and more.” (https://www.synthesia.io/features). This feature has also been integrated into the WebApp to make the experience even more interactive!

3. Design Implementation

The Design Implementation begins from the Streamlit App. Figure 4 shows the design of the interface. The Web App is divided into three columns. On the left panel we have a selection drop box and that affords the user (the primary school student) to select an instrument. Based on this selection the middle column displays the relevant image of the instrument, clearly visible for Norman's principle of visibility. The rightmost column displays a short description of what the instrument is for the student to learn. ChatGPT prompts were tailored to make it sound as if the AI were speaking to a 9 or 10-year-old child. This means that we put ourselves in the shoes of the user (the student), which helped us focus on maximizing their experience from this design. Finally, the middle column also has an attractive center button that reads "LET'S HAVE SOME FUN", which affords the user to click on it and open up the interface for playing the instrument. We must also focus on the simplicity of the interface since this makes it look elegant and user-friendly instead of overwhelming the user with too many features or instructions since after all, this App will mostly be used by young students.

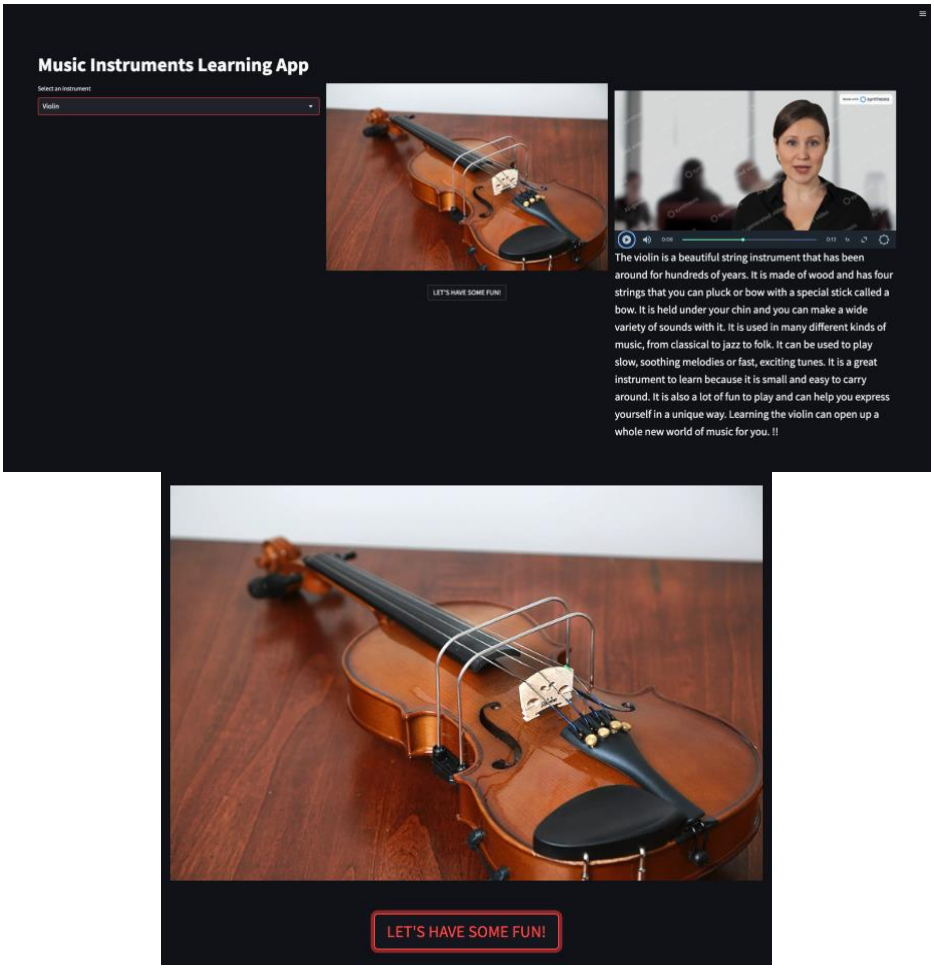

Figure 4 Streamlit Interactive WebApp Interface

Clicking this button takes the user to an interface that helps us play instruments with our hand gestures requiring close to minimal dexterity making this an application that can be easily used by a primary

school student. Figures 5 and 6 show an example to illustrate this interface for playing the chosen musical instrument using hand gestures. It is clear how these two different gestures can be distinguished by the model. Multithreading, a concept of operating systems is used to implement hand tracking and gesture recognition in parallel with the sound generation by creating thread pools to avoid latency between the model classification and sound of the detected music note playing. This was a breakthrough in the design to offer a seamless experience to the user. Sound generation is done using SCAMP package.

