

Motivation:

- 1 之前的 One-shot nas 在 train 完之后随机采样，没法采样出好的结果。
- 2 Darts 这种方法在 train 的时候会造成收敛速度快的网络被不断训练而造成马太效应。

Method:

1 train 的时候均匀随机采样一个操作。

$$\mathbf{H}_t^r = \sum_{r=1}^{\frac{(|\mathcal{O}|+1)}{2}} \hat{\mathbf{h}}_{(r)} \\ \times (\sum_{t=1}^{|\mathcal{I}|} \hat{f}_{(t)} \mathcal{O}_{(r1)}(\mathcal{I}_{(t)}) + \sum_{u=1}^{|\mathcal{I}|} \hat{g}_{(u)} \mathcal{O}_{(r2)}(\mathcal{I}_{(u)})), \quad (10) \\ \text{s.t. calculate } r1 \text{ and } r2 \text{ as Eq. (8),} \quad (11)$$

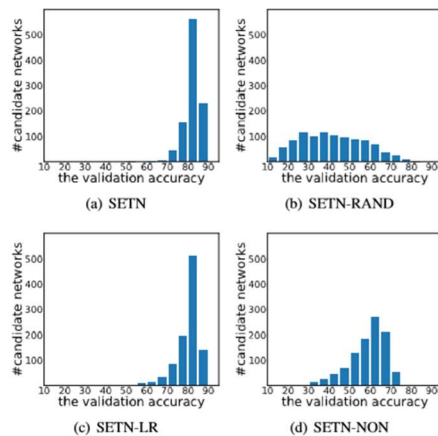
- 2 用 h,f,g 三个分布函数分别控制选择的操作, 选择的两个输入点, 跟 darts 一样优化, train 完之后根据这三个分布函数采样 1000 个网络。

Experiment:

1 Imagenet 实验:

	Method	GPU days	Parameters	+ × (million)	Top-1 Accuracy	Top-5 Accuracy
Human Expert	Inception-v1 [36]	—	6.6 MB	1448	69.8%	89.9%
	ResNet [18]	—	11.7 MB	1814	69.8%	89.1%
	MobileNet-v2 [34]	—	3.4 MB	300	72.0%	—
	ShuffleNet [42]	—	~5 MB	524	73.7%	—
NAS with more than 100 GPU days	Progressive NAS [23]	150	5.1 MB	588	74.2%	91.9%
	NASNet-A [44]	2000	5.3 MB	564	74.0%	91.6%
	NASNet-B [44]	2000	5.3 MB	488	72.8%	91.3%
	NASNet-C [44]	2000	4.9 MB	558	72.5%	91.0%
NAS with less than 5 GPU days	DARTS [25]	4	4.9 MB	595	73.1%	91.0%
	GHN [41]	0.84	6.1 MB	569	73.0%	91.3%
	SNAS [39]	1.5	4.3 MB	522	72.7%	90.8%
	GDAS [14]	0.84	5.3 MB	581	74.0%	91.5%
	SETN (N=1 & C=73)	1.8	5.2 MB	597	73.3%	91.4%
	SETN (N=2 & C=58)	1.8	5.3 MB	600	74.3%	91.6%
	SETN (N=3 & C=49)	1.8	5.3 MB	584	74.1%	91.9%
	SETN (N=4 & C=44)	1.8	5.4 MB	599	74.3%	92.0%

2 Ablation Studies:



A 是原方法, b 是 train 完后随机采样, c 是更少的随机, 每个 cell 选择的 input, operation 都一样, d 是按照 darts 那样去优化