FastLanes for ML: GPU Decoding with PyTorch Images

Yuxin Tang [1], Feng Zhang [1,2], and Peter Boncz [2]

[1] Renmin University of China, [2] CWI, Amsterdam, Netherlands
{yuxintang,fengzhang}@ruc.edu.cn, {feng,boncz}@cwi.nl

FastLanes [1] is a compressed storage format designed for large-scale data that natively supports ML workloads by offering high compression ratios, block-level parallelism, and bandwidth-efficient, parallel-friendly access paths. In this paper, we present a case study on the most common Image AI workload—PyTorch-based image pipelines—showing how GPU-side decoding of FastLanes can directly decompresses data into device memory and overlaps I/O, decoding, and preprocessing. This design alleviates CPU bottlenecks, reduces redundant host–device transfers, and improves end-to-end throughput and GPU utilization.

We show FastLanes' AI-centric, end-to-end solution for PyTorch image data management, as illustrated in Figure 1. First, we introduce FastLanes' **AI-oriented image data format** that identifies model-salient components (e.g., selected frequency bands or structured features) and applies lightweight, highly parallel compression to minimize data movement and processing, thereby accelerating both training and inference. Building on this format, we design a **GPU-native pipeline** that operates directly in the compressed domain: decompression and preprocessing are executed in-place by a massively parallel **GPU decoder**, minimizing host–device data transfers, removing CPU bottlenecks, and sustaining high GPU utilization. The system **integrates seamlessly with mainstream framework** PyTorch, allowing users to incorporate FastLanes datasets into existing DataLoader workflows with minimal code changes, while delivering high-throughput, GPU-accelerated data access. By co-designing the storage format and the data pipeline, we provide an efficient, AI-centric solution for end-to-end image data handling.

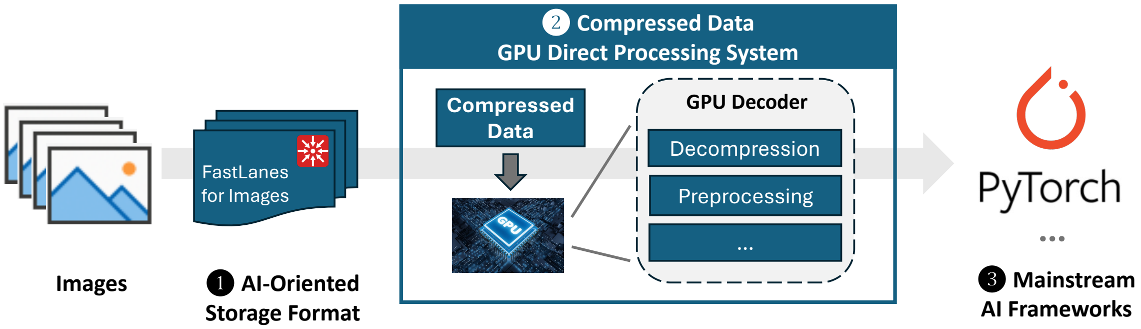

Figure 1: FastLanes Workflow for PyTorch Image Workloads.

References

[1] Azim Afroozeh and Peter Boncz. The fastlanes file format. *PVLDB*, 18(11), 2025.