



中国科学院计算技术研究所

INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES

ComBack: A Versatile Dataset for Enhancing Compiler Backend Development Efficiency

Ming Zhong^{1,2}, Fang Lyu¹, Lulin Wang¹, Hongna Geng^{1,2}, Lei Qiu^{1,2}, Huimin Cui^{1,2} and Xiaobing Feng^{1,2}

¹SKLP, Institute of Computing Technology, CAS

²University of Chinese Academy of Sciences, Beijing, China

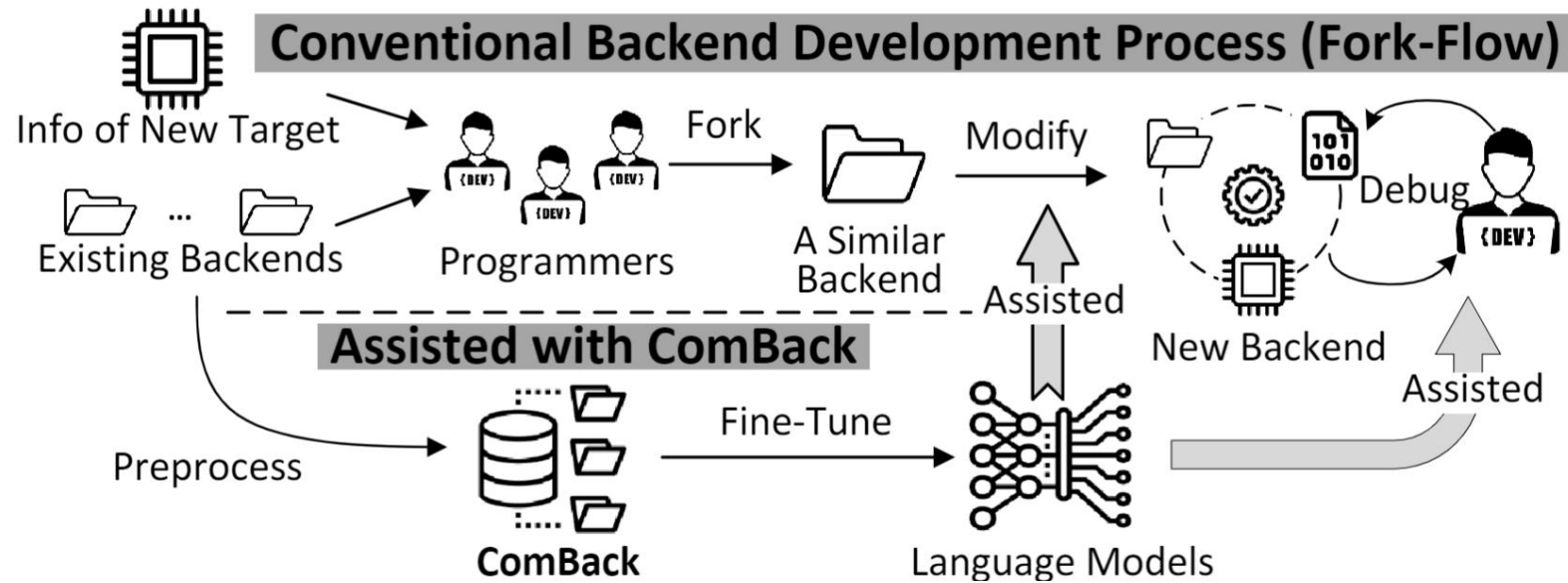
Compiler Backend

- **Compiler: Source Code -> Machine Code.**
 - FrontEnd, MiddleEnd, **BackEnd**.
 - **Backend:** Generating machine code for various processors.

- **Challenges in Backend Development: Slowness and Complexity.**
 - **Slowness:** Necessitating understanding of hardware characteristics and compiler infrastructure.
 - **Complexity:** Magnitude of manual efforts for writing code.

