Wentao Zhang

nsknojj.github.io | (+1)2132551218 | zhan914@usc.edu

# Education

|  |  |
| --- | --- |
| **University of Southern California, Viterbi School of Engineer** M.S. in Computer Science (GPA: 3.82/4.0) | June 2017 — Present |
| **Peking University, School of EECS**  B.S. in Computer Science and Technology (Major GPA: 3.65/4.0) | Sep. 2013 — June 2017 |

# Experience

|  |  |
| --- | --- |
| **Software Engineering Intern** — **Jingchi Corp, Sunnyvale** Autonomous driving startup | May — August 2018 |
| **Research Assistant** — **Computer Systems Lab, Cornell Univ.** Researched on Binarized Neural Network Inference Acceleration on FPGA, and published a paper as third author in FPGA, pp. 15-24. 2017.  • Built a baseline and deployed on general-purpose computing platforms, including Jetson TK1 with embedded GPU and ARM CPU, x86 server with high-performance GPU. To build the baseline, I implemented BNN (Binarized Neural Network) layers in C++, including conv-layer, dense-layer, etc. Reduced the pipeline delay and parallelized computational graph with Halide. This got a 2x speedup on average against the original BNN implementation with Lasagne.  • Collaborated with mentors on speed & energy-efficiency testing, result comparison and reporting. | July — Sept. 2016 |
| **R&D Intern — Nat'l Eng. Lab for Video Technology, Peking Univ.** Developed a video searching demo based on feature matching.  • Introduced Compact Descriptors for Visual Search (CDVS) as an efficient method extracting the video frame features and designed an image-frame matching scheme.  • Experimented different parameters for frame down sampling and video interval sampling to gain speedup without significant loss of accuracy.  • Designed Qt GUI for demo, supported video playing, frame locating, etc. | Apr. 2015 — July 2016 |

# Projects

|  |  |
| --- | --- |
| **Stock Search Web App** | May — August 2018 |
| **Research Assistant** — **Computer Systems Lab, Cornell Univ.** Researched on Binarized Neural Network Inference Acceleration on FPGA, and published a paper as third author in FPGA, pp. 15-24. 2017.  • Built a baseline and deployed on general-purpose computing platforms, including Jetson TK1 with embedded GPU and ARM CPU, x86 server with high-performance GPU. To build the baseline, I implemented BNN (Binarized Neural Network) layers in C++, including conv-layer, dense-layer, etc. Reduced the pipeline delay and parallelized computational graph with Halide. This got a 2x speedup on average against the original BNN implementation with Lasagne.  • Collaborated with mentors on speed & energy-efficiency testing, result comparison and reporting. | July — Sept. 2016 |
| **R&D Intern — Nat'l Eng. Lab for Video Technology, Peking Univ.** Developed a video searching demo based on feature matching.  • Introduced Compact Descriptors for Visual Search (CDVS) as an efficient method extracting the video frame features and designed an image-frame matching scheme.  • Experimented different parameters for frame down sampling and video interval sampling to gain speedup without significant loss of accuracy.  • Designed Qt GUI for demo, supported video playing, frame locating, etc. | Apr. 2015 — July 2016 |

# Awards

The ACM-ICPC Asia Regional Contest Mudanjiang Site 2014 Gold Medal 10th Place

International Olympiad in Informatics (IOI) 2013 China Team Selection Competition 9th Place