25th International Conference
on Pattern Recognition

Research and Innovation Center of Excellence

# Operation and Topology Aware Differentiable Architecture Search


**Shahid Siddiqui[1,2], Christos Kyrkou[1] and Theocharis Theocharides[1,2]**

1. KIOS Research and Innovation Center of Excellence
2. University of Cyprus


University of Cyprus

Imperial College London


ICPR 2020

IAPR

Technically Co-Sponsored by

IEEE

IEEE COMPUTER SOCIETY

**Funded by:**

# Introduction

- Convolutional neural networks (CNNs) are well adopted for computer vision tasks.
- Since AlexNet, the complexity of CNNs is ever increasing i.e. deep ResNet.
- Manual network design demands substantial engineering effort.
- NAS has emerged as a task of automating the network design process.
- Idea: Given any dataset, find the best network.

# Differentiable Architecture Search

- DARTS models search task as differentiable.
- Uses bilevel optimisation with gradient descent.
- P-DARTS suggests to search progressively.
- ‘NAS Evaluation is Frustratingly Hard’ also studies DARTS’ search space.

# Issues

- Progressive searching matters?
- Is it the expertly crafted search space?
- Cell structure is complex.
- Which factor helps discover better results?
- Limited understanding.

# Contributions

- We investigate which components influence resulting cells the most:
  - Search network depth
  - Operation type
  - Cell configurations
- We alleviate skip connection bias problem.
- We significantly speed up search.

# Effect of search depth and operation type

| Search Space | Operations |
|---|---|
| SS1 | sep-3x3, dil-3x3 |
| SS2 | sep-5x5, dil-5x5 |
| SS3 | sep-3x3, sep-5x5 |
| SS4 | dil-3x3, dil-5x5 |
| SS5 | sep-3x3, dil-5x5 |
| SS6 | sep-5x5, dil-3x3 |
| SS7 | sep-3x3, sep-5x5, dil-3x3 |
| SS8 | sep-3x3, sep-5x5, dil-5x5 |

We split the search spaces such that searching within each is possible for deeper networks without running into memory limitations.

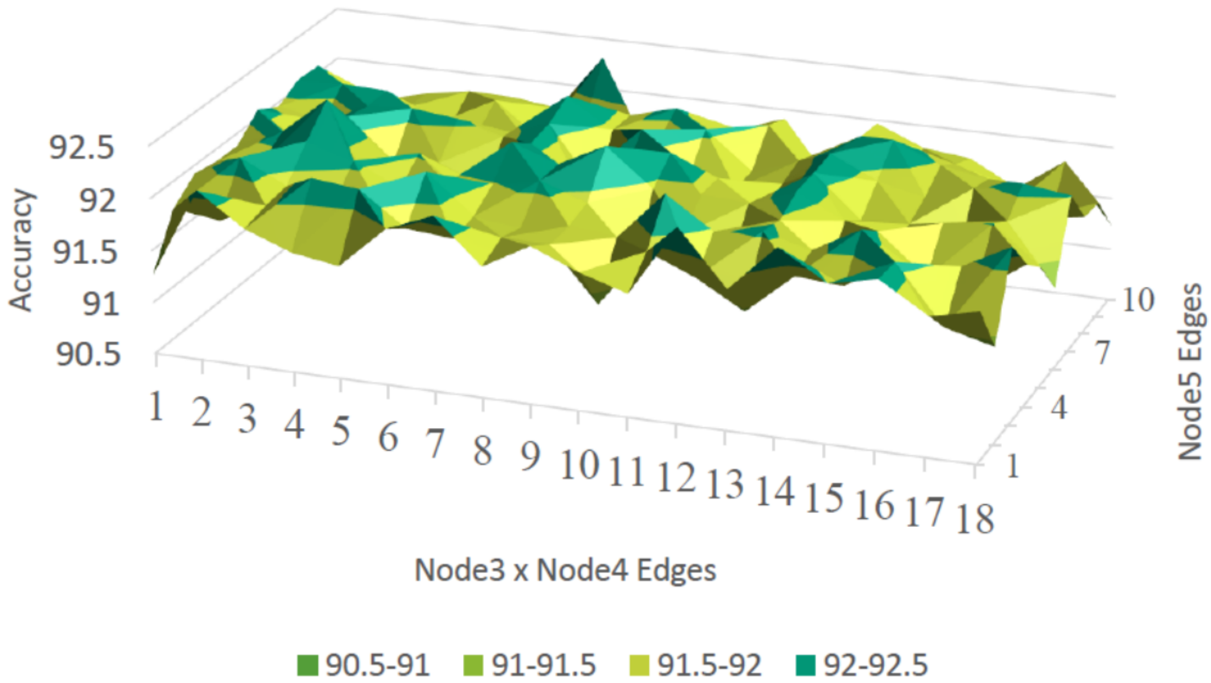
Each search space also has skip connection and pooling operations.

