BLIVIZ - A Blind’s Vision

VASANTH P(CSE C)
SAADHANA G(CSE C)

This technological century has evolutionized a lot because of the advent of AI in the recent past mainly due to the advancements in the field of deep learning in computer vision and natural language processing. Computer Vision is a domain which is used to solve problems related to image and videos and Natural Language Processing for text. BLIVIS uses functionalities of both Computer Vision and Natural Language Processing. BLIVIZ is an app which harnesses the power of AI and brings the vision of a blind person to reality by aiding them in getting aware of what is happening around them. This app takes image data as input and then gives a voice message as output explaining the image (i.e) generates a caption for a given image. This app consists of three main components. They are ViT(Vision Transformer) for Feature extraction from the image on the encoder part and GPT-2 for the decoder part to generate the caption. If a person cant see an image so is text for him so a 3rd Text to Speech component is included where it converts the text input to speech (voice) message. Here for Vision Transfer a model called vit-base-patch16-224-in21k released by Google is used. These models are open-sourced Transformer models from the library called HuggingFace where these models are implemented in various deep learning frameworks but here specifically Pytorch deep learning framework is used and for User Interface streamlit library is used for rapid prototyping. The app flow is as follows: First the image is got as input from which the ViT model extracts the features from it and once it is extracted based on the embeddings(features) generated GPT2 model generates a caption for the specific image which is inturn fed as input for the third component which outputs a voice message reciting the caption generated. This app can be extended by making the model to work for videos and live feed followed by a Translation component for more user convenience.

BLIVIZ - A Blind’s Vision

VASANTH P(CSE C)
SAADHANA G(CSE C)

HARDWARE REQUIREMENTS:

- Laptop or Desktop

SOFTWARE REQUIREMENTS:

- Python
- Pytorch
- Transformers
- OpenCV
- PIL
- Streamlit

SUPPORTING DOCUMENTS:

1. https://arxiv.org/pdf/1502.03044v3.pdf
2. https://openaccess.thecvf.com/content/ACCV2020/papers/He_Image_Captioning_through_Image_Transformer_ACCV_2020_paper.pdf