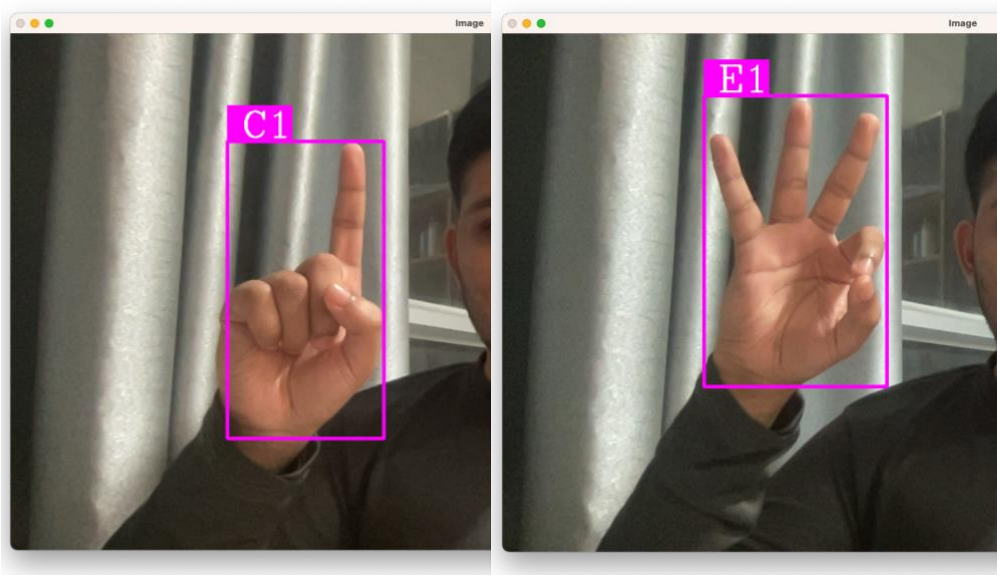

Figure 5 and 6: Hand Gesture identified as the note C1 and E1 by the CNN Classification Model.

4. Analysis of Results and Conclusion

This human-computer interaction design for hand gesture-based music generation (inspired from American Sign Language) has great potential to help primary school students learn more about musical instruments effectively and intrigue them with exciting new technology like hand tracking. The pros include helping them take more interest and actively participate in class to break free from the monotony of classroom study with the help of this real-life application that helps in the holistic growth and development of the individual. Most importantly this design also uses ChatGPT to show how AI can making a great impact in a positive way with a human-centered approach to provide a seamless and enjoyable user experience. The best part about this design is how students at a tender age can get a taste of how these musical instruments sound by playing them with minimal dexterity with their hands. However, the cons of this project are that the implementation pipeline could use some more refining with perhaps some mutex locks or semaphores (concepts of operating systems) for code synchronization before the app and the model are ready for deployment.

The proposed idea extends to ***thousands of instruments*** young students can learn about and play. In conclusion, this project has the potential to bring about a measurable impact on society by directly influencing classroom learning education.

References

Leonard, J., & Ng, K. (2011, August). Music via Motion: A distributed framework for interactive multimedia performance. In *DMS* (pp. 7-8).

Binh, N. D., Shuichi, E., & Ejima, T. (2005). Real-time hand tracking and gesture recognition system. Proc. GVIP, 19-21

Zhang, F., Bazarevsky, V., Vakunov, A., Tkachenka, A., Sung, G., Chang, C. L., & Grundmann, M. (2020). Mediapipe hands: On-device real-time hand tracking. arXiv preprint arXiv:2006.10214.

Hürst, W., & Van Wezel, C. (2013). Gesture-based interaction via finger tracking for mobile augmented reality. Multimedia Tools and Applications, 62, 233-258.

Qadir, J. (2022). Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education.

Zakirova, A. A., Ganiev, B. A., & Mullin, R. I. (2015, November). Using virtual reality technology and hand tracking technology to create software for training surgical skills in 3D game. In AIP Conference Proceedings (Vol. 1688, No. 1, p. 040011). AIP Publishing LLC.

Qadir, J. (2022). Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education.