

Statement of Purpose

of Jun Ye (CSE PhD applicant for Fall—2022)

The field of augmented and virtual reality (AR/VR) has grown by leaps and bounds in recent years. We stand at the cusp of entering a new era, where we can communicate and interact with each other in a purely virtual manner. It has the potential of providing high quality education and healthcare to even the remotest corner of the world. This is validated by companies like Meta and Microsoft investing in creating ecosystems to conduct virtual meetings. These require realistic reproduction of appearances, expressions, and motions. This becomes a harder challenge when we consider the unconstrained settings of the real world. With the rise of deep learning methods, especially neural rendering, we have been able to tackle some of the challenges. I want to study and design solutions as a research scientist to make these platforms feasible and affordable to everyone.

Pursuing doctoral research at the University of California, San Diego, will provide me with the much needed experience to take on such unstructured and under-constrained problems. Democratization of AI has led to the immense progress of the field and empowered people to solve problems in the field of eCommerce, biology, and many more. We stand at the same crossroads for AR/VR and I aim to push the field in the direction where producing “metaverse” content does not require elaborate equipment and setup.

I am pursuing a Master degree program(Computer Science) from Kunming university of Science and Technology - one of the premier research university in China. I am indebted to the same for providing me with the necessary training and inculcating the required attitude and work ethic to take on research statements. Excelling in courses like Computer Vision and Graphics has helped me understand the classical approaches and how deep learning can efficiently solve the same problem. Being nominated for the Dean’s Merit List further validated my decision to pursue research. My three industry internships with startups gave me hands-on experience in designing tools to run deep learning algorithms on real world data.

Since my junior year, I worked under Prof YunZhang at KUST. It was here that I was formally introduced to image-based rendering. My project titled - “Neural view synthesis and appearance editing from unstructured images” focussed on extending neural rendering methods to incorporate appearance editing as well. The goal was to edit the texture of objects which were captured by smartphone cameras. While previous methods baked lighting and material together, the challenge was to effectively disentangle them so that the user could easily edit them. Taking inspiration from “Precomputed radiance transfer”, the local irradiance field and albedo were estimated and projected to a compact basis representation, in this case Spherical Harmonics. Estimating the albedo separately using differentiable rendering techniques gave rise to some structured artifacts on real scenes. Instead of that, the method jointly estimated the local irradiance field along with the albedo which removed those artifacts and improved visual quality. This work was published at ICMIR, 2022 - one of the top computer vision and graphics conferences in China.

My major takeaway was not the rendering equations or the spherical harmonics math but how important it was to keep calm, and rework the details where we overlooked something extremely important. At that critical juncture, my partners guided me to rethink, recalibrate to find solutions. At the 11th hour, I designed and added the missing component in order to meet the submission deadline. Being my first research paper, this was an enthralling experience as it helped me understand how to critically evaluate each experiment and communicate ideas effectively.

In my senior year, I was selected as a research intern at the Institute of Advanced Technology USTC. I interned under Prof. Chuanwen Lin. The project’s objective was to perform indoor room tours, from the omnidirectional images captured of the scene. This was an opportunity for me to extend my neural rendering knowledge and explore the upcoming world of implicit representations. Neural radiance fields were used for encoding objects or small scenes but we tried to encode an entire room. This came with its own challenges as omnidirectional images are spherical in nature and hence every pixel does not correspond to one single point in space. We tried to tackle such issues during post-processing by modeling it as an image-to-image translation task. This improved the sharpness and overall quality of the results. What surprised me most was that despite masking out the subject capturing the images, the network was able to reconstruct the scene effortlessly and model reflection from mirrors accurately. This was a delightful experience as it gave me an insight on how research is being conducted at different labs across the world. Furthermore, it highlighted the importance of software engineering skills in becoming a good Deep Learning researcher.

Currently, for my master's thesis, we are collaborating Prof YunZhang on extending the above mentioned project. In addition to novel view synthesis, the aim is to capture the radiance of the scene. This will enable us to add synthetic objects while accurately modeling the object's interaction with the lighting. An application of such a system would be doing a room tour and adding synthetic furniture into the scene. We aim to publish this project soon to an upcoming venue.

In addition to my research work, I took on the role of Teaching Assistant for Computer Vision and Statistical Methods in AI over three semesters where I designed assignments and projects for more than 200 students. I have taken tutorial sessions on certain topics and mentored a few teams for their final projects. A very intensive evaluation was done based on self-designed rubrics to measure the learning outcomes of the course.

I would like to work with the Visual Computing Center at the University of California, San Diego .They have been instrumental in laying the foundations for performing photo-realistic neural rendering , evident in their contributions reflected at the top vision and graphics conferences.

A significant issue with photorealistic neural rendering is the lack of high-quality data. While we can capture high-quality assets using advanced equipment such as Kinect, lidar, light stages, they are not accessible to the general public. It is thus critical to leverage deep learning in such a way we can interact with scenes that are rendered from unlabeled and unstructured images. Current methods can reconstruct the texture and geometry from images as well as reliably correct the noisy camera parameters. However, considerable work is required to edit these scenes interactively. This is true for both object and human reconstruction. Editing material and lighting independently is a complex problem since the rendering equation combines them together. We can disentangle the two variables using clever representations and deep learning methods. I would like to work on these problems and, hopefully, make it possible to edit the lighting and textures of an entire scene in real-time.

Prof. Hao Su's work on "MVSNeRF" was an essential step towards generalization and fast training of scenes. However, the combination of classical multiview stereo and neural radiance fields does not give us any scope for editing. I want to work on adding relighting capabilities to such methods. His expertise with 3D scene representation would be essential for tackling such problems. Prof. Manmohan Chandraker's work with "Single-Shot Neural Relighting and SVBRDF Estimation" was a major step towards object property acquisition from a single image. I want to incorporate material editing which can convert some part of the object from ceramic to metal. His experience with 3D scene understanding would be crucial for designing such systems. Prof. Ravi Ramamoorthi's work on "Inverse Rendering for Complex Indoor Scenes" was an integral step towards recovering scene parameters and depth from a single image. However, for AR applications, we need a system which can edit the scene in real-time. I would like to work in this direction, and his guidance would be pivotal.

I believe my current project and my past experiences have enabled me to understand how to model problems as inverse rendering statements. My learning has been that it's not about how fast you can solve these problems or how well you can scale but rather how wild you can get. Research for me is an extension of that. While I might get stuck on some sub-problems for weeks, working on these projects propels me to be creative and try unconventional ideas.

The rigorous standards of KUST have taught me that true intellectual contribution is only possible through perseverance, determination, and a ruthless eye for weakness in both experimental design and execution. Balancing laboratory workloads, and teaching assistantships with a full schedule of undergraduate classes has been a taxing endeavor, but this too has been essential to my growth as a researcher. Having worked with appearance editing and image based rendering, I feel that I would be able to make important contributions to the field and develop algorithms that would make it more feasible to create AR/VR content. Today, I look forward to the new intellectual challenges that a doctoral program will provide, and I am certain that I will discover new passions, curiosities, and questions as I prepare for my future as a research scientist.