Graph-based Neural Architecture Search with Operation Embeddings

MICHAIL CHATZIANASTASIS, GEORGE DASOULAS, GEORGIOS SIOLAS, MICHALIS VAZIRGIANNIS

Introduction

Neural Architecture Search (NAS) is a technique that automatically designs a neural network architecture. A crucial part of the NAS pipeline is the *encoding* of the architecture.

Motivation: Most of the existing approaches either fail to capture the structural properties of the architectures or use fixed hand-engineered vectors, that cannot exploit information from data, to encode the operators .

Contributions:

- We propose **operation embeddings**:
  - a continuous representation of the applied operators,
  - integration into various graph autoencoders as parameters.
- We demonstrate that the learnable representations of the operations lead to generation of **state-of-the-art** architectures.
- We observe that the top-performing architectures share **similar structural patterns**:
  - clustering coefficient
  - average path length.

Neural Network as a DAG

Computation graph

$\mathbf{G}_{\mathcal{A}} = (\mathbf{V}, \mathbf{E})$

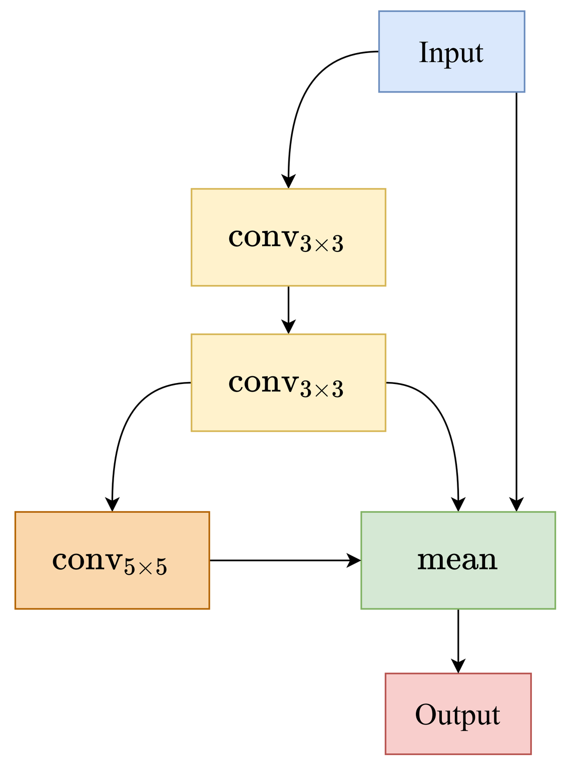
of an architecture $\mathcal{A}$:

· **Nodes V**: the applied operations

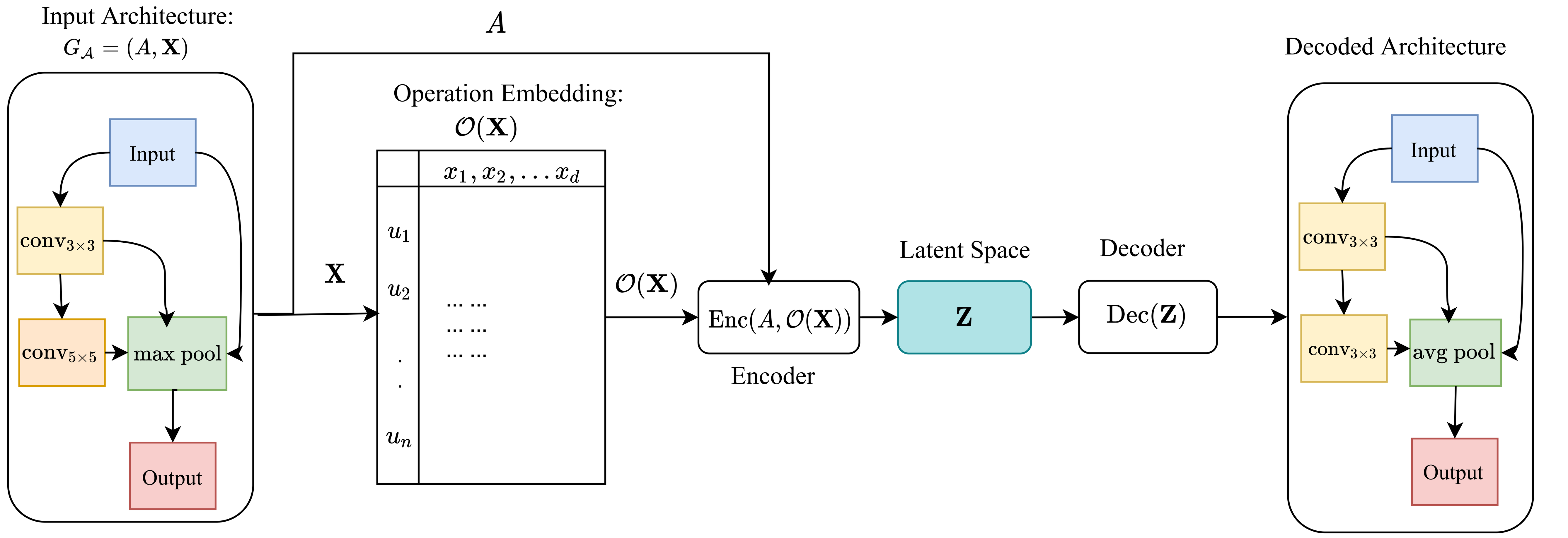
· **Edges E**: the signal flow the applied operations

