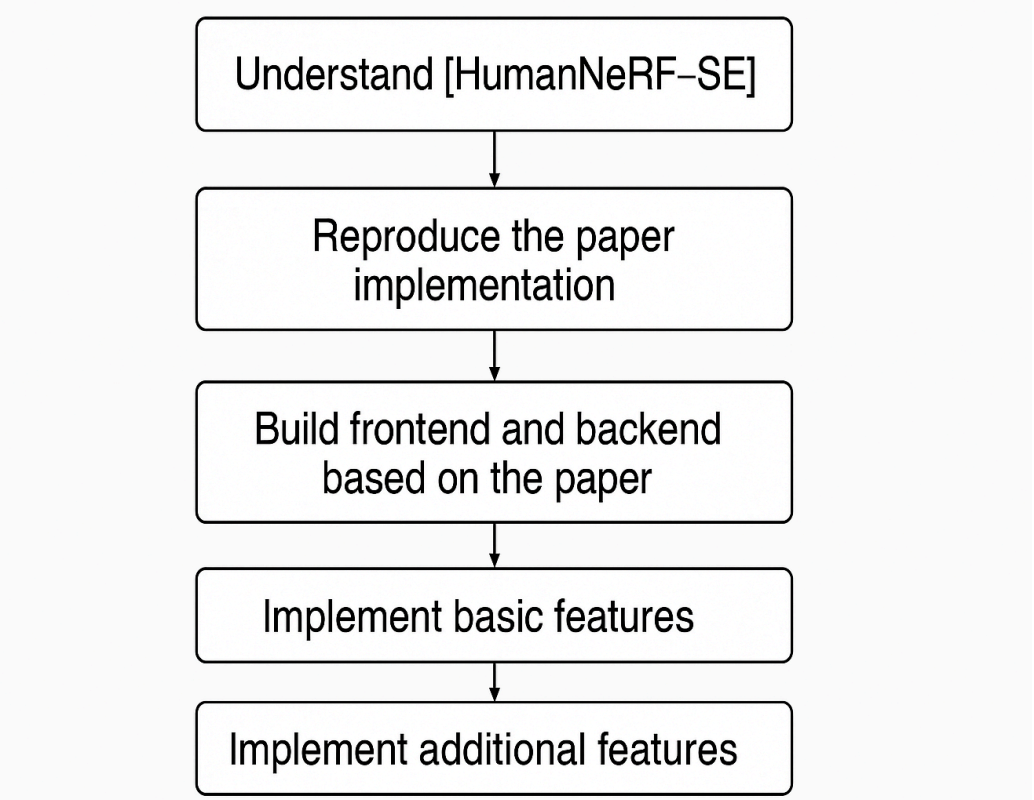
Final Project Proposal

1. Selected Project Topic

Monocular Motion Transfer with NeRF Avatars
We plan to implement a system where users can upload a video and a reference character image, then generate a new video in which the character replaces the original person's appearance while following their motion.

2. Overall Flowchart / Block Diagram



3. Techniques Planned to Use (brief description)

- **Pose and Shape Estimation**: Using ROMP to extract 3D human pose and shape from monocular input video.

- **NeRF Avatar Rendering**: Following methods similar to HumanNeRF-SE paper, creating or reusing a NeRF-based avatar from the target image.

- **Background Replacement**: Using segmentation techniques such as **MediaPipe Selfie Segmentation** or **Mask R-CNN** to separate the background for optional replacement.

- **Person Tracking**: Use Appearance matching (Visual Embedding Matching, from DeepSORT, FairMOT), Location Matching (IoU-based Matching, from DeepSORT), Shape Consistency (SMPL matching, from NeuralBody, MonoHuman)

- **Web Frontend and Backend**:

    - Frontend: React + TailwindCSS.

    - Backend: FastAPI for video/image processing and serving models.

4. Dataset

- People Snapshot Dataset:The People-Snapshot dataset contains 24 monocular video sequences of 11 individuals slowly rotating in an A-pose, captured to enable 3D human model reconstruction from video. It was introduced by Alldieck et al. for research on recovering detailed 3D models, including clothing and hair, using only RGB input.

- Users' uploaded videos and character images.

5. Planned Outcome

Basic features:

- Web-based system (accessible through browsers)

- Allow users to upload a video (input video with a person)

- Allow users to upload a character image (replacement avatar)
- Detect and track the person in the input video
- Replace the person with the uploaded character while preserving the original motion
- Generate and provide a new video as output

Additional features:

- Allow users to select which person to replace if multiple people appear in the input video.
- Allow background replacement.

Demo & Evaluation:

- Visual demo: side-by-side comparison of original and synthesized videos.
- Measure rendering time per frame as a quantitative metric.