GENNAPE: Towards Generalized Neural Architecture Performance Estimators

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

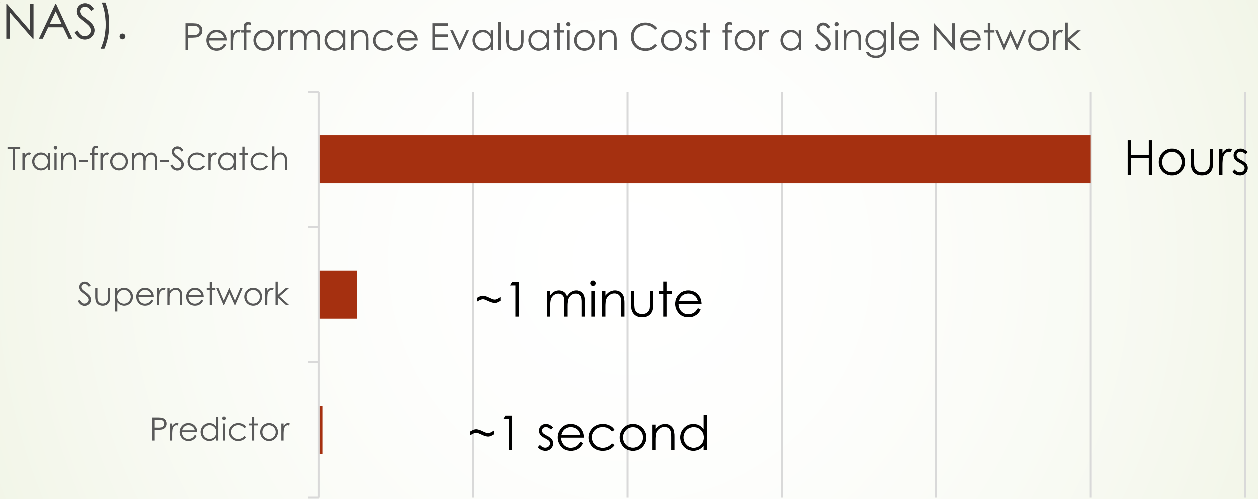
[1]University of Alberta

[2]Huawei Technologies Canada Co., Ltd.

[3]Huawei Kirin Solution, Shanghai, China

Motivation

- Performance Evaluation Strategy is a bottleneck of Neural Architecture Search (NAS).



- Neural predictors are low-cost by learning to estimate performance.
- However, most neural predictors are confined to specific search spaces. They lack generalizability!

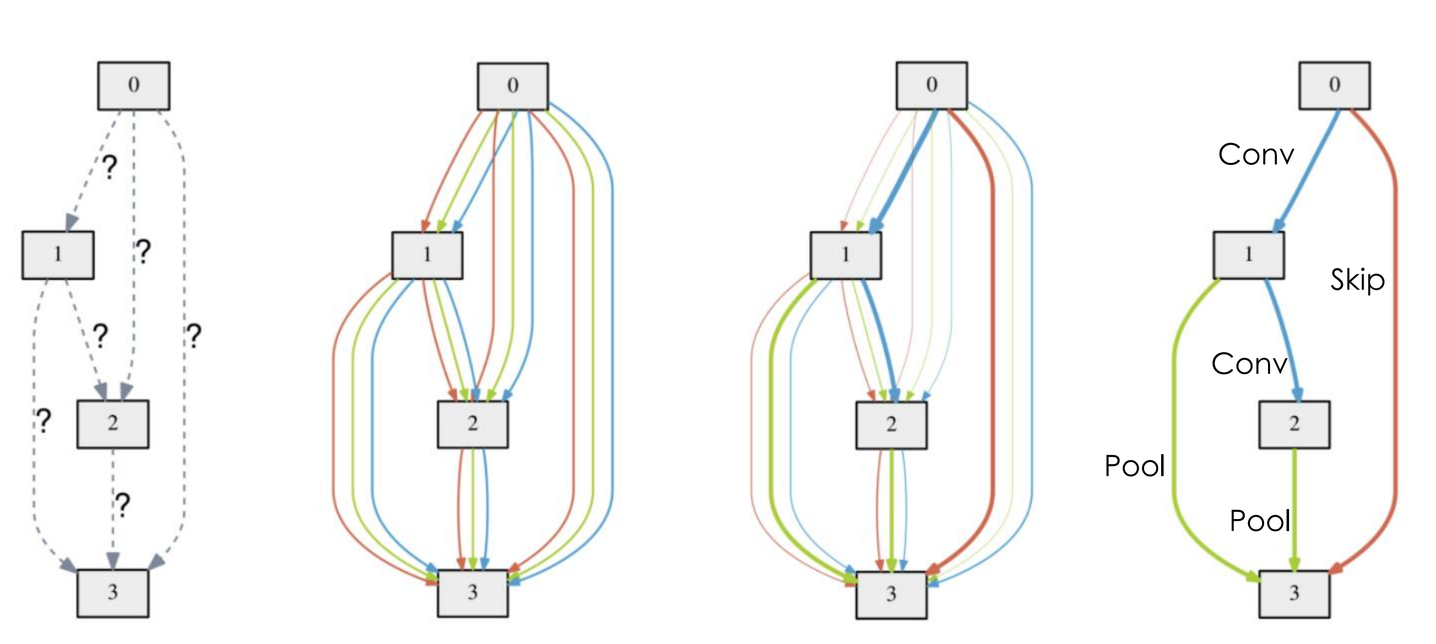
Contributions

We propose **GEN**eralized **N**eural **A**rchitecture **P**erformance **E**stimators; GENNAPE:

- **Generalizable Architecture Representation** using Computational Graphs (CG).
- **Pre-train** a graph encoder using self-supervised Contrastive Learning (CL).
- **Cluster** embeddings using Fuzzy C-Means (FCM) for a weighted ensemble.
- Introduce **open-source benchmark** families, HiAML, Inception and Two-Path.
- Verify the **transferability** of our scheme on known benchmarks.

