Facial Deocculusion - Remove objects covering your face

1. **Domain Background**
   Facial deocclusion is the process of reconstructing occluded (covered) parts of a face in an image using AI techniques. This task holds significant value in various real-world scenarios, such as presentations, livestreaming, and security. By enhancing visibility of facial features, this technology can improve user experiences in situations where parts of the face may be obscured by objects like masks, sunglasses, hands, microphones, or cups. It aligns with existing innovations like Eyes Focus, aiming to bring clearer, more engaging interactions during live or recorded video feeds. Additionally, with further refinement, it has potential applications in security for facial recognition even under partial occlusions.
2. **Problem Statement**
   The goal of this project is to develop an AI-based system capable of removing occlusions from facial images, thereby reconstructing hidden facial parts as naturally and accurately as possible. The project aims to create a model that can restore facial features by intelligently filling in areas obscured by various objects. This would have practical applications in ensuring seamless presentations, livestreams, and video communications, especially when external elements partially cover the face.
3. **Datasets and Inputs**
   The dataset for this project will consist of images of faces with various objects partially covering them, such as masks, sunglasses, hands, microphones, and other accessories. Given the time constraints, the data will be sourced based on availability, and there may be a mix of synthetic and real-world images to ensure the diversity of occlusions. The dataset will include labeled data, where each image has a corresponding mask identifying the occluded regions, which will help in the supervised training of the model.
   Regarding the dataset, I will build my own dataset by using an existing dataset of human faces (Here some datasets I find out: https://github.com/NVlabs/ffhq-dataset, https://github.com/cabani/MaskedFace-Net, https://paperswithcode.com/dataset/celeba-hq). Then, I will use MediaPipe or any model like Yolo for face detection to overlay a transparent image of an object onto the face area (I plan to use transparent images of masks and glasses, depending on the variety I can collect). This will create pairs of original images and images with obscured faces. The Gen model will then attempt to reconstruct the obscured part to match the original image as closely as possible. The original and object-added images will be fed into a Discriminator model to evaluate and improve during the GAN training process.
4. **Solution Statement**
   The proposed solution involves leveraging Generative Adversarial Networks (GANs) with architectures similar to U-Net to reconstruct occluded parts of the face. GANs are well-suited for this problem due to their ability to generate realistic images, making them an ideal choice for tasks that require high-quality image restoration. The model will be trained to understand patterns in facial features and learn to fill in occluded areas based on context, resulting in more natural and accurate reconstructions.
   The key steps in the solution will include:

- Preprocessing the dataset to identify and mask occluded regions.
- Training the GAN-based model on this dataset to learn how to reconstruct these regions.
- Validating the model against a test set to evaluate its performance in reconstructing occluded parts.
- Fine-tuning the model to enhance its ability to handle various types of occlusions and ensure realistic outcomes.

5. **Benchmark Model**
As a benchmark, a U-Net with Skip Connections and Encoder-Decoder Architectures: This architecture helps in retaining essential features from different levels of the image, which is particularly useful for reconstructing occluded parts.
6. **Evaluation Metrics**
The project will primarily evaluate the model's performance using metrics such as Mean Squared Error (MSE), Structural Similarity Index (SSIM). SSIM will help assess the perceptual quality of the deoccluded images, while MSE will provide a more quantitative analysis of pixel-level accuracy. A higher SSIM, coupled with a lower MSE, will indicate better performance in restoring the occluded facial parts.
7. **Project Design**
The project will follow a structured approach:
    - **Data Collection & Preparation**: Compile a diverse dataset of facial images with varying degrees of occlusion. This step will involve data augmentation to ensure robustness.
    - **& Training**: Start by training a simple GAN model, then experiment with more complex variations, such as Conditional GANs or Pix2Pix, which use U-Net architectures to handle context-aware image inpainting
    - **Define Model Architecture and Traning**: Train a GAN, where the Discriminator acts as a classifier to distinguish between real and generated images. The Generator will use a U-Net-like architecture, leveraging skip connections for effective image restoration. Experiment with Generator models with and without attention modules to test if attention improves focus on occluded areas. Implement the GAN using frameworks like **PyTorch** or **TensorFlow**, defining custom loss functions for **adversarial** and **reconstruction loss**.
    - **Testing & Validation**: Measure the model's performance on unseen data to evaluate its generalization capabilities. Adjust hyperparameters as needed.
    - **Integration & Application**: Integrate the trained model into a prototype application that can be used in real-world scenarios like video calls or livestreams, demonstrating the model's effectiveness in reconstructing facial features in real-time.
    - **Deployment**: Explore potential deployment options, such as cloud-based services using AWS, to ensure scalability and efficiency.