Rapid Neural Architecture Search by Learning to Generate Graphs from Datasets

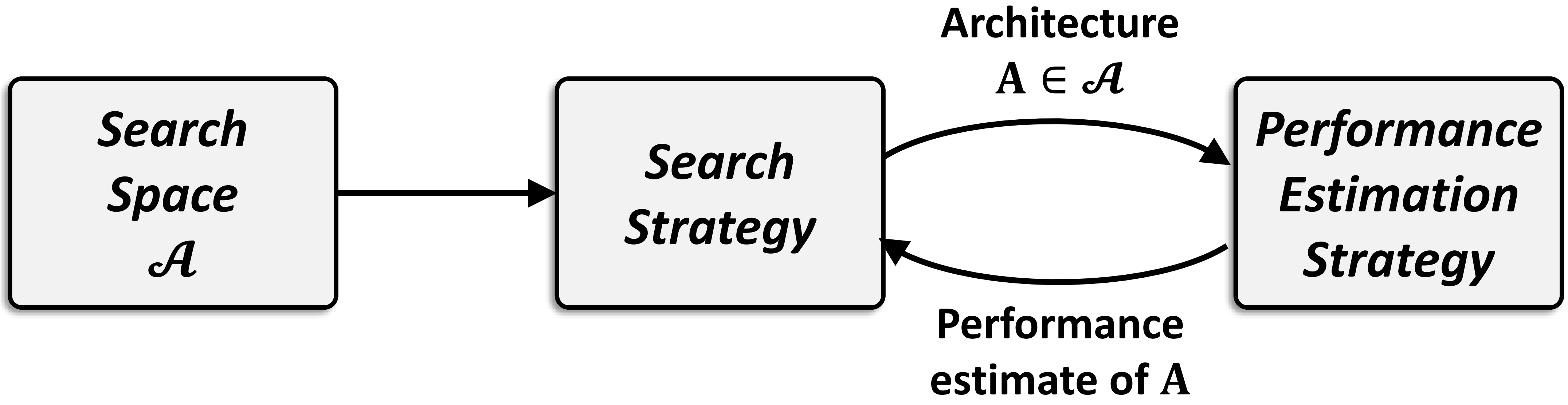
ICLR 2021

Hayeon Lee[1*], Eunyoung Hyung[1*], Sung Ju Hwang[1 2]

KAIST[1], AITRICS[2]

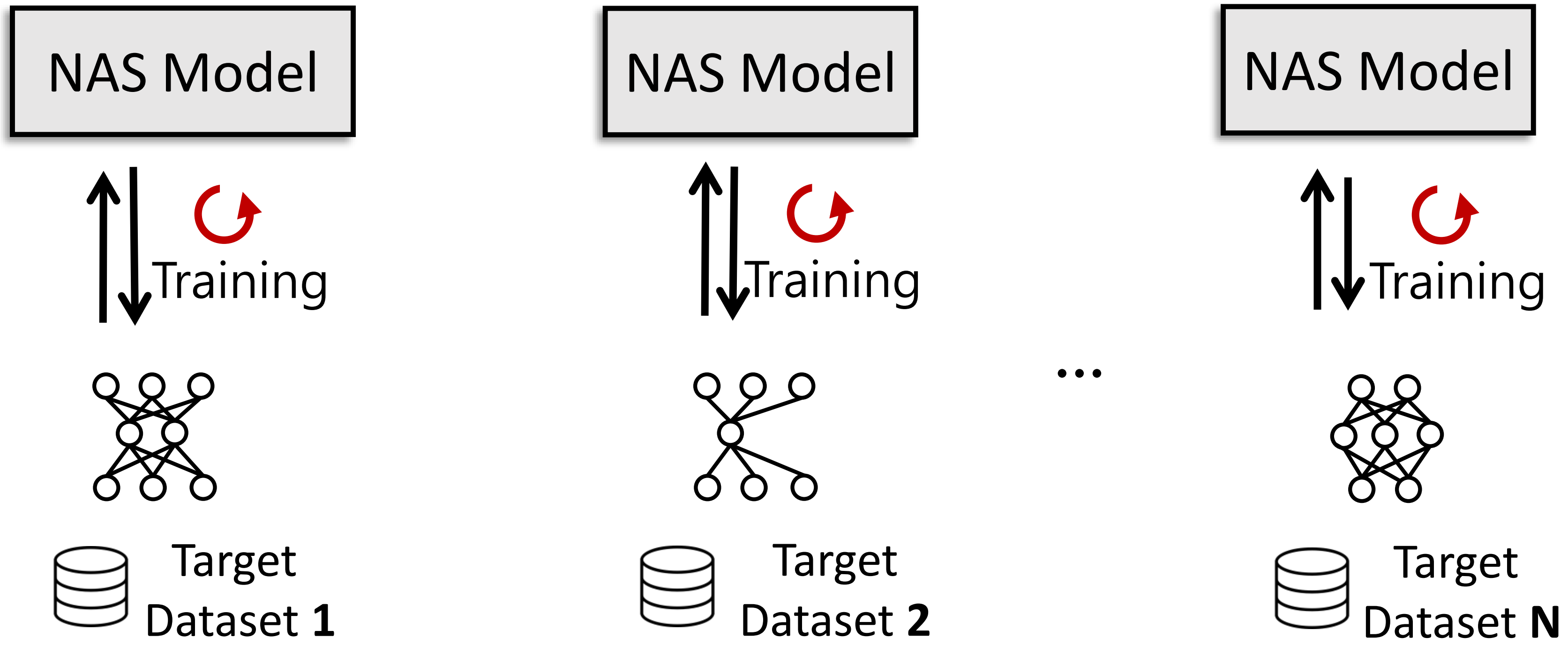
Neural Architecture Search (NAS)