Architecture Representation in Literature

- Micro Search
- Find some Directed Acyclic Graph structure of operations – *the cell.*

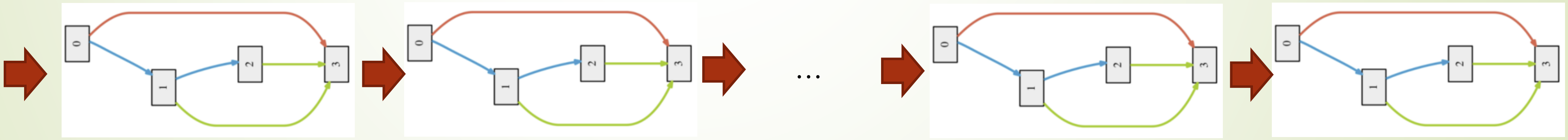


(a) (b) (c) (d)

Image from DARTS: "Differentiable Architecture Search" – ICLR 2019

Architecture Representation in Literature

- Micro Search
- Find some Directed Acyclic Graph structure of operations – *the cell.*
- Repeat many times.

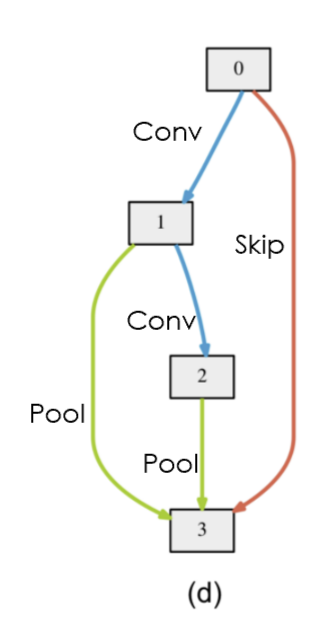


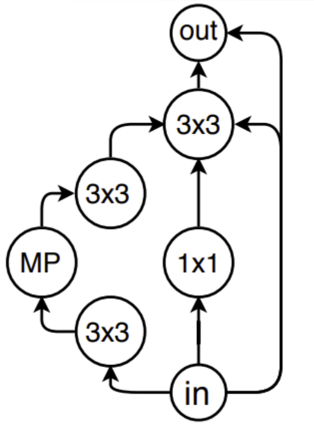
Cannot change one cell on its own!

Image from DARTS: "Differentiable Architecture Search" – ICLR 2019

Architecture Representation in Literature

- Different, fixed rules for grouping operations:
  - 'Conv-BN-ReLU' vs. 'ReLU-Conv-BN' vs. Dil. & Sep. Conv.
- Is an operation grouping an *edge* or a *node*?

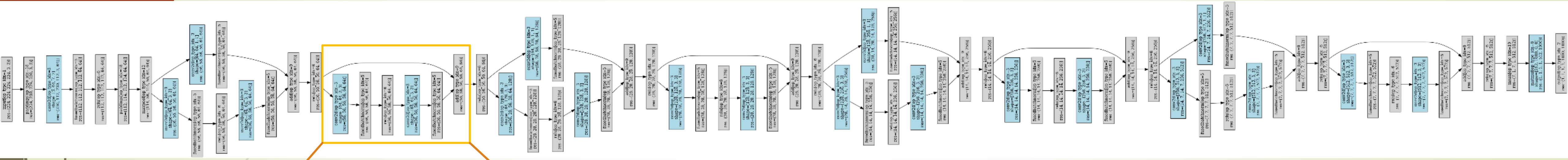


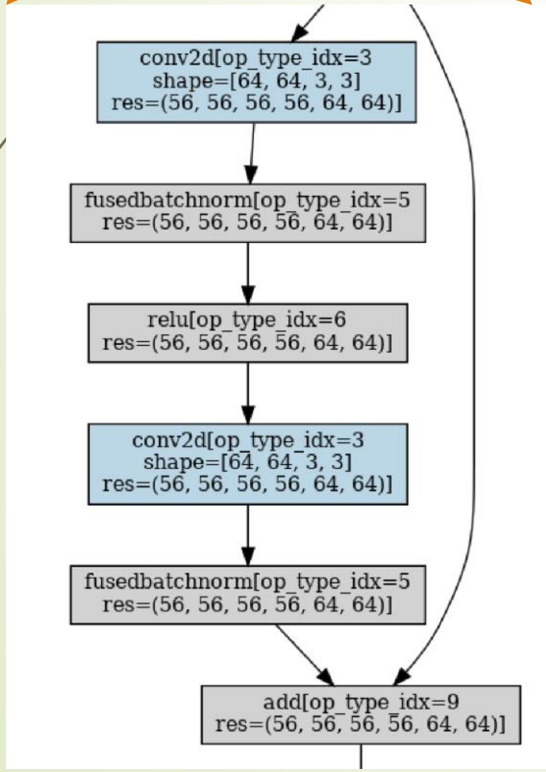


- If I train a predictor on one format, can I infer performance on the other?