· **Adjacency matrix** $\mathbf{A} \in \{0,1\}^{|V|\times|V|}$

· **Feature matrix** $\mathbf{X} \in \mathbb{Z}^{|V|\times|K|}$

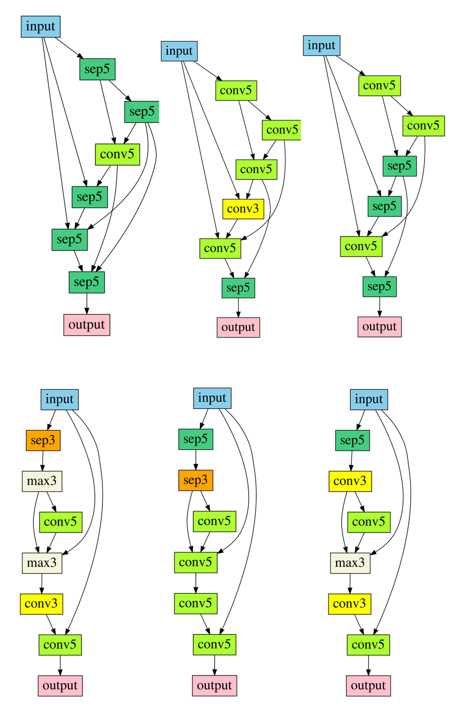


Operation Embeddings in Graph Variational Autoencoders



- We propose the **incorporation of the embedding** $\mathcal{O}: K \rightarrow \mathbb{R}^{|K|\times d_{op}}$ into the autoencoder model. The mapping $\mathcal{O}(\cdot)$ projects the set of available operations into a d_{op}-dimensional continuous space in a differentiable manner.
- **GNN model** $\phi: \mathbb{Z}^{|V|\times|V|} \times \mathbb{Z}^{|V|\times|K|} \rightarrow \mathbb{R}^{|V|\times d}$ that takes as input the architecture graph, and outputs a representation of every node.
- $\psi_1, \psi_2: \mathbb{R}^{|V|\times d} \rightarrow \mathbb{R}^{l}$ denote two differentiable pooling functions that.
- **Encoder**: $\mu_G = \psi_1(\phi(\mathbf{A}, \mathcal{O}(\mathbf{X}))), \sigma_G = \psi_2(\phi(\mathbf{A}, \mathcal{O}(\mathbf{X})))$.
- **Loss function**: $L(\phi, \theta; \mathbf{A}, \mathbf{X}) = \mathbb{E}_{q_\phi(\mathbf{Z}|\mathbf{A},\mathbf{X})}[\log p_\theta(\mathbf{A}, \mathbf{X}|\mathbf{Z})] - \mathrm{KL}[q_\phi(\mathbf{Z}|\mathbf{A}, \mathbf{X})||p(\mathbf{Z})]$, where KL denotes the Kullback–Leibler divergence.

