**University of Washington: Submit a personal statement of ~1000 words (max 500KB) that includes: a) how you became interested in doing research, b) a relevant project or research experience that shows your technical knowledge and skill, and c) your plans for the future in computer science. You may wish to include information about what you feel are the strengths of your application, such as special interests and abilities, or give explanations for what you feel are any weaknesses in your academic record. If you have background that might particularly contribute to the intellectual and social enrichment of the program, please describe it. Examples include unique educational or cultural opportunities (or lack of them), social and economic disadvantages that you may have had to overcome, and interesting or unusual influences on your intellectual development.**

**-- Why I chose CS research --**

Since I was very little, I have dreamed of changing the world to a better place. This desire has only become stronger as I grew up. 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. In the first summer of high school, I attended a filmmaking course at Emerson College, and I made lot of videos and short films ever since. 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. Around the end of the sophomore year, I decided to become a researcher in technology for visual art.

**-- Beginning & eye tracking with prof Katherine Breeden --**

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 joined my first research group in the last summer. It was a group of two students, led by professor Katherine Breeden. We aimed at collecting data of people’s eye movement while watching music videos. I helped build a C++ pipeline that played music videos while reading the eye movements from Gazepoint eye tracker. In this experience, my coding skill reached to another level. I also learned the importance of presentation – how to make your idea or proposal sound more interesting. This lesson was quite useful when I asked other students to become our experiment participants. Most importantly, I learned what research was about in this experience. I fell in love with it and decided to conduct a project on my own.

**-- Facial Recognition with prof Gu, Harvey Mudd College --**

I then self-taught machine learning and computer vision, since the courses in my college were too packed. As I learned the CNN methods for facial recognition, I wondered if sliding a kernel across the whole image was the best way to do so. Psychology studies showed that when humans tried to recognize faces, we would fixate our view on the nose first, and then the eyes and the mouth (Tan et al 2016). I first wondered if we could extract these parts of the image and ran a CNN on the collage. Since the collage would be smaller in size than the original image, the training would be shortened. However, a feature of the face was lost in this extraction – the relative locations of these parts. They are important to the recognition as well, since the task could be done even with the feature alone (Ahdid et al. 2017). My idea was thus to combine the separate predictions based on distance feature and the collage image of eyes, nose and mouth to generate the final prediction. Ideally, the precision of our model would not be significantly lower than that of deep CNNs and the time consumption would be much smaller.

In 2019 Fall, I started this project with professor Weiqing Gu at Harvey Mudd College. The idea I came up was to cluster the 2D face shapes and generate a probability on each cluster for a new face. Then, we multiply this probability with the result of the CNN model. To accelerate the progress, I registered the project to 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. The result was not ideal. We concluded that the plan could have been better if we used a neural network instead of gaussian clustering to classify the shape data. Looking back, I would have chosen the right way if I were to do it now; however, I learned a lot about machine learning during the procedure, and this experience paved the way for my future research.

**-- Reptile-NeRF with Alex Beatson, PhD --**

Earlier in this year, I read about the exciting work of NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis (Mildenhall et al, 2020). It trained a machine learning model that mapped a 5D space coordinate to color and transparency value on that point, and then used a rendering function for scattering volume to sum up the color values along the ray, so as to generate the color value of each pixel in the image. 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 300k iterations, which was 1-2 days on a Tesla V100 GPU. It would be too long for interactive 3D graphics or training a full video. I thought of accelerating the training with meta-learning, but I was not sure at first which method could be applied to neural rendering tasks.

Then, I read about MetaSDF (Sitzmann et al, 2020), a paper that applied a meta-learning method, MAML (Finn et al, 2017), 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.

At first, I tried model-agnostic meta-learning, or MAML in short. In this method, for each step of meta-training, the meta-model copies its parameters to inner models, each of which learns one task for several steps. The meta-model traces the change of parameters of the running inner model, and then takes gradient derived from the loss of the trained inner model on a validation set. In this case, each task would be rendering one scene. After I implemented MAML-NeRF, I found that the trained meta-model showed poor results. While training an unseen task, it performed slightly better than a blank model, but became worse as iteration increased.

I concluded that the failure of this plan was due to the limit on inner steps of MAML. To trace each inner step, the meta-model needed an extra chunk of memory. Therefore, the number of inner steps was quite limited – with 32GB of GPU memory, I could only take 6 inner steps with a reduced batch size. This number could be insufficient for 300k steps of Adam optimization in order to converge.

I went on looking for other meta-learning methods. I found that besides MAML, Reptile (Nichol et al. 2018) 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. As I switched to Reptile-NeRF, I achieved good result.

It was 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 skill 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.

**-- AR Language Learning with prof Sra, UCSB --**

Another thing my Reptile-NeRF project taught me 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 create an AR app for language learning and conduct a user experiment to see if it actually helps language learning.

