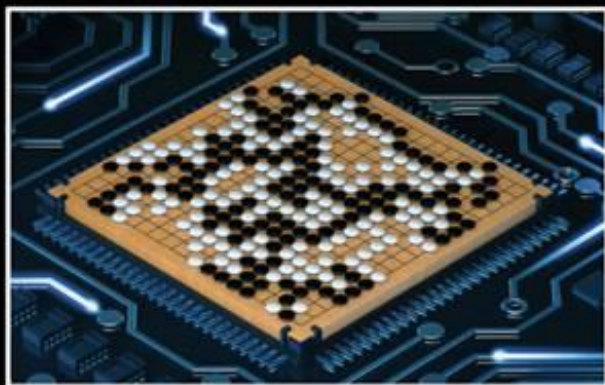


Automatically Determining Hyperparameters

Source of image: <https://medium.com/intuitionmachine/the-brute-force-method-of-deep-learning-innovation-58b497323ae5> (Denny Britz's graphic)

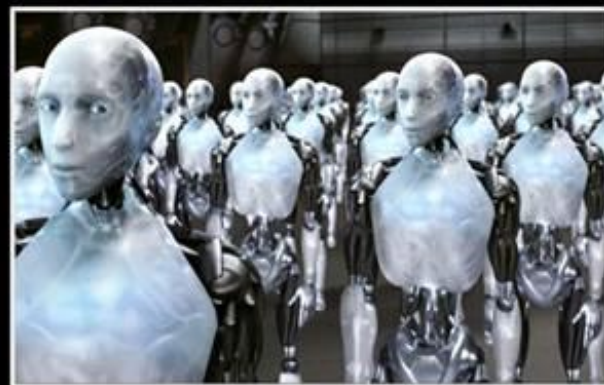
Deep Learning 研究生



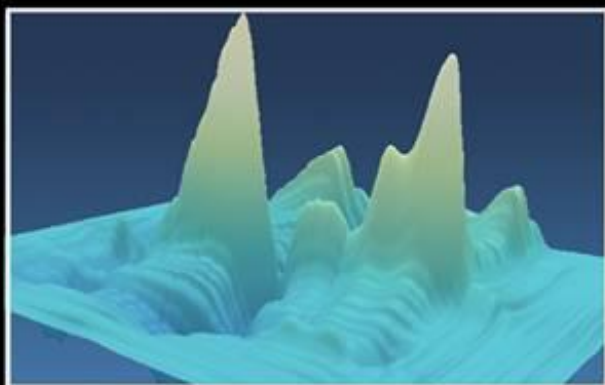
朋友覺得我在



我媽覺得我在



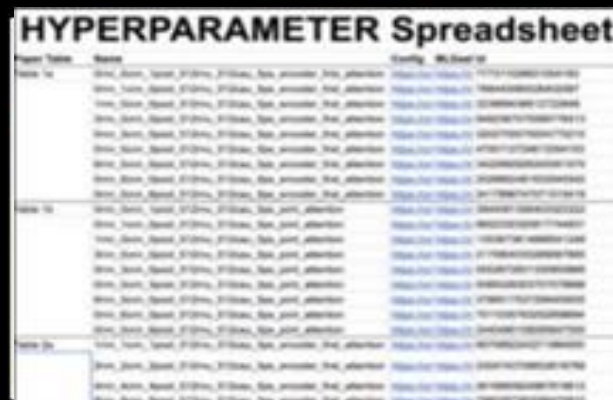
大眾覺得我在



指導教授覺得我在



我以為我在

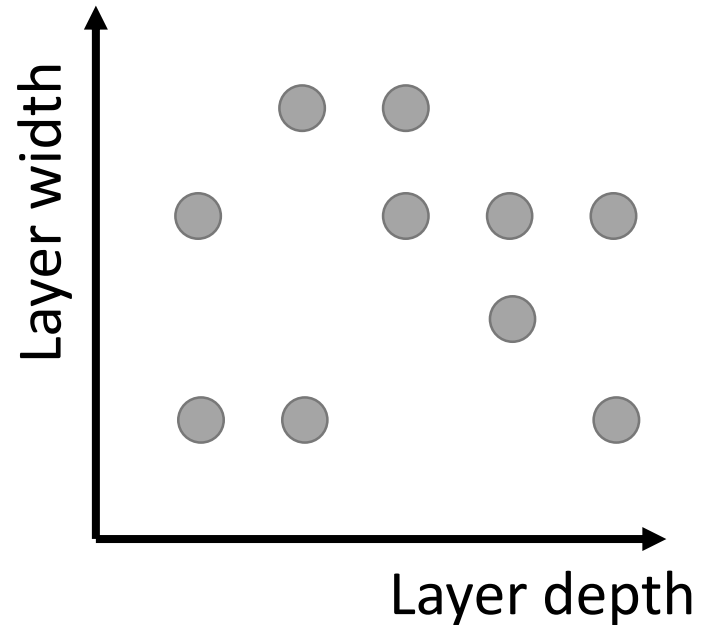
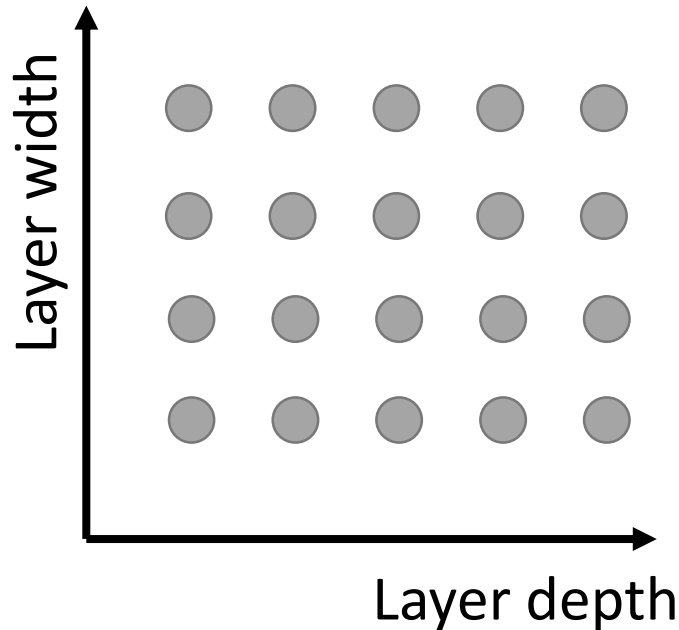


