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[2]

[1]Huawei Technologies Canada Co., Ltd.

[2]University of Alberta

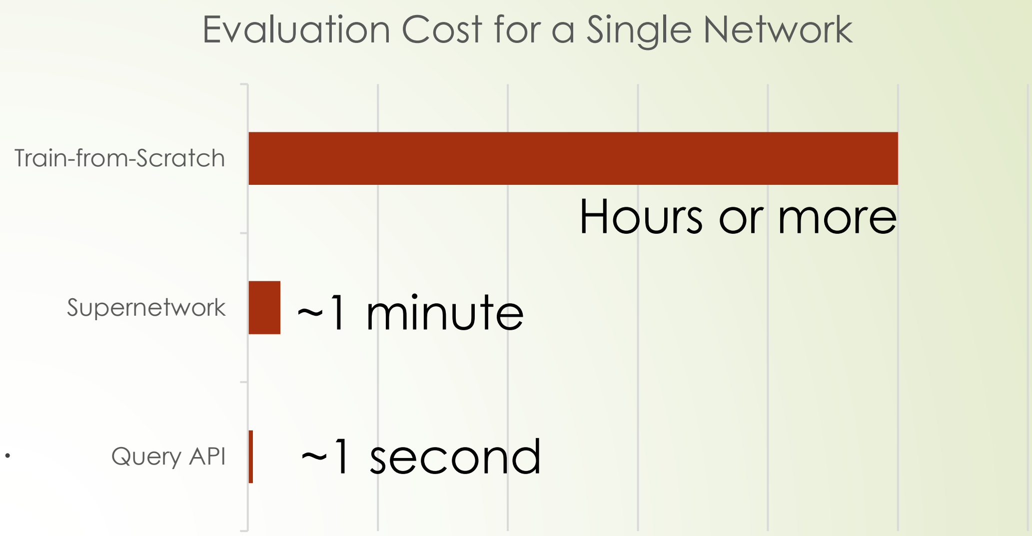
[3]University of British Columbia

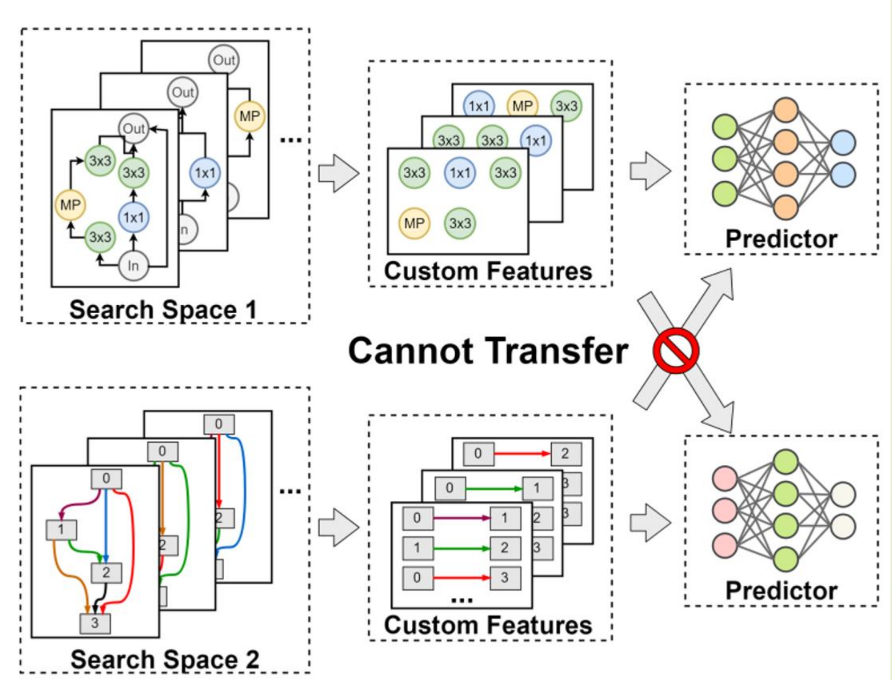
[4]Huawei Kirin Solution, Shanghai, China

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

Motivation

- Neural Architecture Search (NAS) is about finding high-performance neural networks.
- Early NAS methods have high search cost:
- There were several responses to this cost:
  - Supernetworks show significant speed-up.
  - Tabular NAS-Benchmarks with known performance are faster.
  - Low-cost performance-estimating predictors build atop of benchmarks.
- Existing performance predictors have a weakness:
  - Designed around benchmark-specific network encodings.
  - Lack of generalizability.