Neural Architecture Search (NAS) is an **automated** architecture search process that aims to overcome the suboptimality of manual architecture designs.



Conventional Task-specific NAS Approach

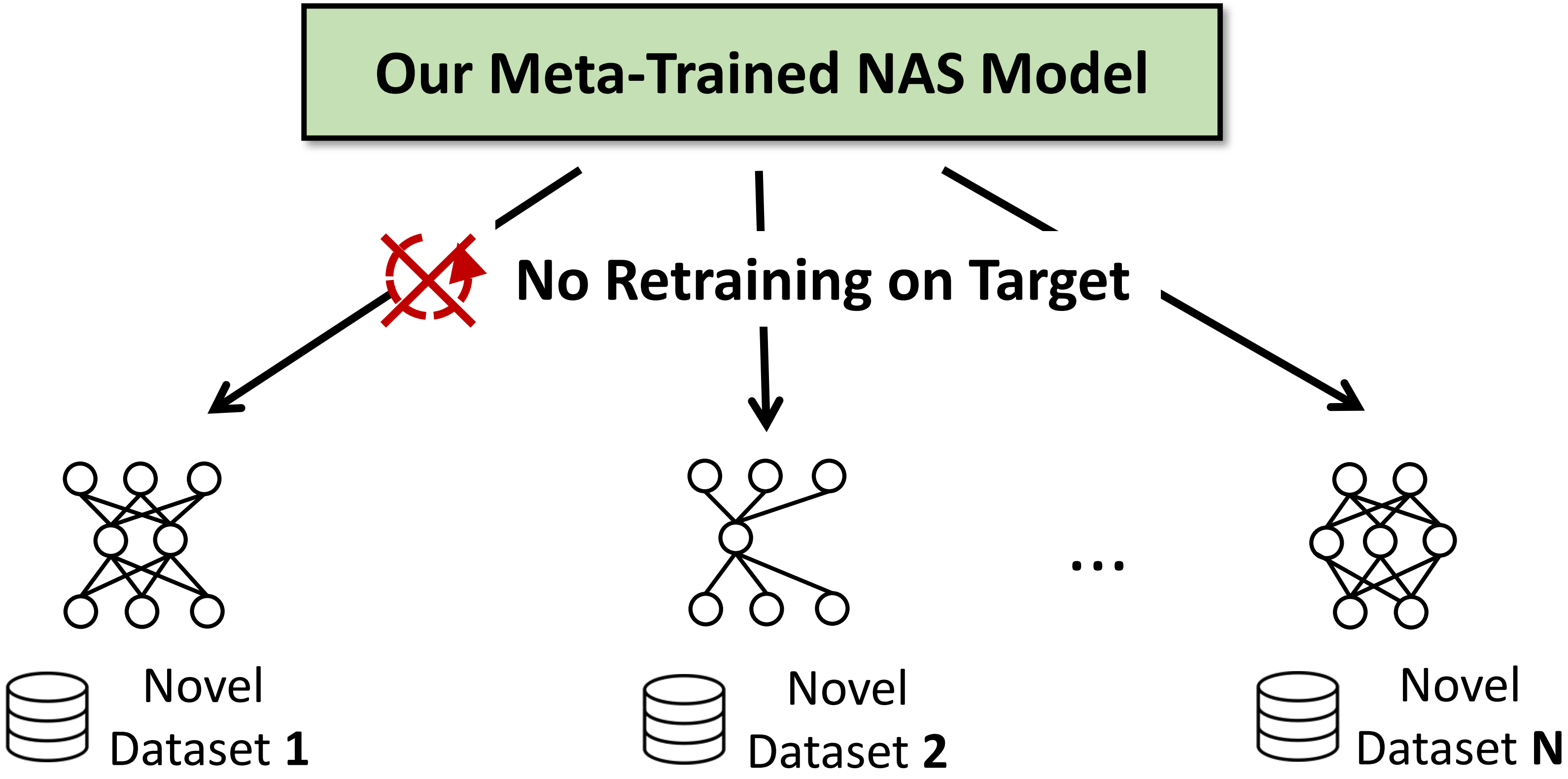
The conventional **task-specific** NAS approaches search from scratch for every given **single** dataset require a **huge** computational cost.



