DOTS: Decoupling Operation and Topology in Differentiable Architecture Search

Yu-Chao Gu[1], Li-Juan Wang[1], Yun Liu[1], Yi Yang[2],
Yu-Huan Wu[1], Shao-Ping Lu[1], Ming-Ming Cheng[1]

[1]Nankai University
[2]Zhejiang University

Introduction

- Revisit DARTS

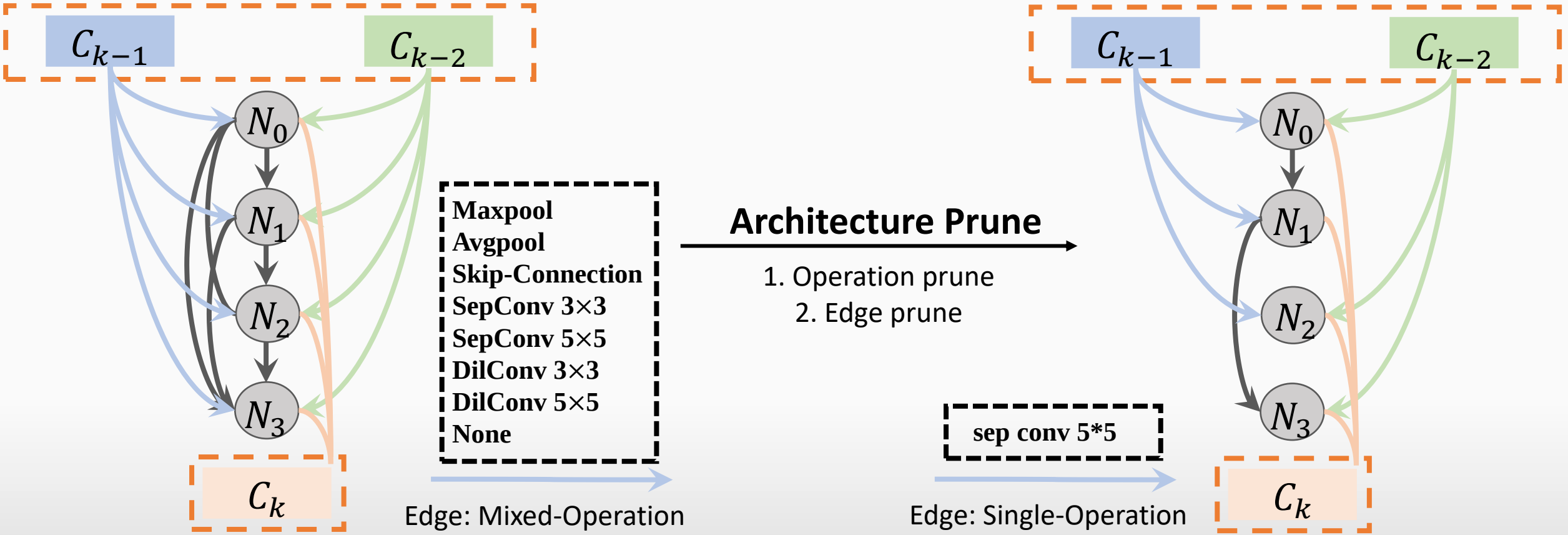


Figure 1. Derive the cell structure in DARTS

Introduction

- Problems in DARTS

1) Select edge based on operation weight (coupling operation and topology search)
2) Handcraft policy of edge numbers (fix to 2)

