Fluidie

(Using Pix2Pix for real-time fluid simulation to create fun camera effects)

Shengzhi Wu, Mimi Wang

PROJECT 4 - Fludie
Shengzhi Wu, Mimi Wang

Concept:
Fluidie explores the artistic expression of selfie into a constantly flowing and changing status, questioning the ephemerality of the concept of "self" and "ego." By leveraging machine learning for fluid simulation, the project transfers a selfie photo into a real-time generated video that demonstrates how our self-portraits morph, reconstruct, and dissolve over time.

The traditional fluid simulation requires algorithms that can reflect the real world fluid physics to obtain realistic simulation, and it requires a lot of domain knowledge. On the contrary, our approach can take any fluid video (e.g., ink in water downloaded from YouTube) as the training dataset, and use Pix2Pix GAN to predict the next frame from a given image. Subsequently, the machine-generated frame will become the new input image to predict the next frame. As a result, we will be able to achieve a self-sufficient loop for infinity simulation.

Technique:
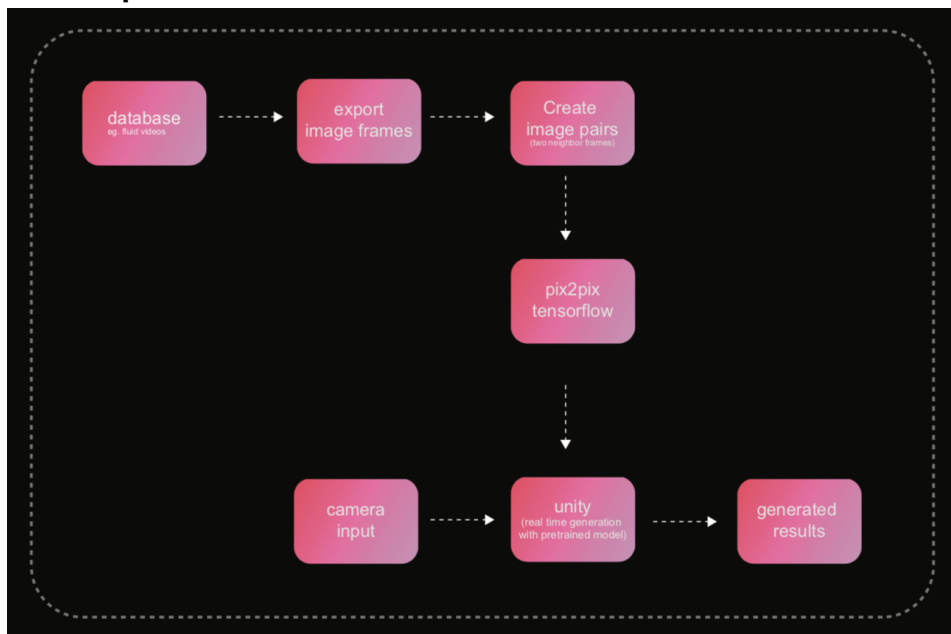
Our technique focuses on fluid simulation using Pix2Pix, and we've tried various fluid dataset to achieve the best results. Additionally, our pipeline allows us to generate the simulation in real-time with Nvidia GTX 1070 or better with a framerate over 35 FPS. Audiences can also interact with the application through the camera to generate results in real time.
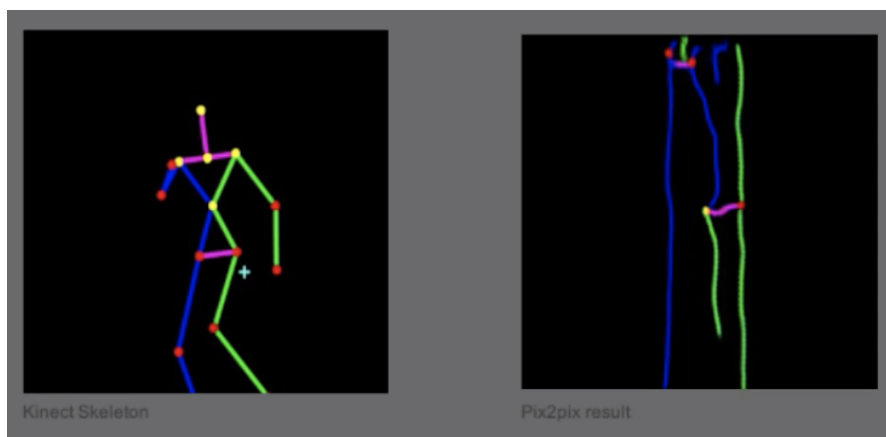
Process:

PROJECT 4 - Fludie
Shengzhi Wu, Mimi Wang
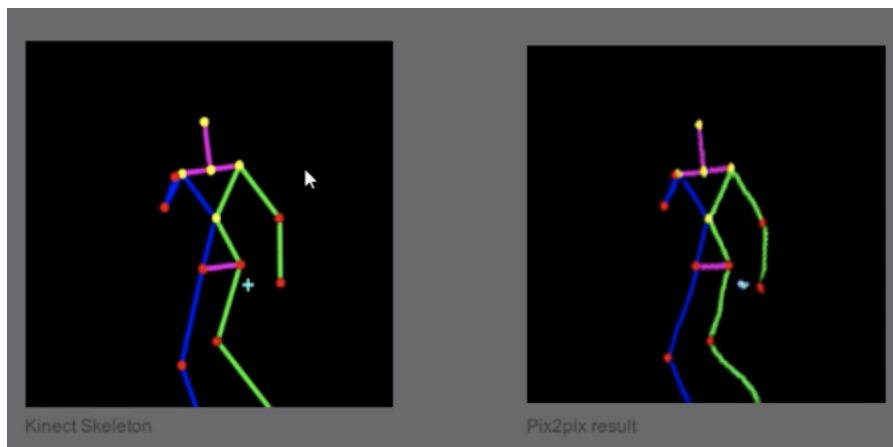
Our initial goal was to use Pix2pix to predict human dancing, and we collected our own dataset using a Kinect sensor to capture the skeleton of dancing. The predicted results are a little bit distorted. Then after using the generated image to predict the next frames. The result soon significantly distorted and no longer looks like a human being (please see a failure demo here https://drive.google.com/file/d/16cZuCMqvFbFd1XZNZTdYYYKT9-JrltDL/view?usp=sharing). So we believe this next frame prediction is better for something for abstract, and it is why we pivoted to fluid simulation.

For the fluid simulation, we collected some data videos from Youtube and transferred the video into image sequences. And we used processing to put the images from A to B together. We used the generated results to train the data using Pix2pix tensorflow. After training, we exported weight data (pre-trained model) and imported it into Unity. We trained with over 10 different styles, including ink in water, fires, jellyfish, color explosion, tunnel, etc. Some results are more interesting, while some looks a little boring. Eventually, we choose 4 different styles to show in the exhibition.

