ACDP: Architecture-aware Cross-Dataset Performance Predictor for NAS

NeurIPS 2026 Presentation

Jiawen Deng, Yuqi Feng, Han Ji, Yanan Sun* | Sichuan University

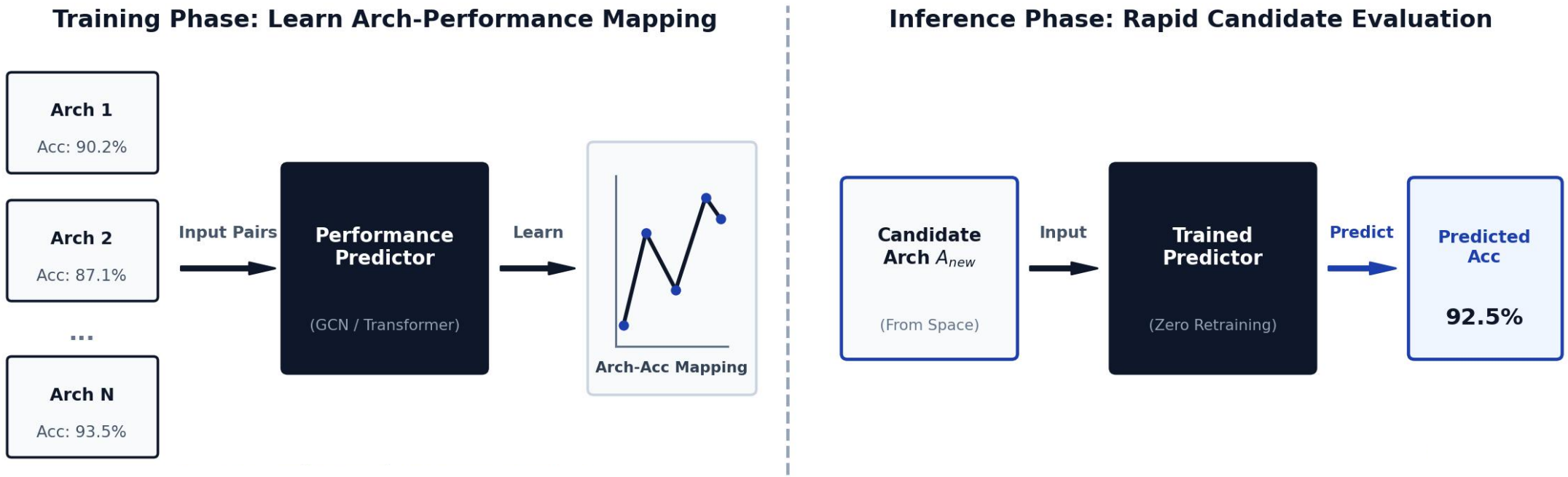
Background: Performance Predictors in NAS

Traditional NAS Evaluation Bottleneck

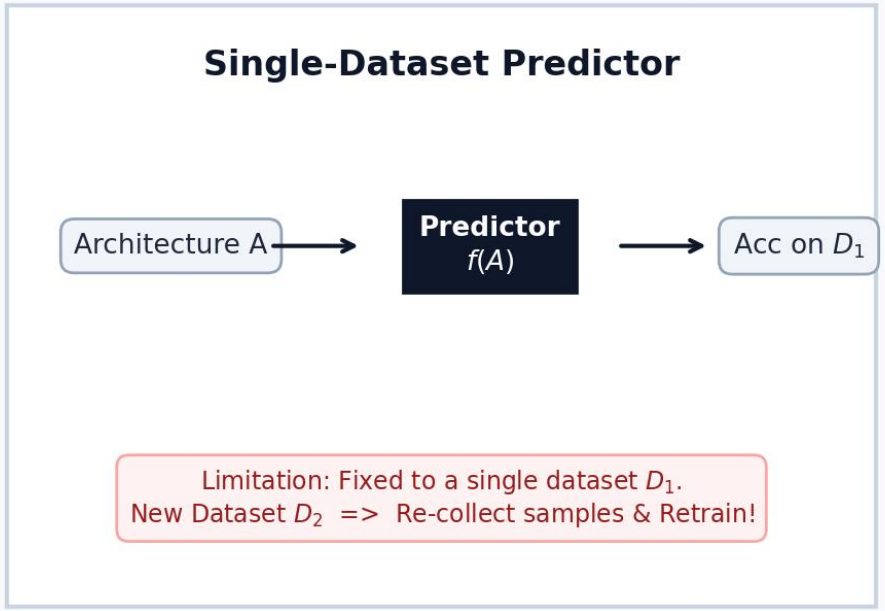
- NAS automates optimal architecture discovery across massive search spaces.
- Full-training each candidate model requires days of GPU compute.
- Infeasible for real-world scaling across large candidate pools.