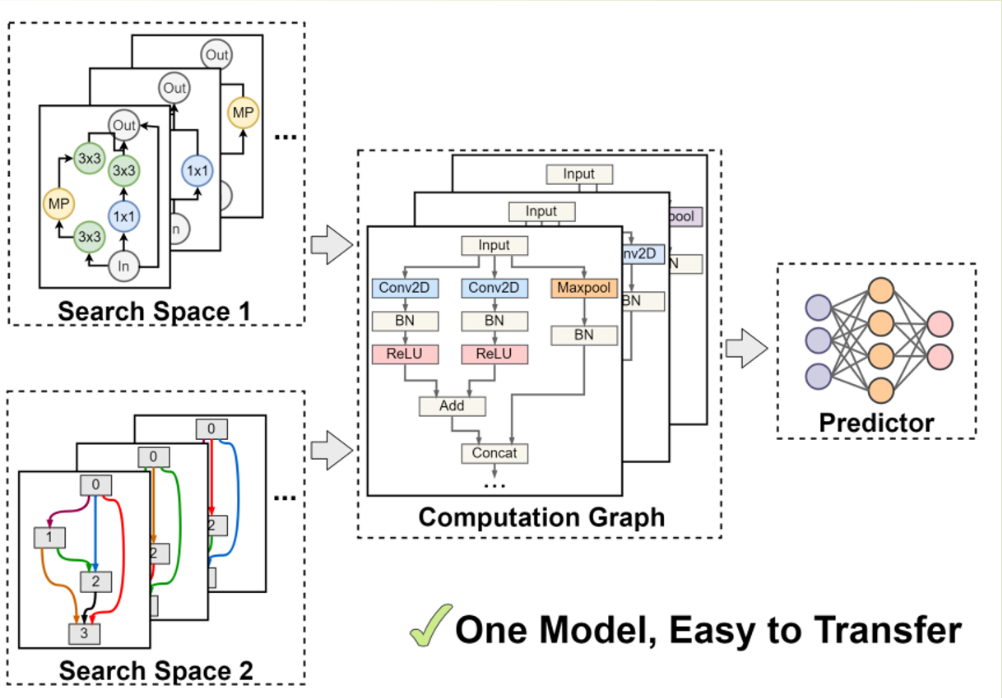




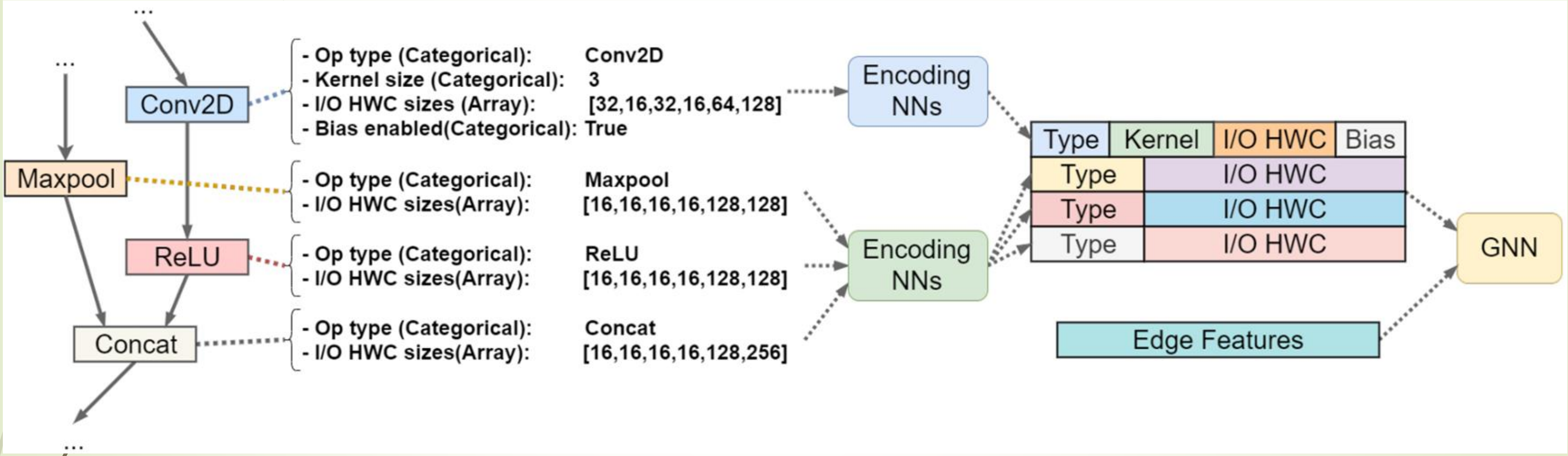
Contributions

A General-Purpose Transferable Predictor NAS:

- Transferable Computational Graph (CG) architecture representation.
- Semi-supervised graph encoder using Contrastive Learning (CL) with Laplacian Eigenvalues.
- Experimental findings demonstrate the benefits of our scheme:
  - Higher rank correlation than other generalizable prediction methods.
  - Find high-performance architectures on several benchmarks.
  - Find better architectures faster; using fewer queries than other neural predictors.
  - Extend beyond CIFAR-10 benchmarks to find a MobileNetV3 architecture with obtains 79.2% top-1 accuracy on ImageNet.