# Effect of search depth and operation type


Searching deeper is only marginally better computationally costly.

Separable convolution operations yield best cells while dilated convolutions tend to deteriorate it.

# Effect of cell topology




180 cell configurations are evaluated without changing operation type.

Cell topology has substantial effect on final accuracy, hence we can search only along good cells.

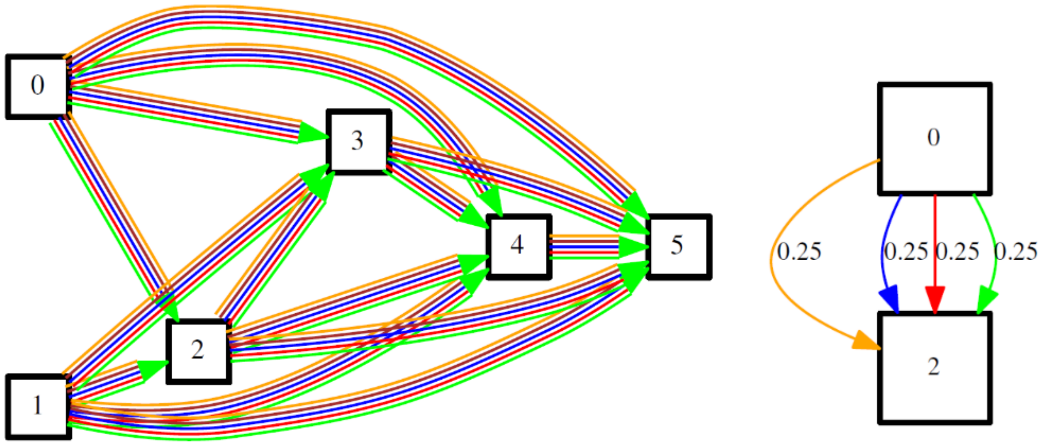
# Unbiased and fast search

| Method | Skip Connections | | Shallow Network Accuracy | Parameters (M) | Search Cost (GPU Hours) |
|---|---|---|---|---|---|
| | Normal | Reduction | | | |
| DARTS | 14 | 19 | 90.25 | 0.11 | 36 |
| DARTS + Data Subset | 4 | 14 | 91.27 | 0.15 | 6 |
| DARTS + Epoch Update | 5 | 11 | 91.33 | 0.16 | 4.6 |
| DARTS + Data Subset + Epoch Update | 5 | 8 | 92.15 | 0.2 | 0.85 |