Search Cost O(N)

MetaD2A: Rapid Neural Architecture Search

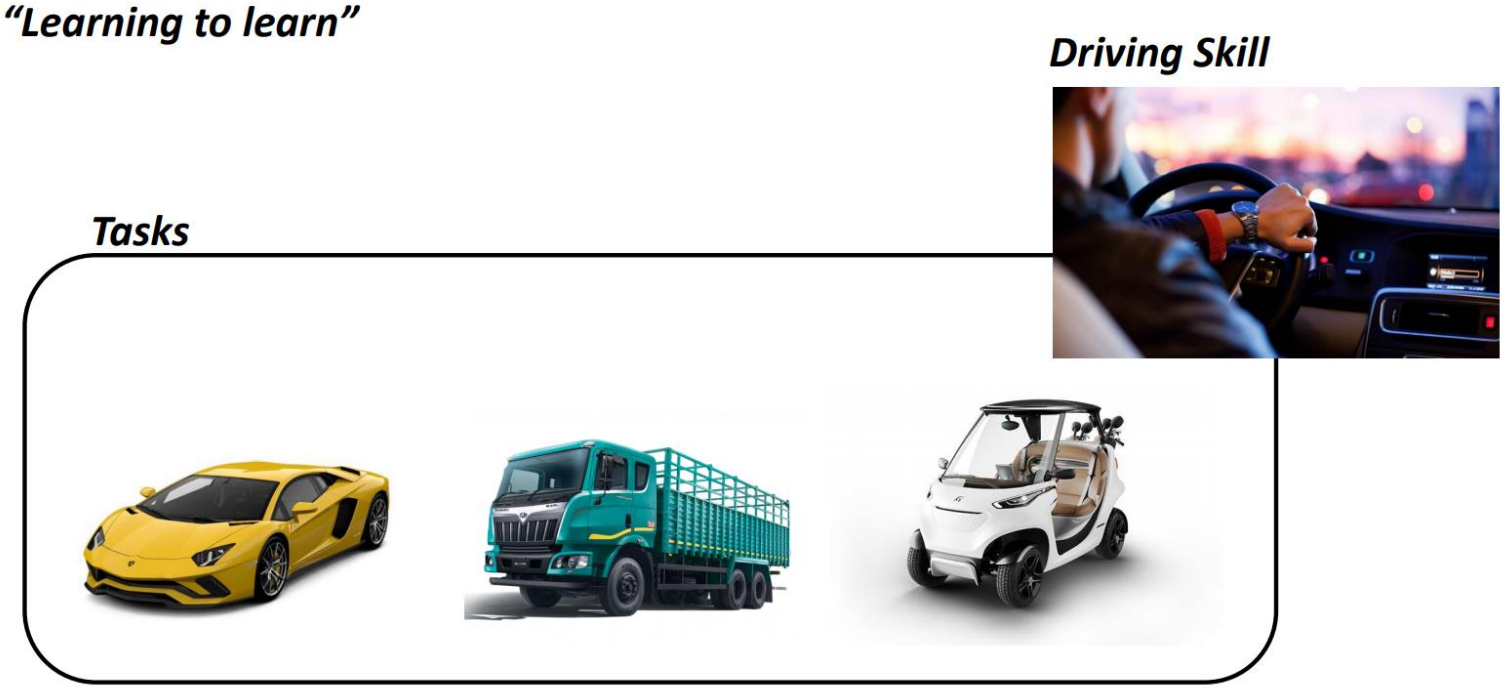
We propose MetaD2A that is meta-trained once on a database and can **rapidly** search for a neural architecture on a **novel** dataset **without additional training**.