HYPERPARAMETER Spreadsheet		
Name	Email	Config. Metadata
John Doe	john.doe@university.edu	lr=0.001, batch_size=32, num_epochs=100
Jane Smith	jane.smith@university.edu	lr=0.01, batch_size=64, num_epochs=50
Mike Johnson	mike.johnson@university.edu	lr=0.0001, batch_size=128, num_epochs=200
Sarah Lee	sarah.lee@university.edu	lr=0.001, batch_size=16, num_epochs=1000
David Kim	david.kim@university.edu	lr=0.0005, batch_size=256, num_epochs=500
Emily White	emily.white@university.edu	lr=0.001, batch_size=32, num_epochs=100
Chris Brown	chris.brown@university.edu	lr=0.0001, batch_size=64, num_epochs=200
Alex Green	alex.green@university.edu	lr=0.001, batch_size=128, num_epochs=100
Olivia Black	olivia.black@university.edu	lr=0.0005, batch_size=32, num_epochs=500
Noah Gray	noah.gray@university.edu	lr=0.001, batch_size=64, num_epochs=100
Isabella Blue	isabella.blue@university.edu	lr=0.0001, batch_size=128, num_epochs=200
Liam Red	liam.red@university.edu	lr=0.001, batch_size=32, num_epochs=100
Mia Purple	mia.purple@university.edu	lr=0.0005, batch_size=64, num_epochs=500
Lucas Yellow	lucas.yellow@university.edu	lr=0.001, batch_size=128, num_epochs=100
Charlotte Orange	charlotte.orange@university.edu	lr=0.0001, batch_size=32, num_epochs=200
Benjamin Silver	benjamin.silver@university.edu	lr=0.001, batch_size=64, num_epochs=100
Abigail Gold	abigail.gold@university.edu	lr=0.0005, batch_size=128, num_epochs=500
Ethan Bronze	ethan.bronze@university.edu	lr=0.001, batch_size=32, num_epochs=100
Sophia Platinum	sophia.platinum@university.edu	lr=0.0001, batch_size=64, num_epochs=200
Matthew Nickel	matthew.nickel@university.edu	lr=0.001, batch_size=128, num_epochs=100
Chloe Copper	chloe.copper@university.edu	lr=0.0005, batch_size=32, num_epochs=500
Andrew Zinc	andrew.zinc@university.edu	lr=0.001, batch_size=64, num_epochs=100
Grace Iron	grace.iron@university.edu	lr=0.0001, batch_size=128, num_epochs=200
Robert Tin	robert.tin@university.edu	lr=0.001, batch_size=32, num_epochs=100
Victoria Lead	victoria.lead@university.edu	lr=0.0005, batch_size=64, num_epochs=500
William Cadmium	william.cadmium@university.edu	lr=0.001, batch_size=128, num_epochs=100
Oliver Mercury	oliver.mercury@university.edu	lr=0.0001, batch_size=32, num_epochs=200
Isabella Selenium	isabella.selenium@university.edu	lr=0.001, batch_size=64, num_epochs=100
Lucas Tellurium	lucas.tellurium@university.edu	lr=0.0005, batch_size=128, num_epochs=500
Chloe Polonium	chloe.polonium@university.edu	lr=0.001, batch_size=32, num_epochs=100
Andrew Astatine	andrew.astatine@university.edu	lr=0.0001, batch_size=64, num_epochs=200
Grace Francium	grace.francium@university.edu	lr=0.001, batch_size=128, num_epochs=100
Robert Actinium	robert.actinium@university.edu	lr=0.0005, batch_size=32, num_epochs=500
Victoria Thorium	victoria.thorium@university.edu	lr=0.001, batch_size=64, num_epochs=100
William Protactinium	william.protactinium@university.edu	lr=0.0001, batch_size=128, num_epochs=200
Oliver Uranium	oliver.uranium@university.edu	lr=0.001, batch_size=32, num_epochs=100
Isabella Neptunium	isabella.neptunium@university.edu	lr=0.0005, batch_size=64, num_epochs=500
Lucas Plutonium	lucas.plutonium@university.edu	lr=0.001, batch_size=128, num_epochs=100
Chloe Americium	chloe.americium@university.edu	lr=0.0001, batch_size=32, num_epochs=200
Andrew Curium	andrew.curium@university.edu	lr=0.001, batch_size=64, num_epochs=100
Grace Berkelium	grace.berkelium@university.edu	lr=0.0005, batch_size=128, num_epochs=500
Robert Californium	robert.californium@university.edu	lr=0.001, batch_size=32, num_epochs=100
Victoria Einsteinium	victoria.einsteinium@university.edu	lr=0.0001, batch_size=64, num_epochs=200
William Fermium	william.fermium@university.edu	lr=0.001, batch_size=128, num_epochs=100
Oliver Mendelevium	oliver.mendelevium@university.edu	lr=0.0005, batch_size=32, num_epochs=500
Isabella Nobelium	isabella.nobelium@university.edu	lr=0.001, batch_size=64, num_epochs=100
Lucas Lawrencium	lucas.lawrencium@university.edu	lr=0.0001, batch_size=128, num_epochs=200
Chloe Rutherfordium	chloe.rutherfordium@university.edu	lr=0.001, batch_size=32, num_epochs=100
Andrew Dubnium	andrew.dubnium@university.edu	lr=0.0005, batch_size=64, num_epochs=500
Grace Seaborgium	grace.seaborgium@university.edu	lr=0.001, batch_size=128, num_epochs=100
Robert Bohrium	robert.bohrium@university.edu	lr=0.0001, batch_size=32, num_epochs=200
Victoria Hassium	victoria.hassium@university.edu	lr=0.001, batch_size=64, num_epochs=100
William Meitnerium	william.meitnerium@university.edu	lr=0.0005, batch_size=128, num_epochs=500
Oliver Darmstadtium	oliver.darmstadtium@university.edu	lr=0.001, batch_size=32, num_epochs=100
Isabella Tennessine	isabella.tennessine@university.edu	lr=0.0001, batch_size=64, num_epochs=200
Lucas Oganesson	lucas.oganesson@university.edu	lr=0.001, batch_size=128, num_epochs=100