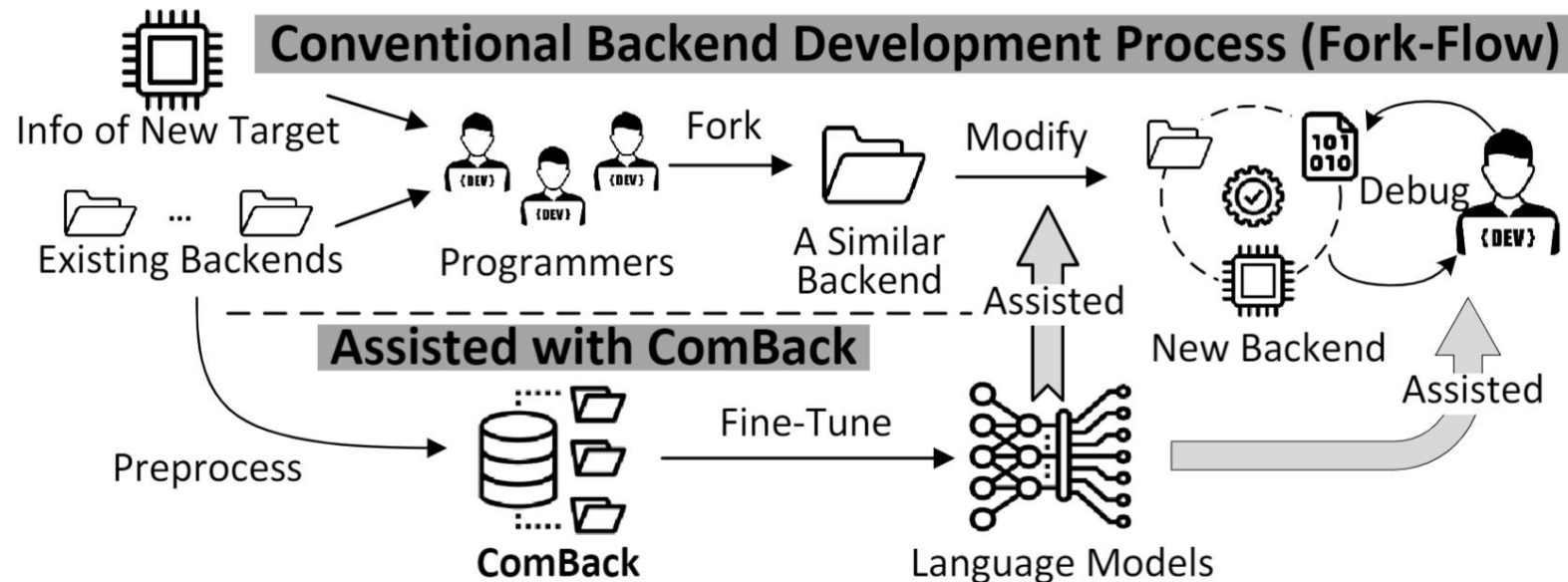
Background

➤ Traditional Manual Approach: Fork-Flow



Background

➤ Traditional Manual Approach: Fork-Flow



➤ Current Problem:

- Limited accuracy in LLMs (e.g., ChatGPT) for backend code.

➤ Solution:

- A domain-specific dataset for compiler backend development.

ComBack: First Public Dataset for Backends

➤ Feature

- **Large-Scale:** 5,700,000+ Lines of Code, 181,000+ Functions.
- **Multi-Targets:** 77 for GCC, 101 for LLVM.

(a) GCC

| Type | Target | Function | KLoC |
|------------|-----------|---------------|--------------|
| CPU | 30 | 35,147 | 647.2 |
| MPU | 33 | 6,010 | 183.9 |
| GPU | 2 | 457 | 11.2 |
| VLIW | 5 | 959 | 25.4 |
| DSP | 3 | 399 | 9.6 |
| Virtual | 4 | 327 | 6.5 |
| Sum | 77 | 43,299 | 883.7 |

(b) LLVM

| Type | Target | Function | KLoC |
|------------|------------|----------------|----------------|
| CPU | 43 | 84,914 | 3,450.4 |
| MPU | 30 | 11,311 | 173.0 |
| GPU | 5 | 22,591 | 768.3 |
| VLIW | 4 | 2,048 | 24.3 |
| DSP | 7 | 9,646 | 263.2 |
| Virtual | 12 | 8,430 | 168.3 |
| Sum | 101 | 138,940 | 4,847.5 |