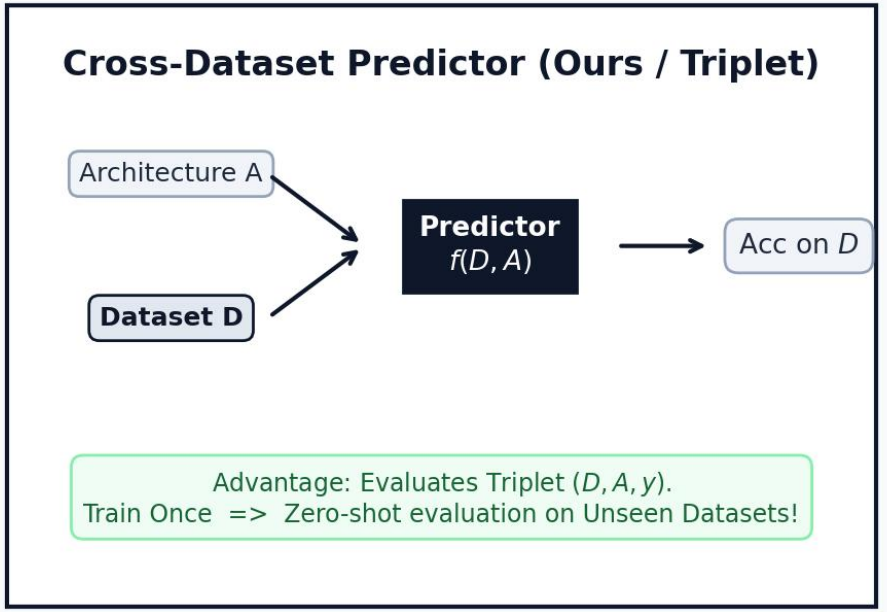
Performance Predictor: Efficient Proxy

- Learns direct mapping function f:
  Architecture -> Performance from offline data.
- Evaluates candidate models in milliseconds.
- Serves as the primary engine to accelerate real-world NAS evaluation.