事實上我在

感謝 沈昇勳 同學提供圖檔

Grid Search v.s. Random Search



Assumption: **top K results are good enough**

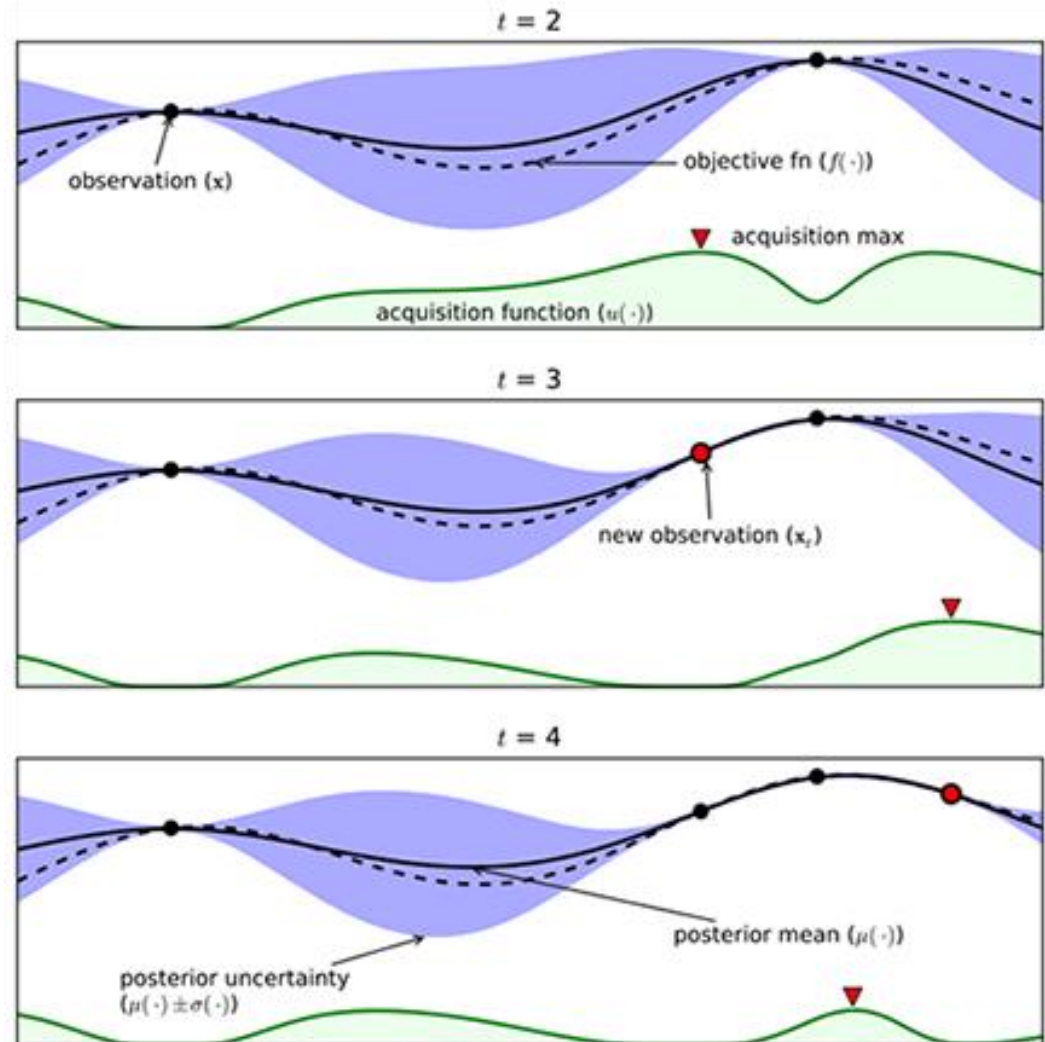
If there are N points, probability K/N that your sample is in top K