Our Approach: Computational Graphs



Two types of nodes: Regular and Weighted

- Regular: Do not have learnable parameters.
  - E.g., ReLU, pooling, concat or add.
  - Extract I/O HWC dimensions, and operation type.
- Weighted: Have weights.
  - Conv, MatMul, etc.
  - Extract further information like weight dimensions, bias Boolean.

HUAWEI HISILICON

Semi-Supervised Contrastive Learning

Idea of Contrastive Learning: Similar data samples should have similar representations:

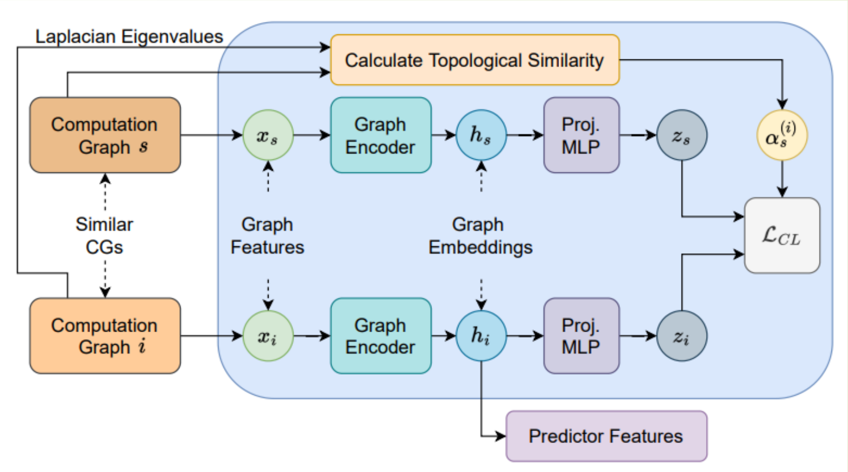
$$\chi_{i,j} = \log \frac{\exp(sim(z_i, z_j))}{\sum_{r \neq i} \exp(sim(z_i, z_r))}$$

Question? How to determine similarity for CGs?

- Our approach: Laplacian Eigenvalues from adjacency matrix, as a spectral distance:

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

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



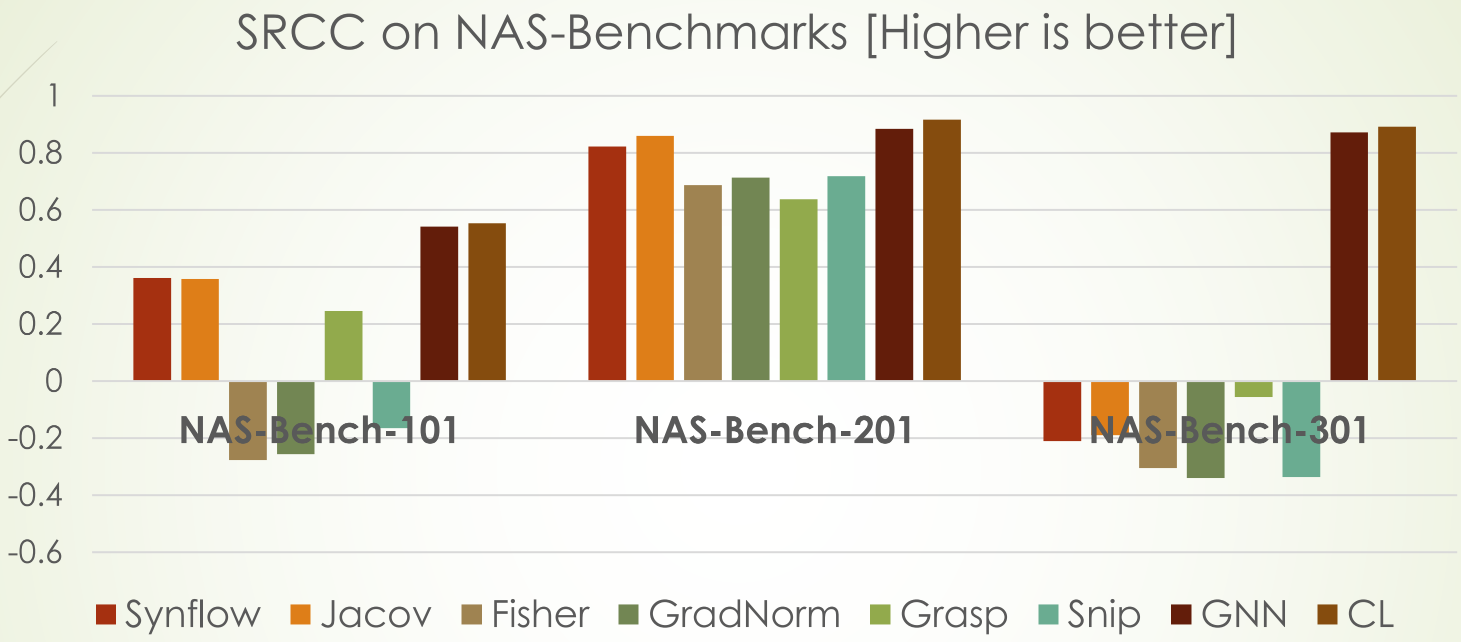
From here we can compute a *Contrastive Loss:*

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

Experimental Setup

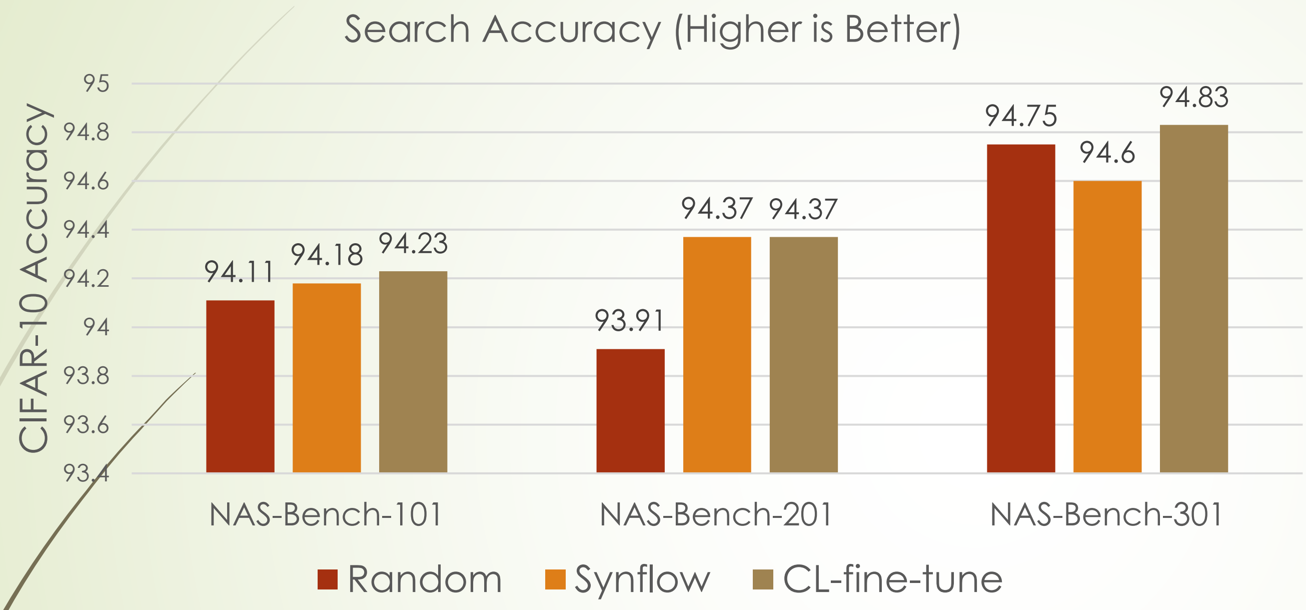
1. **NAS-Benchmark Rank Correlation Evaluation**
   - Train a CL graph encoder on unlabeled target benchmark dataset.
   - Then, train an MLP predictor head on labeled graphs from other benchmarks.
   - Fine-tune on some labeled samples from target dataset.
2. **NAS-Benchmark Search**
   - Pair our predictor with an evolutionary search algorithm.
   - Find high-performance architectures and compare with other predictors.
3. **Generalization to ImageNet Architectures**
   - Train a CL graph encoder on three NAS-Benchmarks.
   - Fine-tune on some ImageNet architectures and perform search.