- **Versatility:** 3 tasks for common scenarios.
- **Availability:** <https://huggingface.co/datasets/docz-ict/ComBack>

Three Tasks in ComBack

Inputs: ... adjustReg(DL, SPReg, FPReg, -StackSize+RVFI->getVarArgsSaveSize() _____)
Ground Truth: MachineInstr::FrameDestroy);

(a) Statement-Level Completion

Inputs: ... maxCallFrameSize = (maxCallFrameSize + AlignMask) & ~AlignMask;
Ground Truth: MFI -> setMaxCallFrameSize(maxCallFrameSize);

(b) Next-Statement Suggestion

Inputs:
getPointerRegClass: Returns a TargetRegisterClass used for pointer values.
Target-Specific Value: Sparc, SP::I64RegsRegClass, SP::IntRegsRegClass.
Ground Truth:
TargetRegisterClass *SparcRegisterInfo::getPointerRegClass(MachineFunction &MF, unsigned Kind) {
 return Subtarget.is64Bit() ? &SP::I64RegsRegClass : &SP::IntRegsRegClass;
}

(c) Code Generation

Evaluation

➤ Setup:

- 6 language models (max 220M parameters) for fine-tuning.
- 2 LLMs (ChatGPT-3.5-turbo and CodeLLaMA-34B) for inferencing.
- **RQ.1 Improvements of backend code accuracy through fine-tuning.**
 - **Train/Validation/Test set: 80%:10%:10%.**

| Model | Stmt. Comp. | | Next. Sugg. | | Code. Gen. | | Stmt. Comp. | | Next. Sugg. | | Code. Gen. | |
|---------------|----------------------------|-------|-------------|-------|------------|-------|-------------------|--------------|--------------|--------------|--------------|--------------|
| | EM (%) | ED | EM (%) | ED | BLEU-4 | ED | EM (%) | ED | EM (%) | ED | BLEU-4 | ED |
| | Without Fine-Tuning | | | | | | Fine-Tuned | | | | | |
| CodeBert | 0.00 | 0.97 | 0.00 | 1.31 | 0.00 | 0.44 | 53.84 | 77.44 | 52.67 | 70.82 | 23.54 | 54.63 |
| GraphCodeBert | 0.00 | 0.35 | 0.00 | 0.54 | 0.00 | 2.41 | 43.00 | 71.89 | 47.10 | 61.31 | 20.73 | 48.83 |
| UniXcoder | 0.07 | 27.56 | 15.93 | 29.11 | 0.00 | 31.81 | 67.84 | 85.06 | 58.51 | 75.31 | 56.24 | 73.45 |
| CodeT5 | 0.65 | 21.45 | 7.23 | 23.50 | 0.00 | 13.57 | 66.47 | 84.34 | 58.52 | 76.03 | 70.87 | 80.45 |
| NatGen | 0.00 | 13.52 | 0.02 | 15.95 | 0.01 | 28.76 | 67.47 | 84.83 | 60.30 | 76.84 | 71.73 | 81.39 |
| CodeT5+ | 0.02 | 7.24 | 0.12 | 9.87 | 0.00 | 12.33 | 66.93 | 84.45 | 59.57 | 76.41 | 75.29 | 82.92 |