Search Cost O(1)

Meta-learning

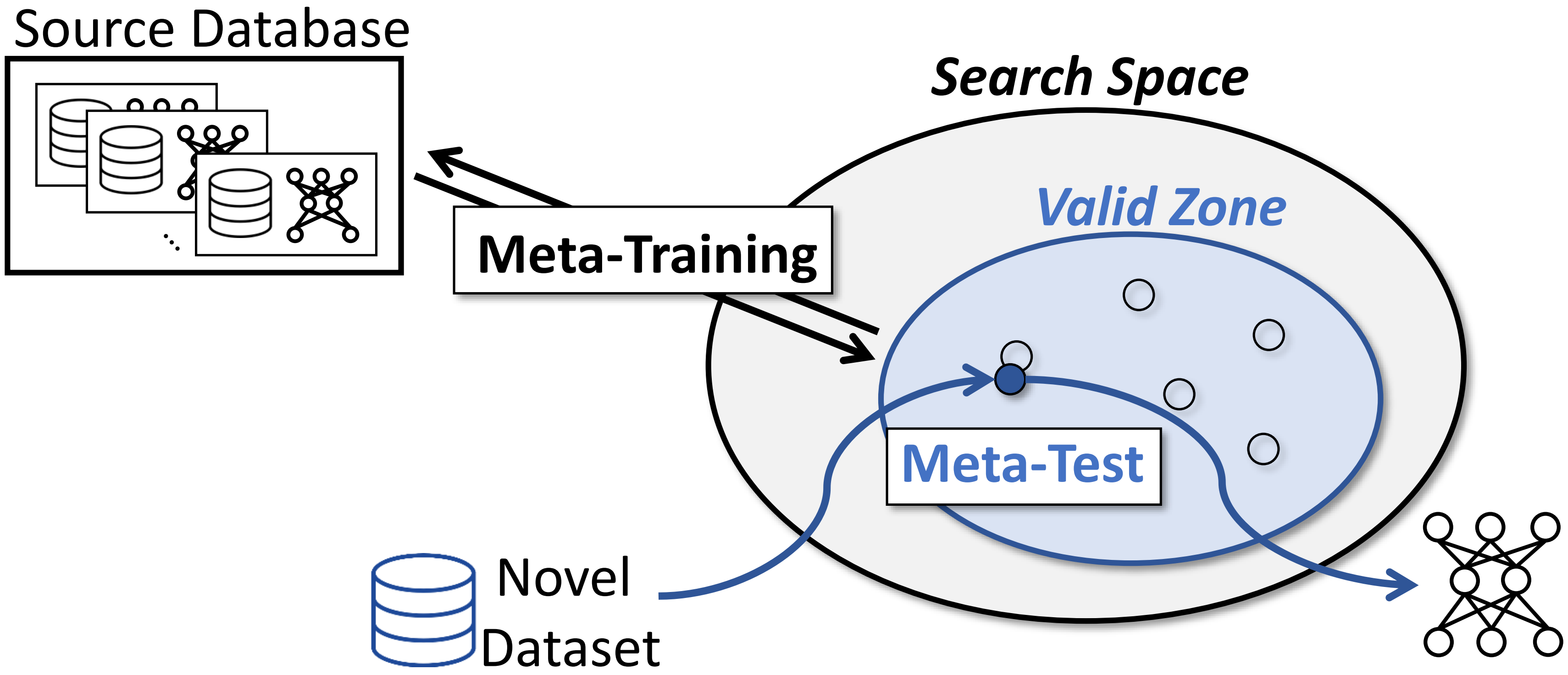
Meta-learning learns a model to **generalizes** over multiple tasks.