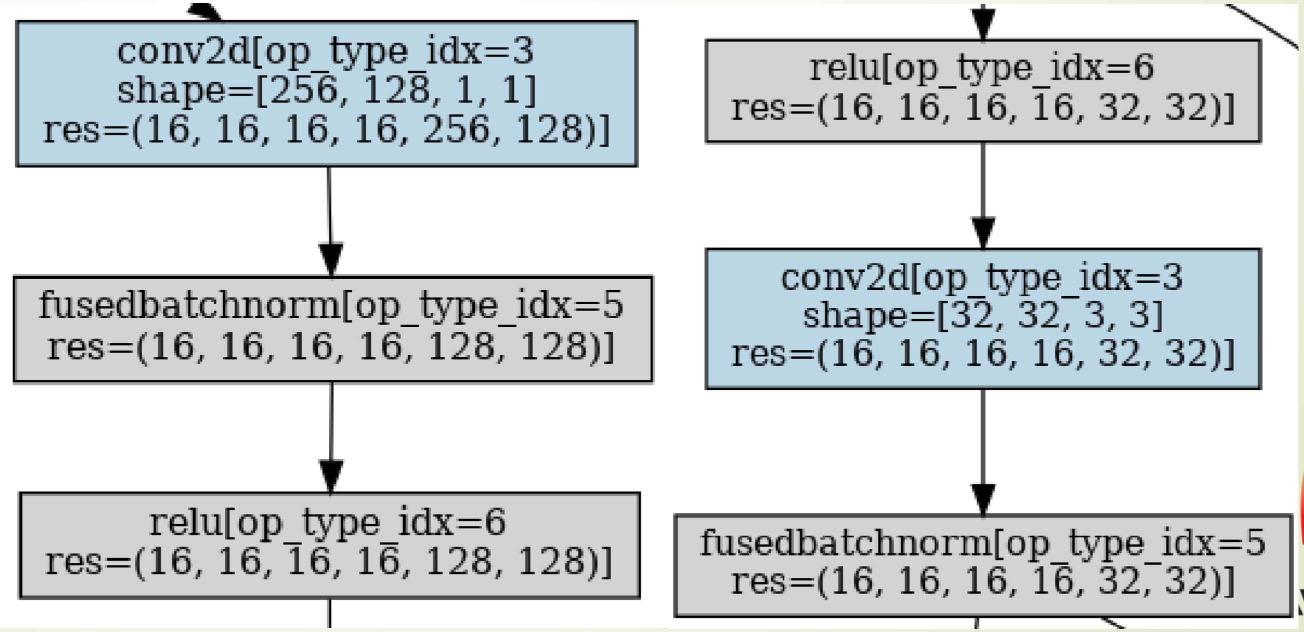
Images from DARTS: "Differentiable Architecture Search" – ICLR 2019 and "NAS-Bench-101: Towards Reproducible Neural Architecture Search" – ICML 2019

Our Approach: Computational Graphs





- Graph of the entire network.
- Each primitive op is a node.
- Transferability!



New Benchmark Family - HiAML

- 4.6k architectures.
  - Comparable FLOPS to NB-201.
- 4 stages; 2 identical blocks/stage.
- Pool of 14 blocks to choose from.
- Spanning [91.11%, 93.44%] on CIFAR-10.
  - Many ties make rank correlation difficult.
- Feature extractors in Huawei Mobile Facial Landmark Detection Application.

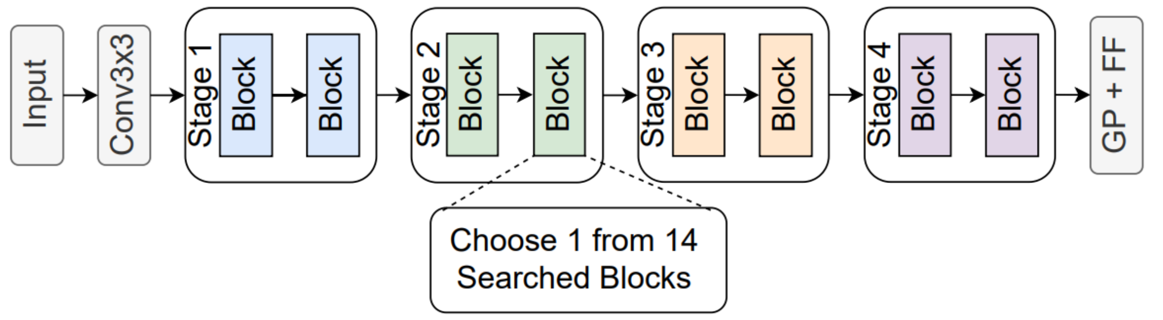
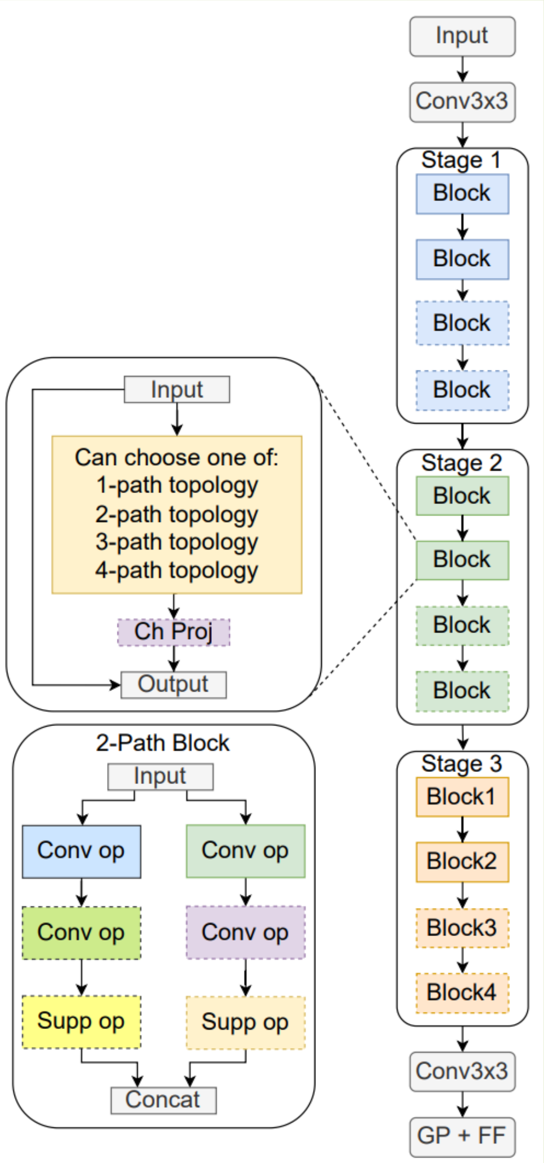