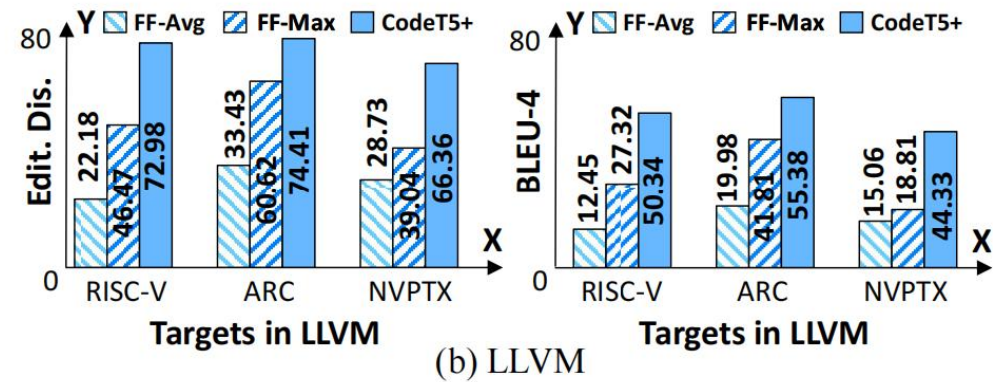
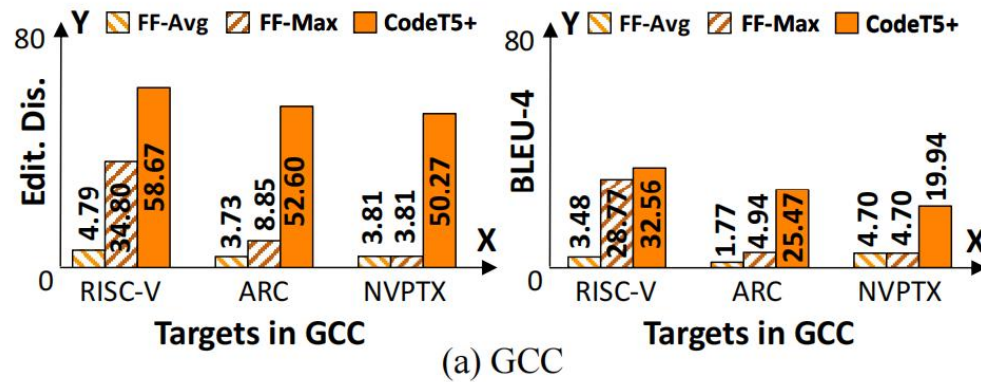
Evaluation

- RQ.2 Backend code accuracy for new targets of **existing types**.
 - **Testset:** (RISC-V (CPU), ARC (MPU), NVPTX (GPU)).
 - **Train/Validation set:** Remaining data (90%:10%).
 - CodeT5+ (220M) for fine-tuning, ChatGPT and Code-LLaMA for Inferencing.

| Model | Stmt. Comp. | | | | | | Next. Sugg. | | | | | | Code. Gen. | | | | | |
|----------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | RISC-V | | ARC | | NVPTX | | RISC-V | | ARC | | NVPTX | | RISC-V | | ARC | | NVPTX | |
| | EM (%) | ED | EM (%) | ED | EM (%) | ED | EM (%) | ED | EM (%) | ED | EM (%) | ED | BLEU-4 | ED | BLEU-4 | ED | BLEU-4 | ED |
| GCC | | | | | | | | | | | | | | | | | | |
| ChatGPT | 10.34 | 38.41 | 15.35 | 42.94 | 12.01 | 41.47 | 6.44 | 12.90 | 9.75 | 20.79 | 7.97 | 17.79 | 1.37 | 24.12 | 1.67 | 28.26 | 1.57 | 26.97 |
| Code-LLaMA | 0.41 | 19.07 | 0.85 | 16.77 | 0.56 | 18.22 | 1.58 | 13.54 | 2.66 | 17.95 | 2.47 | 16.59 | 1.67 | 27.89 | 1.71 | 30.49 | 1.57 | 27.65 |
| CodeT5+ | 51.16 | 75.32 | 52.45 | 74.57 | 50.56 | 75.52 | 49.11 | 67.84 | 38.26 | 59.21 | 38.33 | 56.31 | 32.56 | 58.67 | 19.94 | 50.27 | 25.47 | 52.60 |
| LLVM | | | | | | | | | | | | | | | | | | |
| ChatGPT | 12.08 | 41.39 | 16.77 | 42.02 | 14.73 | 43.72 | 9.80 | 21.86 | 10.81 | 20.66 | 11.39 | 22.82 | 1.23 | 25.12 | 1.30 | 27.19 | 1.43 | 25.45 |
| Code-LLaMA | 0.45 | 17.61 | 0.61 | 17.21 | 0.99 | 17.23 | 1.75 | 15.04 | 0.42 | 11.27 | 2.42 | 16.25 | 1.43 | 27.24 | 1.61 | 32.12 | 1.59 | 28.08 |
| CodeT5+ | 62.68 | 82.02 | 71.34 | 85.98 | 64.45 | 81.53 | 48.71 | 68.95 | 58.68 | 74.57 | 47.81 | 65.51 | 50.34 | 72.98 | 55.38 | 74.41 | 44.33 | 66.36 |