We adopt amortized meta-learning which utilizes a set encoder to perform learning at the task-level, considering each task as a data instance and minimizing loss over it.

Amortized Meta-learning Framework for NAS

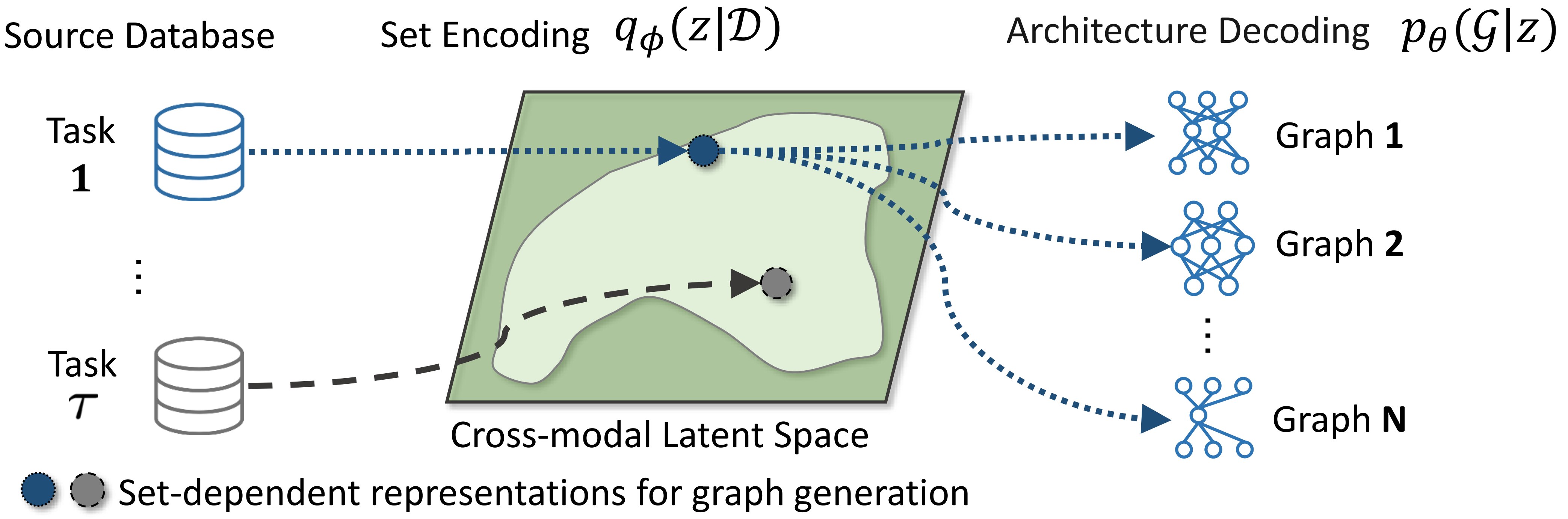
MetaD2A learns a **cross-modal latent space** of datasets and architectures via **amortized meta-learning** on source database.



For a novel dataset, we generalize amortized meta-knowledge of the cross-modal latent space to search for an architecture.

