

Submission Summary

Conference Name

IEEE MYSORE SUBSECTION FLAGSHIP CONFERENCE MYSURUCON 2024

Track Name

Track-5: Technology for Education and Entertainment

Paper ID

703

Paper Title

Image to speech conversion using Machine Learning

Abstract

we present a novel approach to image-to-speech conversion using machine learning techniques. Our system integrates a Generative Image-to-text Transformer (GIT) model, which unifies various vision-language tasks such as image/video captioning and question answering. Unlike previous methods that rely on complex structures and external modules, our approach simplifies the architecture into a single image encoder and text decoder, enhancing performance through scaled-up pre-training data and model size. Notably, our GIT model achieves state-of-the-art results on multiple benchmarks, including surpassing human performance on TextCaps. Additionally, we implement text-to-speech (TTS) functionality and provide camera access, enabling direct conversion of real images to speech. This comprehensive system offers a streamlined and effective solution for image-to-speech conversion, demonstrating promising results in both research and practical applications.

Created

5/28/2024, 9:56:08 AM

Last Modified

5/28/2024, 9:56:08 AM

Authors**Aditi Sharma** (SVCE) <aditisharma@svcengg.edu.in> ✓

Ghanashree B N (SVCE) <shreeghana02@gmail.com> ✓

Mohith Gowda BM (Sri venkateshwara college of engineering)

<mohithbmgowda@gmail.com> ✓

Manoj K S (SVCE) <ksmanojmanu@gmail.com> ✓

Padmavathy B (Sri Venkateshwara College of Engineering) <padmavathys82@gmail.com>

**Submission Files**

Image to speech__paper1.pdf (581.4 Kb, 5/28/2024, 9:55:15 AM)