Figure 1: The architecture backbone of HiAML, containing 4 stages. Each stage contains 2 identical blocks.

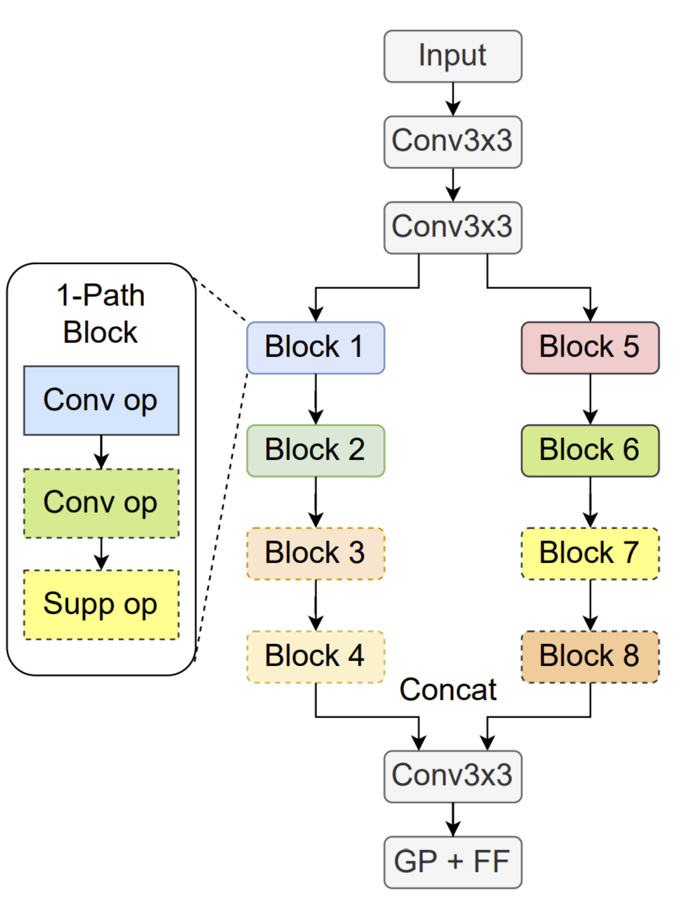
New Benchmark Families - Inception

- 580 architectures.
  - Inspired by classical Inception-v4.
- Branching paths, with channel splits.
  - Largest benchmark by FLOPs.
- Each block can have 1-4 branches.
- Used in mobile facial recognition.



New Benchmark Families – Two-Path

- 6.9k architectures.
  - Most lightweight benchmark.
- Applied in mobile 4k LivePhoto and Super Resolution.
- Meta-structure is 2 branching paths.
  - Each block is single-path.
  - Complement of Inception.



Contrastive Learning for Graph Encoder Pre-Training

- Combine aspects of the NT-Xent loss from SimCLR [Chen et al. 2020] and class-awareness of SupCon [Khosla et al. 2020]:

$$\mathcal{L}_{CL} = -\sum_{i \in I} \sum_{\ell \neq i} \alpha_\ell^{(i)} \log \frac{\exp(sim(z_i, z_\ell))}{\sum_{r \neq i} \exp(sim(z_i, z_r))},$$

- Semi-supervised, $\alpha_l^{(i)}$ is the structural similarity of CGs *i* and *l*.
  - Calculate using Laplacian Eigenvalues.
  - Like a continuous class similarity
- Use NB-101 as encoder training family.
  - Contains many architectures
  - Is topologically diverse.
- Encoder separates NB-101 CGs into groups.
- Divides other families into distinct clusters.

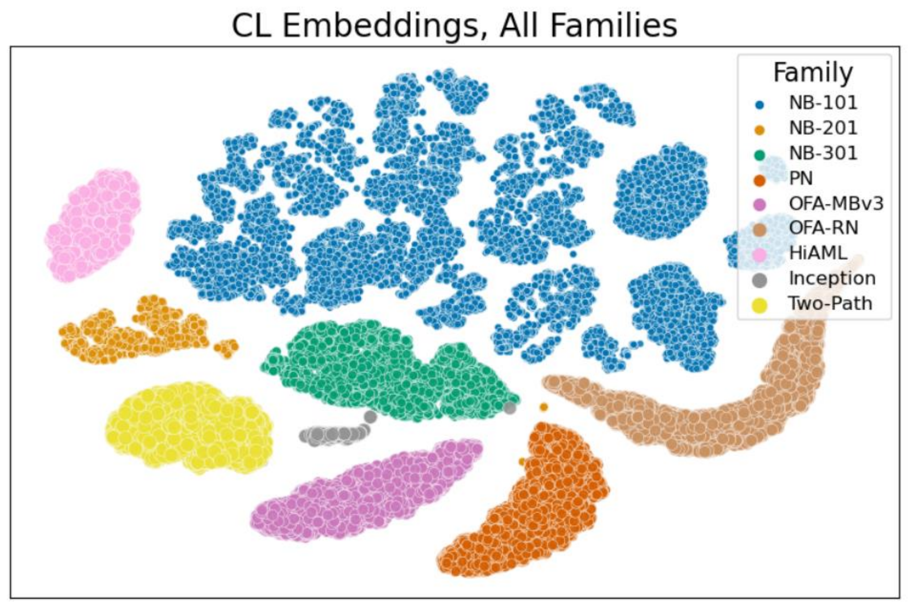


Figure 5: t-SNE scatterplot of the Contrastive Learning embeddings for all architecture families. Best viewed in color.