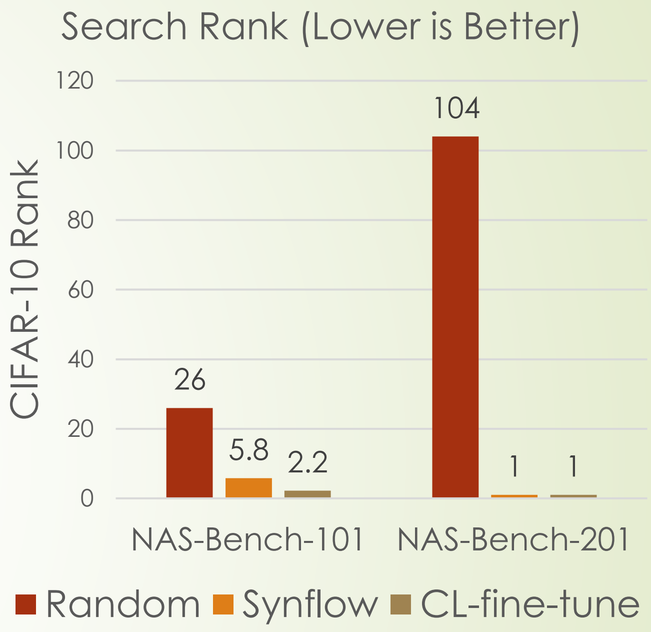
Ranking Correlation Test



- Compare to Zero-Cost Proxies (ZCP) [Abdelfattah et al., 2021]
  - ZCPs calculate gradient heuristics which can correlate with performance.
  - Also consider a simple end-to-end GNN baseline.
- Result: Our CL-based predictor finds higher rank correlation than ZCPs and the GNN.

