Deep Learning: State of the Art (2019)

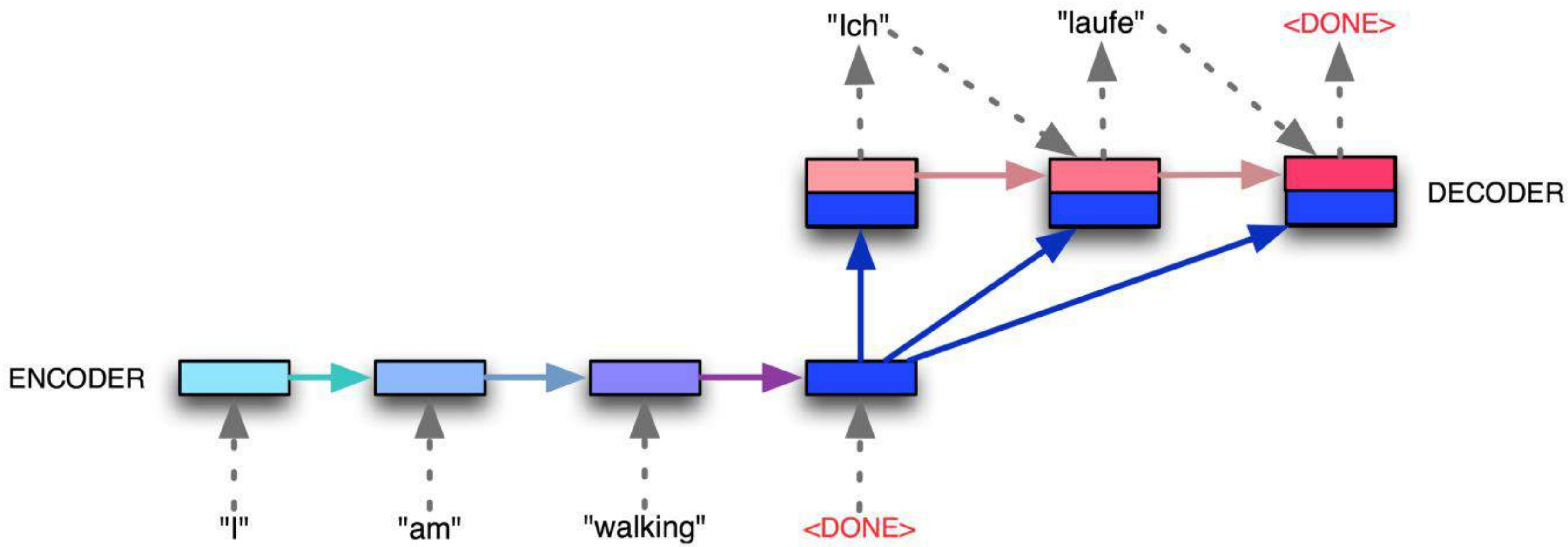
Deep Learning: State of the Art*
(Breakthrough Developments in 2017 & 2018)

- **BERT and Natural Language Processing**
- Tesla Autopilot Hardware v2+: Neural Networks at Scale
- AdaNet: AutoML with Ensembles
- AutoAugment: Deep RL Data Augmentation
- Training Deep Networks with Synthetic Data
- Segmentation Annotation with Polygon-RNN++
- DAWNBench: Training Fast and Cheap
- BigGAN: State of the Art in Image Synthesis
- Video-to-Video Synthesis
- Semantic Segmentation
- AlphaZero & OpenAI Five
- Deep Learning Frameworks

* This is not a list of state-of-the-art results on main machine learning benchmark datasets. It's an overview of exciting recent developments.