Evaluation

- RQ.2 Code accuracy for new targets of **existing types**.
 - Comparing with traditional method (Fork-Flow).
 - **Simulating Fork-Flow**: Calculate the evaluation metrics between test set functions and train/valid set functions.



Evaluation

- RQ.2 Code accuracy for new targets of **new types**.
 - **Testset:** (ARC (MPU), NVPTX (GPU)).
 - **Train/Validation set:** CPU data (90%:10%).
 - **Fine-Tuning CodeT5+ (220M)** from scratch.

| Dataset | Stmt. Comp. | | | | Next. Sugg. | | | | Code. Gen. | | | |
|------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | ARC (MPU) | | NVPTX (GPU) | | ARC (MPU) | | NVPTX (GPU) | | ARC (MPU) | | NVPTX (GPU) | |
| | EM (%) | ED | EM (%) | ED | EM (%) | ED | EM (%) | ED | BLEU-4 | ED | BLEU-4 | ED |
| GCC | | | | | | | | | | | | |
| -w/o GPU and MPU | 50.53 | 74.09 | 46.37 | 72.45 | 37.22 | 58.21 | 38.33 | 56.83 | 19.29 | 49.12 | 22.46 | 50.33 |
| -w GPU and MPU | 52.45 | 74.57 | 50.56 | 75.52 | 38.26 | 59.21 | 38.33 | 56.31 | 19.94 | 50.27 | 25.47 | 52.60 |
| Diff | -1.92 | -0.48 | -4.19 | -3.07 | -1.04 | -1.00 | 0.00 | +0.52 | -0.65 | -1.15 | -3.01 | -3.37 |
| LLVM | | | | | | | | | | | | |
| -w/o GPU and MPU | 69.82 | 85.59 | 60.04 | 79.85 | 58.26 | 73.75 | 46.28 | 63.92 | 49.62 | 70.26 | 42.94 | 65.43 |
| -w GPU and MPU | 71.34 | 85.98 | 64.45 | 81.53 | 58.68 | 74.57 | 47.81 | 65.5 | 55.38 | 74.41 | 44.33 | 66.36 |
| Diff | -1.52 | -0.39 | -4.41 | -1.68 | -0.42 | -0.82 | -1.53 | -1.58 | -5.76 | -4.15 | -1.39 | -0.93 |

Evaluation

➤ RQ.3 Iterative Expansion Ability.

➤ **Testset:** (RI5CY (Customized RISC-V Processor)).

➤ **Train/Validation set:**

➤ Set 1: CPU data (**without RISC-V**) (90%:10%).

➤ Set 2: CPU data (**with RISC-V**) (90%:10%).

➤ Fine-Tuning CodeT5+ (220M) from scratch.

| Dataset | Stmt-Level. Comp. | | Next-Stmt. Sugg. | | Code. Gen. | |
|-------------|-------------------|--------------|------------------|--------------|---------------|---------------|
| | EM (%) | ED | EM (%) | ED | BLEU-4 | ED |
| -w RISC-V | 74.06 | 87.91 | 67.25 | 81.28 | 79.46 | 89.92 |
| -w/o RISC-V | 66.16 | 83.79 | 57.29 | 74.73 | 54.41 | 75.41 |
| Diff | +7.90 | +4.12 | +9.96 | +6.55 | +25.05 | +14.51 |

Thanks!

zhongming21s@ict.ac.cn