Indian Institute of Science Education and Research Bhopal

Computer Vision (DSE/EECS-312)

Bonus Assignment-1

Deadline: 14-10-2024, 11:59 PM Max mark: 10

Please follow the instructions carefully.

1. **Plagiarism and copying from anywhere (similar submission) can debar you from this course and invite the academic dishonesty policy.**
2. **The length of the video should be between 3-5 minutes.**
3. **The video should not be abusive and/or harmful in any sense to anyone.**
4. **The purpose is to learn the recent advancements of AI in the Computer Vision world only. Therefore, no personal interest will be entertained.**

Submission:

1. Video
2. A 1-2 page report only in the given overleaf template explaining the topic of your video, the crux, an explanation of the tool used in both questions, and its link.

1. Using recent advancements in generative AI and computer vision technologies, create a simulated debate video featuring two prominent historical or contemporary figures. The debate should center around a critical and timely issue, such as climate change, technological ethics, international relations, or social justice. Additionally, consider the ethical implications and technical challenges of creating realistic AI-generated representations of real individuals.

   Here is one example: https://www.linkedin.com/feed/update/urn:li:activity:7118570023987683328?updateEntityUrn=urn%3Ali%3Afs_feedUpdate%3A%28V2%2Curn%3Ali%3Aactivity%3A7118570023987683328%29

   Answer:
   Name: Shrey Srivastava
   Roll No.: 22303
   Topic of Video:
   In this video, Rohit Sharma (Indian Cricket Team Captain) and Gautam Gambhir (coach of the Indian Cricket Team) are discussing the selection of the playing XI for the upcoming World Cup in 2027.

   Famous Personalities in the Video:
   Rohit Sharma and Gautam Gambhir

   Steps and Tools Used to Generate the Video:

   - Firstly, I generated the conversation between Rohit Sharma and Gautam Gambhir using ChatGPT.

- Then, I separated the text of Rohit Sharma and Gautam Gambhir.
- After that, I generated the speech of Rohit Sharma and Gautam Gambhir using the tool ElevenLabs.
- Then, I generated AI avatars of Rohit Sharma and Gautam Gambhir, and provided the audio to create an AI-generated lip-sync video using D-ID. It worked, but it doesn't support famous personalities' images; that's why I had to create AI avatars of them. ("Although I can't get a very matching avatar of Rohit Sharma and Gautam Gambhir because they are highly famous personalities and there are significant restrictions on creating AI avatars of them, I obtained some look-alike avatars that I used in my video.")
- I also tried to include the original voices of Rohit Sharma and Gautam Gambhir in the video, but I couldn't find a download option for their original voice.
- Now, I have separate videos of Rohit Sharma and Gautam Gambhir. Using VEED, I concatenated these two videos side by side.
- To show which tools I used, I did not erase the watermark, which I can remove using 123apps.

Explanations of the Tools:

- **ChatGPT:** ChatGPT is an advanced AI language model developed by OpenAI that can generate human-like text based on prompts.
- **ElevenLabs:** ElevenLabs specializes in AI voice synthesis, providing high-quality text-to-speech capabilities that closely mimic human voice characteristics.
- **D-ID:** D-ID offers AI-driven tools for generating realistic, lip-synced videos from audio input. Its technology creates animated avatars that can simulate facial expressions and lip movements, making it useful for creating engaging video content.
- **VEED:** VEED is an online video editing platform that allows users to create, edit, and share videos easily. It includes features such as video merging, subtitle generation, and various effects, catering to content creators and businesses.
- **123apps:** 123apps provides a suite of online tools for various media tasks, including audio and video editing, conversion, and compression.

2. Explore the various open-source toolboxes and applications used in the field of computer vision and come up with some interesting applications or toolboxes related to the Computer Vision field (Inpainting, Pose Estimation, DeepFakes (Image, Audio, or Video), Adversarial Example Generation, Style Transfer). Select any toolboxes or applications and use them to test their effectiveness in solving the specific problem you are addressing, and discuss their core functionalities, advantages, and real-world applications.

Answer:
1. Inpainting:

- **Toolbox/Application:** LaMa by SAIC AI

- **Core Functionalities:** Deep learning-based tool for filling large missing regions in images using CNNs.
- **Advantages:** Handles complex textures, uses attention mechanisms, open-source.
- **Real-World Applications:** Photo restoration, object removal, image smoothing.
- **Testing Effectiveness:** Maintains image context while filling gaps, ideal for restoration.

2. Pose Estimation:

- **Toolbox/Application:** OpenPose by CMU Perceptual Computing Lab
- **Core Functionalities:** Detects keypoints for body, hand, and face using a bottom-up approach.
- **Advantages:** Real-time, high accuracy, pre-trained models.
- **Real-World Applications:** Sports analysis, VR/AR, gesture recognition.
- **Testing Effectiveness:** Accurate tracking of body movements for various applications.

3. DeepFakes:

- **Toolbox/Application:** DeepFaceLab
- **Core Functionalities:** Deep learning framework for face swapping in videos.
- **Advantages:** High-quality results, customizable models, large community support.
- **Real-World Applications:** Movie production, meme creation, and educational content.
- **Testing Effectiveness:** Achieves realistic face swaps; raises ethical concerns.