Sample x times: $1 - (1 - K/N)^x > 90\%$

If $N = 1000$, $K = 10$ \longrightarrow $x = 230$

$K = 100$ \longrightarrow $x = 22$

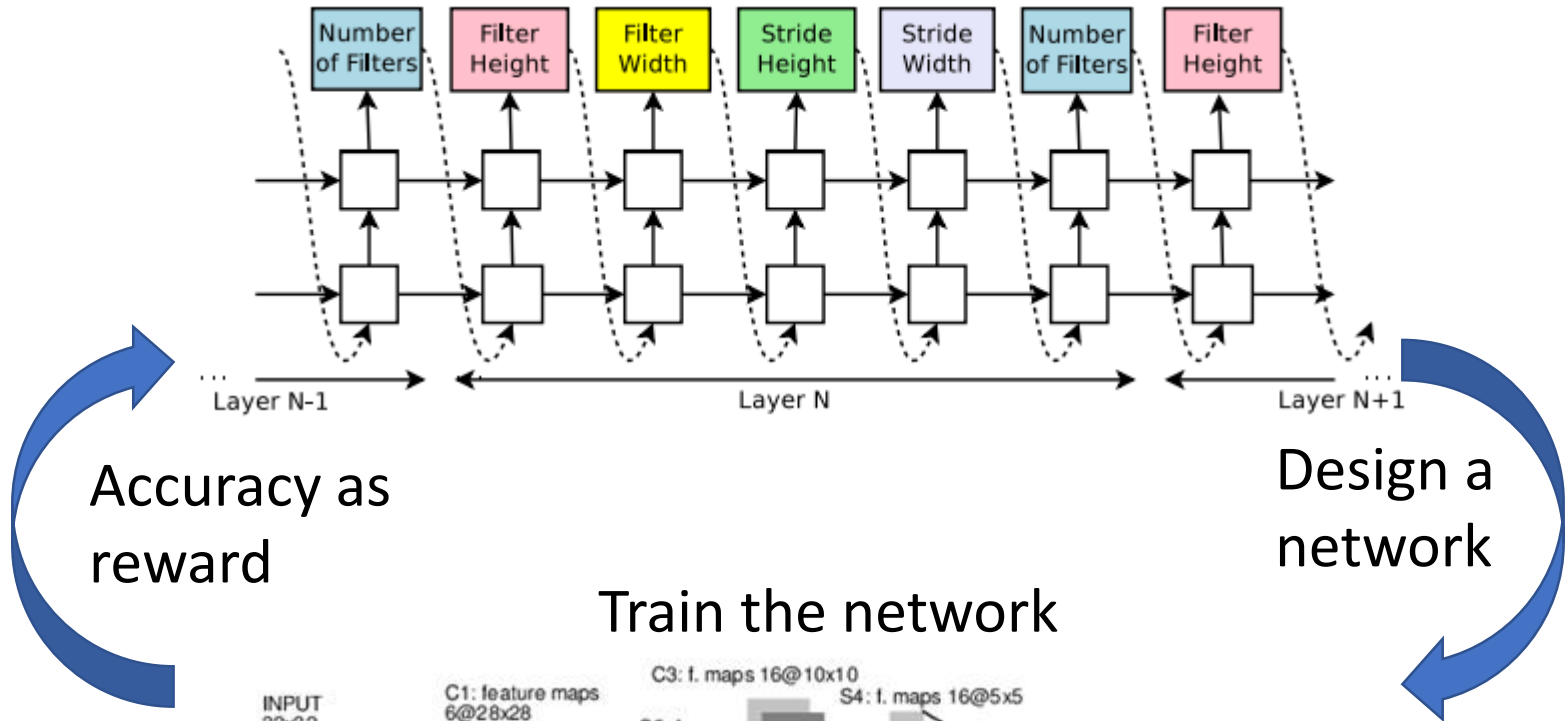
Model-based Hyperparameter Optimization



<https://cloud.google.com/blog/big-data/2017/08/hyperparameter-tuning-in-cloud-machine-learning-engine-using-bayesian-optimization>

Reinforcement Learning

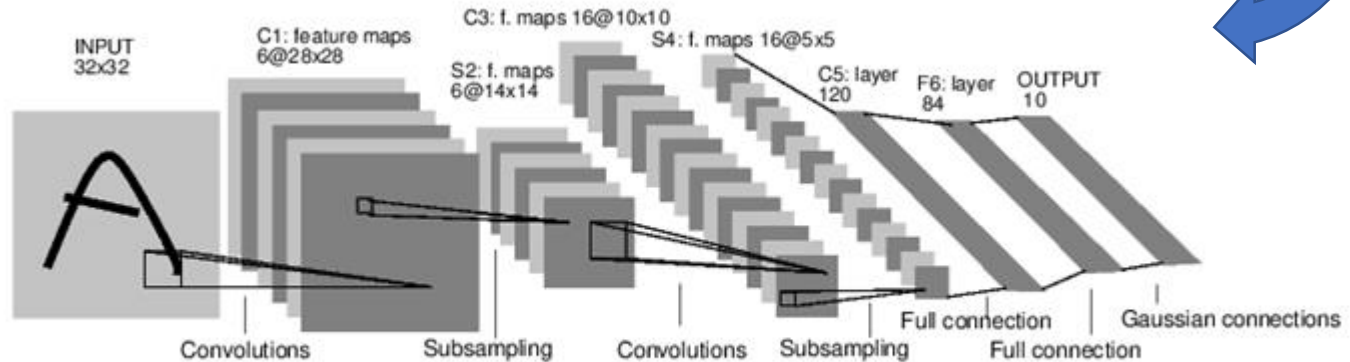
One kind of meta learning (or learn to learn)



Accuracy as reward

Design a network

Train the network



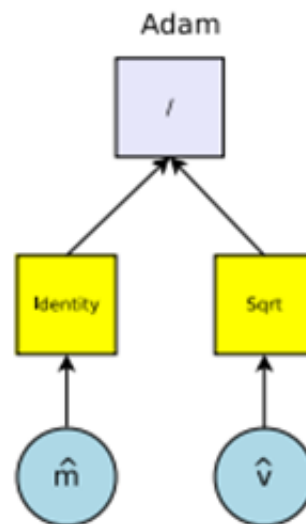
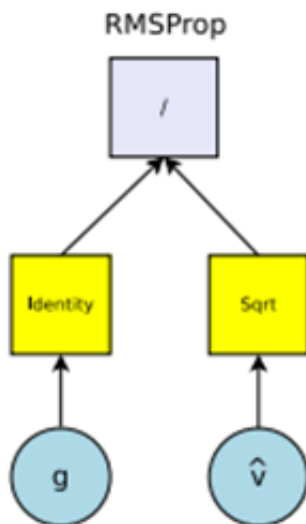
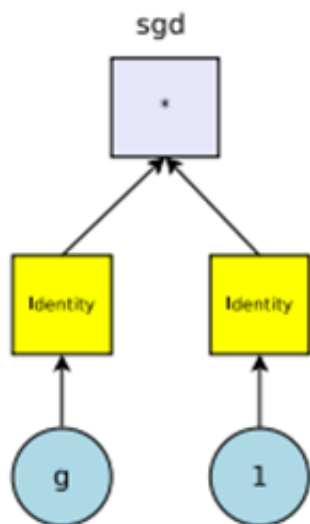
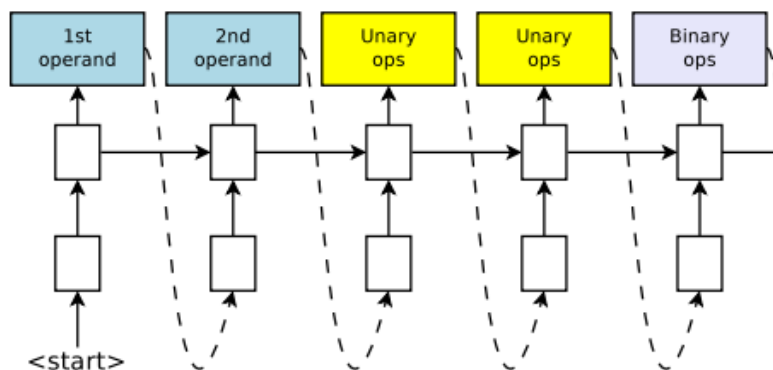
A Full Convolutional Neural Network (LeNet)

