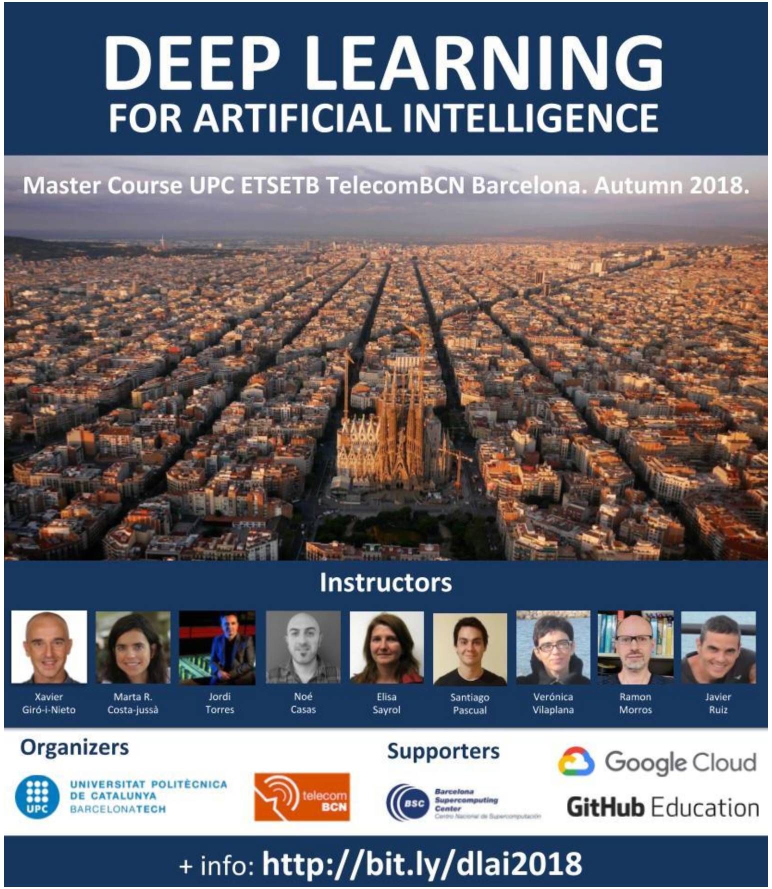


http://bit.ly/dlai2018

#DLBCN

Day 8 Lecture 2

Neural Architecture Search

Xavier Giro-i-Nieto
xavier.giro@upc.edu

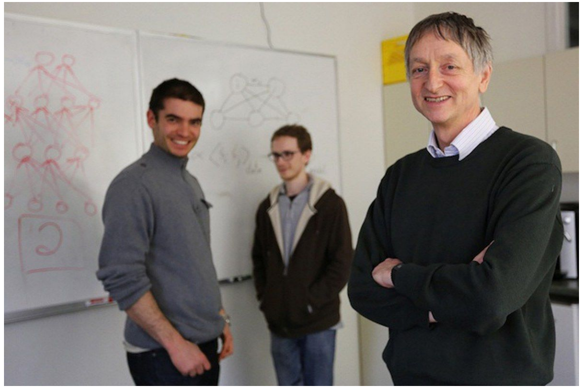
Associate Professor
Universitat Politècnica de Catalunya
Technical University of Catalonia

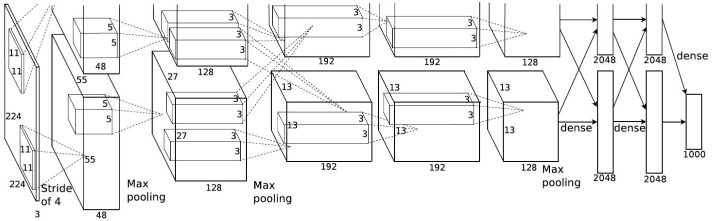
telecom BCN

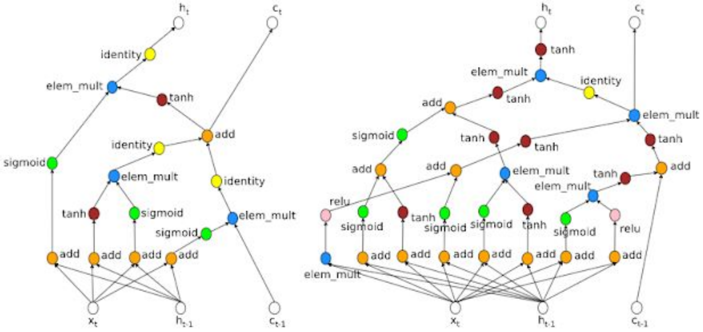
UPC

Neural Architecture Search (NAS)

Neural Architecture Search (AutoML): Instead of manually defining the architectures, allow software to search among a set of plausible architectures.







Neural Architecture Search (NAS)