Application Search

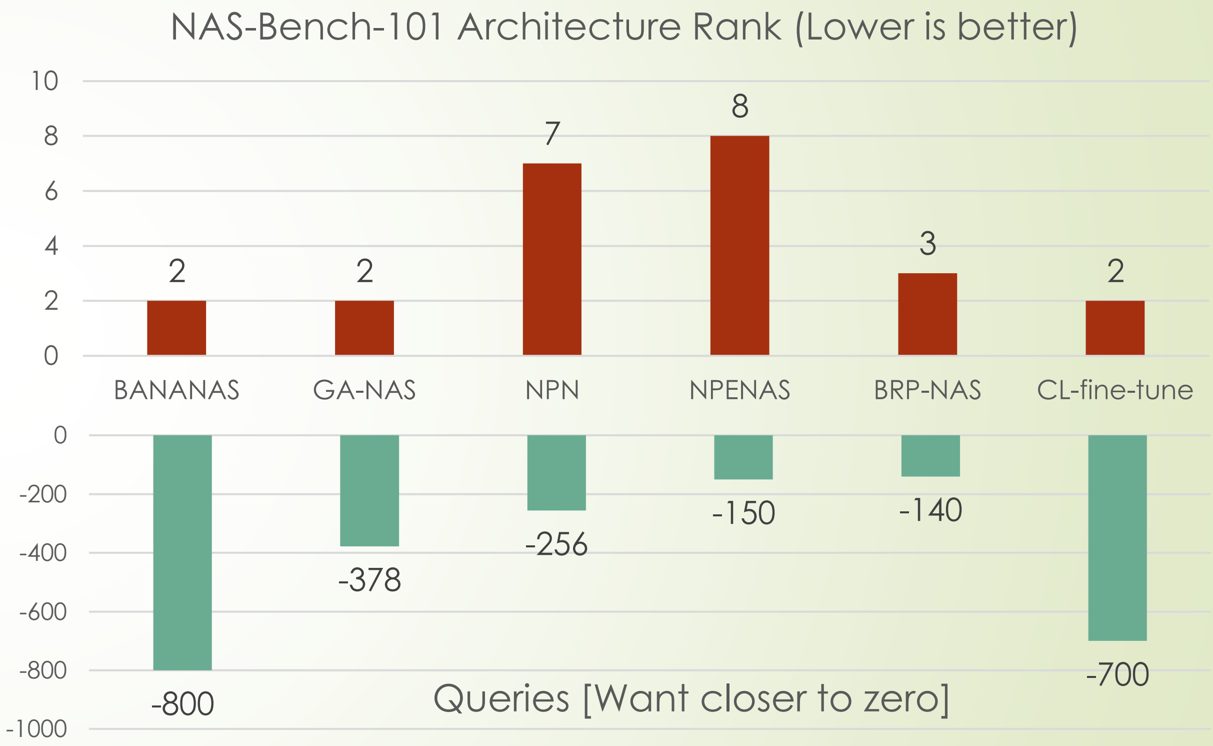




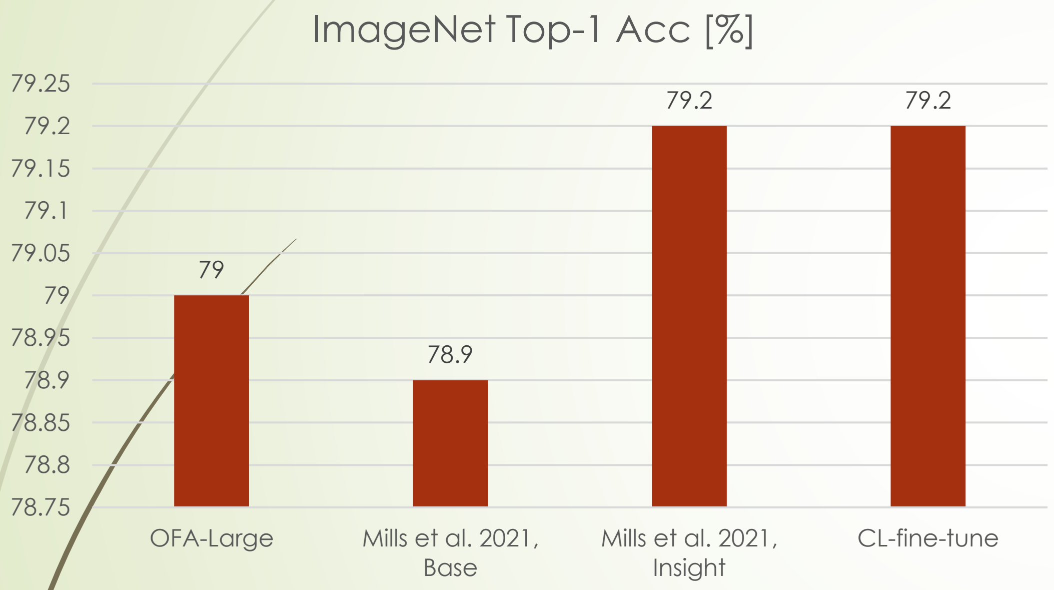
- Pair our CL graph encoder with a simple evo search algorithm and find high-performance architectures on NAS-Benchmarks.
- Compare to Random Search baseline and Synflow ZCP.
  - Find 94.23% acc/Rank 2 architecture on NAS-Bench-101.
  - Find the best NAS-Bench-201 architecture.
  - Achieves over 94.8% on NAS-Bench-301.

