

(a) Benchmark Construction

Open-Source Sources

- Edge Models
- Operator Libraries
- Kernel Repositories

Static-Graph Adaptation

- PyTorch Re-expression
- Semantic-Shape Decoupling

Expert Review

- Graph Consistency
- Semantic Fidelity

EdgeKernelBench

Multi-Level Tasks

Level 1 Level 2 Level 3

Scenario Profiles

Mobile Smart AIoT Robotics

(b) LLM-based Kernel Generation



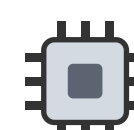
Reference Task



LLM

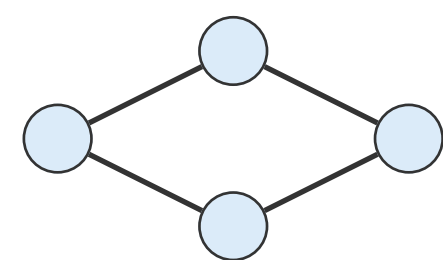


Device Description

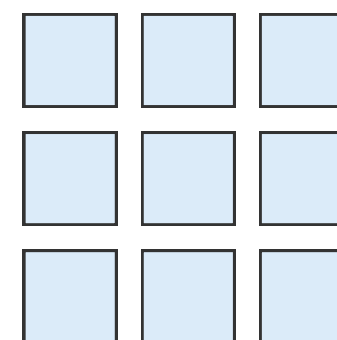


Unified Kernel IR *backend-independent*

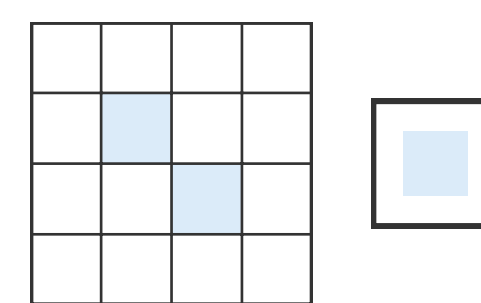
Graph-Level



Tensor-Level



Schedule-Level



LLVM Adapter



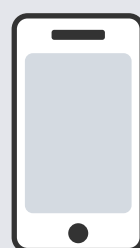
OpenCL Adapter



CUDA Adapter

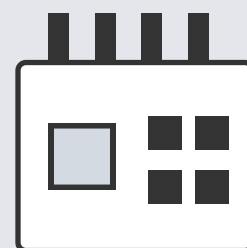


Heterogeneous Edge Hardware



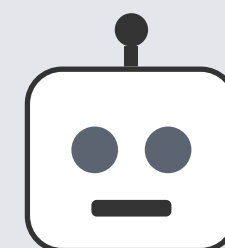
Mobile

CPU · LLVM



Smart AIoT

GPU · OpenCL



Robotics

GPU · CUDA

(c) Evaluation Pipeline

1 Build & Execute

compile and run on the target device



2 Correctness Verification

output equivalence vs. PyTorch reference

Candidate Output



PyTorch Reference



3 Performance Comparison

warm-up, then repeated paired runs

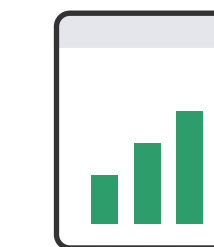


Candidate vs. Baseline



4 Benchmark Report

aggregated per backend and level



- Generation Success
- Correctness Rate
- Speedup