Background: Single-Dataset vs. Cross-Dataset Predictors





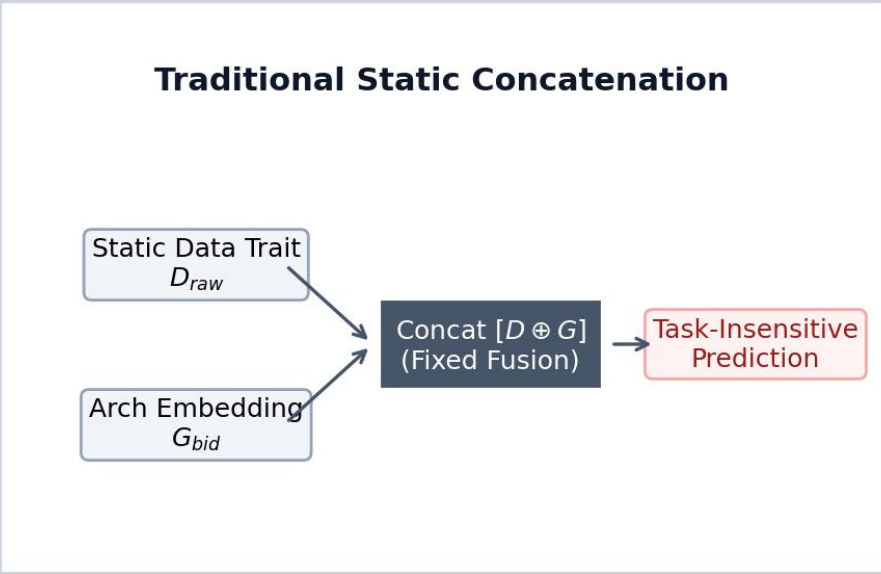
Traditional Single-Dataset Predictor

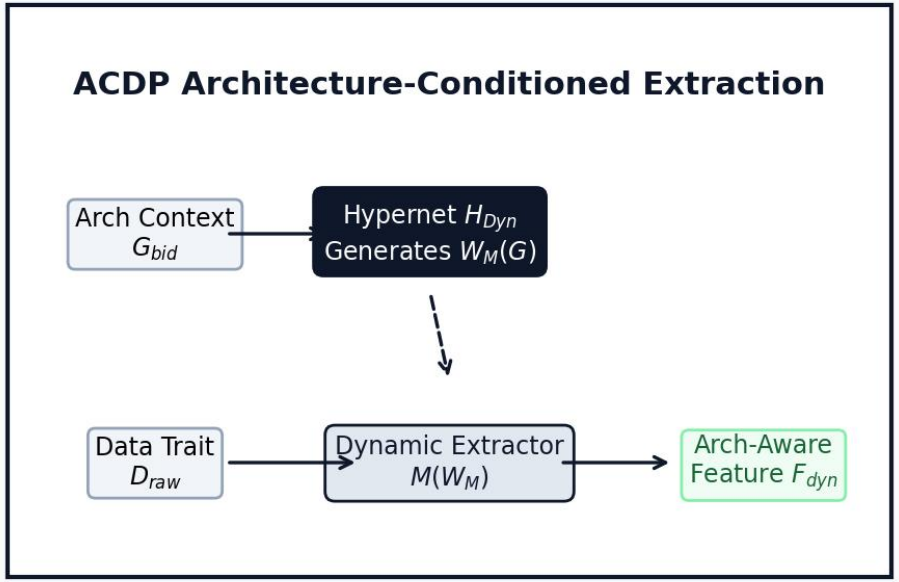
- Input: Candidate Architecture A only -> Output: Acc on target dataset
- Bottleneck: Isolated to 1 dataset. Switching datasets requires collecting new training tuples and retraining from scratch.

Cross-Dataset Predictor (Triplet Formulation)

- Input: Triplet Tuple (Dataset D, Architecture A, Performance y)
- Advantage: Incorporates dataset traits as explicit input. Enables zero-shot transfer and evaluation on unseen target datasets without retraining.