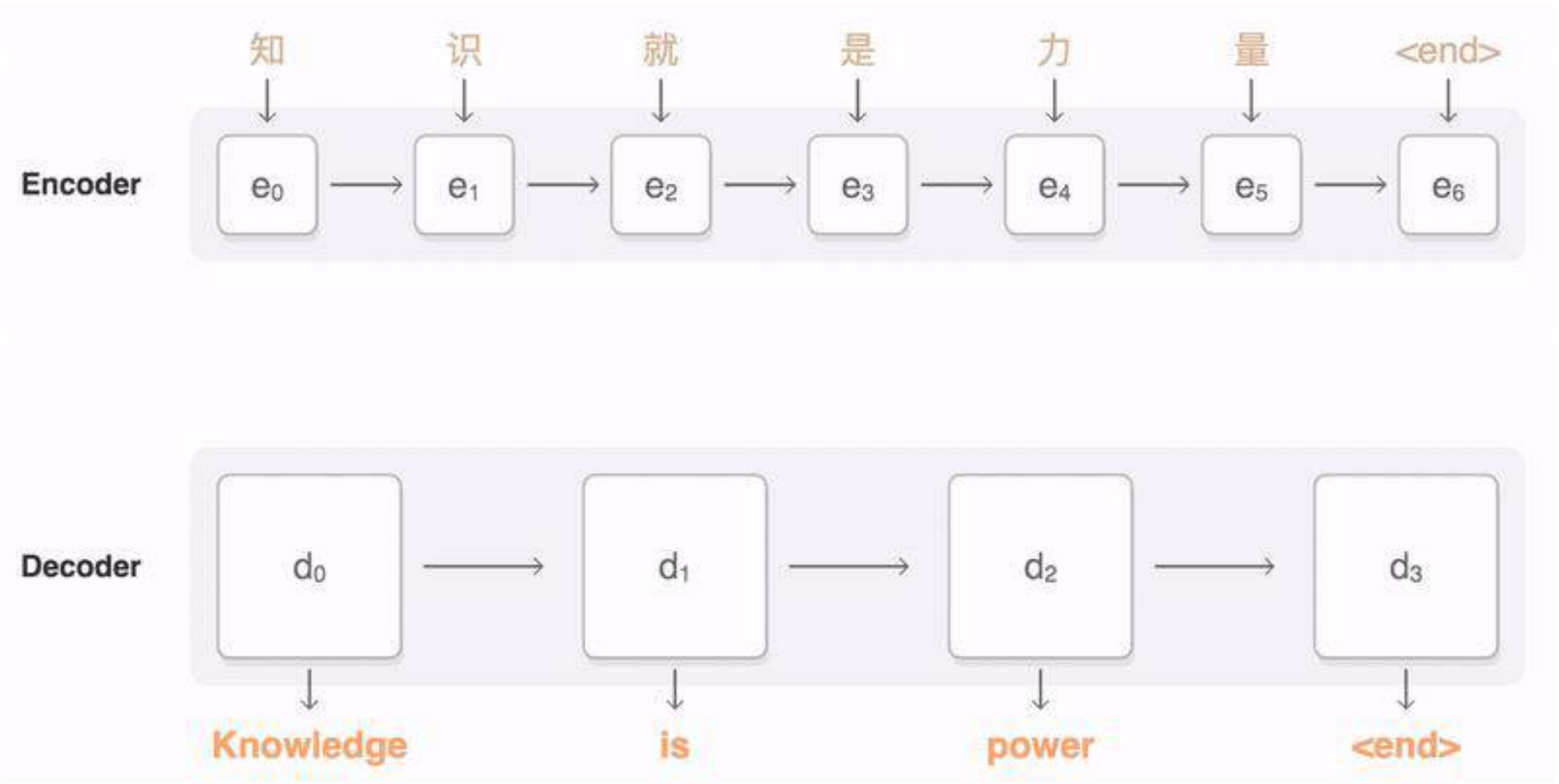
Encoder-Decoder Architecture

Sequence-to-Sequence Model - Neural Machine Translation



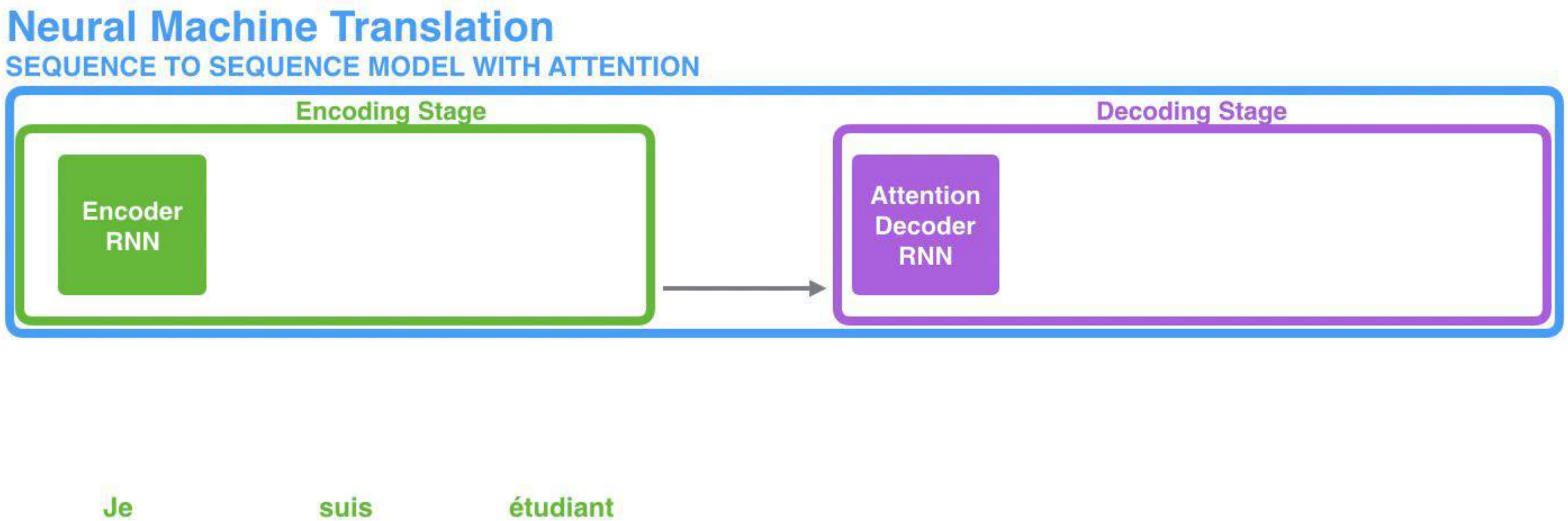
Encoder RNN encodes input sequence into a fixed size vector, and then is passed repeatedly to decoder RNN.