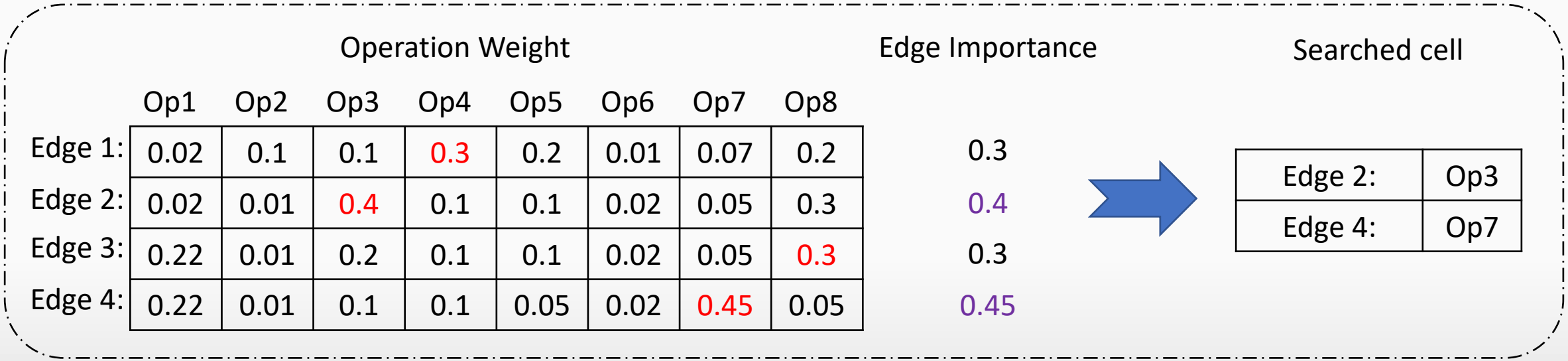


Figure 2. Architecture search via continuous relaxation

Introduction

- Can operation weight represents the edge importance?

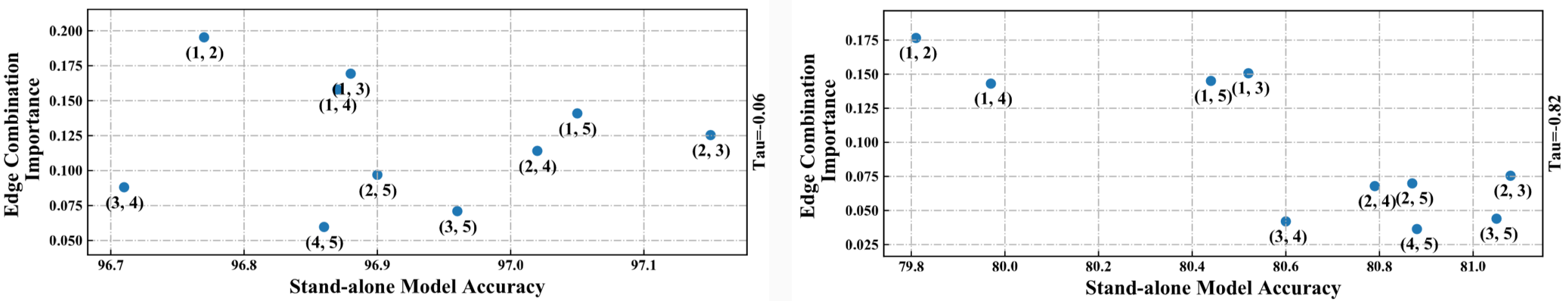
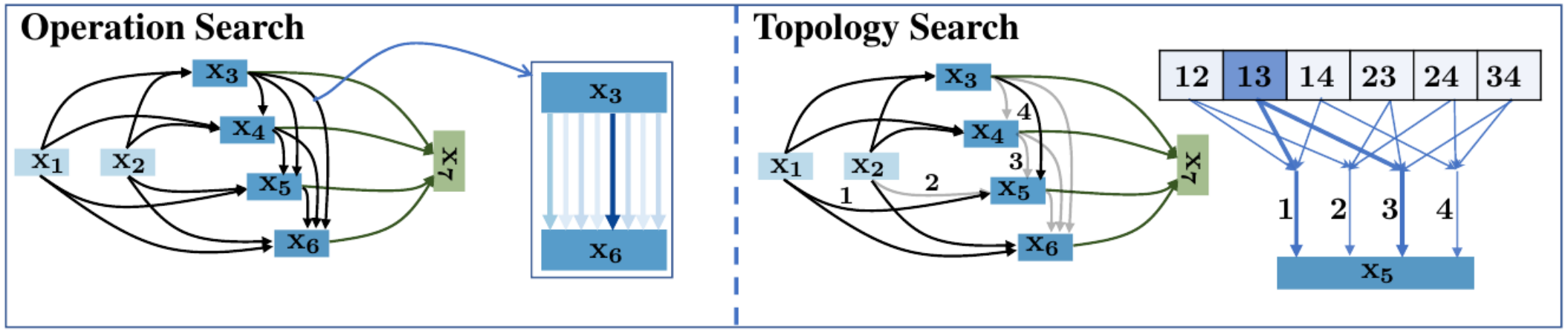