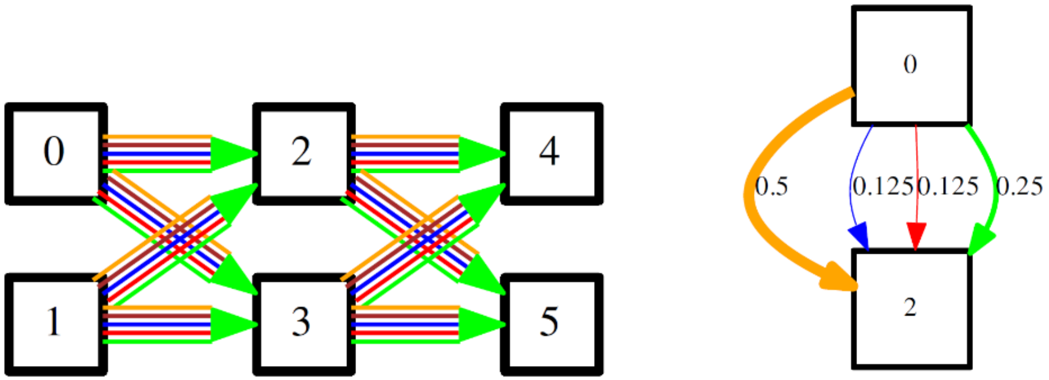
Searching with a subset of target dataset reduces the skip connection bias yet speeds up the search significantly.

Using ”sparse-update” through out the search is also effective at reducing the skip connection aggregation problem.

# eDARTS: Putting it all together




DARTS search supernet with equal attention to all operations at each edge.




eDARTS has a fixed search path with more attention to important operations.

# Comparison with state-of-the-art

| Architecture | Test Err. (%) C10 | Test Err. (%) C100 | Params (M) | Search Cost (GPU-days) | Search Method |
|---|---|---|---|---|---|
| DenseNet-BC [24] | 3.46 | 17.18 | 25.6 | - | manual |
| NASNet-A + cutout [7] | 2.65 | - | 3.3 | 1800 | RL |
| AmoebaNet-A + cutout [8] | 3.34 | - | 3.2 | 3150 | evolution |
| AmoebaNet-B + cutout [8] | 2.55 | - | 2.8 | 3150 | evolution |
| Hierarchical evolution [9] | 3.75 | - | 15.7 | 300 | evolution |
| PNAS [37] | 3.41 | - | 3.2 | 225 | SMBO |
| ENAS + cutout [27] | 2.89 | - | 4.6 | 0.5 | RL |
| DARTS (first order) + cutout [10] | 3.00 | 17.76† | 3.3 | 1.5 | gradient-based |
| DARTS (second order) + cutout [10] | 2.76 | 17.54† | 3.3 | 4 | gradient-based |
| SNAS + mild constraint + cutout [38] | 2.98 | - | 2.9 | 1.5 | gradient-based |
| SNAS + moderate constraint + cutout [38] | 2.85 | - | 2.8 | 1.5 | gradient-based |
| SNAS + aggressive constraint + cutout [38] | 3.10 | - | 2.3 | 1.5 | gradient-based |
| ProxylessNAS + cutout [39] | 2.08 | - | 5.7 | 4 | gradient-based |
| PC-DARTS + cutout [12] | 2.57 | - | 3.6 | 0.1 | gradient-based |
| Fair DARTS [28] | 2.54 | - | 2.8 | 0.3 | gradient-based |
| P-DARTS CIFAR10 + cutout [11] | 2.50 | 16.55 | 3.4 | 0.3 | gradient-based |
| P-DARTS CIFAR100 + cutout [11] | 2.62 | 15.92 | 3.6 | 0.3 | gradient-based |
| eDARTS CIFAR10 + cutout | 2.53 | 17.00 | 3.1 | **0.015‡** | gradient-based |
| eDARTS CIFAR100 + cutout | 2.72 | 16.83 | 3.5 | 0.016 | gradient-based |

eDARTS is 6x faster than PC-DARTS while achieving almost same error rate.

# Conclusion

- Good architectures can be discovered regardless of the search network depth.
- Separable convolution is the most effective operation in DARTS' search space.
- 3x3 kernel size yields better cells.
- Cell configuration also affects accuracy.
- eDARTS alleviates skip connection problem of DARTS.
- Search cost is less than 30 minutes.
- Design a new search space and improve search by feedback.

# Thank you :)