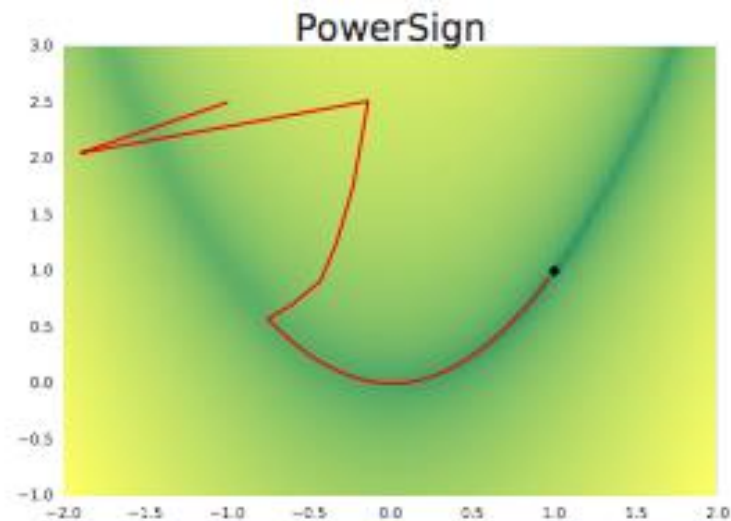
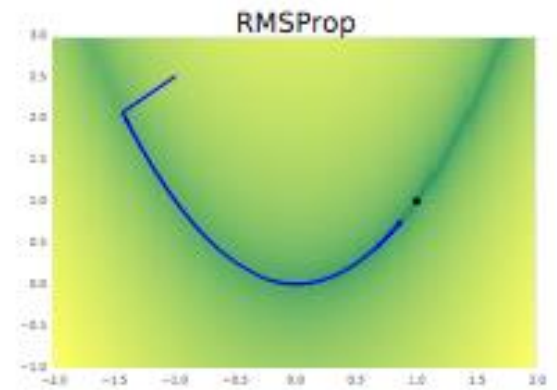
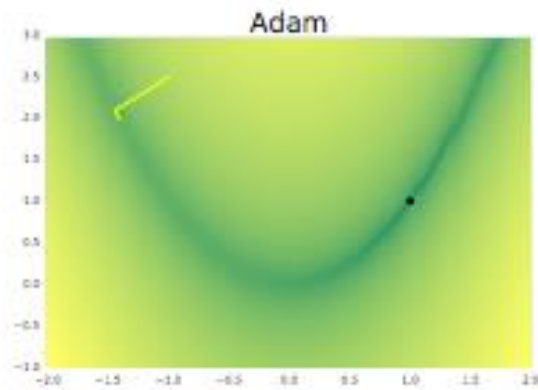
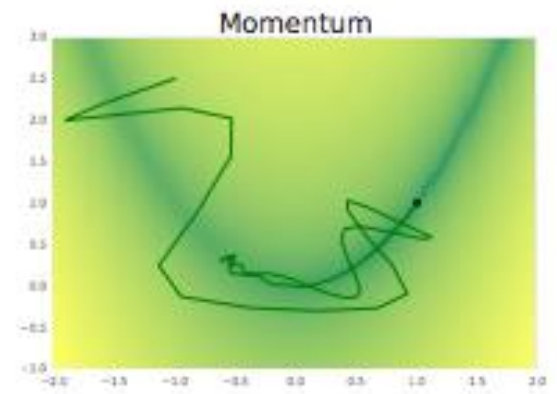
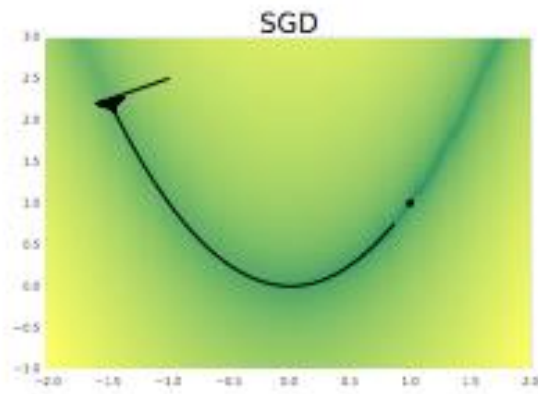
Learning Rate

- **Operands:** g , g^2 , g^3 , \hat{m} , \hat{v} , $\hat{\gamma}$, $\text{sign}(g)$, $\text{sign}(\hat{m})$, 1, 2, $\epsilon \sim N(0, 0.01)$, $10^{-4}w$, $10^{-3}w$, $10^{-2}w$, $10^{-1}w$, Adam and RMSProp.

- **Unary functions** which map input x to: x , $-x$, e^x , $\log|x|$, $\sqrt{|x|}$, $\text{clip}(x, 10^{-5})$, $\text{clip}(x, 10^{-4})$, $\text{clip}(x, 10^{-3})$, $\text{drop}(x, 0.1)$, $\text{drop}(x, 0.3)$, $\text{drop}(x, 0.5)$ and $\text{sign}(x)$.

- **Binary functions** which map (x, y) to $x + y$ (addition), $x - y$ (subtraction), $x * y$ (multiplication), $\frac{x}{y+\delta}$ (division), x^y (exponentiation) or x (keep left).