Learning to Generate Graphs from Datasets

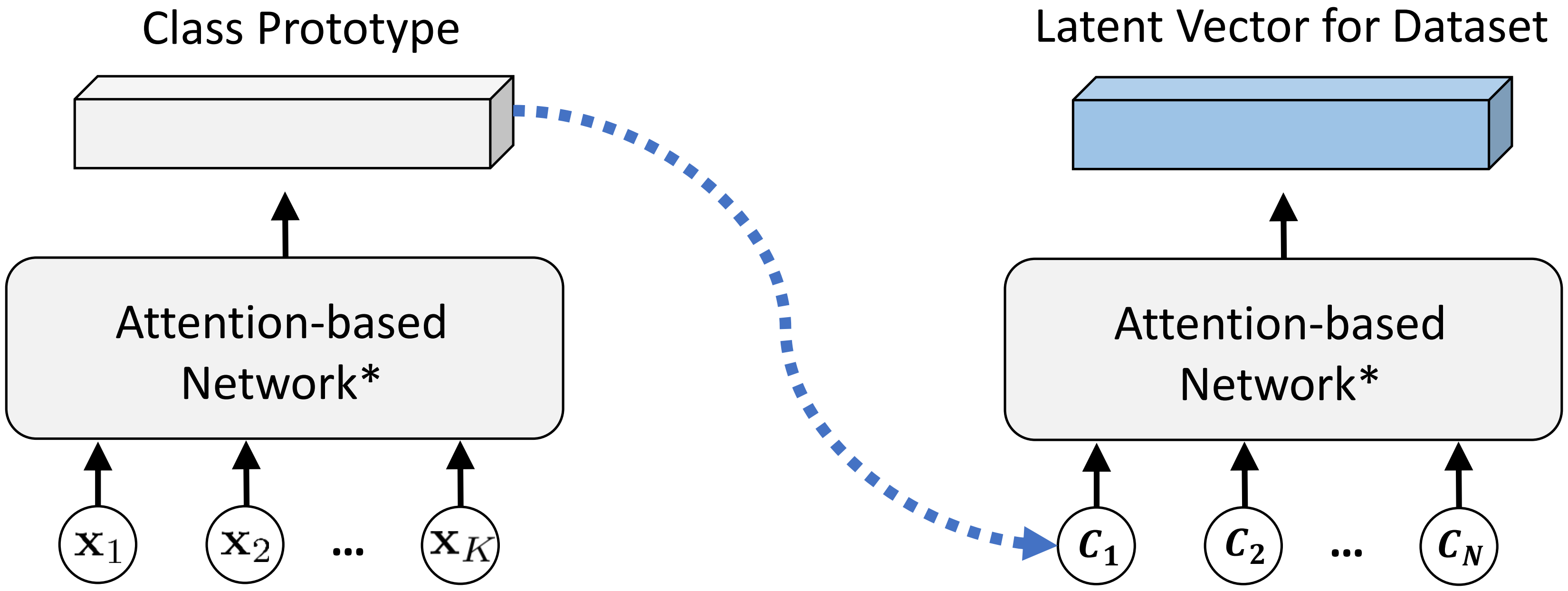
The set-to-architecture generator encodes datasets as **set-dependent** representations for **graph generation** and stochastically decodes it as graphs (architectures).



Objective: $$\max_{\phi,\theta} \sum_{\tau\sim p(\tau)} \mathcal{L}^{\tau}_{\phi,\theta}(\mathcal{D},\mathcal{G}) = \max_{\phi,\theta} \sum_{\tau\sim p(\tau)} \mathbb{E}_{z\sim q_\phi(z|\mathcal{D})}[\log p_\theta(\mathcal{G}|z)] - \lambda \cdot L^{T}_{KL}[q_\phi(z|\mathcal{D})||p(z)]$$

Hierarchical Set Encoder

We introduce a novel set encoder which is **permutation-invariant** with attention-based learnable parameters.