Theoretical works proposed that AR language learning could improve learning performance and increase motivation (Parmaxi & Demetriou et al), but existing AR language learning technology is quite limited. The first approach is creating AR objects located in real places and form a scenario for dialogues (Lee & Park et al, 2019). In the users’ study, the learners’ experience largely depends on the quality of the scene, and thus the creators must take huge time and effort to create engaging and educational contents. I think it is not plausible to be used on a daily basis. Another approach is using a machine-learning-based object detection model to label the objects in live view (Huynh et al, 2019; Torelli et al 2020). The advantage is that no human efforts are required to generate the contents. However, the users’ experience is limited by the precision and pool of the model, since object detection is still under development in the status quo.

Professor Sra’s original idea was to develop a detect-and-show app that is similar to the second kind of work. As I found the limitation of this plan, I proposed that we combined the two approaches by letting the users create the AR words by themselves. This approach would totally circumvent the limitations of object detection technology and thus increase the freedom and communication in a community; also, creators don’t have to take much energy in just creating word tags.

Eventually, professor Sra accepted the change. Our final Android app works like this: users can enter creator mode or learner mode. In the creator mode, users can create AR tags and drag them onto the objects in real life; then, the location and contents of the tags will be uploaded to the cloud; in the learner mode, users can download the words and view them. We are conducting a user study in late November. Right now, we are writing the paper, and we plan to submit the work in January.

I had no experience with Android development before this project, but the idea attracted me, so I learned everything in a short time. 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. I was able to draw conclusions from the non-quantitative data, and I believed that was an important part in HCI research.) 括号里的内容是11月末即将发生的。

-- Conclusion on research experience --

Not all periods of my research experience were fruitful, yet I learned a lot in each of them. 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.

-- 我觉得我这段写得很有文采，但实在找不到好的地方加进去。删掉吗？--

(Over 2000 years ago, the Greek philosopher Plato argued that the theater was a futile activity since it was an imprecise representation of real subjects. For instance, those who enjoyed a work about war could learn about it much better by consulting a general. Nonetheless, his disciple Aristotle contended that theater was not meant to offer knowledge, but a source of emotional catharsis. I have a very similar view on technology. While we develop more capable machines and algorithms, we should also give people better ways to express their feelings and build communications. Life is not all about functionality.)

-- 我目前在以下部分写的是只给Ira组的版本。现在情况有变——有两个组都招人而且我都感兴趣，一个是做比较综合的Ira组，他们设计Graphics, Vision, AR, VR；另一个是更多Graphics的facial expression group，他们更多和graphics和filmmaking有关。但很遗憾PS只能交一份。我该如何同时拍两个组的马屁并且express interest呢？

**-- Why I want to switch to AR / VR --**

I envision the future entertainment in full-view live VR and AR, because they are the best platforms to combine the knowledge in the three fields I have worked on - computer vision, computer graphics and HCI. When the live-like appearance is combined with the capacity to be viewed from different angles, a truly new reality is created. Furthermore, we want to encourage communication between people. That requires a careful design in user study.

**-- Complement the school and the group --**

Paul Allen school in University of Washington is my top choice among all PhD programs. I appreciate its richness in communication and wide range of academic possibilities. Each student gets to work with two advisors in their PhD career instead of one. I could have more research possibilities and learn more from both knowledgeable professors. Academic-wise, take my favorite group led by professor Ira Kemelmacher-Shlizerman as an example: it is an extraordinary graphics group that doesn’t confine itself in one field; instead, it combines the knowledge of computer graphics, computer vision, AR and VR to make develop creative hybrid technologies. This pursuit matches my perspective in the future of visual art. In all, I believe I will become an outstanding researcher in visual technology by joining UW.

**Reference:**

Chrystalle B. Y. Tan, Elizabeth Sheppard, and Ian D. Stephen. Recognizing dynamic faces in Malaysian Chinese participants. Perception, 45(3):300–314, 2016.

Rachid Ahdid, Khaddouj TAIFI, S Said, and Bouzid MANAUT. Euclidean & geodesic distance between a facial feature points in two-dimensional face recognition system. *human-computer interaction*, 1:5, 2017.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis, 2020.

Vincent Sitzmann, Eric R Chan, Richard Tucker, Noah Snavely, and Gordon Wetzstein. MetaSDF: Meta-learning signed distance functions. June 2020.

Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* June 2019.

Finn, C.; Abbeel, P.; and Levine, S. 2017. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks.

Nichol, A.; and Schulman, J. 2018. Reptile: A Scalable Metalearning Algorithm. *arXiv: Learning*.

Parmaxi, A., & Demetriou, A. A. (2020). Augmented reality in language learning: A state-of-the-art review of 2014–2019. Journal of Computer Assisted Learning (n/a).

Lee, S.-M., & Park, M. (2019). Reconceptualization of the context in language learning with a location-based ar app. Computer Assisted Language Learning, 0(0), 1-24.

Huynh, B., Orlosky, J., & Hollerer, T. (2019). In-situ labeling for augmented reality language learning. In 2019 ieee conference on virtual reality and3d user interfaces (vr) (p. 1606-1611).

Torelli, E., Shemy, I. E., Silleresi, S., Moskwa, L., Cosentino, G., & Garzotto, F. (2020). Do you speak holo? A mixed reality application for foreign language learning in children with language disorders. In 2020 ieee conference on virtual reality and 3d user interfaces abstracts and workshops(vrw) (p. 608-609).