

# OVERVIEW

Differentiable Neural Architecture Search

DARTS as Surrogate

Beyond Finite Search Spaces

# OVERVIEW

Differentiable Neural Architecture Search

DARTS as Surrogate

Beyond Finite Search Spaces

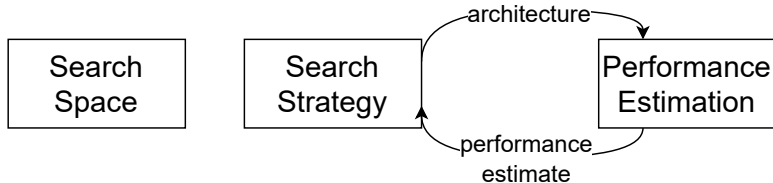
# NEURAL ARCHITECTURE SEARCH (NAS)

- ▶ Automatize choice of neural network architecture

# NEURAL ARCHITECTURE SEARCH (NAS)

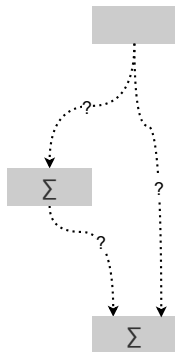
- ▶ Automate choice of neural network architecture
- ▶ Discover new architectures

- ▶ Automate choice of neural network architecture
- ▶ Discover new architectures



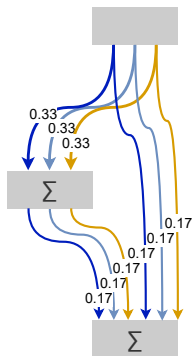
# DIFFERENTIABLE NAS

*DARTS* [Liu et al., 2018] considered as pioneer work



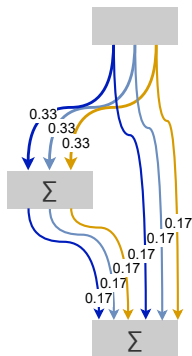
# DIFFERENTIABLE NAS

*DARTS* [Liu et al., 2018] considered as pioneer work

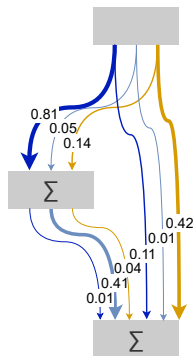


# DARTS

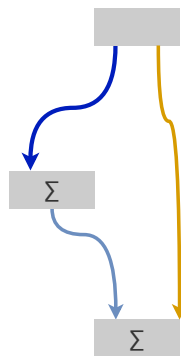
*DARTS* [Liu et al., 2018] considered as pioneer work



Training start



Training end



Obtain best architecture



# GUMBEL-SOFTMAX SAMPLING

We define the *Standard Gumbel* probability density as

$$g : \mathbb{R} \rightarrow [0, 1], x \mapsto \exp^{-(x + \exp^{-x})}$$

For  $k \in \mathbb{N}$ ,  $G \sim P_g^k$  and architecture parameters  $a \in \mathbb{R}^k$  it holds:

$$\text{Softmax}(a + G, 0) \sim \text{Multinomial}(1, \text{Softmax}(a))$$

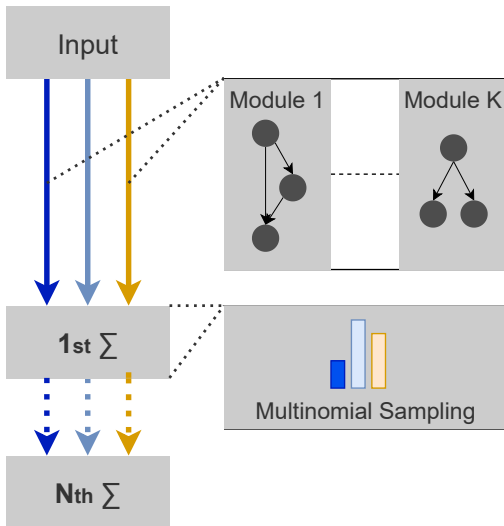
# OVERVIEW

Differentiable Neural Architecture Search

DARTS as Surrogate

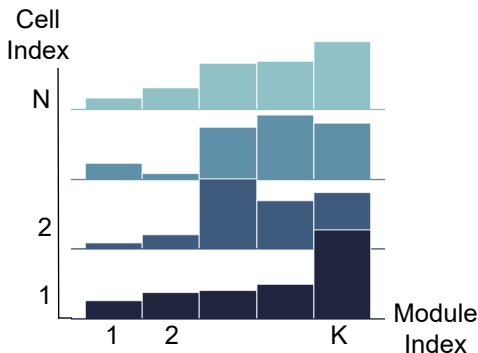
Beyond Finite Search Spaces

# SEARCH SPACE



# RELATIVE SURROGATE

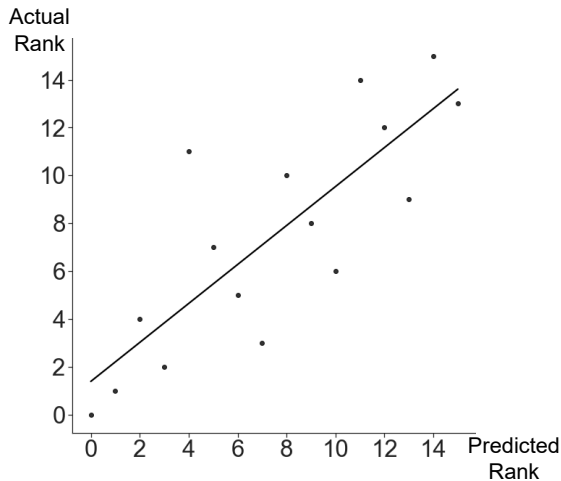
Joint trained multinomials induce ranking on search space