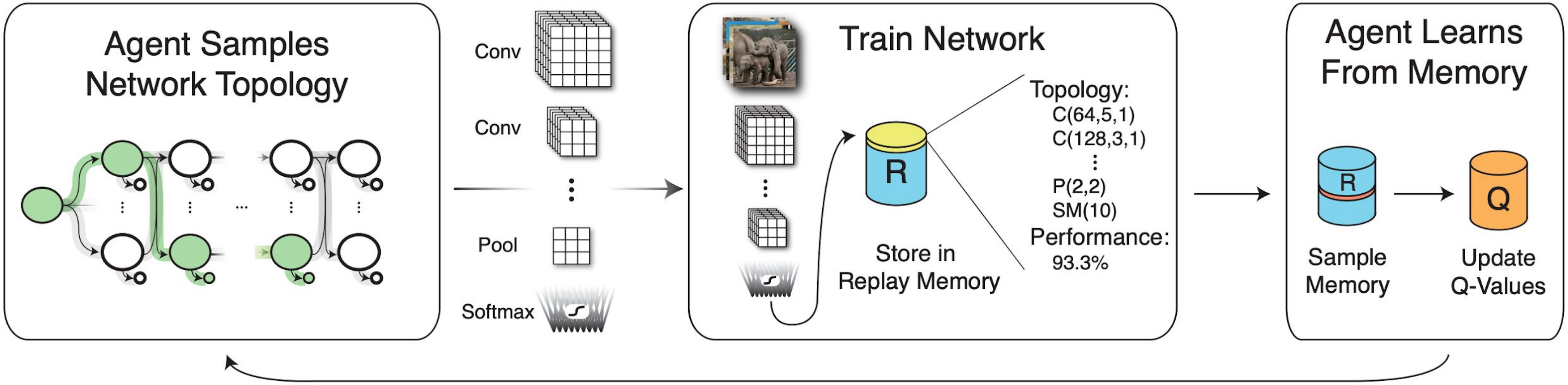
Reinforcement Learning

Evolution

Random Search

Neural Architecture Search (NAS) - Reinforcement Learning

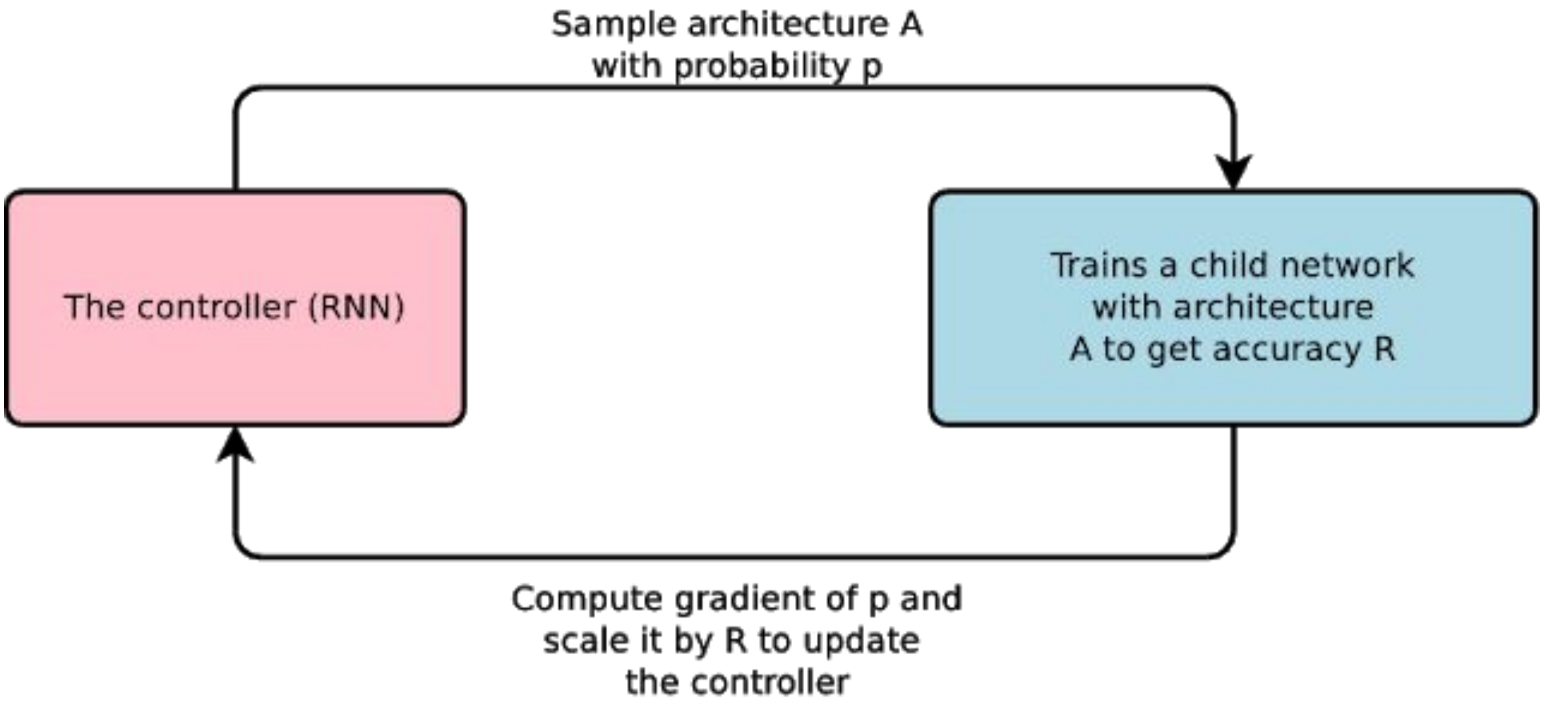
RL Neural Architecture Search: A controller to generate architectural hyperparameters of neural networks



Baker, B., Gupta, O., Naik, N., & Raskar, R. [Designing neural network architectures using reinforcement learning](). ICLR 2017.

Neural Architecture Search (NAS) - Reinforcement Learning

RL Neural Architecture Search: A controller to generate architectural hyperparameters of neural networks



Zoph, Barret, and Quoc V. Le. "Neural architecture search with reinforcement learning." ICLR 2017.

Neural Architecture Search (NAS) - Reinforcement Learning

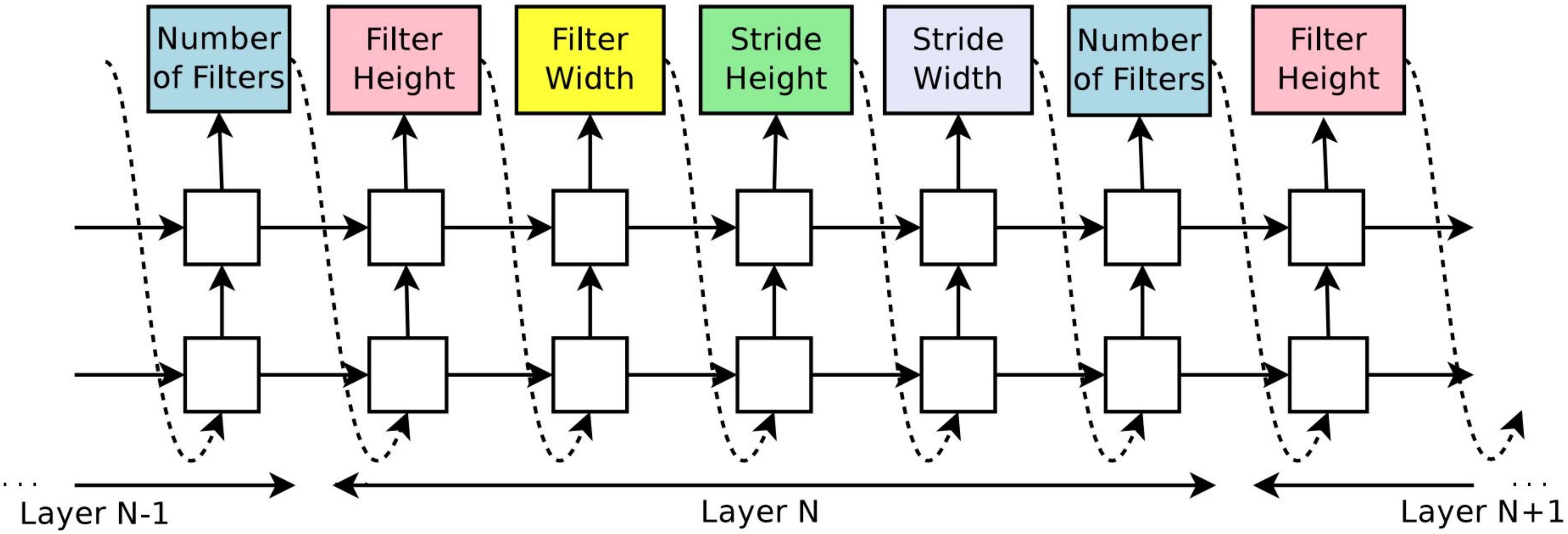
Model	Depth	Parameters	Error rate (%)
Network in Network (Lin et al., 2013)	-	-	8.81
All-CNN (Springenberg et al., 2014)	-	-	7.25
Deeply Supervised Net (Lee et al., 2015)	-	-	7.97
Highway Network (Srivastava et al., 2015)	-	-	7.72
Scalable Bayesian Optimization (Snoek et al., 2015)	-	-	6.37
FractalNet (Larsson et al., 2016)	21	38.6M	5.22
with Dropout/Drop-path	21	38.6M	4.60
ResNet (He et al., 2016a)	110	1.7M	6.61
ResNet (reported by Huang et al. (2016c))	110	1.7M	6.41
ResNet with Stochastic Depth (Huang et al., 2016c)	110	1.7M	5.23
	1202	10.2M	4.91
Wide ResNet (Zagoruyko & Komodakis, 2016)	16	11.0M	4.81
	28	36.5M	4.17
ResNet (pre-activation) (He et al., 2016b)	164	1.7M	5.46
	1001	10.2M	4.62
DenseNet ($L = 40, k = 12$) Huang et al. (2016a)	40	1.0M	5.24
DenseNet($L = 100, k = 12$) Huang et al. (2016a)	100	7.0M	4.10
DenseNet ($L = 100, k = 24$) Huang et al. (2016a)	100	27.2M	3.74
DenseNet-BC ($L = 100, k = 40$) Huang et al. (2016b)	190	25.6M	3.46
Neural Architecture Search v1 no stride or pooling	15	4.2M	5.50
Neural Architecture Search v2 predicting strides	20	2.5M	6.01
Neural Architecture Search v3 max pooling	39	7.1M	4.47
Neural Architecture Search v3 max pooling + more filters	39	37.4M	3.65

Table 1: Performance of Neural Architecture Search and other state-of-the-art models on CIFAR-10.

Zoph, Barret, and Quoc V. Le. "Neural architecture search with reinforcement learning." ICLR 2017.

