

SHUAI SHI

+44 07749831095 | shishuaicareer@gmail.com | Edinburgh
[Linkedin](#) | [Personal Website](#) | [Github Profile](#)

EDUCATION

University of Edinburgh

2017–2022

BSc Artificial Intelligence, First-Class Honours degree

Courses: Data Structure & Algorithms, Software Engineering, Machine Learning and Deep Learning, Deep Learning Systems

SKILLS

Languages: C/C++ == Python > Java == SQL > Rust, JavaScript, HTML/CSS

Developer Tools: Cuda, PyTorch, TensorFlow, TVM, Git, AWS, Docker

PROJECTS

End-to-End Pedometer Android App Using IoT Device

[Thesis](#) | *Python, Java, SQL, TensorFlow, FFT, Wavelet Transform*

- **High-Accuracy Pedometer Development:** Engineered a pedometer app using a two-step system, Python/SQL data pipeline, and LSTM model with TensorFlow for 97% accurate activity detection.
- **Step Counting Algorithms:** Implemented FFT and wavelet transform algorithms for 91% accurate step-counting, surpassing traditional methods by 11%.
- **Model Optimization:** Integrated machine learning models into the Android app using TensorFlow Lite and Chaquopy to run Python in the JVM, optimizing efficiency.

Optimization and Localization of DNN Inference Models

[GitHub](#) | *CUDA, GPU Optimization, Nsight Tools, Inference Acceleration*

- **Convolution Layer Optimization:** Merged techniques including tiled shared memory, shared memory matrix multiplication, input matrix unrolling, and kernel fusion for efficient convolution.
- **Performance Enhancement:** Combined performance tuning through constant memory, loop unrolling, parameter optimization, and specialized kernel implementations with computational efficiency methods like input channel reduction and Tensor Core acceleration.
- **GPU Utilization Boost:** Leveraged Nsight tools to enhance GPU usage, employing advanced matrix multiplication algorithms and overlapping computation with data transfer to maximize efficiency.

Needle: A Deep Learning Framework

[GitHub](#) | *C++, Python, Framework Development*

- **Framework Architecture:** Designed and implemented core components like automatic differentiation, layers, activation functions, and optimizers.
- **Performance Optimization:** Engineered a high performance NumPy-like ndarray on CPU and GPU using techniques like multi-threading, SIMD vectors, and CUDA kernels.
- **Computer Vision & NLP Models:** Constructed models including ResNet, LSTM, and CNNs achieving state-of-the-art accuracy on CIFAR-10 and Penn Treebank.

OPEN SOURCE CONTRIBUTIONS

PaddlePaddle - PyTorch Code Conversion Toolkits

[GitHub](#) | *PaddlePaddle, PyTorch, AST, GitHub Collaboration, Test-Driven Development (TDD)*

- **API Conversion and Documentation:** Implemented and tested rules for converting PyTorch APIs (e.g., torch.nn.Module) into PaddlePaddle, authored guides on API mapping, and utilized AST-based conversion methodologies.
- **Model Migration Initiative:** Contributed to a seamless and efficient migration process between AI frameworks, reflecting expertise in code conversion and testing methodologies.