ART AND MACHINE LEARNING, CMU 2019 SPRING
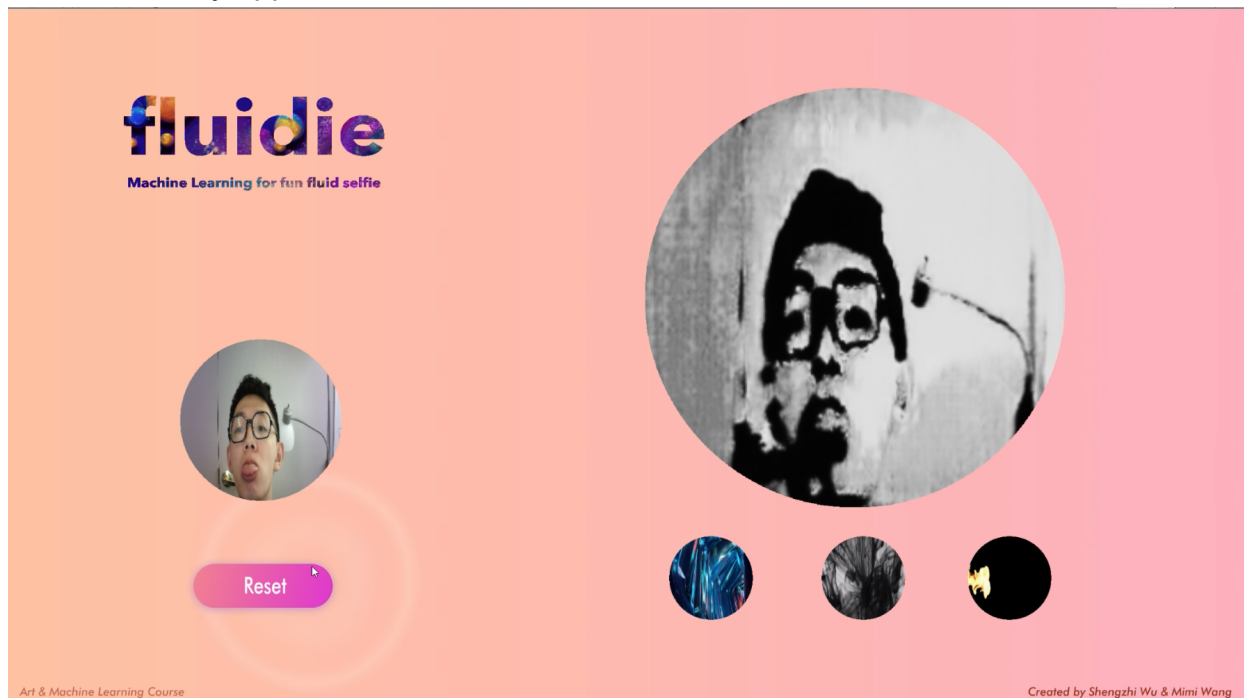**PROJECT 4 - Fludie**
Shengzhi Wu, Mimi Wang

To make it more interactive for the exhibition, we build an interactive demo in Unity. The demo uses a camera as input. Once a user presses the capture button, the captured image becomes the first frame and using it to predict the next frame based on 4 different trained styles.

Result:
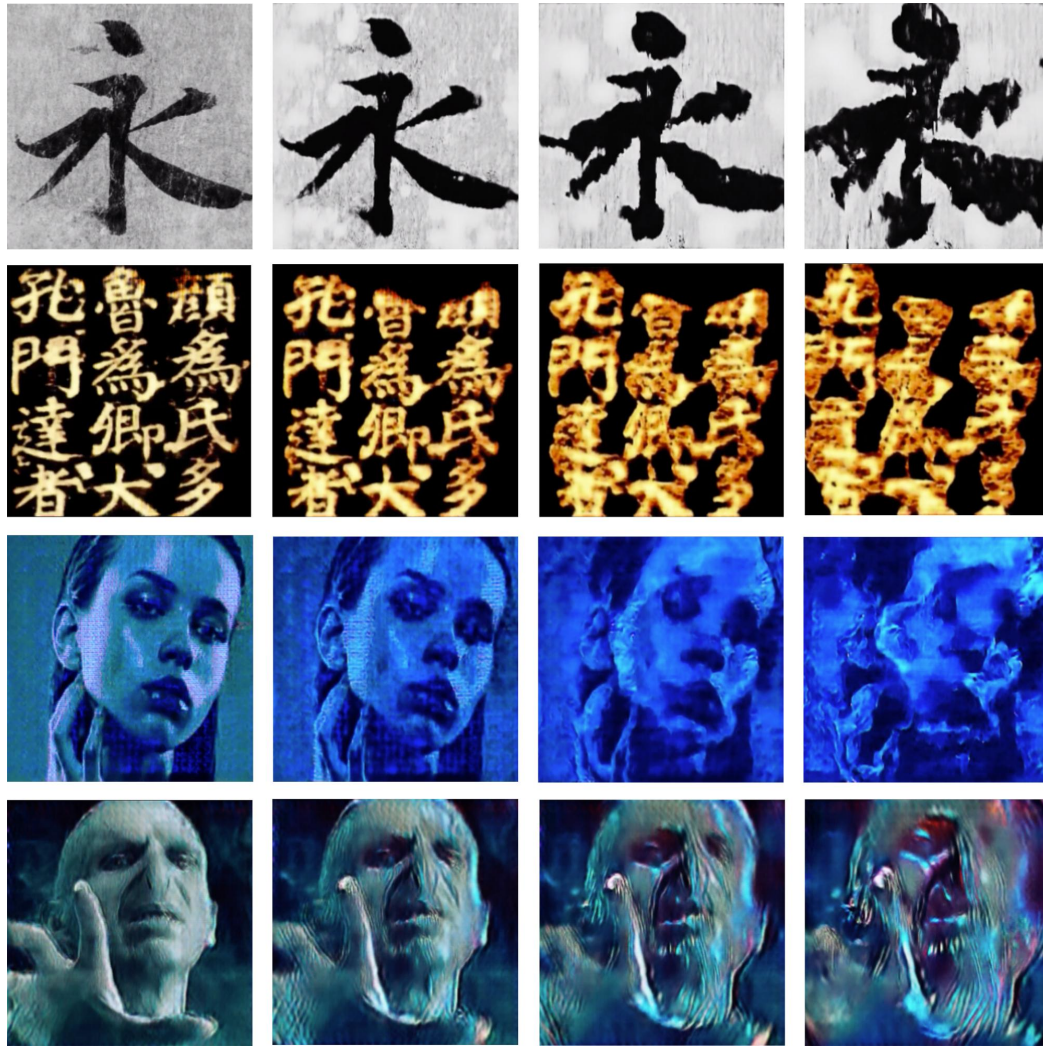Our demo video: https://youtu.be/5EjXgIV9H_w
Real-time Unity application:

A screenshot of the demo

PROJECT 4 - Fludie
Shengzhi Wu, Mimi Wang

Reflection:

First, our concept is to predict people's next frame of dancing movements using machine learning methods to explore new possible dance styles. We wanted to use Kinect open pose to extract human's pose and using pix2pix to predict the next frame of the human's body movement. After we put the training model into Unity, the result is not as good as we expected. The probable reason is that every frame is predicted using the previous trained frame so that it kept losing quality on and on. In the end, we can't distinguish the humanoid at all.

Additionally, without the first failure, we further discovered the result had a very smooth fluid effect so we pivoted our idea of using Pix2Pix for real-time fluid simulation prediction to create fun camera effects. The method is very effective for generating abstract content.

At last, we also realize that this method requires the dataset to have a consistent color tone and texture style. Through our training process, we had a lot of failures, those failures usually either have too much color or texture in it, or the frames are not very consistent. Eventually, we selected four styles that have the best effects(ink, burning, tunnel, and jellyfish). We learned that the best effect's original data all have a uniform color tone. And the texture is transferred very well with this method as well.

Code:
Processing code to pair images A to B
https://drive.google.com/drive/folders/1OQBAs-HjazKDB0weM7Px1M60-Eoz0rqS?usp=sharing

Tensorflow Pix2pix training
https://github.com/affinelayer/pix2pix-tensorflow

Unity real-time Pix2pix generation
https://drive.google.com/drive/folders/1sR3ykNYv6dpzUJZsT8tNuwGST3LSblBQ?usp=sharing

Reference:
[1] Pix2Pix Tensorflow
https://github.com/affinelayer/pix2pix-tensorflow

[2] COLORS – Experimental Video by Thomas Blanchard Using Colorful Liquids (training data source)
https://www.youtube.com/watch?v=VjnTS7JphVE&list=PLehuHqXGJ8KzgLhQ9HywsX7jN1go8Eq6s&index=4

[3] Color Explosion (training data source)
https://www.youtube.com/watch?v=aSrgaclHsZM

[4]4K Color Explosion on WHITE Background (training data source)
https://www.youtube.com/watch?v=ZosfEUpi1Xc

[5]4K Ink Drop/Drip in water Black_01 - Free Stock Footage(training data source)
https://www.youtube.com/watch?v=vOxmsqd8f58(data images)

PROJECT 4 - Fludie
Shengzhi Wu, Mimi Wang

[6]Unity Pix2pix
https://github.com/keijiro/Pix2Pix