Neural Architecture Search (NAS) - Reinforcement Learning

A controller recurrent neural network samples a simple convolutional network:



Zoph, Barret, and Quoc V. Le. "Neural architecture search with reinforcement learning." ICLR 2017.

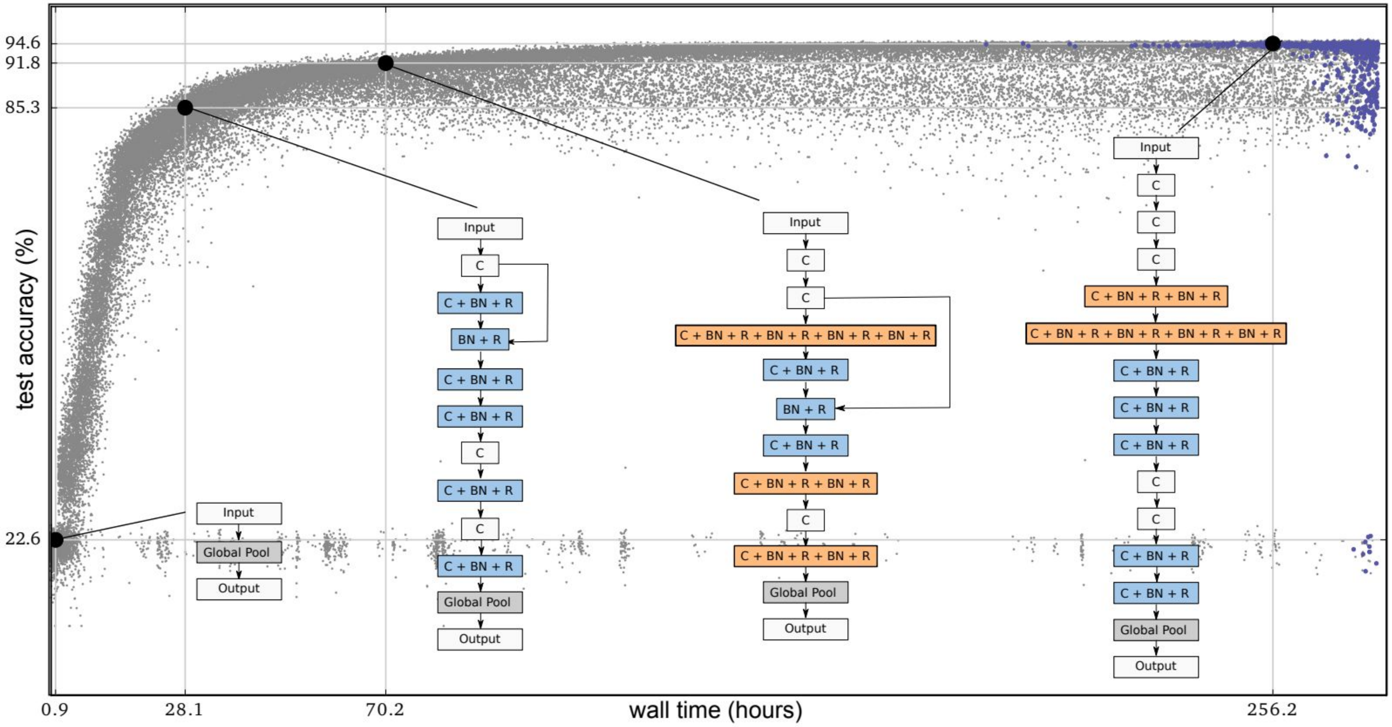
Neural Architecture Search (NAS)

Reinforcement Learning

Evolution

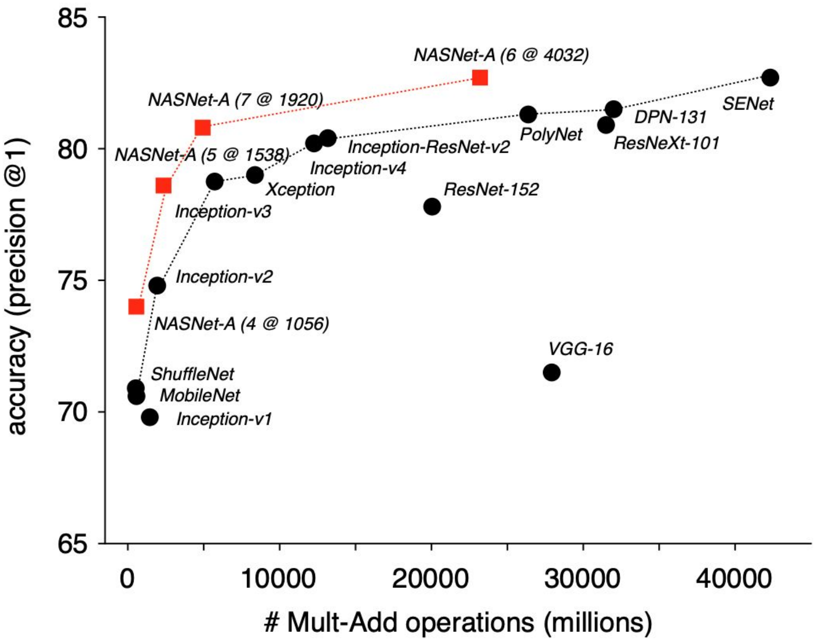
Random Search

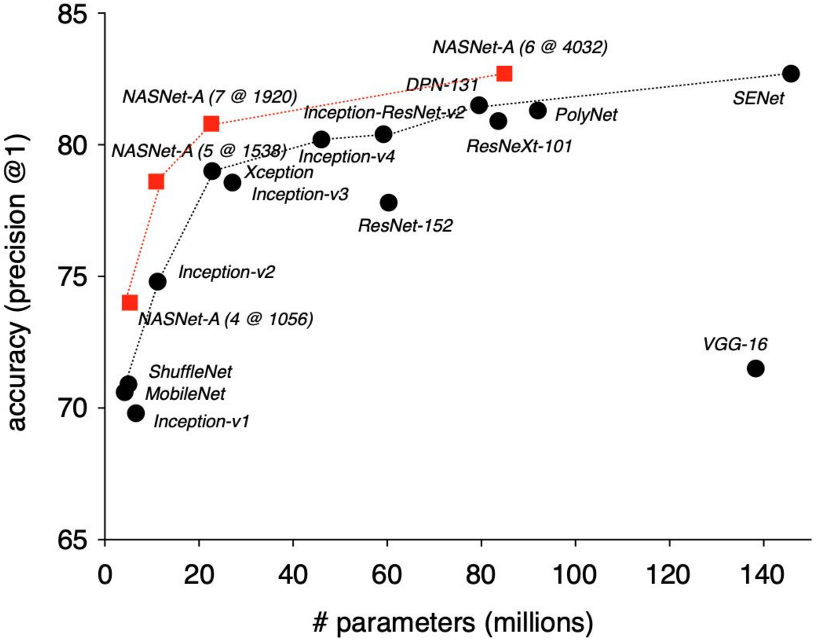
Neural Architecture Search (NAS)- Evolutionary



#AutoML Real, Esteban, Sherry Moore, Andrew Selle, Saurabh Saxena, Yutaka Leon Suematsu, Jie Tan, Quoc V. Le, and Alexey Kurakin. "Large-scale evolution of image classifiers." ICML 2017. [blog]

Neural Architecture Search (NAS)



