A General-Purpose Transferable Predictor for Neural Architecture Search

Fred X. Han[1], Keith G. Mills[2], Fabian Chudak[1], Parsa Riahi[3], Mohammad Salameh[1], Jialin Zhang[4], Wei Lu[1], Shangling Jui[4] and Di Niu[1]

[1]Huawei Technologies Canada, [2]University of Alberta, [3]University of British Columbia, [4]Huawei Kirin Solution

Understanding and modelling the performance of neural architectures is key to Neural Architecture Search (NAS).

Recently, performance predictors have seen widespread use in low-cost NAS by learning to predict neural architecture performance.

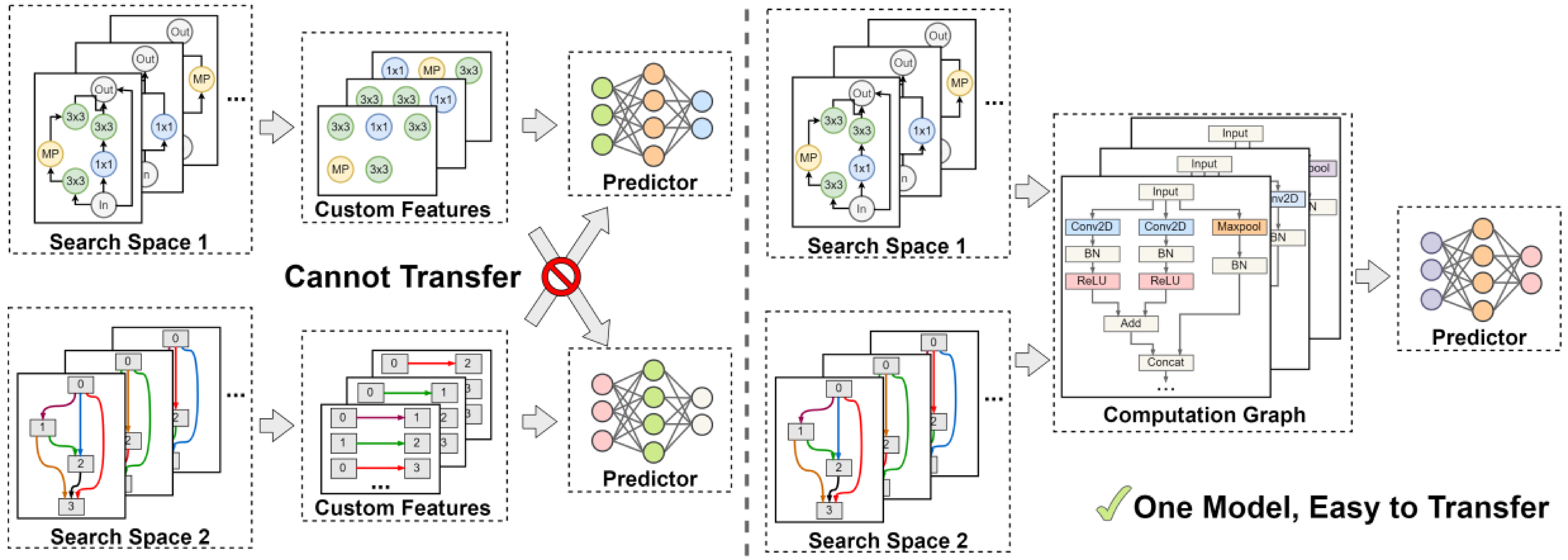
However, existing predictors are designed based on network encodings specific to predefined search spaces and are not transferable outside of specific NAS benchmark datasets.

