Vision to Sound: Enhancing accessibility to the visually impaired

GITAM SCHOOL OF TECHNOLOGY BENGALURU

DEPARTMENT OF EECE
GITAM SCHOOL OF TECHNOLOGY BENGALURU

[Alli Gopi – BU21EECE0100161, Bhoomika N – BU21EECE0100494
Supervisor : Dr. Arvind Kumar

Abstract

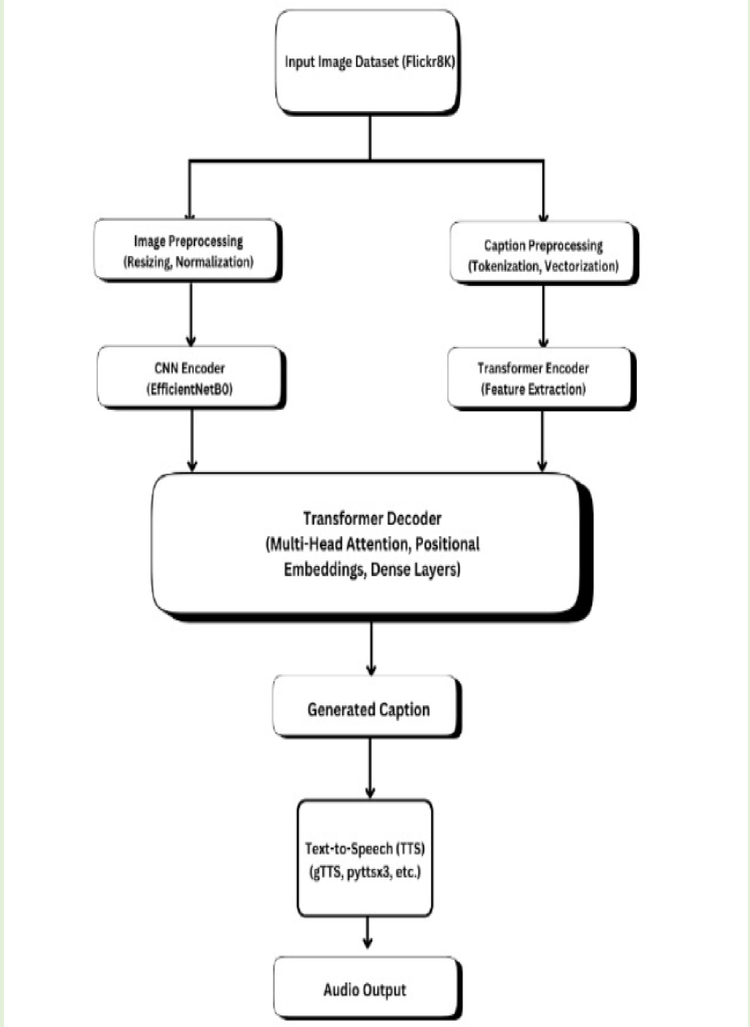
This project develops an **AI-powered image captioning system with Text-to-Speech (TTS)** to generate and vocalize image descriptions, enhancing accessibility.

Using **EfficientNetB0, a Transformer-based model, and Gotts**, it achieves **47.77% accuracy**, with future improvements in **real-time processing and multi-language support**.

Background

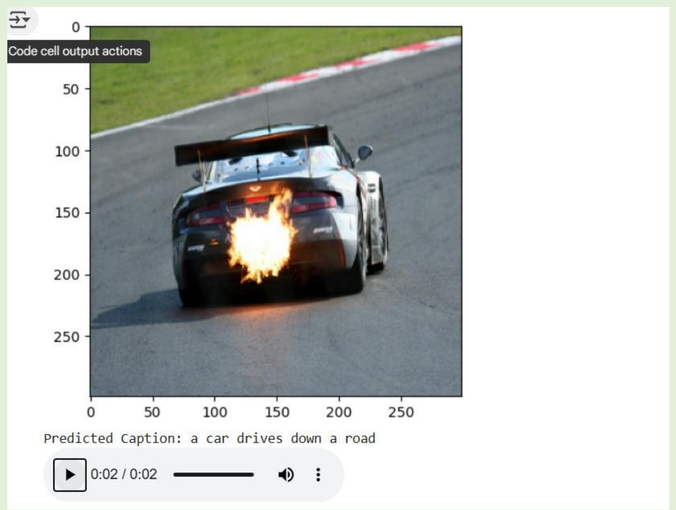
- **Evolution of Image Captioning** – Traditional models used **RNNs and LSTMs**, but they struggled with long-range dependencies and efficiency.
- **Modern Approaches – CNNs (EfficientNetB0) for feature extraction** and **Transformers for caption generation** improve accuracy and speed.
- **Accessibility Enhancement** – Integrating **gTTS for speech output** makes image descriptions more interactive and inclusive for visually impaired users.

Methods



Results

Vision to Sound



Conclusion

This project successfully integrates **AI-based image captioning with TTS**, enhancing accessibility by generating and vocalizing accurate image descriptions.

Future Perspectives

- Multi-Language Support
- Mobile & Web Deployment
- Real-Time Captioning

Impact on Society

- **Multilingual Communication** – Expanding TTS to support multiple languages will help bridge language barriers and make image descriptions accessible to a global audience.
- **Visually Impaired Support** – AI-powered image captioning with speech output enhances accessibility, allowing visually impaired individuals to understand visual content through audio descriptions.

To know more

GitHub link: https://github.com/Bhoomika-7/Vision-to-Sound-Enhancing-accessibility-for-Visually-impaired

Video link: https://drive.google.com/file/d/1ArjcjurFy5zin3JUgu4Pp7oyJLZp1fnJ/view?usp=sharing