Attention

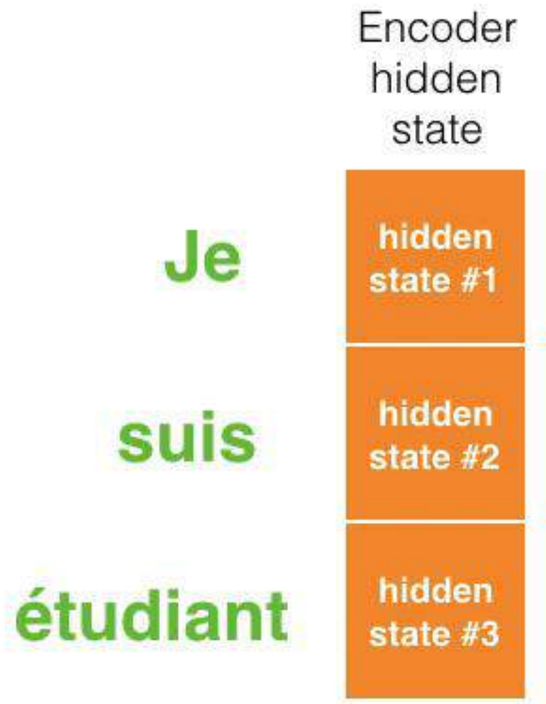


Attention mechanism allows the network to refer back to the input sequence, instead of forcing it to encode all information into one fixed-length vector.