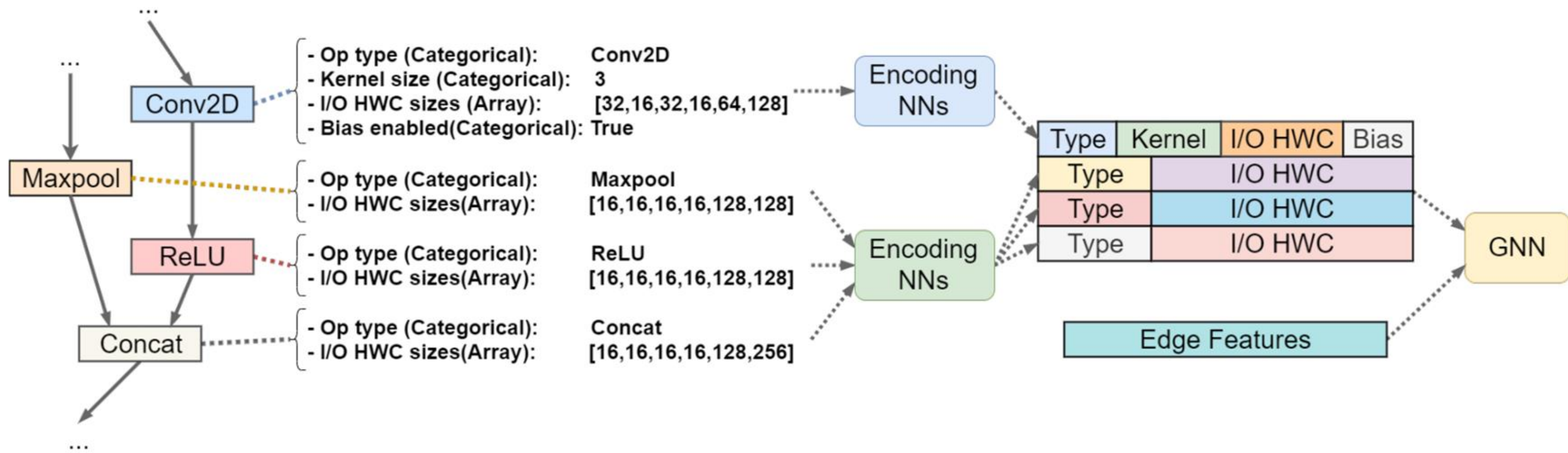
We propose a general-purpose neural predictor for NAS that can transfer across search spaces, by representing any given candidate CNN with a Computational Graph (CG) of primitive operations.

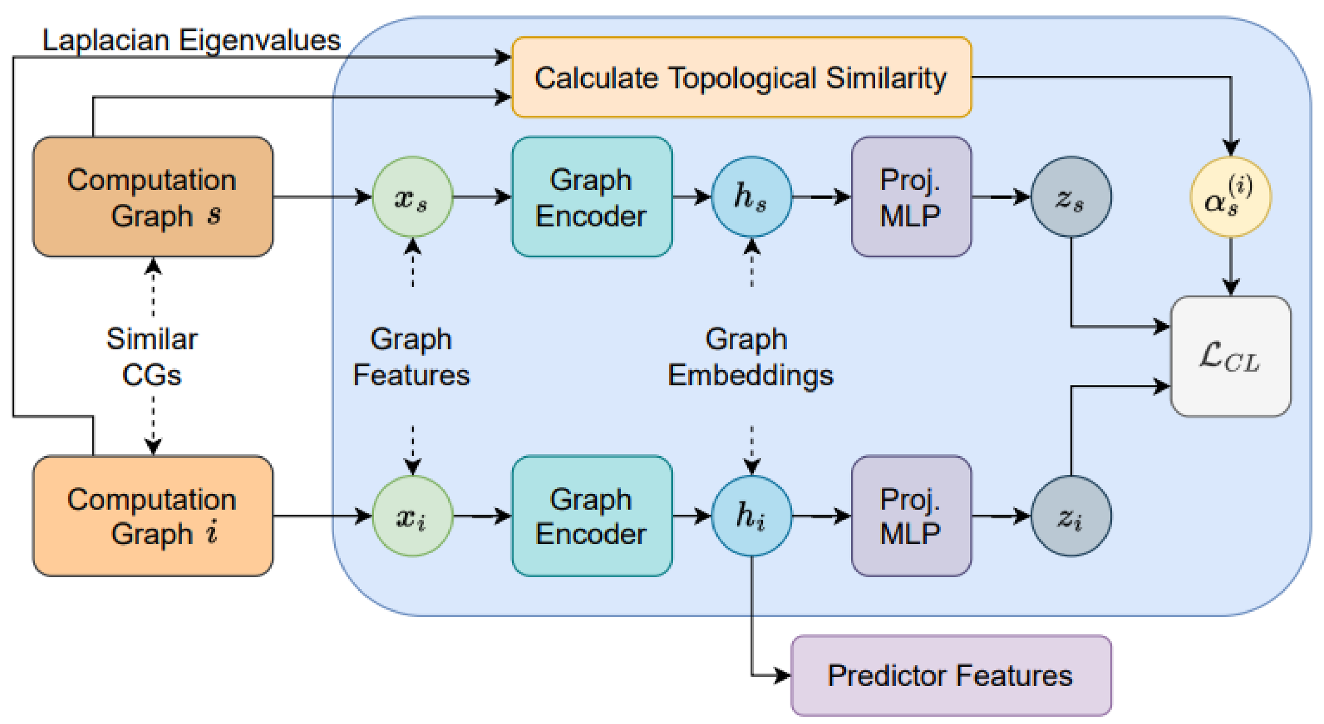
We further combine our CG network with semi-supervised Contrastive Learning (CL) graph encoder that leverages structural information from Laplacian Eigenvalues of unlabeled CGs.