Motivation: Why Dataset Feature Must Be Architecture-Conditioned





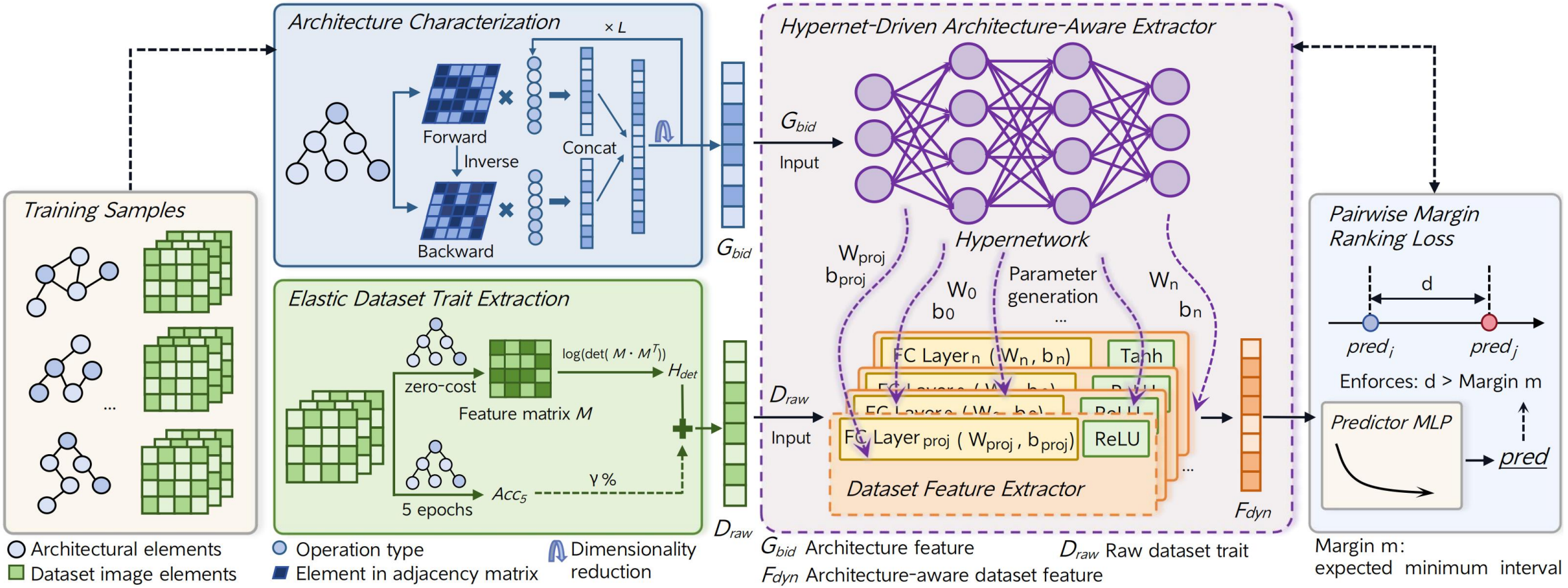
The Static Concat Bottleneck

Existing cross-dataset methods (MetaD2A, DiffusionNAG) naively concatenate static dataset features [D ⊕ G]. This ignores dynamic coupling between models and data.

Dynamic Architecture-Conditioned Extraction

Dataset feature extraction must be dynamically modulated by candidate architecture properties (G_bid) via Hypernetwork parameters W_M, creating task-aware representations.