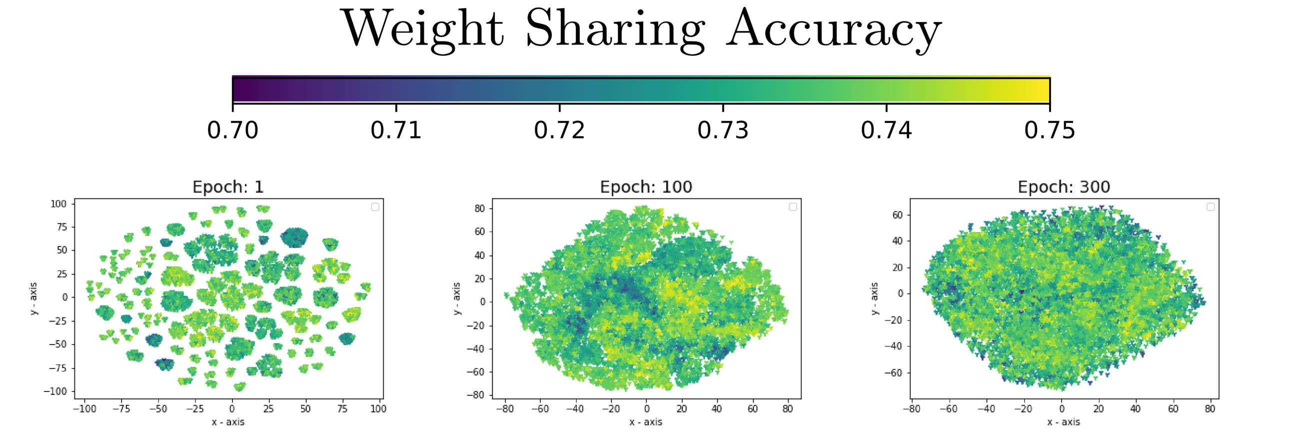
Best Generated Architectures



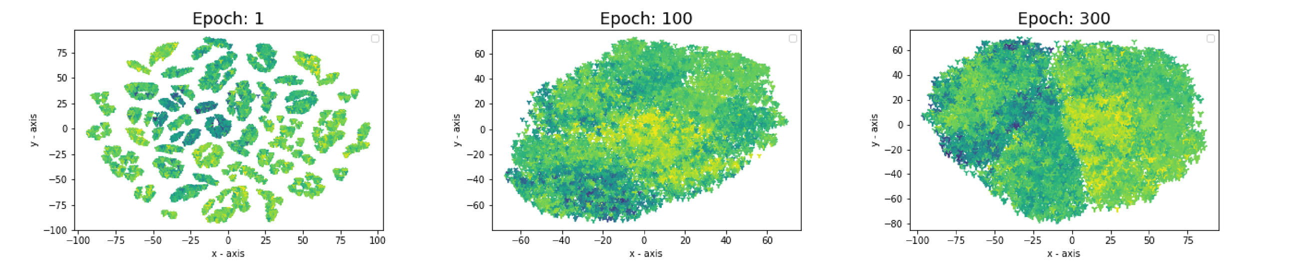
- **DVAE-EMB** [1] architectures Acc%: **95.35**, 95.33, 95.17 (above)
- **DVAE** [2] architectures Acc%: **94.8**, 94.74, 94.7 (below)
- Exhibit **similar structural** characteristics.

- DVAE-EMB present a **smoother** operation transition than those generated from D-VAE.

Latent Space Visualization



D-VAE's latent space during training.



DVAE-EMB's latent space during training.

Experiments

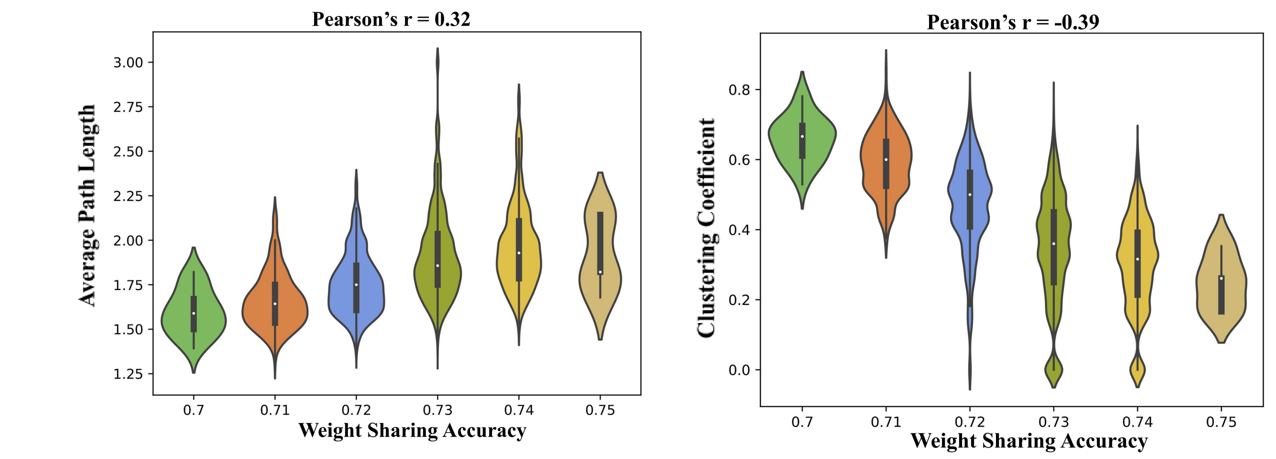
Basic abilities of Variational Graph Auto-Encoders

Model	Accuracy	Validity	Uniqueness
D-VAE	99.96	100.00	37.26
GCN	98.70	99.53	34.00
S-VAE	99.98	100.00	37.03
GraphRNN	99.85	99.84	29.77
DVAE-EMB	**99.99**	**100.00**	**39.15**
GCN-EMB	98.87	99.95	32.63

Predictive Performance of Encoded Latent Embeddings

Model	RMSE	Pearson's r
D-VAE	0.384 ±0.002	0.920 ±0.001
GCN	0.485 ±0.006	0.870 ±0.001
S-VAE	0.478 ±0.002	0.873 ±0.001
GraphRNN	0.726 ±0.002	0.669 ±0.001
DVAE-EMB	**0.371 ±0.003**	**0.925 ±0.001**
GCN-EMB	0.441 ±0.002	0.892 ±0.001

Architecture Performance with respect to Graph Properties



References

[1] Chatzianastasis, M., Dasoulas, G., Siolas, G., Vazirgiannis, M. (2021). *Graph-Based Neural Architecture Search With Operation Embeddings*. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops.

[2] Zhang, M., Jiang, S., Cui, Z., Garnett, R., Chen, Y. *D-VAE: A Variational Autoencoder for Directed Acyclic Graphs*, NeurIPS (2019)