Figure 3. Correlation between edge importance and stand-alone accuracy in DARTS (Left: CIFAR10, Right: CIFAR100)

Method

- Decouple operation and topology search



① Decouple the topology representation from operation weight
- Relax combinatorial edge selection into continuous

② Decouple the search process
- Incorporate previous cell-based into operation search
- Group-based operation search

Method - topology search

1. Define topology search space + continuous relaxation

Two edges (follow DARTS)

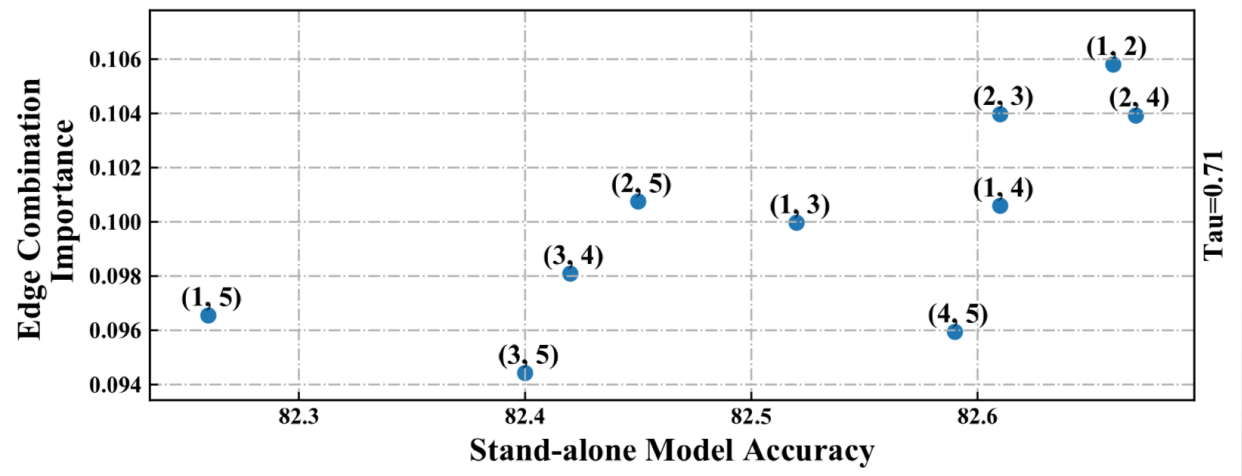
$$\mathcal{E}_{x_j} = \{\langle (i_1, j), (i_2, j) \rangle \,|0 < i_1 < i_2 < j\}$$