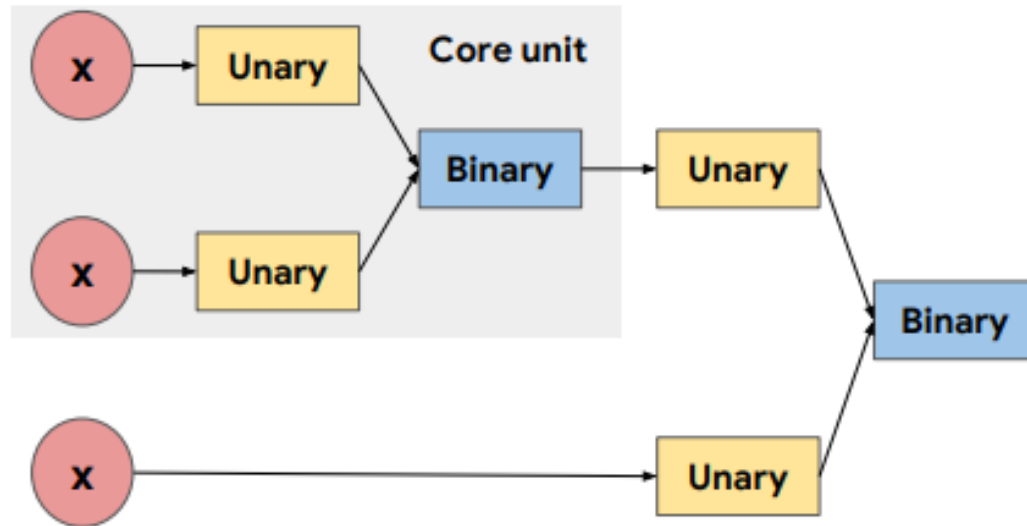




$$e^{\text{sign}(g) * \text{sign}(m)} * g$$

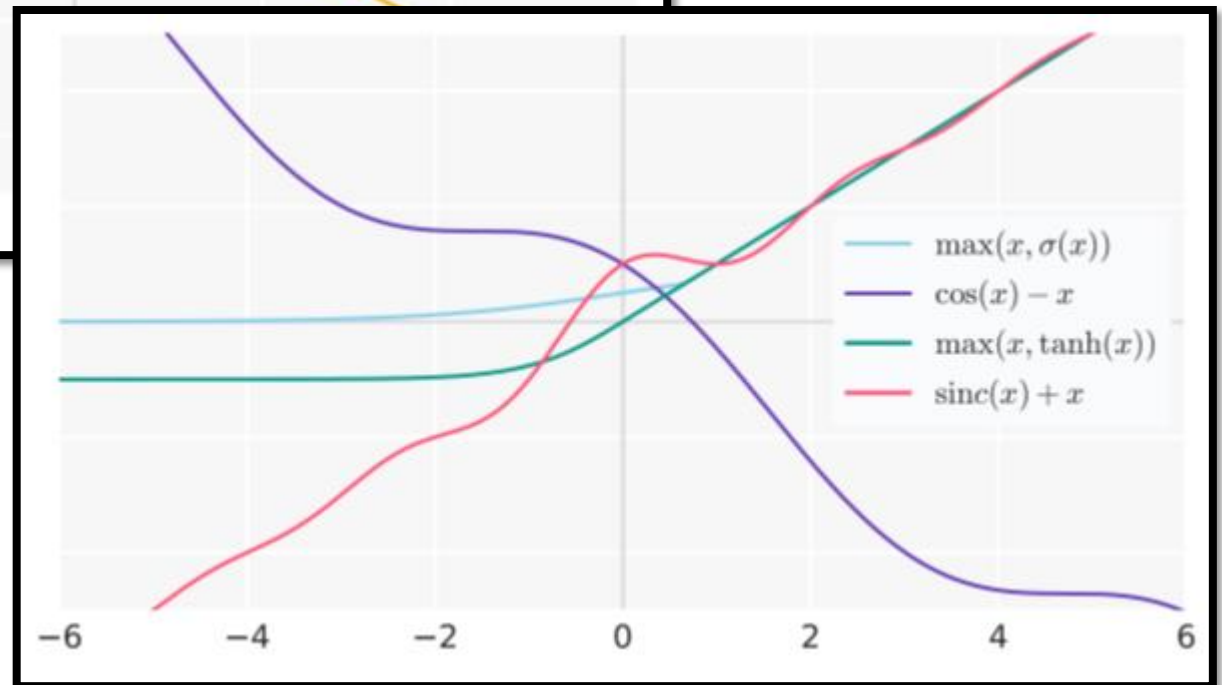
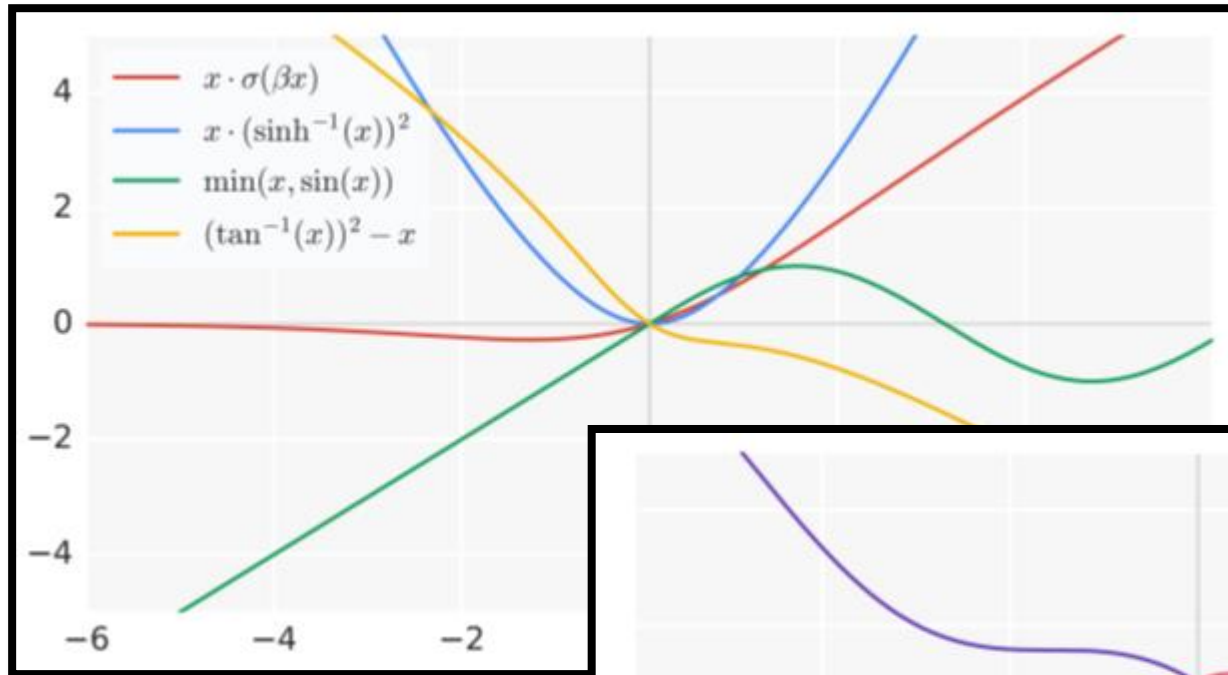
Can transfer to
new tasks