Fuzzy C-Means (FCM) Soft Clustering Ensemble

- Perform FCM on NB-101 graph embeddings.
  - Clusters overlap.
  - Continuous membership.
- Predictor ensemble: 1 head per cluster.
  - Weighted summation.
  - Cluster membership is the weight.

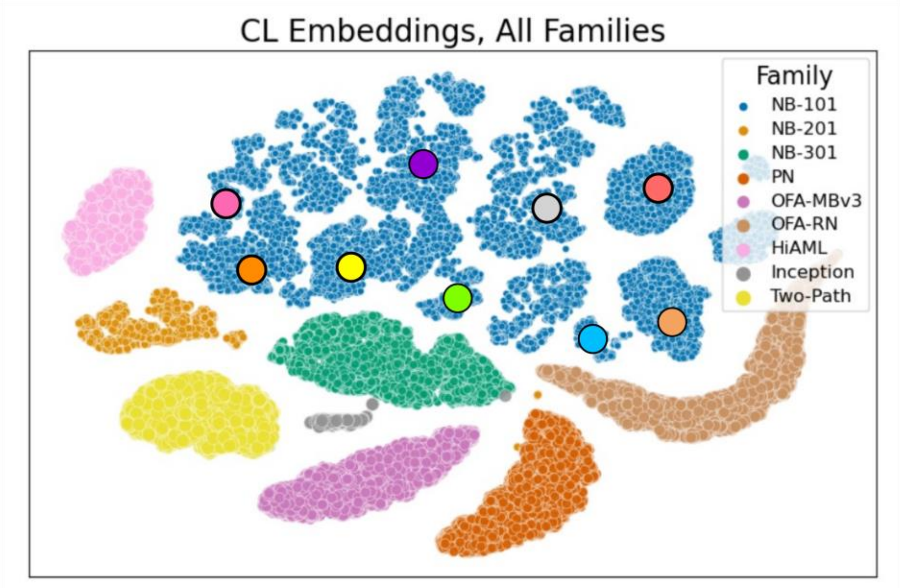


Figure 5: t-SNE scatterplot of the Contrastive Learning embeddings for all architecture families. Best viewed in color.

Fuzzy C-Means (FCM) Soft Clustering Ensemble

- Perform FCM on NB-101 graph embeddings.
  - Clusters overlap.
  - Continuous membership.
- Predictor ensemble: 1 head per cluster.
  - Weighted summation.
  - Cluster membership is the weight.
- Each ensemble head represents a different region of the latent space.
  - Each family lies in a distinct region.

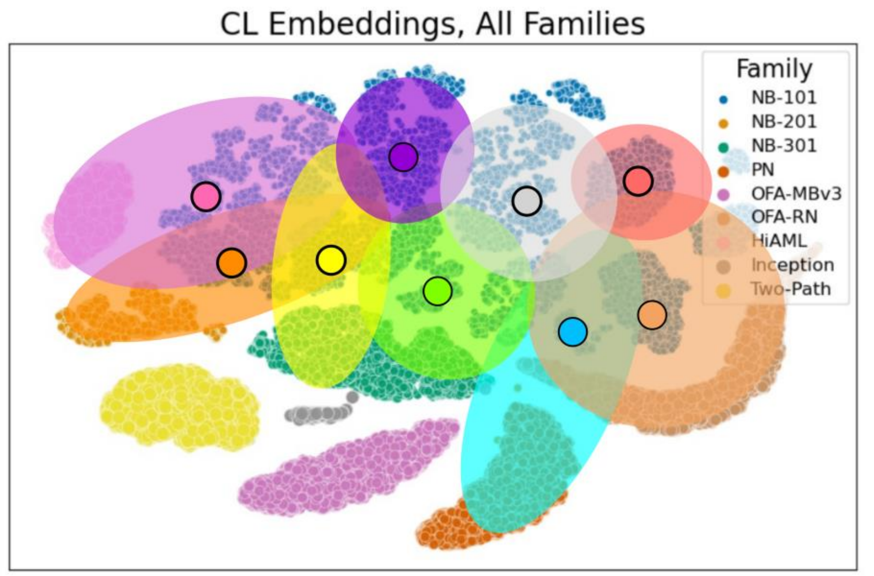


Figure 5: t-SNE scatterplot of the Contrastive Learning embeddings for all architecture families. Best viewed in color.