Any number of edges

$$c_m = \{e_1, e_2, \ldots, e_n\}$$

$$\mathcal{E}_{x_j} = \{c^1, c^2, \ldots, c^M\}$$

2. Correlation analysis



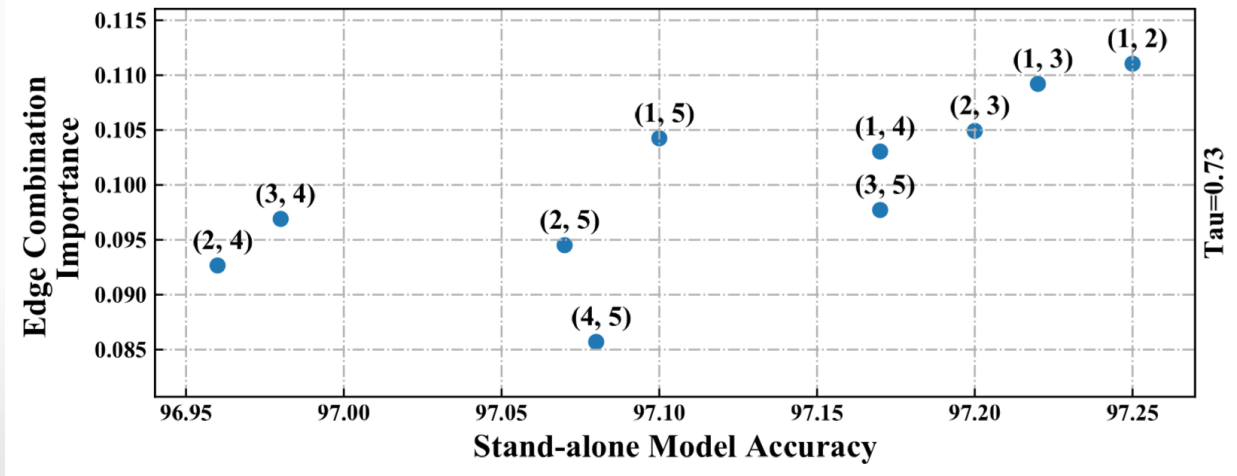


Figure 4. Correlation between edge importance and stand-alone accuracy in DOTS (Left: CIFAR10, Right: CIFAR100)

Method - operation search

- Method 1: Incorporate existing cell-based operation search

Operation Weight

	Op1	Op2	Op3	Op4	Op5	Op6	Op7	Op8
Edge 1:	0.02	0.1	0.1	0.3	0.2	0.01	0.07	0.2
Edge 2:	0.02	0.01	0.4	0.1	0.1	0.02	0.05	0.3
Edge 3:	0.22	0.01	0.2	0.1	0.1	0.02	0.05	0.3
Edge 4:	0.22	0.01	0.1	0.1	0.05	0.02	0.45	0.05

Intermediate results for further topology search :

Edge 1:	Op4
Edge 2:	Op3
Edge 3:	Op8
Edge 4:	Op7

Architecture	TS	CIFAR10	CIFAR100
DARTS [35]	✗	97.02±0.12	80.74
	✓	97.40±0.09	83.07
DARTS (2nd) [35]	✗	97.01±0.15	81.37
	✓	97.12±0.11	83.28
GDAS [11]	✗	96.84±0.06	82.75
	✓	97.06±0.08	83.01
SNAS [50]	✗	97.05±0.10	81.92
	✓	97.26±0.12	83.25
PC-DARTS [51]	✗	97.28±0.08	81.74
	✓	97.45±0.06	82.36

Table 1. Improve cell-based NAS by topology search

Method - operation search

- Method 2: Group-based operation search

Group Criterion: Parameterized Op: Op1-Op4
Non-Parameterized Op: Op5-Op8

	Op1	Op2	Op3	Op4	Op5	Op6	Op7	Op8
Edge 1:	0.2	0.1	0.4	0.3	0.5	0.1	0.2	0.2
Edge 2:	0.1	0.5	0.1	0.3	0.1	0.3	0.2	0.4
Edge 3:	0.1	0.2	0.2	0.5	0.1	0.2	0.1	0.6
Edge 4:	0.4	0.1	0.3	0.2	0.2	0.2	0.45	0.15

Operation Weight

Intermediate results for further topology search :