Activation Function

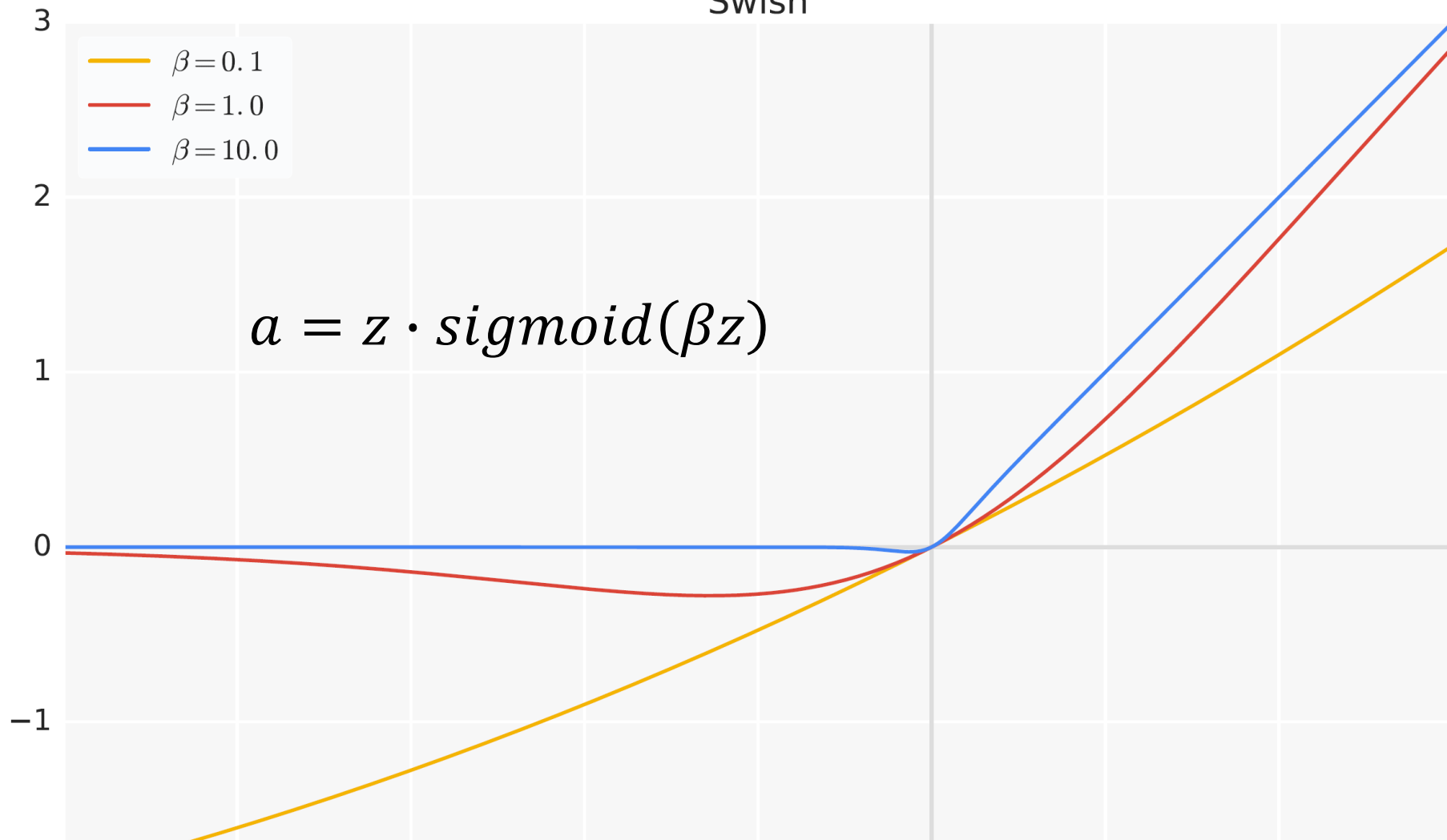


- **Unary functions:** x , $-x$, $|x|$, x^2 , x^3 , \sqrt{x} , βx , $x + \beta$, $\log(|x| + \epsilon)$, $\exp(x)$, $\sin(x)$, $\cos(x)$, $\sinh(x)$, $\cosh(x)$, $\tanh(x)$, $\sinh^{-1}(x)$, $\tan^{-1}(x)$, $\text{sinc}(x)$, $\max(x, 0)$, $\min(x, 0)$, $\sigma(x)$, $\log(1 + \exp(x))$, $\exp(-x^2)$, $\text{erf}(x)$, β
- **Binary functions:** $x_1 + x_2$, $x_1 \cdot x_2$, $x_1 - x_2$, $\frac{x_1}{x_2 + \epsilon}$, $\max(x_1, x_2)$, $\min(x_1, x_2)$, $\sigma(x_1) \cdot x_2$, $\exp(-\beta(x_1 - x_2)^2)$, $\exp(-\beta|x_1 - x_2|)$, $\beta x_1 + (1 - \beta)x_2$