Incorporating FLOPs Into Predictions

- FLOPs indicate model size.
  - Cheap to compute.
- Enjoy positive correlation with performance.
  - Correlation can be strong.
- Augment accuracy labels using FLOPs and standardization:

$$y_i = \mathcal{Z}\left(\frac{A_i}{\text{Log}_{10}(F_i + 1) + 1}\right),$$

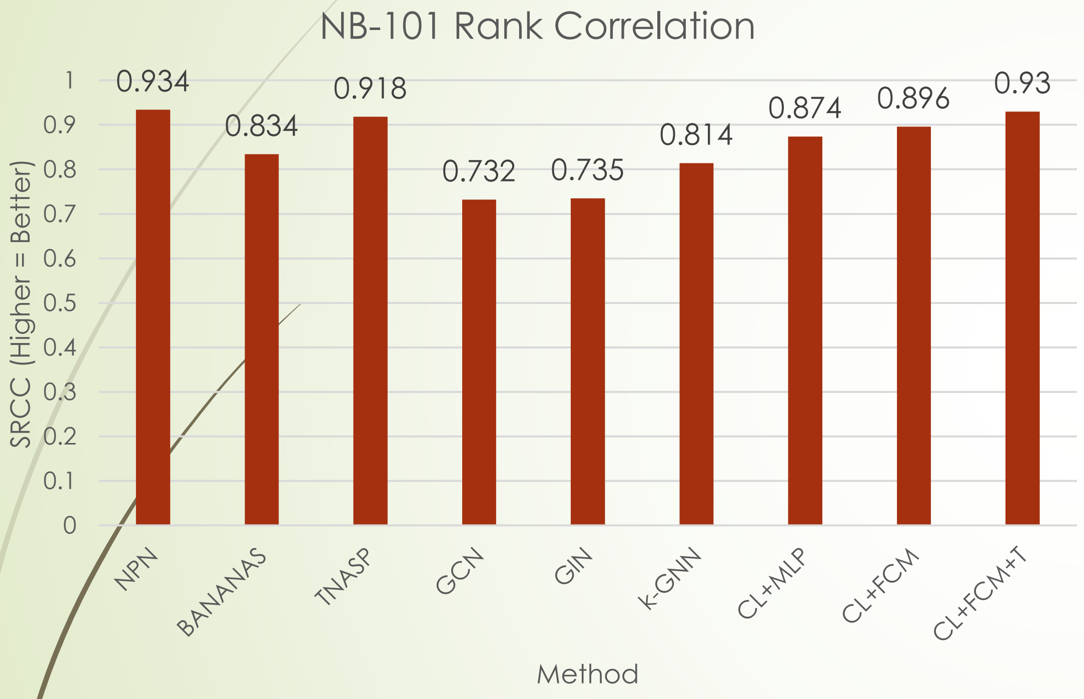
Family	FLOPs
NB-201	0.0002
NB-301	**0.5778**
OFA-PN	0.6886
OFA-MBv3	0.6141
OFA-RN	**0.7850**
HiAML	**0.2767**
Inception	**0.4115**
Two-Path	**0.3332**

Correlation between accuracy performance and FLOPs.

Experimental Setup

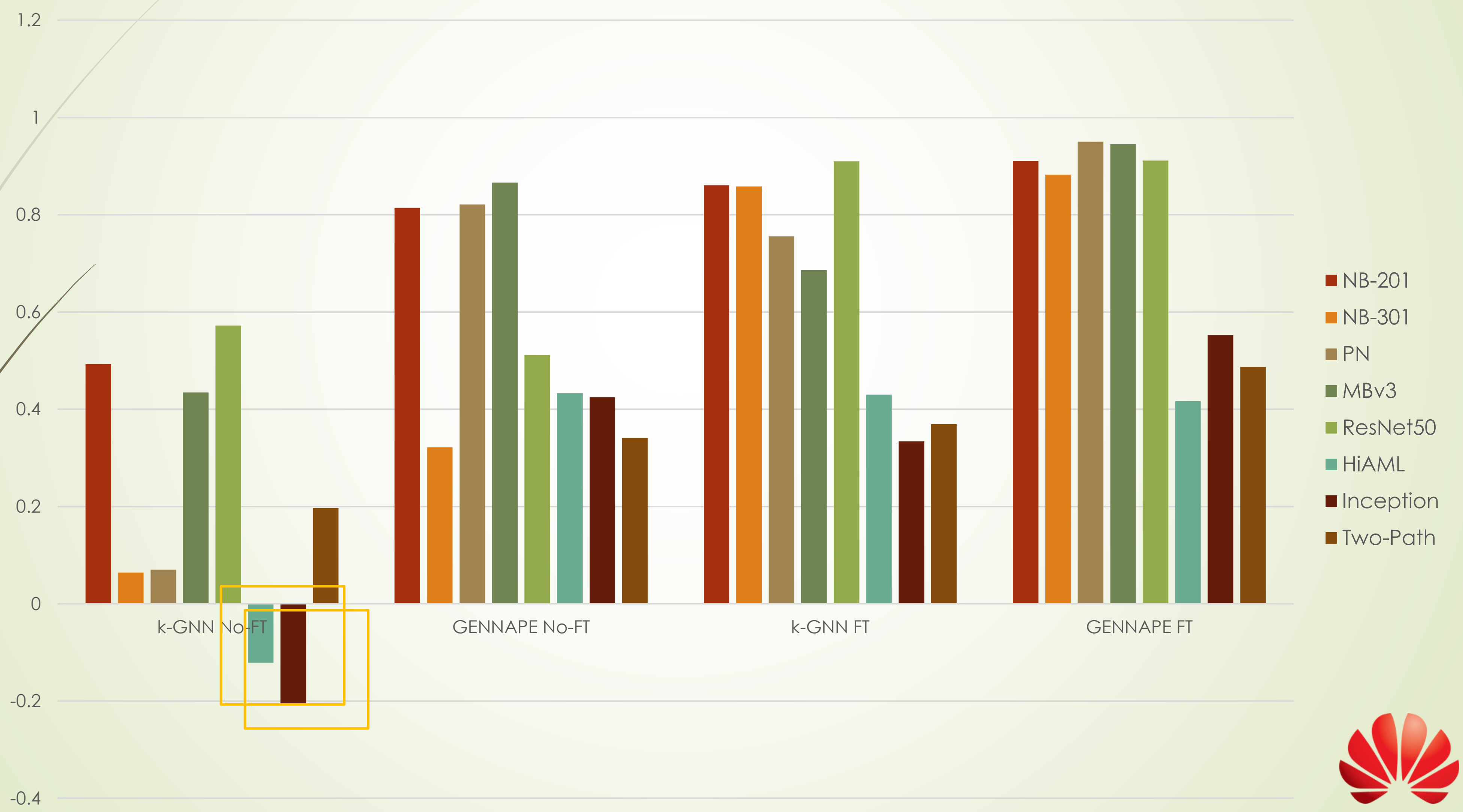
1. Single search space
2. Transferability test
3. **Application to NAS**

