Lab Report

Title: 3D Image Composer Using Deep Learning and OpenCV

Introduction

The goal of this project was to develop a tool that segments a person from an input image, incorporates the segmented person into stereo images at various depth levels, and creates an anaglyph 3D image. This tool was implemented using a pre-trained deep learning model for segmentation, image manipulation techniques from OpenCV, and an interactive Gradio interface for user interaction.

The main accomplishments include:

1. Segmenting a person from an input image using DeepLabV3.
2. Overlaying the segmented person into stereo images with customizable depth levels.
3. Generating an anaglyph image for viewing in 3D with red-cyan glasses.
4. Building a user-friendly Gradio interface to integrate these functionalities.

Methodology

Step 1: Segmentation with DeepLabV3

- **Objective:** Extract the foreground (person) from the input image.
- **Approach:**
    - A pre-trained DeepLabV3 model with a ResNet-101 backbone was used for semantic segmentation.
    - Input images were preprocessed (resized, normalized, and converted to tensors).
    - The model output a segmentation mask, which was resized to match the original image size.
    - The segmentation mask was used to create an RGBA image with the segmented person and a transparent background.

Step 2: Overlaying the Segmented Person

- **Objective:** Integrate the segmented person into stereo images at varying depths.
- **Approach:**

- OpenCV functions were used to overlay the segmented person onto left and right stereo images.
- Depth was controlled by varying the horizontal disparity between the left and right positions.
- Three depth levels (close, medium, far) were implemented, each corresponding to a specific disparity value.

Step 3: Creating an Anaglyph Image

- **Objective:** Combine the stereo images into a single anaglyph 3D image.
- **Approach:**
  - Color channels were extracted from the left (red) and right (green and blue) images.
  - These channels were merged to create an anaglyph image.
  - The result was saved as a PNG file, viewable in 3D using red-cyan glasses.

Step 4: Building the Gradio Interface

- **Objective:** Make the tool accessible and interactive for users.
- **Approach:**
  - Gradio was used to build a web interface where users can upload an image, select a depth level, and view the resulting 3D image.
  - Outputs include the segmented image, stereo images, and the final anaglyph image.

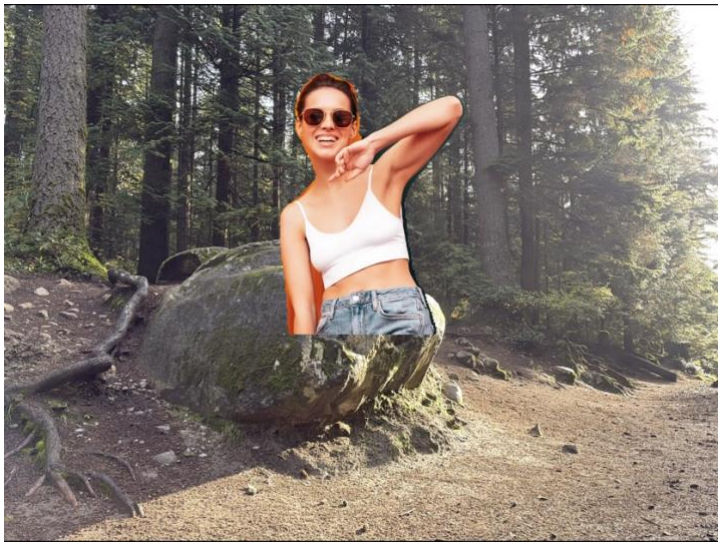
Results

Segmentation Output

- Successfully segmented the person from various input images with a clear mask and transparency.
- Example screenshot:

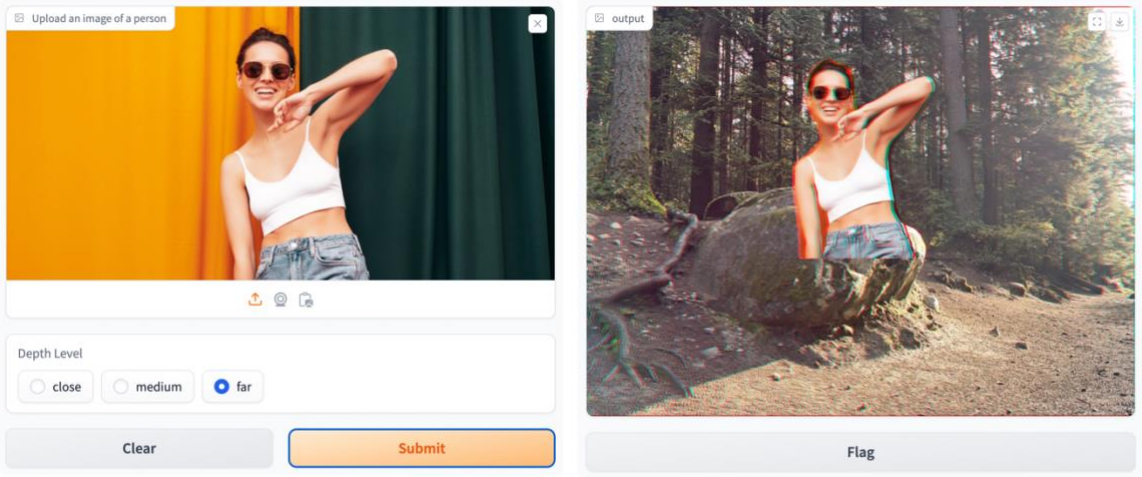
Stereo Image Outputs

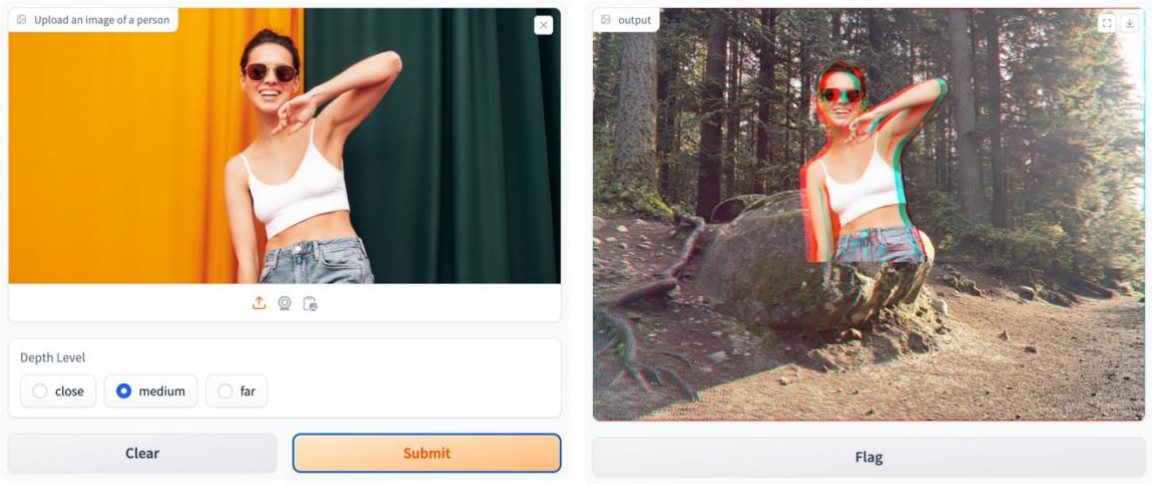
- The segmented person was inserted into the stereo images at the specified depth levels.
- Depth adjustment effectively changed the perceived 3D effect.
- Example screenshots:
  - Left Image: 

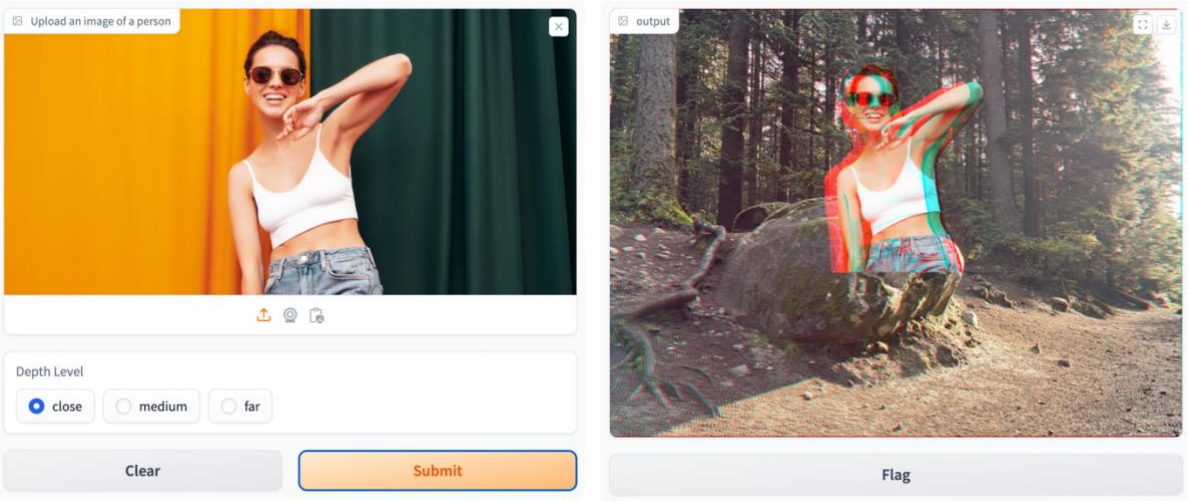
- Right Image: 

Anaglyph Image Output

- The generated anaglyph images were visually coherent, with a clear 3D effect when viewed with red-cyan glasses.
- Example screenshot:







Discussion

Challenges Faced

1. **Segmentation Precision:** The DeepLabV3 model sometimes failed to perfectly delineate the person, especially with cluttered backgrounds or low-resolution images.
    - **Solution:** Experimented with image resizing and preprocessing to improve the segmentation results.
2. **Overlay Artifacts:** Aligning the segmented person onto stereo images introduced minor blending artifacts at edges.
    - **Solution:** Applied nearest-neighbor interpolation during resizing and added a mask to control transparency.

3. **Depth Perception Calibration:** Finding the optimal disparity values for realistic depth perception was non-trivial.
    - **Solution:** Iteratively tested and fine-tuned the disparity values.

Lessons Learned

- Pre-trained models like DeepLabV3 can be effectively adapted for real-world tasks with minimal fine-tuning.
- OpenCV's flexibility enables complex image manipulations, but precise control over transformations is essential for high-quality results.
- Gradio provides a robust platform for deploying interactive AI tools, streamlining the development-to-deployment pipeline.

Conclusion

This project successfully demonstrated the use of deep learning and image processing techniques to create a 3D image composer. The tool allows users to:

1. Segment a person from an image.
2. Insert the segmented person into stereo images with customizable depth.
3. Generate an anaglyph image for 3D viewing.

Future Work

- **Enhance Segmentation:** Fine-tune the DeepLabV3 model with custom datasets for improved segmentation accuracy.
- **Dynamic Positioning:** Allow users to specify insertion positions interactively.

This project highlights the synergy between deep learning, traditional computer vision, and user-friendly interfaces, paving the way for practical applications in fields like AR/VR and creative media.