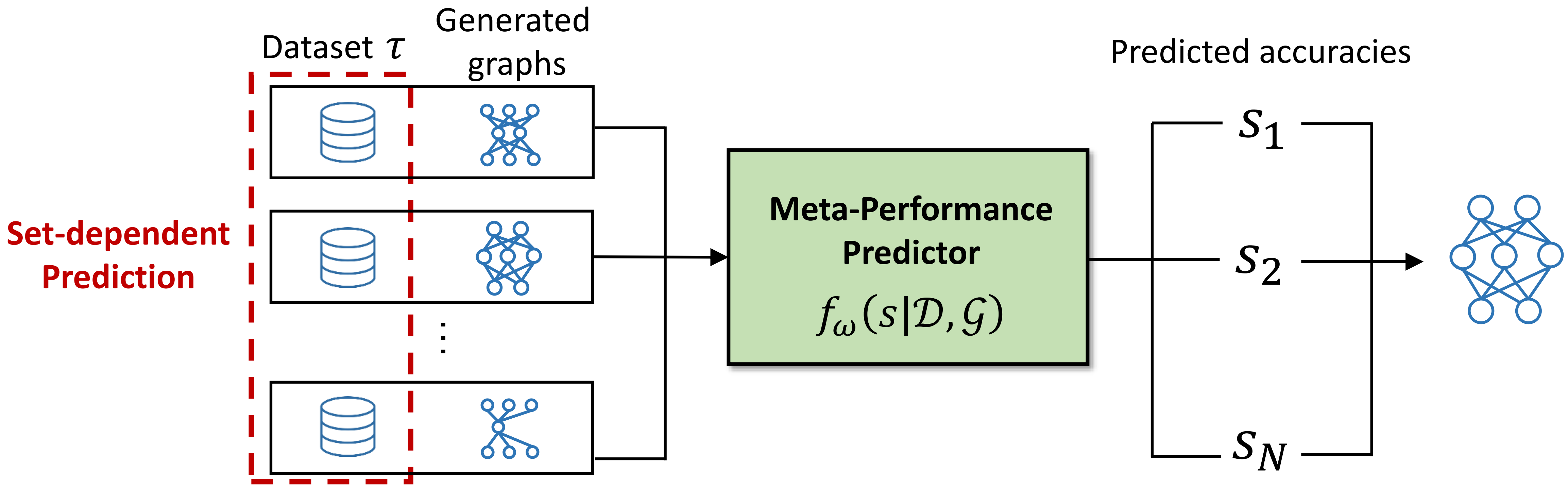


It can capture both the class prototypes that reflect label information and the high-level inter-class relationship between class prototypes.

*[SetTransformer] Lee, J., et al. Set transformer: A framework for attention-based permutation-invariant neural networks. ICML 2019.

Meta-Performance Predictor

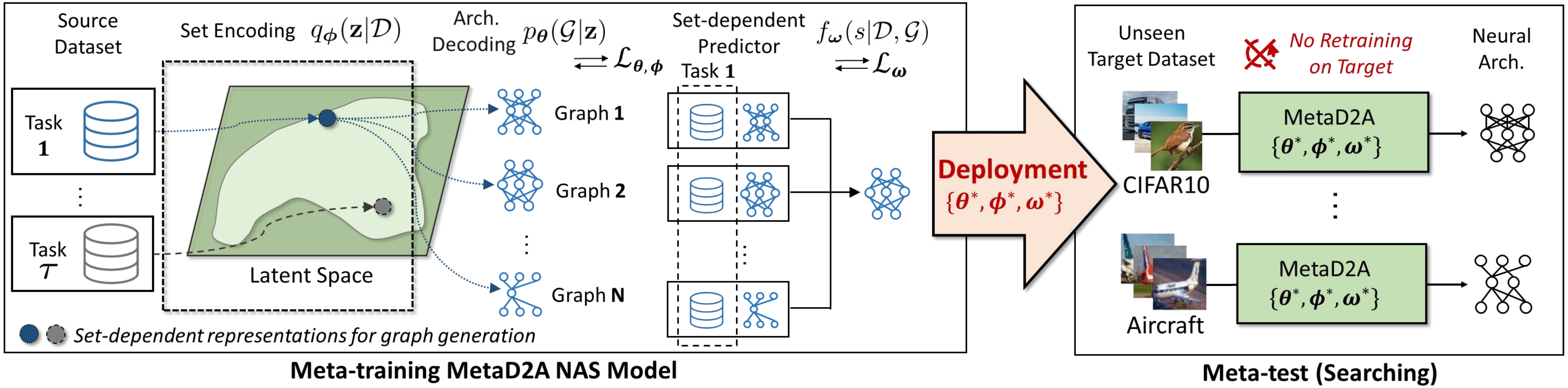
While the existing performance predictor takes a graph only, the proposed meta-performance predictor takes a **dataset** as well as graph to **support multiple datasets**.



Objective: $\min_{\omega} \sum_{\tau \sim p(\tau)} \mathcal{L}_{\omega}^{\tau}(s,\mathcal{D},\mathcal{G}) = \sum_{\tau \sim p(\tau)} (s - f_\omega(\mathcal{D},\mathcal{G})^2)$

Overview of MetaD2A Framework

In the meta-test (searching) phase, we can **deploy** the meta-learned MetaD2A to output **set-specialized** neural architecture for new target datasets without additional training.