Attention

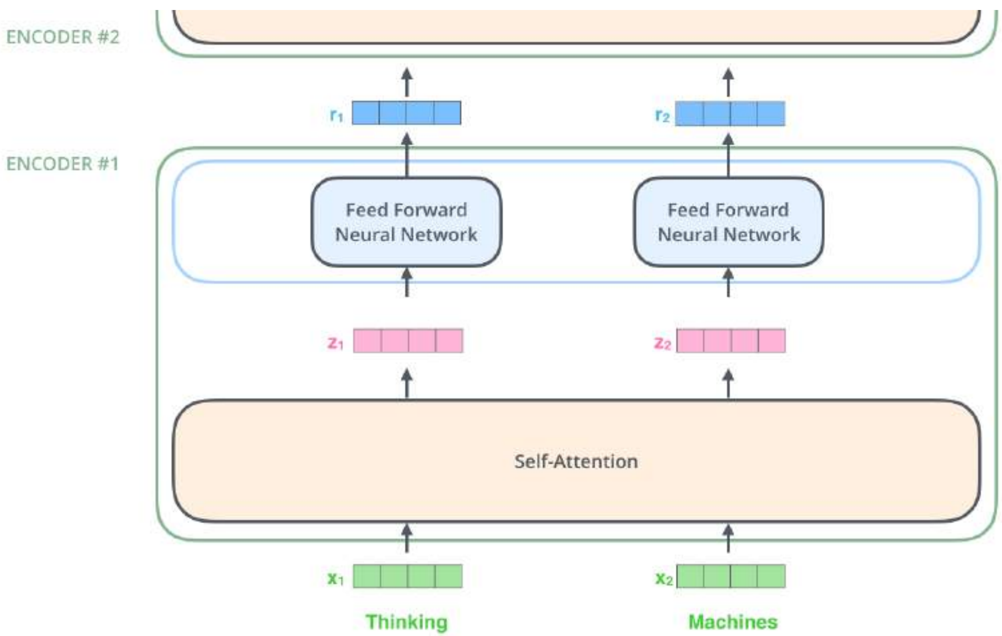

Neural Machine Translation
SEQUENCE TO SEQUENCE MODEL WITH ATTENTION
Encoding Stage
Decoding Stage
Encoder RNN
Attention Decoder RNN
Je
suis
étudiant

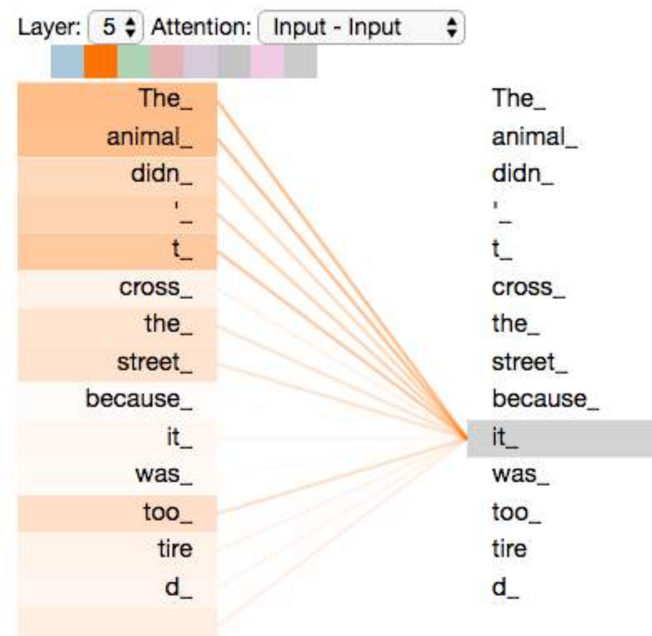
Attention


Encoder hidden state
Je
suis
étudiant
hidden state #1
hidden state #2
hidden state #3

