Yinqi (Bill) Sun Website: https://billsun.dev Email: sunyinqi0508@gmail.com

Personal Statement

I love exploring. Not just because of the excitement upon discovering something new, something beautiful or useful, but also because uncovering reasons underneath those phenomena gives me a closer peek at the principles of the world. This is why I am enthusiastic about **Data Science** and **Visualization**, especially the ways to distill and organize useful information from the endless world of data and deliver them so that people from different fields can easily explore and analyze the data. I'm also interested in Software System Engineering, especially **Databases** Systems. I'm fascinated by such magnificent yet delicate systems that are the building blocks behind many higher-level applications, including Data Science and Visualization. I believe innovations in these fields will often fundamentally improve their applications and ultimately enable new industry-changing possibilities.

In college, I joined the Visualization Research Lab led by Dr. **Yunhai Wang**. In one of our research projects, we enhanced the stress-majorization (MDS) algorithm to support specifying various forms of constraints to the visualization and is both fast and has stable convergence (**Vis'17**). My primary contributions were implementing the algorithm with CUDA to run in parallel on GPUs and mathematically proving our algorithm's convergence. I then proposed to use stress majorization to reduce distortions introduced in Fisheye Views while using edge ratios as custom constraints to enforce the zoom ratio (**Vis'18**). As my first step into research, it was exciting and gratifying to see all the knowledge, linear algebra I learned in class, optimization methods I learned from the internet, and the programming skills I learned through practice all come together in the effort of solving a problem.

My work caught the eye of Dr. **Eugene Wu**, who offered to sponsor me as a visiting scholar at Columbia University. I participated in a data provenance project (SMOKE, VLDB'18) to find ways to increase lineage capture performance during query execution using the patterns that individual physical operator reorganizes data. This project was quite different from previous research, but I saw it as an adventure to explore a new field, and everything I learned will benefit my future journey.

In late 2019, I returned to China due to family issues. Then COVID hits. My city Wuhan then became the first global epicenter of the pandemic. My parents and I were confined in our little apartment with minimal supplies to survive while grieving the loss of our loved ones.

When facing crisis, I fight while accepting the inevitable. Right after this disaster, I started pursuing a master's study at NYU. The coursework took on a renewed sense of relevance given my research experience at Columbia, where I learned first-hand the importance of a solid academic foundation to the research process. I also enjoyed the course projects, especially the weekly projects in Computer Graphics class, where I implemented an interactive recursive ray tracer with refraction using GLSL. My professor **Ken Perlin** was impressed by my work and invited me to his lab. In the Compiler Construction class, I built a compiler that compiles a subset of Python into RISC-V assembly. I also took Prof. Margaret Wright's Advanced Numeric Optimizations class, where I thoroughly studied and implemented optimization methods, such as CG and BFGS, which are used everywhere in computer science.

Because of my great work in the Advanced Database class, Dr. **Dennis Shasha** introduced me to his lab. Thanks to my research experience in Columbia, I was able to independently set my research objectives and progressively push my project forward, solving obstacles in the way. In one project (AQuery++, In Submission to VLDB'23), I designed and implemented a column-store database system featuring support for time-series queries and extensibility via UDFs. Because it's compiling queries into native C++ code, recurring queries will be very fast, which makes it excellent for incremental analysis. The C++ code generated is simple and succinct thanks to my C++ template-based library that offers declarative style data manipulations at native speed. This allows users to easily profile and tune query performances or customize query behavior by editing generated code. The database I wrote is also robust enough to be used as a platform for student assignments in future Advanced Database classes at NYU.

I plan to enroll in a PhD program because I like exploring the world with explanations and proofs, and I look forward to working with people from different backgrounds and being exposed to fresh perspectives. My experiences collectively shaped my research interests. Noticing the ever-rising scale of data and the immense need for data analysis, I think Human Data Interaction is a great topic for my PhD research. With techniques I learned about Database, Visualization and Machine learning, I would like to further explore data-driven applications that catalyze people's exploration in the world of data.

Following my PhD program, I would like to pursue a career in academia as a professor at a university or a researcher in a tech company, so I can continue my exploration of Data Science and other mysteries of the world while uncovering knowledge that benefits humanity.