Performance on Unseen Datasets (Meta-Test)

Without direct NAS model training on target datasets, our model meta-learned on the source database can successfully **generalize** to 6 **unseen datasets**.

Meta-training	Subsets of ImageNet-1K and Architectures of NAS-Bench-201					
Meta-test	**CIFAR-10**	**CIFAR-100**	**MNIST**	**SVHN**	**Aircraft**	**Oxford-IIIT Pets**
ResNet	93.97	70.86	99.67	96.13	47.01	25.58
SETN	87.64	59.09	99.69	96.02	44.84	25.17
GDAS	93.61	70.70	99.64	95.57	53.52	24.02
PC-DARTS	93.66	66.64	99.66	95.40	26.33	25.31
DrNAS	94.36	**73.51**	99.59	96.30	46.08	26.73
MetaD2A (Ours)	**94.37**	**73.51**	**99.71**	**96.34**	**58.43**	**41.50**

MetaD2A clearly outperforms baseline NAS models on multiple unseen datasets.

Search Time (GPU sec) on Unseen Datasets

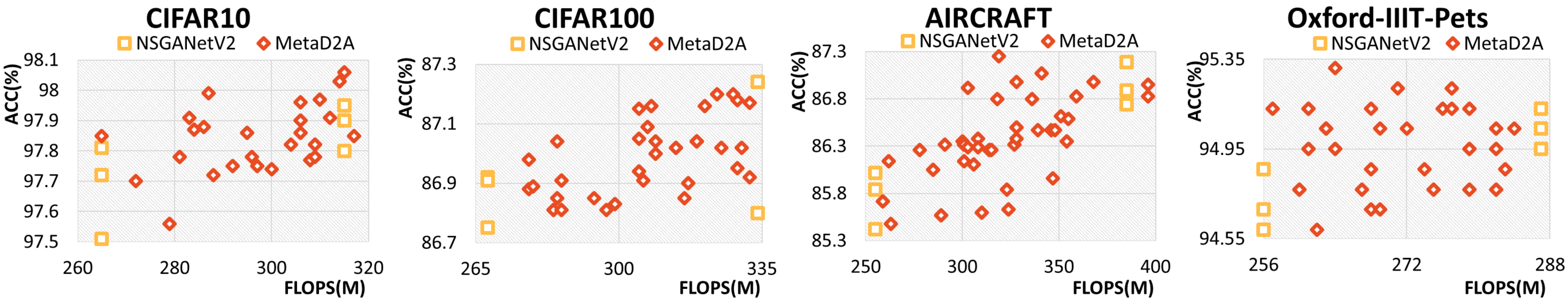
The search time of MetaD2A on unseen datasets is within **33 GPU seconds** on average with a single 2080ti GPU.

Meta-training	Subsets of ImageNet-1K and Architectures of NAS-Bench-201					
Meta-test	CIFAR-10	CIFAR-100	MNIST	SVHN	Aircraft	Oxford-IIIT Pets
RSPS	10200	18841	22457	27962	18697	3360
SETN	30200	58808	9950	85189	18564	8625
GDAS	25077	51580	60186	71595	18508	6965
PC-DARTS	10395	19951	24857	31124	3524	2844
DrNAS	21760	34529	44131	52791	34529	6019
MetaD2A (Ours)	**69**	**96**	**7**	**7**	**10**	**8**

This supports **realistic scenarios** that users with lack of computing resource get architectures suitable for **their own datasets** rapidly by using MetaD2A.

Evaluation in MobileNetV3 Search Space

MetaD2A reduces search time **5,523 times** on average while showing competitive performance compared with transfer NAS (NSGANetV2) on MobileNetV3 search space.



Summary

- We propose an efficient NAS framework (MetaD2A) which **rapidly** searches for a neural architecture on a **new dataset**.
- To this end, we propose to learn a **cross-modal latent space** of datasets and architectures by **amortized meta-learning** of it on subsets of ImageNet 1K.
- The meta-learned our model successfully **generalizes** to search for architectures on **unseen** datasets and shows rapid search time on them.