Edge 1:	Op3	Op5
Edge 2:	Op2	Op8
Edge 3:	Op4	Op8
Edge 4:	Op1	Op7

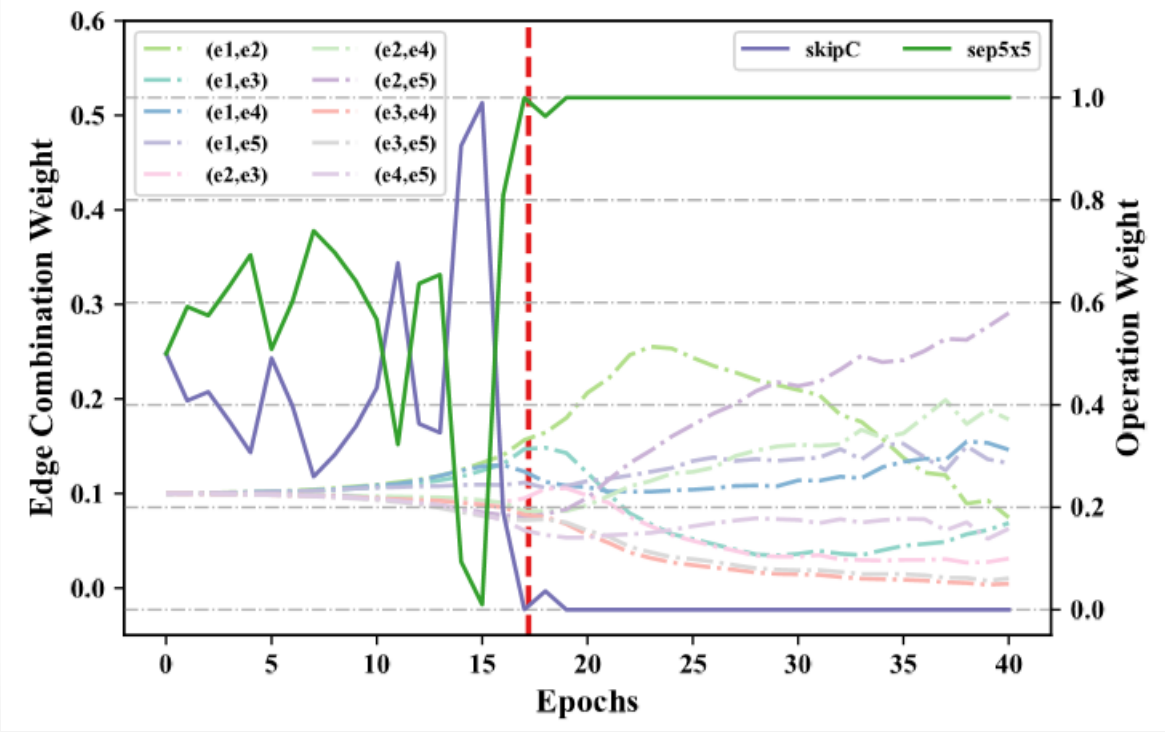


Figure 5. Decouple the grouped operation in topology search by annealing

Experiments

- Comparison with state-of-the-art models on CIFAR/ImageNet

Architecture	Top-1 Acc. (%) CIFAR10	Params (M) CIFAR10	Top-1 Acc. (%) CIFAR100	Params (M) CIFAR100	Search Cost (GPU-days)	Search Method
DARTS [35]	97.00	3.4	82.46†	3.4	0.4	GD
SNAS [50]	97.15	2.8	82.45	2.8	1.5	GD
GDAS [11]	97.07	2.5	81.62†	3.4	0.2	GD
P-DARTS [6]	97.50	3.4	82.51†	3.6	0.3	GD
FairDARTS [9]	97.46	2.8	82.39	2.8	0.4	GD
PC-DARTS [51]	97.43 ± 0.07	3.6	83.10	3.6	0.1	GD
DropNAS [20]	97.42 ± 0.14	4.1	83.13	4.0	0.6	GD
MergeNAS [45]	97.27 ± 0.02	2.9	82.42	2.9	0.2	GD
ASAP [38]	97.32 ± 0.11	2.5	82.69	2.5	0.2	GD
SDARTS-ADV [5]	97.39 ± 0.02	3.3	83.27	3.3	1.3	GD
DARTS- [7]	97.41 ± 0.08	3.5	82.84†	3.4	0.4	GD
DOTS (best)	97.63	3.5	83.72	4.1	0.26	GD
DOTS (avg)*	97.51±0.06	3.5	83.52±0.13	4.1	0.26	GD