Activation Function



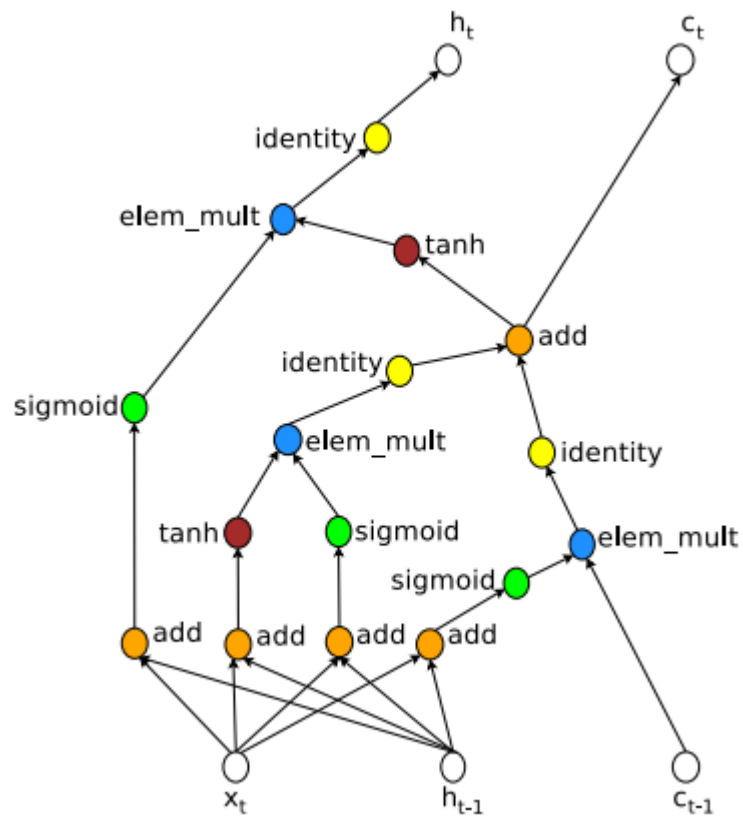
Swish



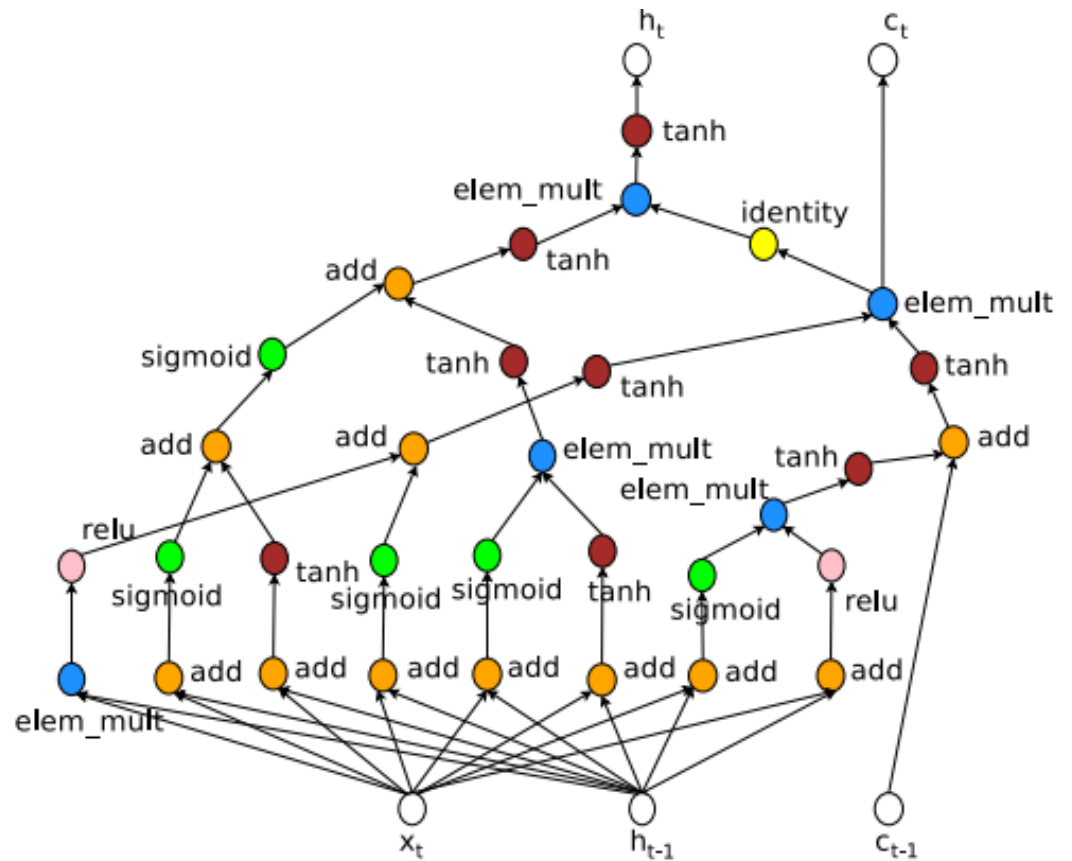
Baselines	ReLU	LReLU	PReLU	Softplus	ELU	SELU	GELU
Swish > Baseline	9	7	6	6	8	8	8
Swish = Baseline	0	1	3	2	0	1	1
Swish < Baseline	0	1	0	1	1	0	0

Neural Architecture Search with Reinforcement Learning

LSTM



From Reinforcement Learning



Computation Issue?

- Original version: 450 GPUs for 3-4 days (32,400-43,200 GPU hours).
- New version: Nvidia GTX 1080Ti GPU takes less than 16 hours.
- Main idea: forcing all child models to share weights to instead of training from scratch.

Hieu Pham, Melody Y. Guan, Barret Zoph, Quoc V. Le, Jeff Dean, "Efficient Neural Architecture Search via Parameter Sharing", arXiv, 2018