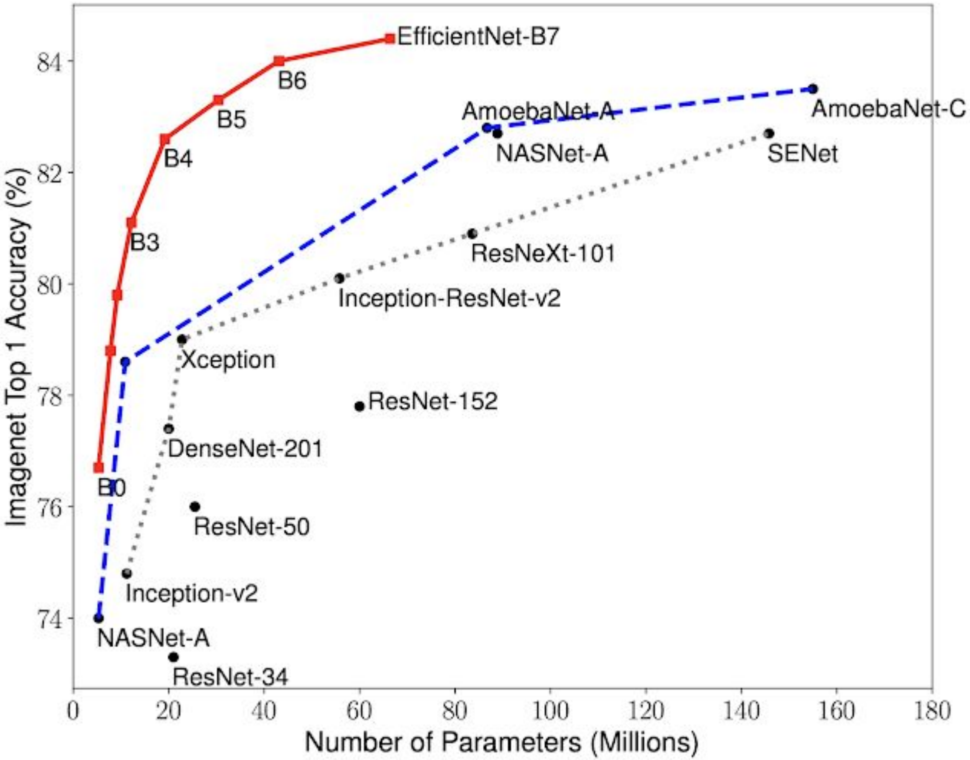


#NasNet Zoph, Barret, Vijay Vasudevan, Jonathon Shlens, and Quoc V. Le. ["Learning transferable architectures for scalable image recognition."]() CVPR 2018.

Neural Architecture Search (NAS) - Ensembles

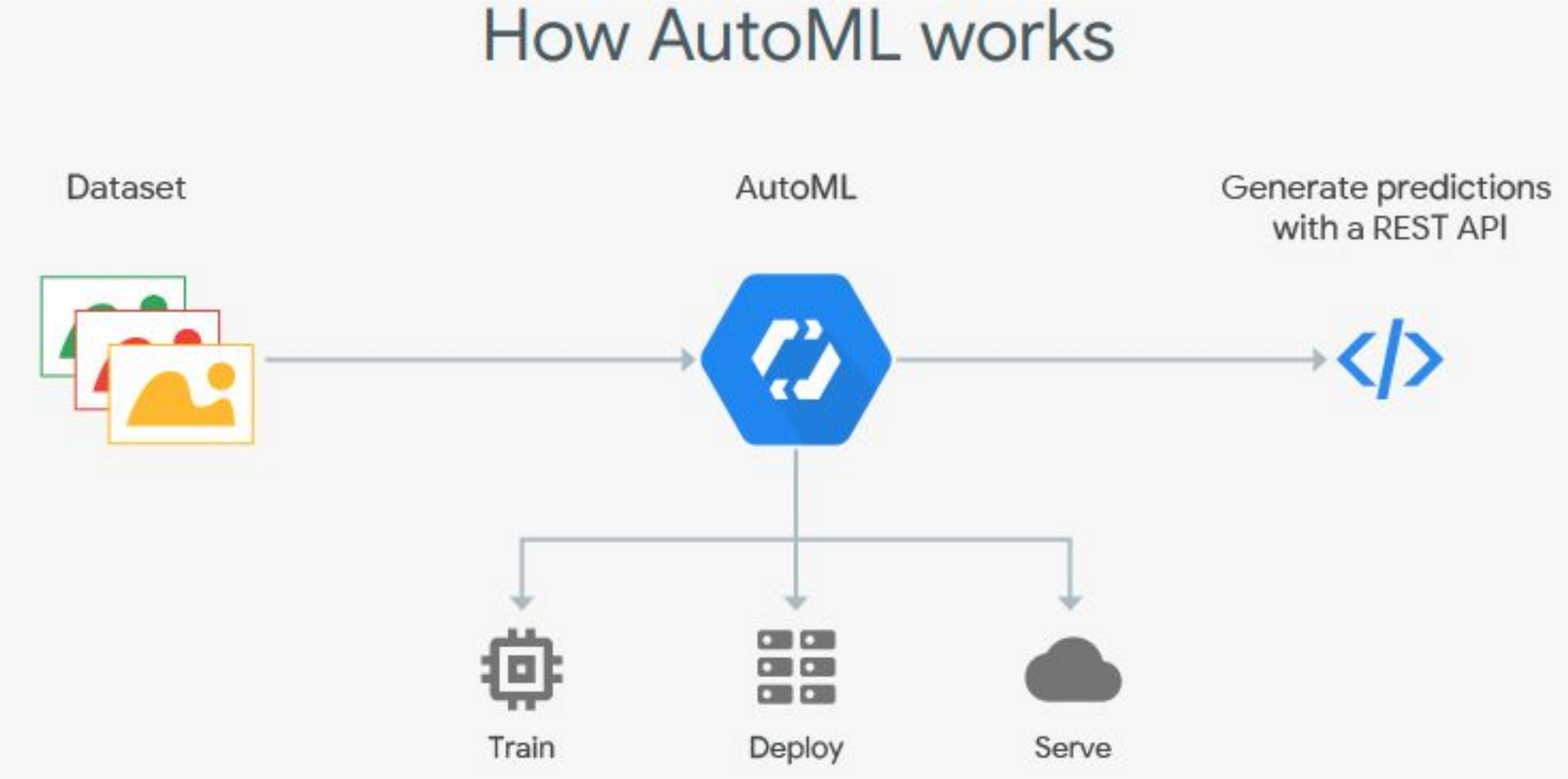
#AdaNet Cortes, Corinna, Xavier Gonzalvo, Vitaly Kuznetsov, Mehryar Mohri, and Scott Yang. "Adanet: Adaptive structural learning of artificial neural networks." ICML 2017. [blog]

Neural Architecture Search (NAS)



#EfficientNet Tan, Mingxing, and Quoc V. Le. "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks." arXiv preprint arXiv:1905.11946 (2019). [blog]

Neural Architecture Search (NAS)



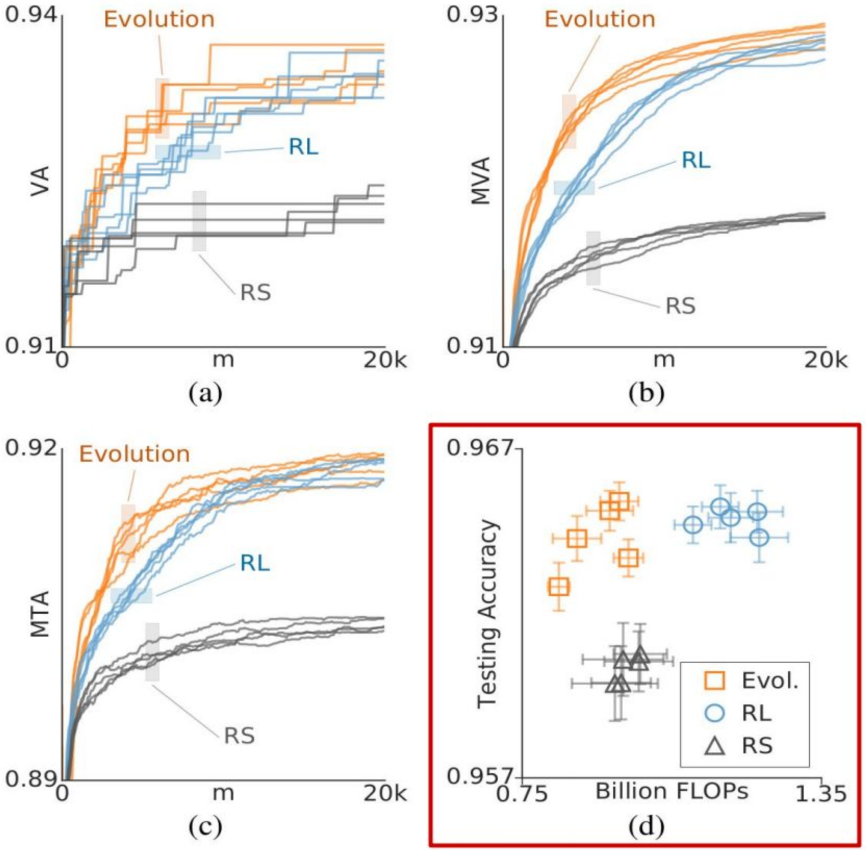
Neural Architecture Search (NAS)