Sampling probability per module per cell

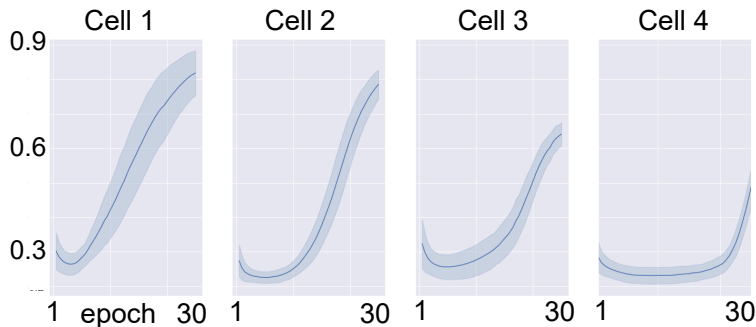
# RELATIVE SURROGATE

Validate surrogate ranking on actual architecture performances



# ARCHITECTURE REGULARIZATION

Control speed of convergence dependent on cell index



Maximum norm of architecture parameter vector per epoch per cell

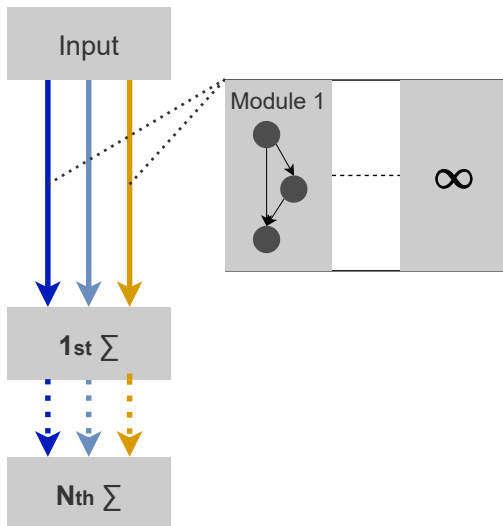
# OVERVIEW

Differentiable Neural Architecture Search

DARTS as Surrogate

Beyond Finite Search Spaces

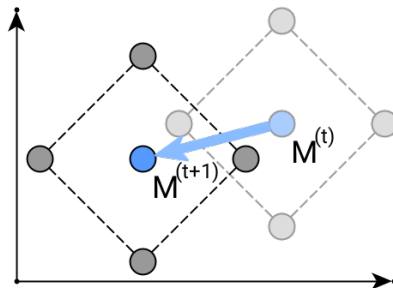
# SEARCH SPACE EXTENSION





# FINITE DIFFERENCE DESCENT

Finite difference descent on pseudo environment in euclidean search space



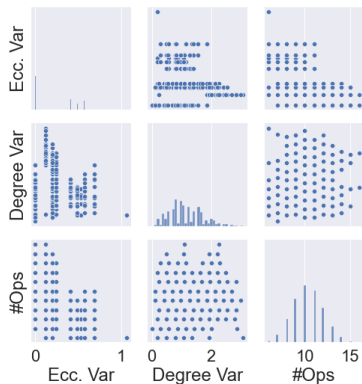
2-dim euclidean search space

# EXPERIMENTAL SEARCH SPACE

We model architectures with directed acyclic graphs (DAG)

# EXPERIMENTAL SEARCH SPACE

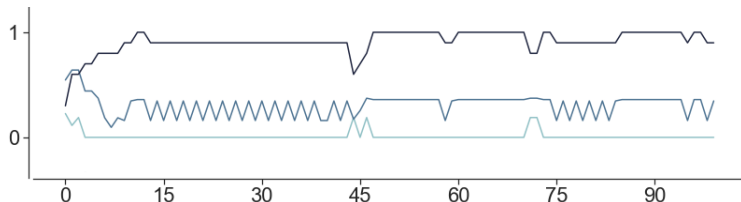
We model architectures with directed acyclic graphs (DAG)



Eccentricity variance, degree variance and # edges for 6-vertex DAGs

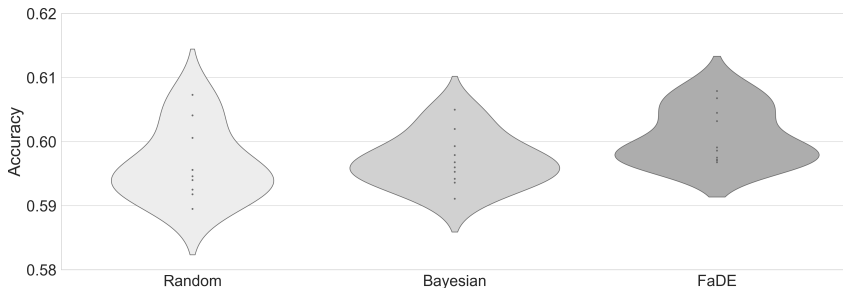
# EXPERIMENT RESULTS

Search space trajectories (per dimension) for one exemplary cell over 100 epochs of finite difference descent



# EXPERIMENT RESULTS

Comparing performance of top 10 architectures found by Random Search, Bayesian Search and our approach



# REFERENCES

[Liu et al., 2018] Liu, H., Simonyan, K., and Yang, Y. (2018).  
Darts: Differentiable architecture search.