1) Single Search Space Evaluation

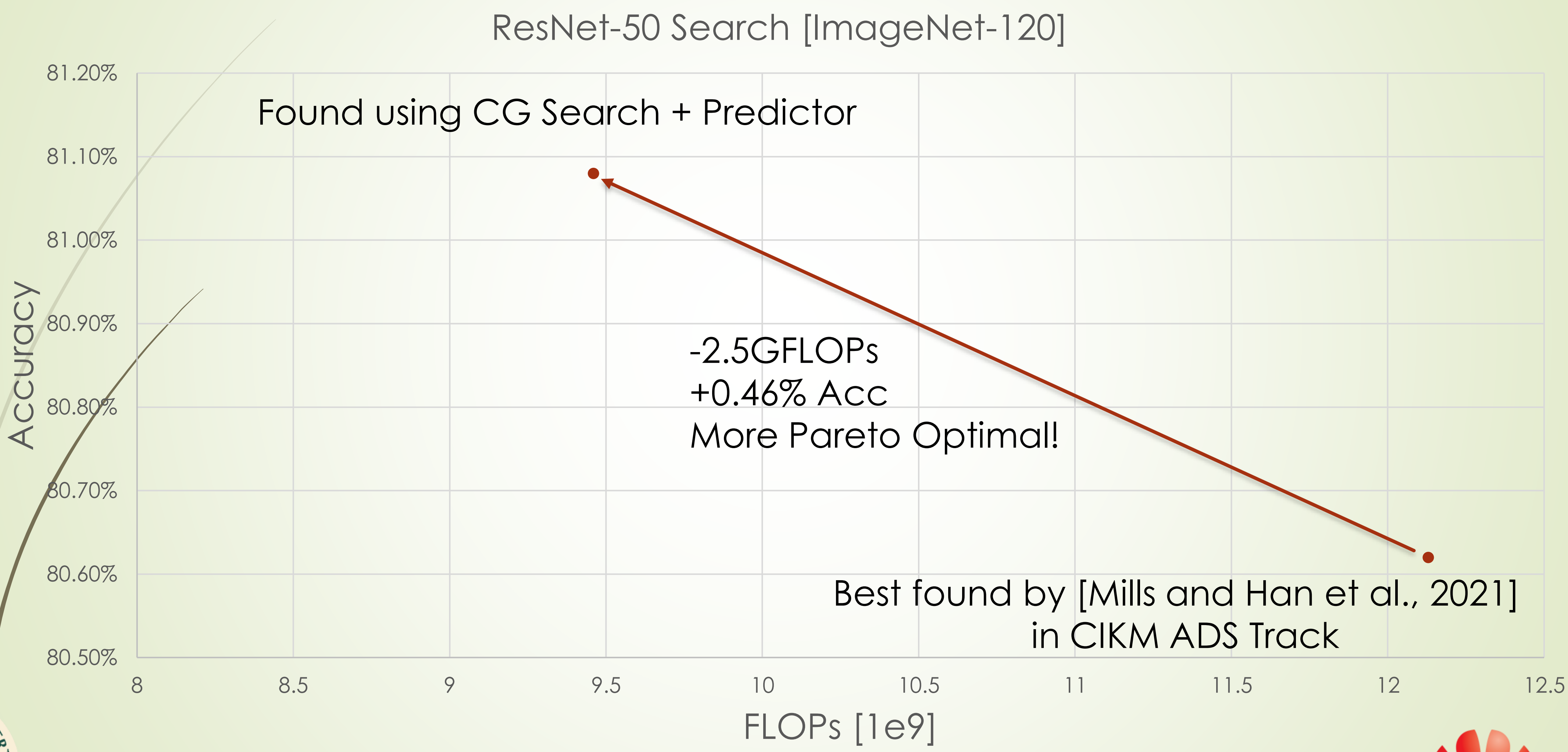


- Compare rank correlation (SRCC).
- Consider several single search space predictors from literature.
- Simple GNNs variants that use CGs.
- Instead, using a predictor built using our contributions… achieve **SRCC** > 0.9, achieving performance of best single-space predictor.

