When I was 14, I fell in love with visual arts. I greedily consumed the great works of cinema, TV series, storytelling video games and VR shorts. I believed visual art was the best way to reach to our insides and bring us together. Then, I entered Pomona College as a film studies major. I am grateful for the freedom I had in here in exploring my interest. As I delved deeper into the field, I gradually realized the importance of technology on the development of visual art. As I saw the visual effects in movies like *The Curious Case of Benjamin Button* and *Avengers*, I was attracted by how graphics technology could be waived into storytelling. I realized that technology and art are related rather than separated. Thus, I wished to explore new ways to enrich storytelling with graphics.

My exploration in academics was not smooth. I first looked into computer graphics, and yet there was no professor in my college who led projects in this field. I then self-taught machine learning and computer vision, with the help of knowledge in probability, advanced linear algebra and differential equations learned at college. As I learned the CNN methods for facial recognition, I wondered if sliding a kernel across the whole image was the best approach for this task. I then started a project in facial recognition with professor Weiqing Gu at Harvey Mudd College. I proposed to use a joint network combining the prediction of face shape recognition and CNN recognition on extracted areas. The reduced image size largely shortened the total training time of CNN, while still preserving the distance information between the extracted areas. Our goal was to maintain the precision of the model while reducing the time consumption in training.

To accelerate the progress, I registered the project in P-AI, a college-wise machine learning club, and recruited three students through the organization. As the group leader, I led the research direction, held group discussions and distributed research works. Based on what we learned, we implemented the distance recognition with Gaussian model clustering. We concluded that the precision could be improved if we used a neural network instead of gaussian clustering to classify the shape data. Although it did not reach the initial project goals, I learned a lot about machine learning during the procedure, and this experience paved the way for my future research.

Earlier in this year, I read about the exciting work of NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. I was impressed by the high-quality renderings from the trained model, as well as the elegance of this pipeline. Nonetheless, the training on each scene took 1-2 days on a high-end GPU. It would be too long for interactive 3D graphics or training on a full video. I thought of accelerating the training with meta-learning, so I began to look for relevant articles. Then, I read about the paper MetaSDF, which applied a Model-Agnostic Meta-Learning (MAML) model to a neural rendering function for computer-generated scenes. It confirmed the idea that first-order meta-learning methods could have great effects on accelerating neural rendering tasks.

After I implemented MAML-NeRF, it performed only slightly better than a blank model, but became worse as the iteration increased. Hence, I went on looking for other meta-learning methods. I found that Reptile was another optimization-based meta-learning algorithm, a type of meta-learning aimed at accelerating the learning procedure. Since this method could have much more inner steps, I let Reptile-NeRF treat each scene as an inner task, and thus shuffle the rays to decrease the variance of inner training data. This approach improved the result of the original training and enabled me to achieve better result.

I experienced the whole process of research alone in this project. It was also the first time for me to write a paper as first author. To learn from the best, I carefully read through NeRF and MetaSDF quite a few times and learned their writing structures. For details, I received help from my research partner Alex Beatson, a PhD at Princeton University. My writing was also trained in my thesis project – a survey of rendering functions for neural rendering. Currently, the paper is under review for a conference. Lastly, I also learned to alleviate my frustrations at each obstacle and come back with fresh energy and hope.

I also discovered in my Reptile-NeRF project that visual technology was broader than I thought. I decided to try another emerging visual field – AR. In August, I started working with professor Sra from UCSB. Our goal was to investigate the effect of location-based memory in AR language learning. We let our participants walk around an outdoor area holding a smartphone, through which they would see word tags in the foreign language attached to the real objects.

Designing the first user study in my life was quite challenging, so I went through every detail in relevant studies and learned the HCI study principles behind them. In this way, I finished experiment setup and testing metrics.

However, several procedures had to be changed drastically. First, the machine-learning-based object recognition algorithm functioned poorly in the wild. As a result, I proposed to let researchers create the word tags as cloud anchors. In such a way, the researchers would have more control on what the participants see, and the interface would be noise-free. Another challenge is timing. Since participants in our experiment must walk around to see the words, they would expectedly take more time to view the same number of words than both the Arbis group and the flashcard group. Also, walking a long distance while holding a phone could be exhausting. To tackle with these two problems, I first proposed to restrict the walking zone in a relatively smaller area with a diameter of 40 meters; then, I broke the learning procedure into three phases, so the timing difference in each phase would be shortened; third, I let the participants walking around the area before the learning trial to accelerate their navigation. The actual experiment is postponed due to COVID, but we have all the details set at this point. We plan to conduct the user study in a college in late January, and then publish the paper in February.

I had no experience with Android development before this project, but I learned everything in a short time and was able to finish the app for experiment by myself. I am grateful for prof Sra as she guided me through the design of experiments, writing interview questions and analyzing users’ reviews. This HCI research experience is invaluable to my career.

For my PhD career, I want to help expand the possibilities of visual art. I wish to contribute to the development in cinematic special effects, animation, AR and VR. I think a promising field in the near future would be extracting, reconstructing and expanding real-life scenes in virtual settings in a higher level of precision. I am thus highly interested in Professor Linda Shapiro’s group and Professor Ira Kemelmacher-Shlizerman’s group. The PhD of Paul Allen School of UW is my top choice among all programs.