Methodology: Hypernetwork-Driven Architecture-Aware Predictor



Hierarchical Evaluation & Core Contributions

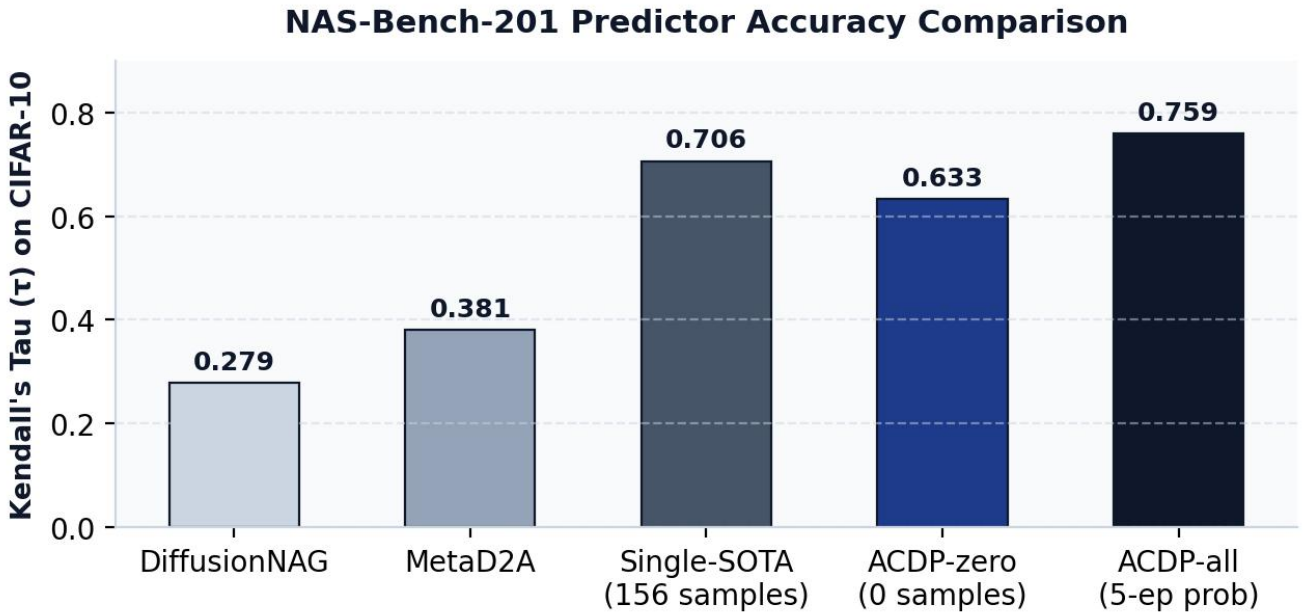
Two-Stage Hierarchical Strategy

- **Stage 1: ACDP-zero (Global Pruning)**
  Rapidly prunes the bottom (1-γ)% of unpromising candidate models at near-zero compute cost via intrinsic expressivity H_{det}.
- **Stage 2: ACDP-γ% (Local Re-ranking)**
  Allocates 5-epoch low-fidelity probing (Acc_5) to the top γ% remaining candidates for refined local re-ranking.

Core Technical Contributions

- **Dynamic Feature Modulation**
  First framework to introduce hypernet-driven architecture-conditioned parameter generation for dataset traits.
- **Elastic Compute Control**
  Provides a flexible control lever between computational budget and predictive accuracy.
- **State-of-the-Art Performance**
  Demonstrates superior generalization and search efficiency across 6 unseen target datasets.

Experimental Validation Across Benchmarks & Spaces



- **NAS-Bench-201 (Zero-Sample Target)**
  ACDP-zero reaches Kendall's τ = 0.633 on CIFAR-10 without touching target dataset training samples (+66% over MetaD2A).
- **NAS-Bench-201 (Full Hybrid)**
  ACDP-all achieves τ = 0.759 (CIFAR-10) and 0.721 (CIFAR-100), outperforming single-dataset SOTAs while saving 3.5 GPU days.