Comparison With Related Neural Predictors

- Many other single search space neural predictors in the literature.
- We compare with our method on NAS-Bench-101.
- Our method matches others like GA-NAS and BANANAS by finding the 2nd best architecture.
- When #queries are taken into consideration, we beat BANANAS. Only GA-NAS [Changiz Rezaei et al., 2021], our previous work, uses fewer queries.



Application to ImageNet

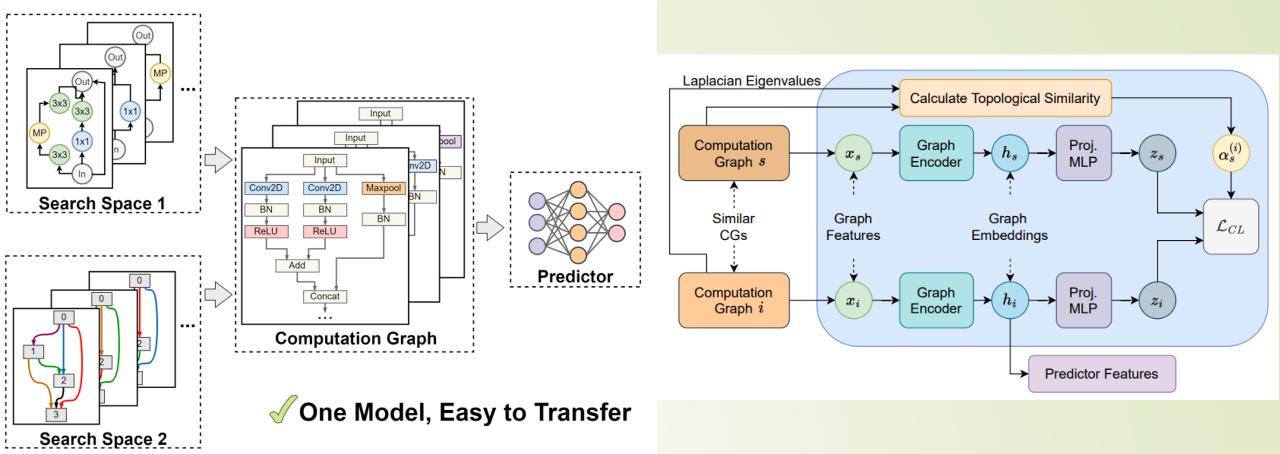


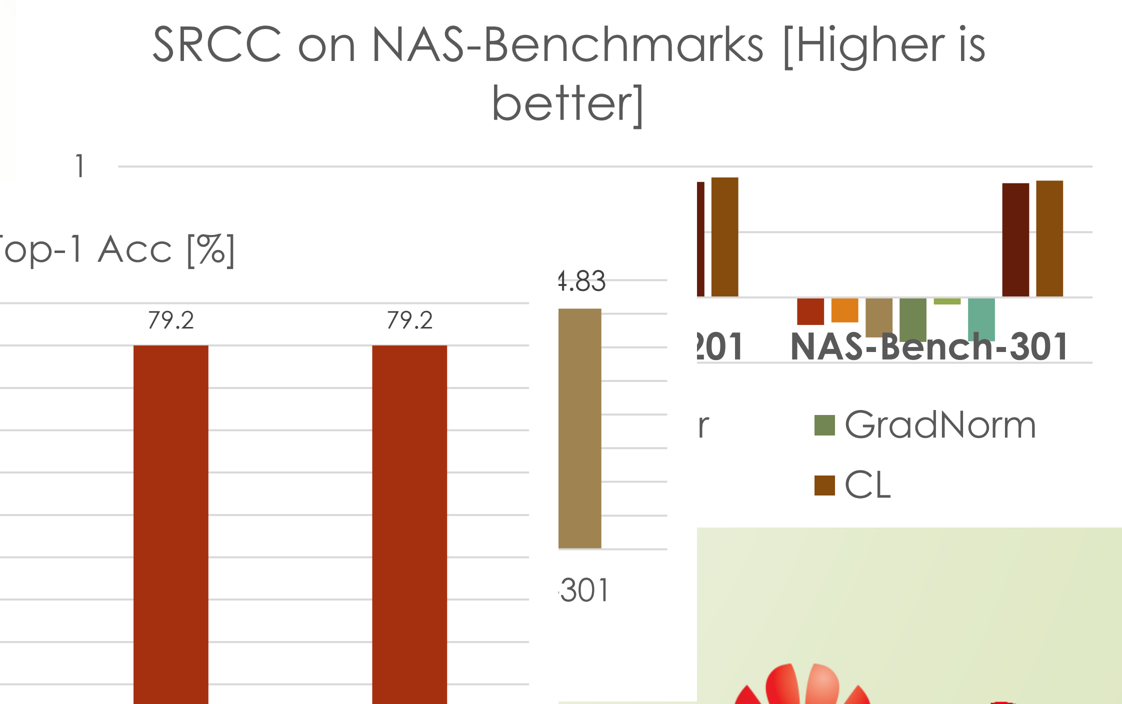
- Furthermore, our CG-based architecture representation is not limited to CIFAR-10.
- Perform search on Once-for-All (OFA) [Cai et al., 2020], MobileNetV3 (MBv3) for ImageNet.
- We train a CL graph encoder on samples from all 3 NAS-Benchmarks.
    - Then fine-tune on a small number of OFA-MBv3 CGs.
- Compare to our previous ADS track paper in CIKM-21 which evaluated using a supernetwork (1minute per arch).
- Results: We find an OFA-MBv3 architecture with 79.2% top-1 accuracy, outperforming the original OFA.