Table 2. Benchmark Results on CIFAR10/100

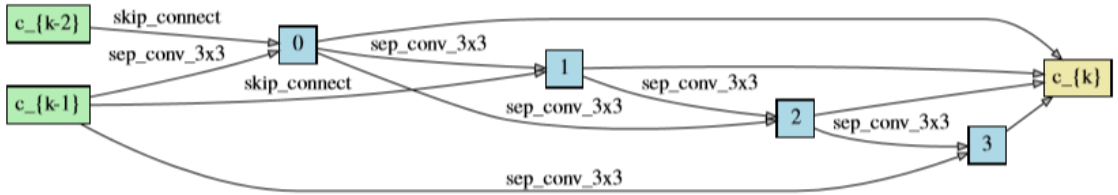
Architecture	Acc. (%) top-1	Acc. (%) top-5	Params (M)	Multi-Add (M)	Search Cost (GPU-days)	Search Method
ProxylessNAS (ImageNet) [4]	75.1	92.5	7.1	465	8.3	GD
FairDARTS (ImageNet) [9]	75.6	92.6	4.3	440	3	GD
PC-DARTS (ImageNet) [51]	75.8	92.7	5.3	597	3.8	GD
DOTS (ImageNet)	76.0	92.8	5.3	596	1.3	GD

Table 3. Benchmark Results on ImageNet

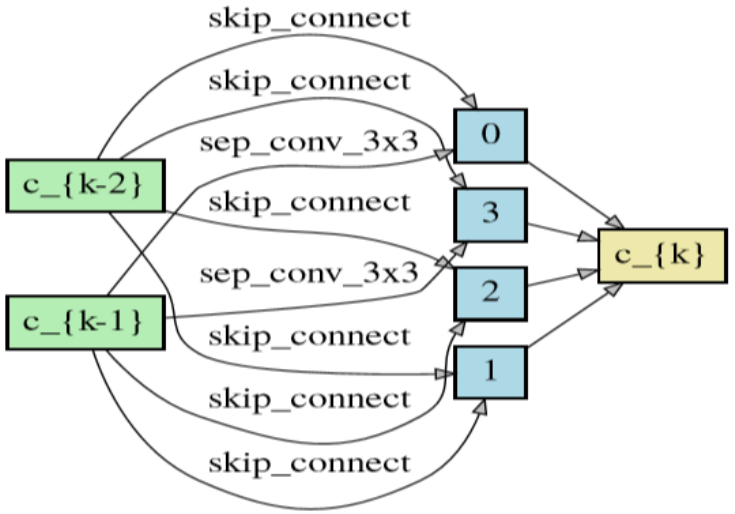
Experiments

EdgeID	1	2	3	4	5	6	7
OP	SkipC	Sep3x3	SkipC	SkipC	Sep3x3	SkipC	SkipC
EdgeID	8	9	10	11	12	13	14
OP	Sep3x3	Sep3x3	SkipC	Sep3x3	Sep3x3	Sep3x3	Sep3x3

(a) Operation searched results



(b) DOTS topology search



(c) DARTS' policy

DOTS: 25% skip-connection
CIFAR100: 83.07% top-1 Acc

DARTS: 75% skip-connection
CIFAR100: 80.74% top-1 Acc

Figure 6. Comparison between DARTS and DOTS topology derivation on the same operation results

Experiments

- Transfer to downstream tasks

Backbone	#Param (M)	FLOPs (M)	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
ResNet-50 [7]	25.6	4120	0.363	0.553	0.386	0.193	0.400	0.488
MobileNet-V2 [14]	3.4	300	0.283	0.467	0.293	0.148	0.307	0.381
SinglePath NAS [6]	4.3	365	0.307	0.498	0.322	0.154	0.339	0.416
MobileNet-V3 [9]	5.4	219	0.299	0.493	0.308	0.149	0.333	0.411
MnasNet [15]	4.8	340	0.305	0.502	0.320	0.166	0.341	0.411
FairDARTSC [4]	5.0	386	0.319	0.519	0.330	0.174	0.353	0.430
DOTS	5.3	596	0.357	0.552	0.378	0.199	0.393	0.478

Table 4. Evaluation of object detection on the MS-COCO 2017 dataset

Backbone	#Param (M)	FLOPs (G)	mIOU(%) val	mIOU(%) test
ResNet-18	14.1	20.1	74.8	74.7
Xception-39	1.9	4.1	69.0	68.4
MnasNet	6.8	11.0	76.8	74.2
DOTS	8.0	12.9	**79.3**	**77.6**

Table 5: Evaluation of semantic image segmentation on the Cityscapes dataset