Experimental results demonstrates that our method can find better, and more consistent ranking performance metrics than Zero-Cost Proxies (ZCP) on several benchmarks.

Moreover, we can find high-performance architectures on several benchmarks, including a MobileNetV3 architecture with 79.2% ImageNet







GRAPH REPRESENTATION WITH CONTRASTIVE LEARNING

Contrastive Learning trains a data encoder such that similar input data samples (e.g., *i and j*) have similar latent embeddings:

$$\chi_{i,j} = \log \frac{\exp(sim(z_i, z_j))}{\sum_{r \neq i} \exp(sim(z_i, z_r))}$$

We use Laplacian Eigenvalues, computed from a graph's adjacency matric, to compute a *spectral* between CGs:

$$Distance(i,j) = \sigma_s(i,j)$$

$$\alpha_*^{(i)} = Softmax(\sigma_s(i,*))$$

From here we can compute a *Contrastive Loss:*

$$\mathcal{L}_{CL} = -\sum_{i \in I} \sum_{s \in P(i)} \alpha_s^{(i)} \chi_{i,s}$$

TRANSFERABILITY TO MOBILENETS

We consider a transferability test to MobileNets, specifically Once-for-All (OFA) MobileNetV3 (MBv3).

MobileNets train and evaluate on ImageNet, whereas other NAS-Bench-{101, 201, 301} consider CIFAR-10.

For this experiment, we train a CL graph encoder and predictor MLP head on NAS-Bench-{101, 201, 301} and then fine-tune it on 50 labeled samples from the OFA-MBv3 supernetwork.

Then, we pair our predictor with a simple, mutation-driven search algorithm that uses the OFA-MBv3 search space.

Using this method, can find an architecture that obtains 79.2% top-1 accuracy, outperforming the original OFA.

Model	Top-1 Acc. (%)
MobileNetV2 [24]	72.0
MobileNetV2 #1200 [24]	73.5
MobileNetV3-Large [12]	75.2
NASNet-A [36]	74.0
DARTS [16]	73.1
SinglePathNAS [11]	74.7
OFA-Large [3]	79.0
OFA-Base [19]	78.9
OFA-MBv3-CL	**79.2**

RANK CORRELATION AND NAS-BENCHMARK SEARCH PERFORMANCE

Method	NAS-Bench-101	NAS-Bench-201	NAS-Bench-301
Synflow [25]	0.361	0.823	-0.210
Jacov [18]	0.358	0.859	-0.190
Fisher [27]	-0.277	0.687	-0.305
GradNorm [2]	-0.256	0.714	-0.339
Grasp [29]	0.245	0.637	-0.055
Snip [14]	-0.165	0.718	-0.336
GNN-fine-tune	0.542 ± 0.14	0.884 ± 0.03	0.872 ± 0.01
CL-fine-tune	**0.553** ± 0.09	**0.917** ± 0.01	**0.892** ± 0.01

For each target benchmark, we train an encoder on unlabeled data, then train a predictor MLP using data from other benchmarks, then fine-tune on some limited target data.

We compare to Zero-Cost Proxies (ZCP), which are generalizable, and a simple end-to-end GNN predictor baseline.

Our semi-supervised CL encoder achieves higher rank correlation.

Method	NAS-Bench-101			NAS-Bench-201			NAS-Bench-301	
	#Q	Acc. (%)	Rank	#Q	Acc. (%)	Rank	#Q	Acc. (%)
Random	700	94.11 ± 0.10	26.0	90	93.91 ± 0.2	104	800	94.75 ± 0.08
Synflow	700	94.18 ± 0.05	5.8	90	**94.37** ± 0.0	**1.0**	800	94.60 ± 0.11
CL-fine-tune	700	**94.23** ± 0.01	**2.2**	90	**94.37** ± 0.0	**1.0**	800	**94.83** ± 0.06

We apply our method with an evolutionary search algorithm to find good architectures.

We find high-performance architectures on each NAS benchmark.

Registration and travel support for this presentation was provided by the Society for Industrial and Applied Mathematics.