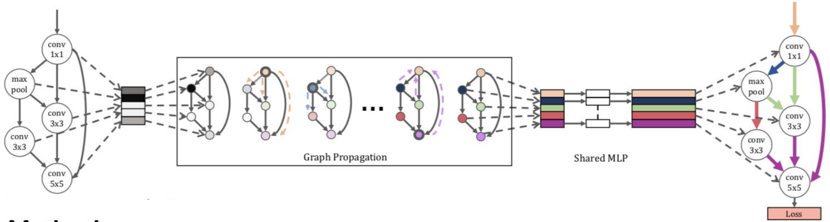
Graph HyperNetworks for Neural Architecture Search

Chris J. Zhang[1,2], Mengye Ren[1,3], Raquel Urtasun[1,3]

[1] Uber Advanced Technologies Group [2] University of Waterloo, [3] University of Toronto

Graph HyperNetworks



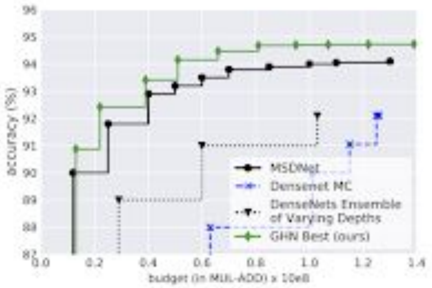
Motivation:

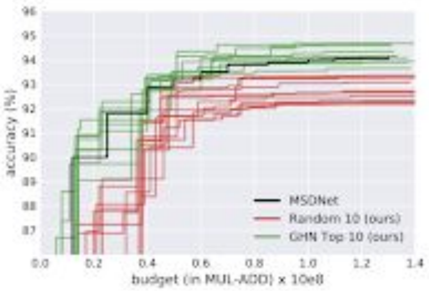
- Neural architecture search is an expensive nested optimization

$$a^* = \arg\min_a \mathcal{L}_{val}(w^*(a), a), \quad w^*(a) = \arg\min_w \mathcal{L}_{train}(w, a)$$

- Instead of using SGD to learn weights, use trained hypernetwork to generate weights
- Graph HyperNetworks (GHN) explicitly model the topology of architectures by learning on a computation graph representation

Anytime Prediction





NAS Benchmarks

CIFAR-10: Comparison with NAS methods which employ random search (top half) and advanced search methods (e.g. RL) (bottom half)

Method	Search Cost (GPU days)	Param $\times 10^6$	Accuracy
SMASHv1 (Brock et al., 2018)	?	4.6	94.5
SMASHv2 (Brock et al., 2018)	3	16.0	96.0
One-Shot Top (F=32) (Bender et al., 2018)	4	2.7 ± 0.3	95.5 ± 0.1
One-Shot Top (F=64) (Bender et al., 2018)	4	10.4 ± 1.0	95.9 ± 0.2
Random (F=32)	-	4.6 ± 0.6	94.6 ± 0.3
GHN Top (F=32)	0.42	5.1 ± 0.6	95.7 ± 0.1
NASNet-A (Zoph et al., 2018)	1800	3.3	97.35
ENAS Cell search (Pham et al., 2018)	0.45	4.6	97.11
DARTS (first order) (Liu et al., 2018b)	1.5	2.9	97.06
DARTS (second order) (Liu et al., 2018b)	4	3.4	97.17 ± 0.06
GHN Top-Best, 1K (F=32)	0.84	5.7	97.16 ± 0.07

ImageNet Mobile: Comparison with NAS methods which employ advanced search methods (e.g. RL)

Method	Search Cost	Param	FLOPs	Accuracy	
	(GPU days)	$\times 10^6$	$\times 10^6$	Top 1	Top 5
NASNet-A (Zoph et al., 2018)	1800	5.3	564	74.0	91.6
NASNet-C (Zoph et al., 2018)	1800	4.9	558	72.5	91.0
AmoebaNet-A (Real et al., 2018)	3150	5.1	555	74.5	92.0
AmoebaNet-C (Real et al., 2018)	3150	6.4	570	75.7	92.4
PNAS (Liu et al., 2018a)	225	5.1	588	74.2	91.9
DARTS (second order) (Liu et al., 2018b)	4	4.9	595	73.1	91.0
GHN Top-Best, 1K	0.84	6.1	569	73.0	91.3