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. As a result, I 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. 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. 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, DeepSDF. It confirmed the idea that first-order meta-learning methods could have great effects on accelerating neural rendering tasks, so I decided to work on a similar approach with NeRF.

After I implemented MAML-NeRF, I found that the trained meta-model showed poor results. While training on a new task, it performed 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. By subtracting the original meta-model from the trained inner model in each step, Reptile avoided tracing the inner gradients, and thus did not have to stack up memories. Furthermore, Reptile enabled each inner task to train on a larger batch of rays per inner step. Consequently, instead of treating each image as an inner task as in MAML, Reptile-NeRF could treat each scene as an inner task, and thus shuffle the rays to decrease the variance of inner training data. Shuffling the rays from different images in one scene improved the result of the original training, and thus as I applied this approach to my meta-learning model, I achieved better result as well.

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 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. My senior advisor prof Radunskaya took great effort in teaching me to write in a mathematically rigorous way. Currently, the paper is under review for a conference. For this project, I did all the work by myself. I almost gave up at a lot of points due to frustration. Admittedly, I could have finished the project earlier if I made some decisions right. However, they were valuable lessons to me. I learned to take a break when I have a bad mood and come back later with a fresh mind. And most importantly, I learned to alleviate my frustrations and come back with energy and hope.

Another thing I discovered in my Reptile-NeRF project was that visual technology was broader than that I thought of. 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. Different from the previous work of Arbis by Ibrahim et al in 2018, where they asked each participant to learn the words of objects placed nicely on a table, we let our users learn in an outdoor setting. They should walk in an area to view and learn the words. We would then draw the conclusion based on the qualitative and quantitative results between our experiment, the Arbis AR experiment and the flashcard comparison group.

The study design was a difficult process. While we kept most of the settings in Arbis for a fair comparison, some procedures had to be changed. 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 have to walk around, 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 decided to set an equal amount of time spent in each group, instead of equalizing the number of words viewed. Furthermore, I proposed to restrict the walking zone in a relatively smaller area with diameter of 40 meters. Lastly, I added the measure of exhaustion in our interview questions to make sure that the participants’ learning experience was not affected by their tiredness. The actual experiment is postponed due to COVID. 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 also learned a lot about how to conduct user study. 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.

I learned a lot in each period of my research experience. By constantly exploring, I gained a better understanding on what was required to be a researcher. It needs a well-rounded knowledge in the field, a collaborative spirit and leadership, a tenacious mind and a constant willingness to learn new things. I have harnessed these merits in my experience, and thus I believe I will be a qualified researcher.

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.