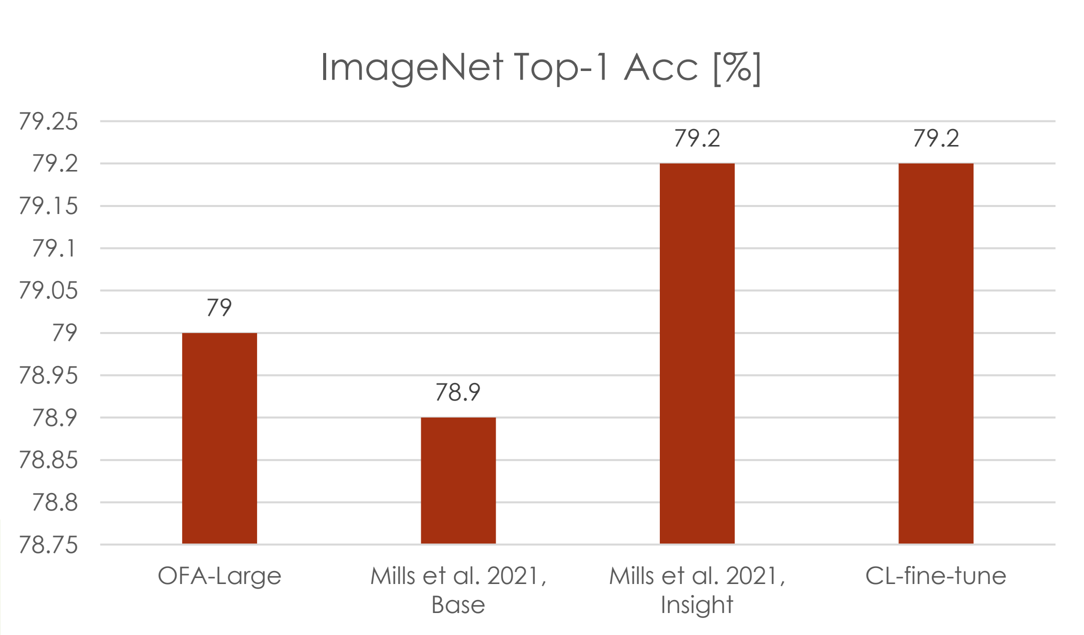
Conclusion

We propose a general-purpose predictor for NAS.

- Computational Graphs.
- Contrastive Learning
- Experiments:
  - SRCC
  - CIFAR-10 Search
  - ImageNet Search







References

- Abdelfattah et al., "Zero-Cost Proxies for Lightweight NAS." In ICLR 2021.
- Changiz Rezaei et al., "Generative Adversarial Neural Architecture Search" – IJCAI-21.
- Cai et al., "Once-for-All: Train One Network and Specialize it for Efficient Deployment" – ICLR 2020.
- Mills et al., "Profiling Neural Blocks and Design Spaces for Mobile Neural Architecture Search" – CIKM'21.

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