Reinforcement Learning

Evolution

Random Search

Neural Architecture Search (NAS)



Real et al. (2018)
The difference in accuracy between best models found by random search, RL, and Evolution is **less than 1%** on CIFAR-10

Nikhil Nail (Salesforce Research), [“Neural Architecture Search: A Tutorial”]() CVPR 2019

Neural Architecture Search (object detection)

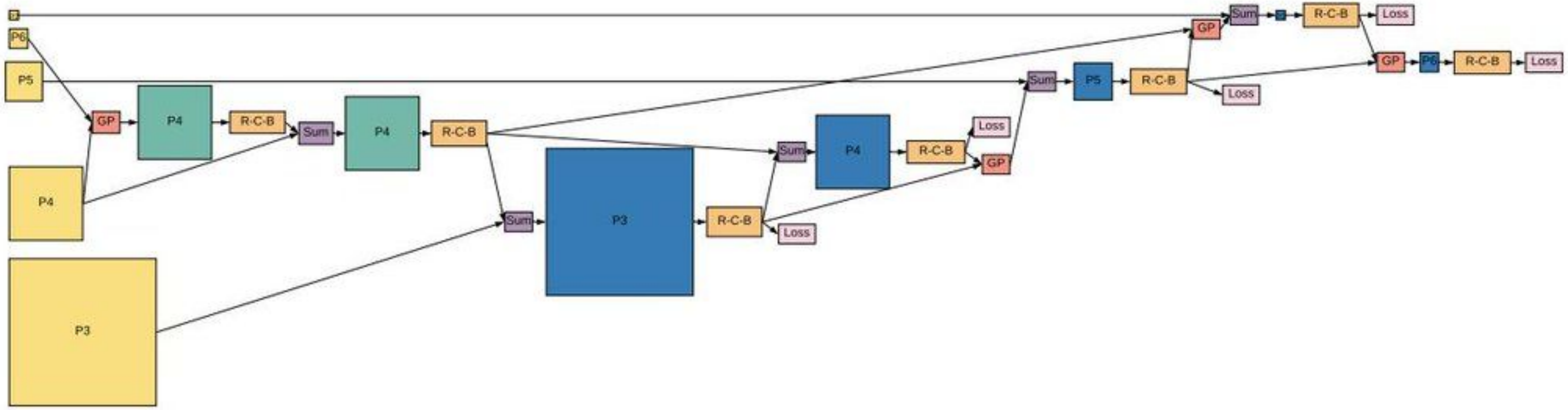
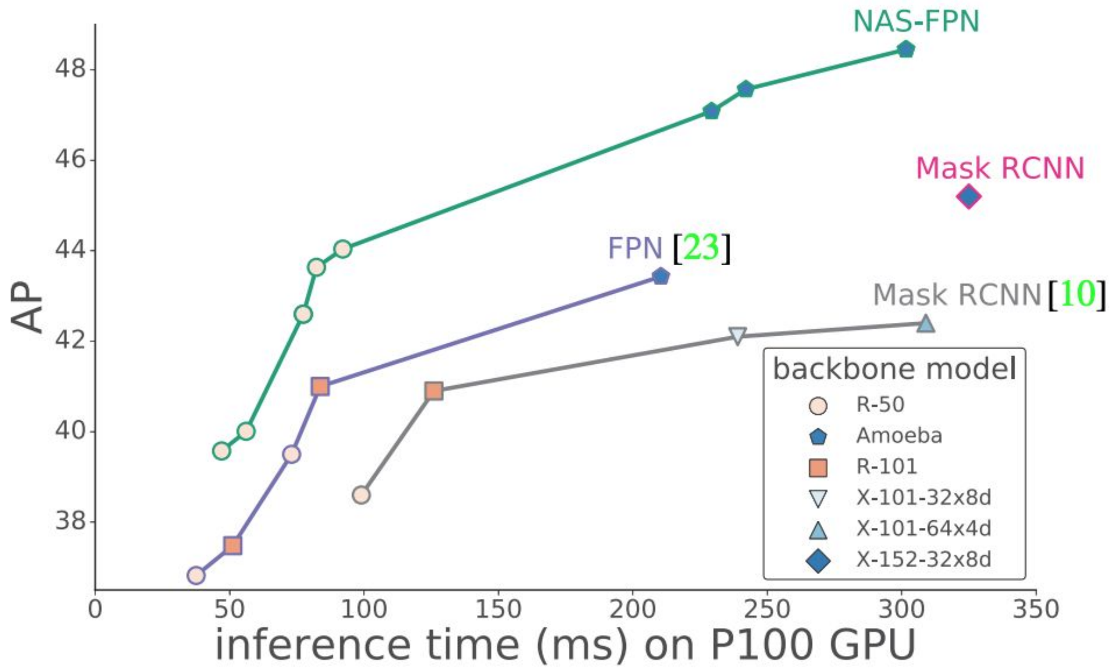


Figure 6: Architecture of the discovered 7-merging-cell pyramid network in NAS-FPN with 5 input layers (yellow) and 5 output feature layers (blue). GP and R-C-B are stands for Global Pooling and ReLU-Conv-BatchNorm, respectively.

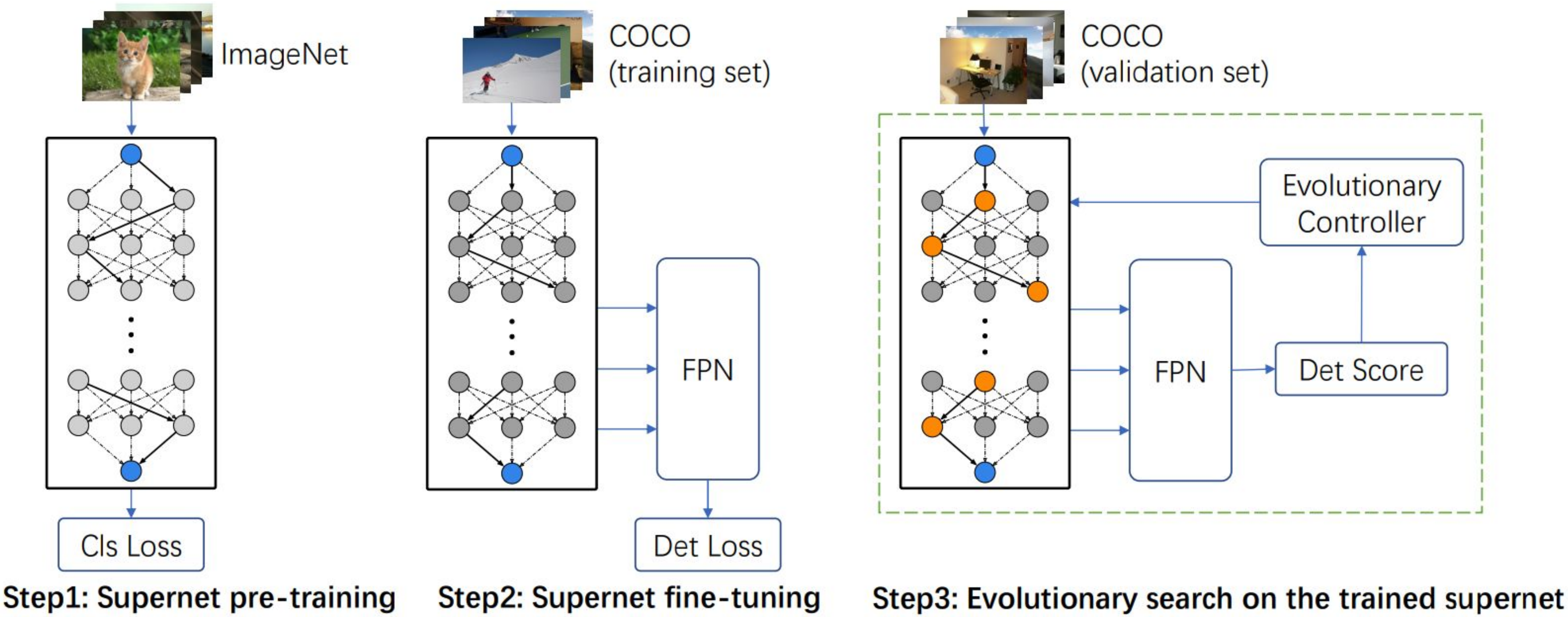
#NAS-FPN Golnaz Ghiasi, Tsung-Yi Lin, Ruoming Pang, Quoc V. Le, [“NAS-FPN: Learning Scalable Feature Pyramid Architecture for Object Detection”]() CVPR 2019