Experimental Validation Across Benchmarks & Spaces

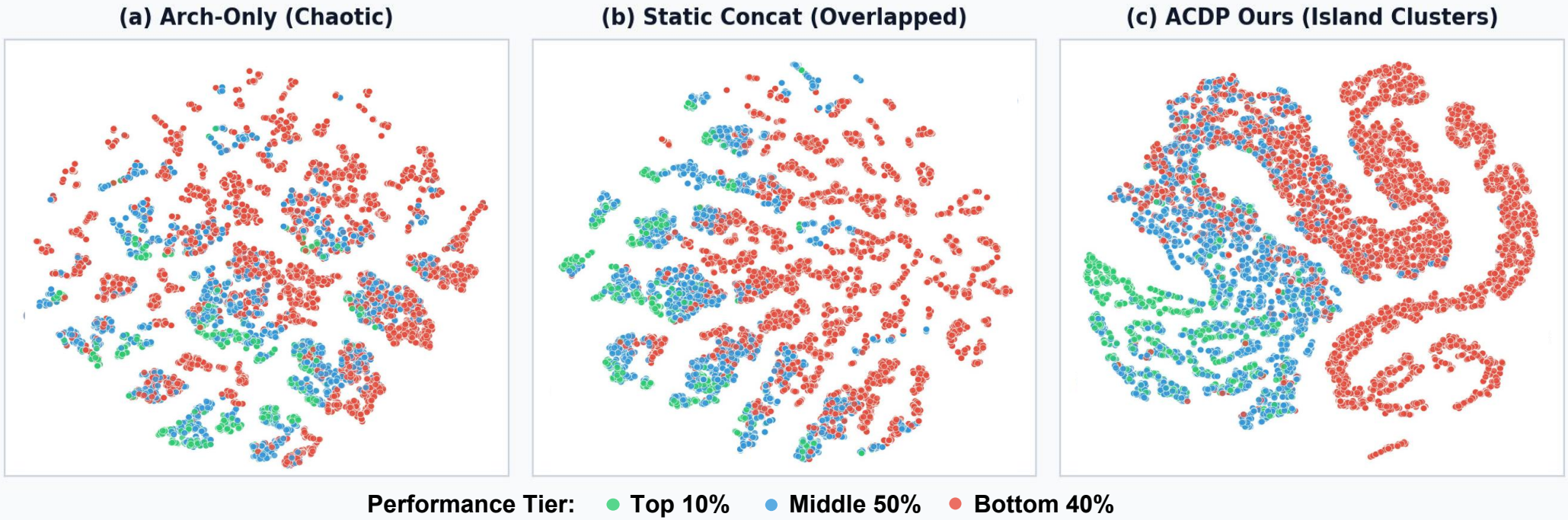
Table 1: Evolutionary search across six datasets in MobileNetV3.

Dataset	MetaD2A		DiffusionNAG		ACDP-zero		ACDP-10%	
	T (s)	A (%)	T (s)	A (%)	T (s)	A (%)	T (s)	A (%)
CIFAR-10	71	97.70±0.11	54	97.88±0.11	**51**	97.92±0.06	22k	**98.04±0.04**
CIFAR-100	82	85.57±0.21	61	85.93±0.18	**55**	86.22±0.15	24k	**87.45±0.12**
MNIST	68	99.54±0.05	41	99.56±0.04	**42**	99.63±0.04	8.2k	**99.66±0.08**
SVHN	73	96.74±0.08	50	96.82±0.10	**49**	96.98±0.14	26k	**97.11±0.12**
Aircraft	89	77.47±0.35	66	77.53±0.41	**61**	78.68±0.07	44k	**79.57±0.02**
Pets	62	93.96±0.15	54	94.14±0.22	**56**	94.51±0.13	24k	**94.77±0.11**

- **MobileNetV3 Space**

Evaluated on the ultra-large MobileNetV3 space across six datasets, ACDP-zero outperforms cross-dataset SOTAs with minimal search latency. ACDP-10% establishes a new state of the art and gains +2.04% top-1 accuracy on FGVC-Aircraft over cross-dataset SOTAs across evolutionary trajectories.

Qualitative Visualization: Progressive t-SNE Evolution



(a) Arch-Only Features

Chaotic and tangled embedding distribution across accuracy tiers.

(b) Static Concat

Weak linear stratification with heavy overlapping class boundaries.

(c) ACDP (Ours)

Reconstructs latent space into distinct, 'island-like' performance clusters.

Conclusion: Enabling Efficient Cross-Dataset NAS

1. Resolves Task-Insensitivity

Dynamic parameter generation conditioned on network topology overcomes static feature fusion limitations.

2. Drastic Cost Savings

Two-stage hierarchical strategy reduces search time from GPU-days to seconds while preserving high ranking accuracy.

3. Universal Applicability

Demonstrates true 'Train Once, Generalize Everywhere' capacity across diverse vision and graph search spaces.

Open Source Code & Benchmark Datasets: https://github.com/Thorn-222/ACDP

Thank you for watching.