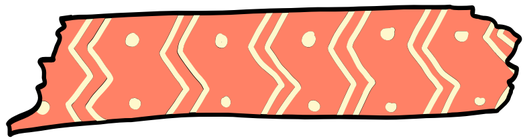
Image-Driven Facial Animation with PaddleGAN: A Web-Based System and Comparative Study on Face Angles and Styles

Team : B6
111550040 曾紹幃
111550143 林彥佑
111550149 林悅揚

Motivation

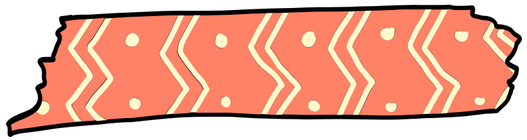
- We built a lightweight animation system that brings static images—especially faces—to life using a driving video.
- Our approach is fast, efficient, and easy to use, requiring only a single image and a short video clip.
- We developed a user-friendly website where users can:
  - Upload a reference image and driving video
  - Generate a motion-transferred animation
  - Preview the result and download the final output



Techniques

We adopted **PaddleGAN's First-Order Motion Model**, which animates a source image by transferring motion from a driving video. It leverages:

- Keypoint detection: Identifies essential facial/body landmarks.
- Motion estimation: Predicts how the source image should deform, including spatial warping that mimics 3D-like transformations and introduces subtle 3D-aware dynamics.
- Occlusion handling: Maintains coherence even when parts become hidden.



Tools

Frontend

- React + Vite: Fast and responsive user interface development
- TypeScript: Improved code maintainability and type safety
- Axios: Handles API communication with the backend
- Tailwind CSS: Utility-first framework for clean and responsive UI design

Backend

- FastAPI: Lightweight and high-performance Python web framework for handling uploads, processing, and animation generation
- FFmpeg: Used for video decoding and frame extraction

Experiment : Frontal/Side Face/Animation

Since the First-Order Motion Model lacks objective evaluation metrics, we conducted a qualitative evaluation based on human visual judgment.
We tested 9 combinations using 3 types of source images and 3 types of driving videos:

Source Images

1. Frontal face
2. Side face
3. Animated face

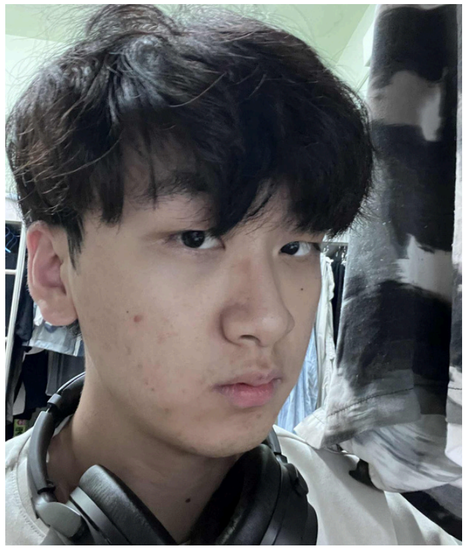
Driving Videos

1. Frontal face with motion
2. Side face with motion
3. Animated face video

Animation

Frontal

Side



Result:Animation image

Video type:

Animation

Frontal

Side

Result:Front image

Video type:

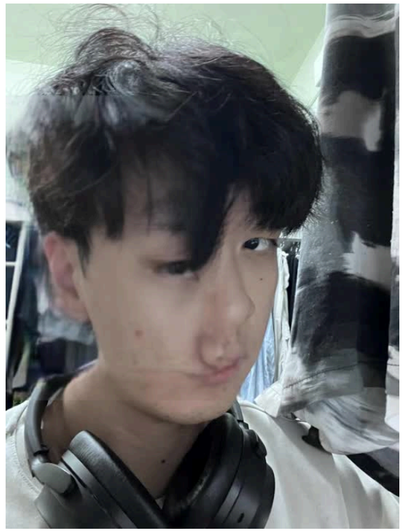
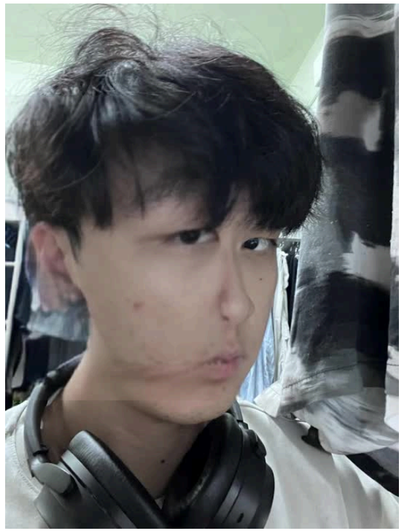
Animation

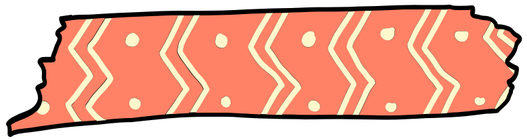
Frontal

Side

Result:Side image

Video type:

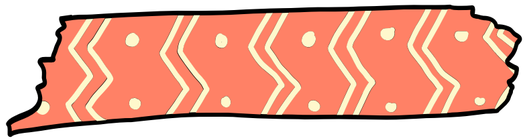
Animation	Frontal	Side
		



Analysis

Through visual inspection of 9 input combinations, we observed the following:

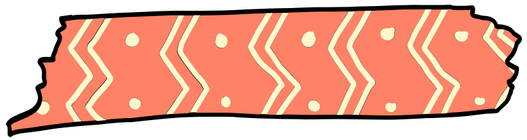
- Frontal image + frontal driving video delivers the most stable and realistic results
- Animated images (e.g., illustrations) perform surprisingly well, nearly matching real frontal photos
- Animated driving videos often produce poor motion transfer, lacking natural movement
- Side-view images or videos consistently yield distorted or unstable outputs, showing clear limitations in PaddleGAN’s handling of non-frontal angles



Futrue work

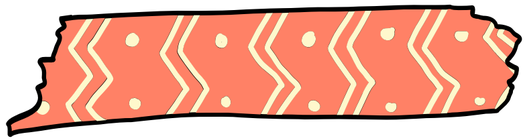
To improve generalizability and animation quality, we propose the following directions:

- Enhance 3D spatial awareness to better handle side views and depth variation
- Improve occlusion handling and keypoint stability to reduce facial distortion and jitter
- Introduce style-adaptive training to support cartoon or illustration-style inputs more effectively
- Add interactive features like motion segment selection or background replacement, to boost user control and creativity



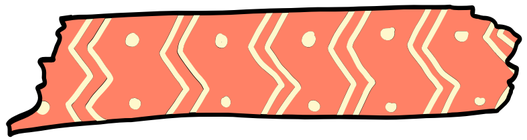
Conclusion

These findings suggest that PaddleGAN’s First-Order Motion Model is highly sensitive to input alignment and viewpoint, performing best with frontal, real-world inputs. Improving performance for side views or stylized inputs may require better 3D awareness and occlusion handling.



What we have learned

- We observed that the alignment and viewpoint of both the source image and the driving video significantly impact animation quality. Frontal faces yield the most stable and natural results
- The First-Order Motion Model struggles with side-view images and highly stylized or animated inputs, often producing distortions or unstable motion.
- The absence of standardized quantitative metrics for animation accuracy needs human-centric evaluation. This highlights an open problem in benchmarking generative motion transfer models.



DEMO

Thank's For Watching