2) Transferability Test – SRCC



Application to NAS



Conclusion

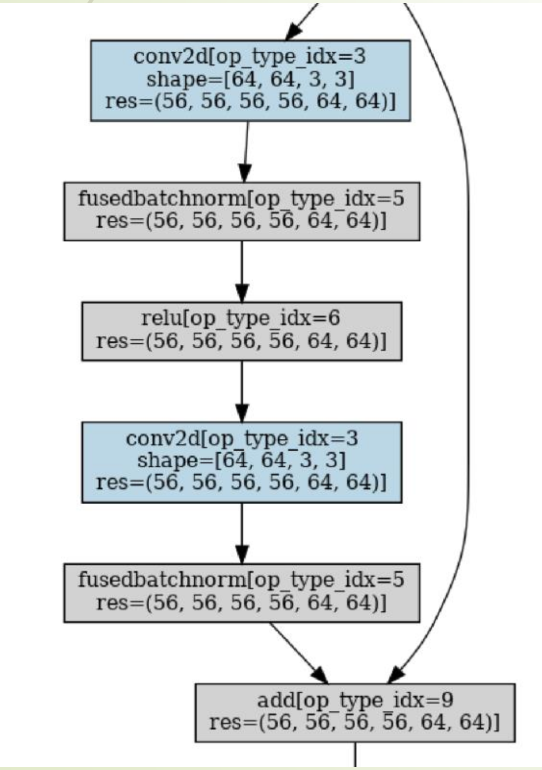
We propose GENNAPE, or **GEN**eralized **N**eural **A**rchitecture **P**erformance **E**stimators:

- Meaning?

Conclusion

We propose GENNAPE, or **GEN**eralized **N**eural **A**rchitecture **P**erformance **E**stimators:

- Meaning?



$$\mathcal{L}_{CL} = -\sum_{i\in I}\sum_{\ell\neq i}\alpha_{\ell}^{(i)}\log\frac{\exp(sim(z_i, z_\ell))}{\sum_{r\neq i}\exp(sim(z_i, z_r))},$$

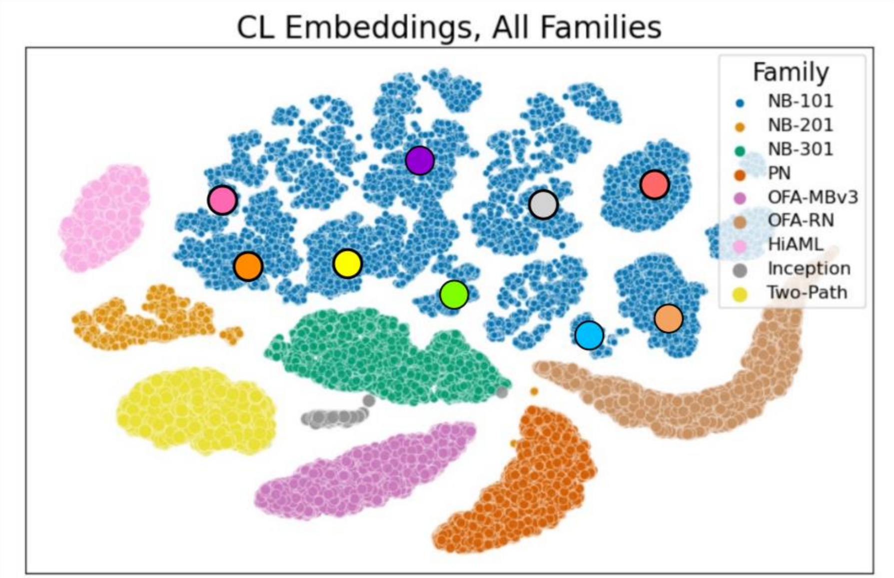
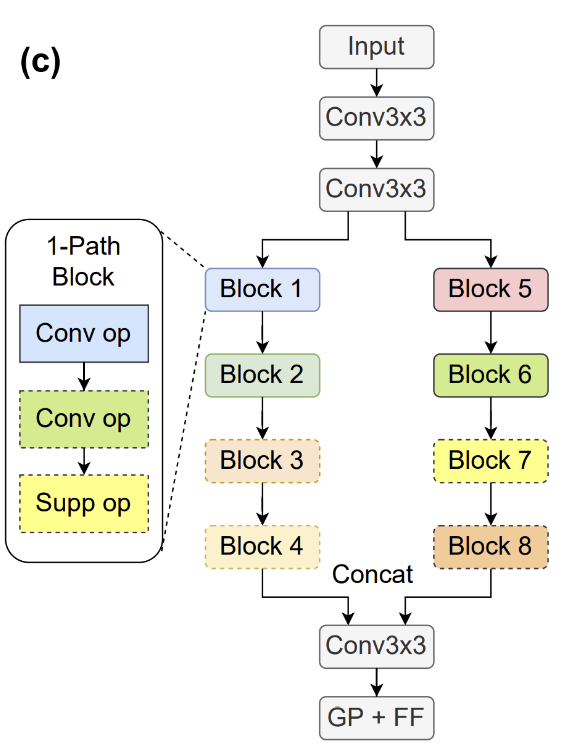


Figure 5: t-SNE scatterplot of the Contrastive Learning embeddings for all architecture families. Best viewed in color.



References

- Ying et al., "NAS-Bench-101: Towards Reproducible Neural Architecture Search." In ICML 2019.
- Dong and Yang, "NAS-Bench-201: Extending the Scope of Reproducible Neural Architecture Search." In ICLR 2020.
- Liu et al., "DARTS: Differentiable Architecture Search." In ICLR 2019.
- Zela et al., "Surrogate NAS Benchmarks: Going Beyond the Limited Search Spaces of Tabular NAS Benchmarks." In ICLR 2022.
- Chen et al., "A Simple Framework for Contrastive Learning of Visual Representations." In ICML 2020.
- Khosla et al., "Supervised Contrastive Learning." In NeurIPS 2020.
- Mills et al., "Profiling Neural Blocks and Design Spaces for Mobile Neural Architecture Search." In CIKM'21.