Neural Architecture Search (object detection)



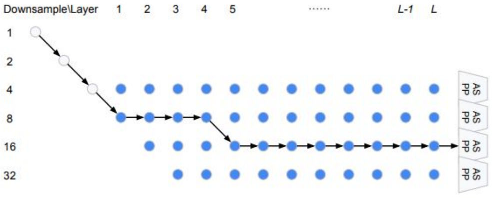
#NAS-FPN Golnaz Ghiasi, Tsung-Yi Lin, Ruoming Pang, Quoc V. Le, [“NAS-FPN: Learning Scalable Feature Pyramid Architecture for Object Detection”]() CVPR 2019

Neural Architecture Search (object detection)

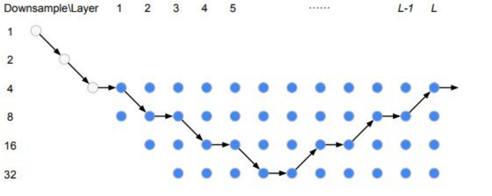


#DetNAS Chen, Yukang, Tong Yang, Xiangyu Zhang, Gaofeng Meng, Chunhong Pan, and Jian Sun. ["Detnas: Neural architecture search on object detection."]() arXiv preprint arXiv:1903.10979 (2019).

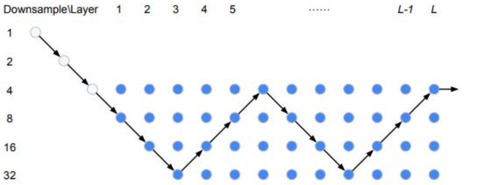
Neural Architecture Search (segmentation)



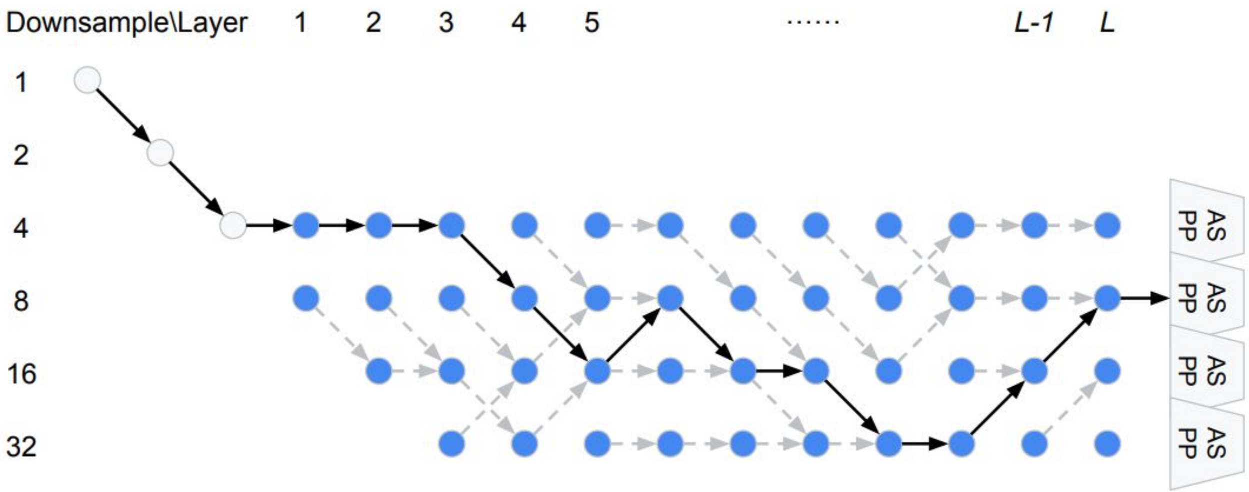
(a) Network level architecture used in DeepLabv3 [9].



(b) Network level architecture used in Conv-Deconv [56].



(c) Network level architecture used in Stacked Hourglass [55].



#Auto-DeepLab Liu, C., Chen, L. C., Schroff, F., Adam, H., Hua, W., Yuille, A. L., & Fei-Fei, L. (2019). [Auto-deeplab: Hierarchical neural architecture search for semantic image segmentation](). CVPR 2019

Neural Architecture Search (segmentation)

Method	ImageNet	F	Multi-Adds	Params	mIOU (%)
Auto-DeepLab-S		20	333.25B	10.15M	79.74
Auto-DeepLab-M		32	460.93B	21.62M	80.04
Auto-DeepLab-L		48	695.03B	44.42M	80.33
FRRN-A [60]		-	-	17.76M	65.7
FRRN-B [60]		-	-	24.78M	-
DeepLabv3+ [11]	✓	-	1551.05B	43.48M	79.55

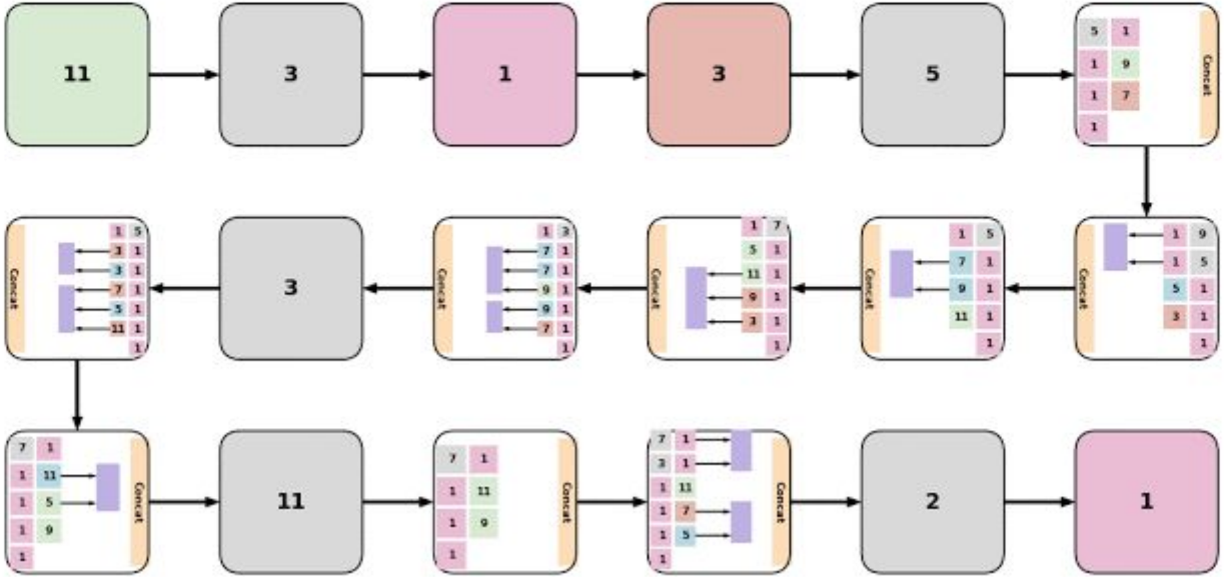
Table 2: Cityscapes validation set results with different Auto-DeepLab model variants. F: the filter multiplier controlling the model capacity. All our models are trained from *scratch* and with *single-scale* input during inference.

Method	**ImageNet**	**Coarse**	mIOU (%)
FRRN-A [60]			63.0
GridNet [17]			69.5
FRRN-B [60]			71.8
Auto-DeepLab-S			79.9
Auto-DeepLab-L			80.4
Auto-DeepLab-S		✓	80.9
Auto-DeepLab-L		✓	82.1
ResNet-38 [82]	✓	✓	80.6
PSPNet [88]	✓	✓	81.2
Mapillary [4]	✓	✓	82.0
DeepLabv3+ [11]	✓	✓	82.1
DPC [6]	✓	✓	82.7
DRN_CRL_Coarse [91]	✓	✓	82.8

Table 4: Cityscapes test set results with *multi-scale* inputs during inference. **ImageNet:** Models pretrained on ImageNet. **Coarse:** Models exploit coarse annotations.

#Auto-DeepLab Liu, C., Chen, L. C., Schroff, F., Adam, H., Hua, W., Yuille, A. L., & Fei-Fei, L. (2019). Auto-deeplab: Hierarchical neural architecture search for semantic image segmentation. CVPR 2019

Neural Architecture Search (video)



#EvaNet Piergiovanni, A. J., et al. "Evolving space-time neural architectures for videos." ICCV 2019. [blog] [code]

Neural Architecture Search (video)

Table 4. Charades classification results against state-of-the-arts.

	mAP
Two-Stream [25]	18.6
Two-Stream + LSTM [25]	17.8
Async-TF [25]	22.4
TRN [40]	25.2
Dicrim. Pooling [35]	26.7
Non-local NN [36]	37.5
3D-Ensemble (baseline)	35.2
iTGM-Ensemble (baseline)	35.7
Top 1 (Individual, ours)	**37.3**
Top 2 (Individual, ours)	36.8
Top 3 (Individual, ours)	36.6
EvaNet (Ensemble, ours)	**38.1**

Table 7. Runtime measured on a V100 GPU. Accuracy numbers on Kinetics-400 are added for context. These numbers are evaluation time for 1 128 frame clip at 224x224.

Method	Accuracy	Runtime
I3D	72.6	337ms
S3D	75.2	439ms
ResNet-50	71.9	526ms
ResNet-50 + Non-local	73.5	572ms
I3D iTGM (ours)	74.4	274ms
Individual learned model (ours)	75.5	**108ms**
EvaNet (Ensemble, ours)	77.2	**258ms**

#EvaNet Piergiovanni, A. J., et al. "Evolving space-time neural architectures for videos." ICCV 2019. [blog] [code]

Neural Architecture Search (video)

Table 4. Charades classification results against state-of-the-arts.

	mAP
Two-Stream [25]	18.6
Two-Stream + LSTM [25]	17.8
Async-TF [25]	22.4
TRN [40]	25.2
Dicrim. Pooling [35]	26.7
Non-local NN [36]	37.5
3D-Ensemble (baseline)	35.2
iTGM-Ensemble (baseline)	35.7
Top 1 (Individual, ours)	**37.3**
Top 2 (Individual, ours)	36.8
Top 3 (Individual, ours)	36.6
EvaNet (Ensemble, ours)	**38.1**

Table 7. Runtime measured on a V100 GPU. Accuracy numbers on Kinetics-400 are added for context. These numbers are evaluation time for 1 128 frame clip at 224x224.

Method	Accuracy	Runtime
I3D	72.6	337ms
S3D	75.2	439ms
ResNet-50	71.9	526ms
ResNet-50 + Non-local	73.5	572ms
I3D iTGM (ours)	74.4	274ms
Individual learned model (ours)	75.5	**108ms**
EvaNet (Ensemble, ours)	77.2	**258ms**

#EvaNet Piergiovanni, A. J., et al. "Evolving space-time neural architectures for videos." ICCV 2019. [blog] [code]

Neural Architecture Search (translation)

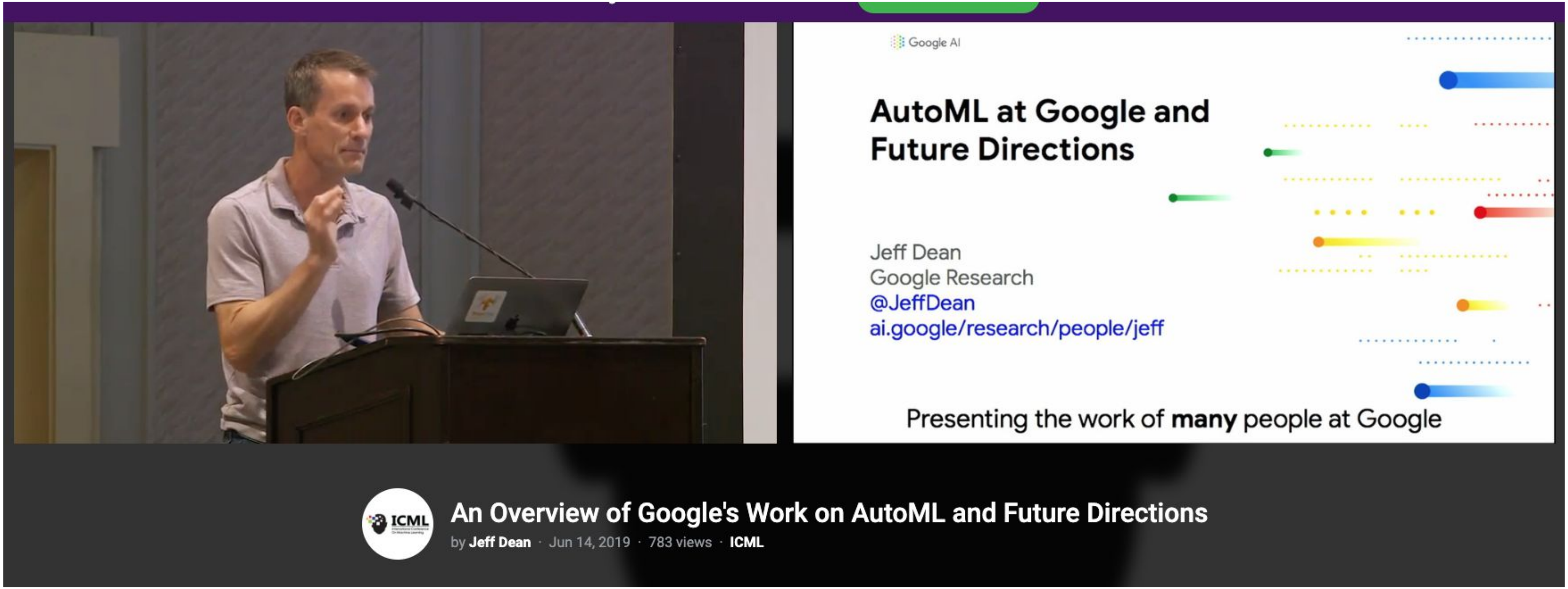
Emma Strubell
@strubell

Are you interested in deep learning for NLP but also concerned about the CO2 footprint of training? You should be! Excited to share our work "Energy and Policy Considerations for Deep Learning in NLP" at @ACL2019_Italy! With @ananya__g and @andrewmccallum. Preprint coming soon.

Consumption	**CO_2e (lbs)**
Air travel, 1 passenger, NY↔SF	1984
Human life, avg, 1 year	11,023
American life, avg, 1 year	36,156
Car, avg incl. fuel, 1 lifetime	126,000
Training one model	
SOTA NLP model (tagging)	13
w/ tuning & experimentation	33,486
Transformer (large)	121
w/ neural architecture search	394,863

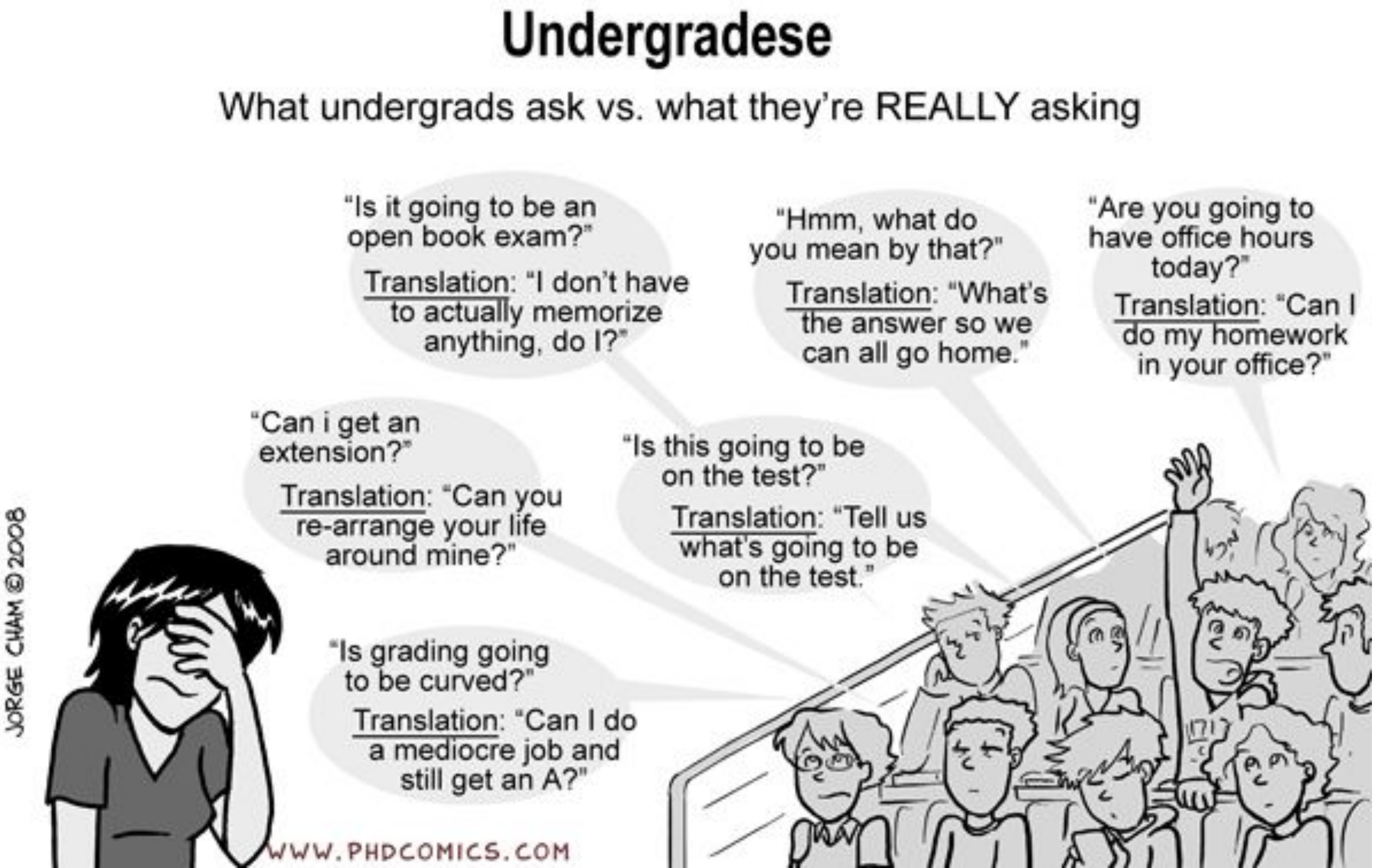
Strubell, Emma, Ananya Ganesh, and Andrew McCallum. "Energy and Policy Considerations for Deep Learning in NLP." ACL 2019. [tweet]

Learn more



Jeff Dean (Google AI), [“AutoML at Google and Future Directions”](). ICML 2019.

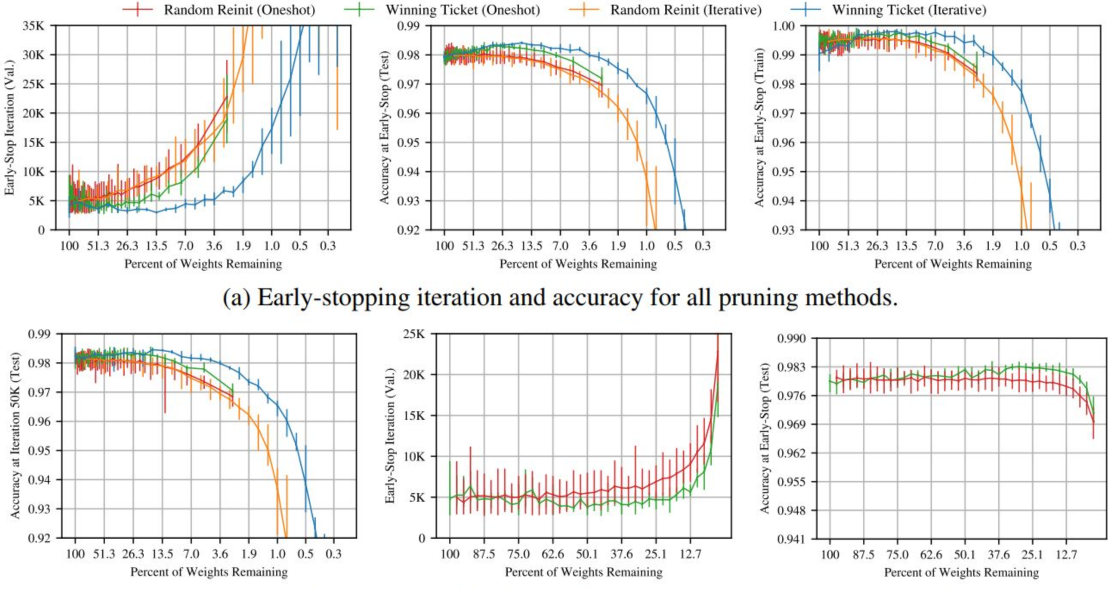
Questions ?


Undergradese
What undergrads ask vs. what they're REALLY asking
"Is it going to be an open book exam?"
Translation: "I don't have to actually memorize anything, do I?"
"Hmm, what do you mean by that?"
Translation: "What's the answer so we can all go home."
"Are you going to have office hours today?"
Translation: "Can I do my homework in your office?"
"Can i get an extension?"
Translation: "Can you re-arrange your life around mine?"
"Is this going to be on the test?"
Translation: "Tell us what's going to be on the test."
"Is grading going to be curved?"
Translation: "Can I do a mediocre job and still get an A?"
JORGE CHAM © 2008
WWW.PHDCOMICS.COM

Progressive NAS

ReadAI [slides]

Jonathan Frankle, Michael Carbin. The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks. ICLR 2019 (oral)



(a) Early-stopping iteration and accuracy for all pruning methods.

(b) Accuracy at end of training.

(c) Early-stopping iteration and accuracy for one-shot pruning.