Self-Attention

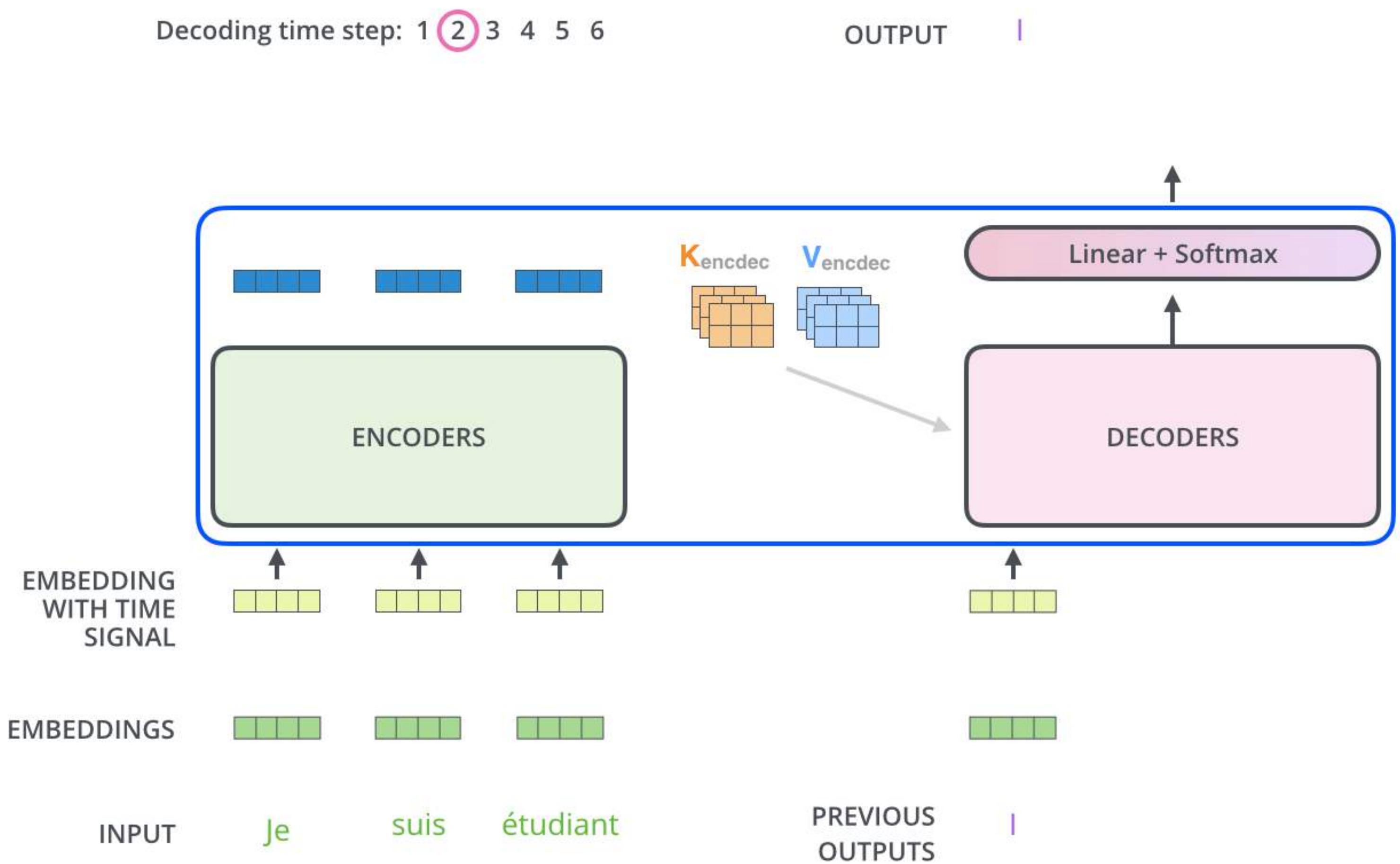
"The animal didn't cross the street because it was too tired"





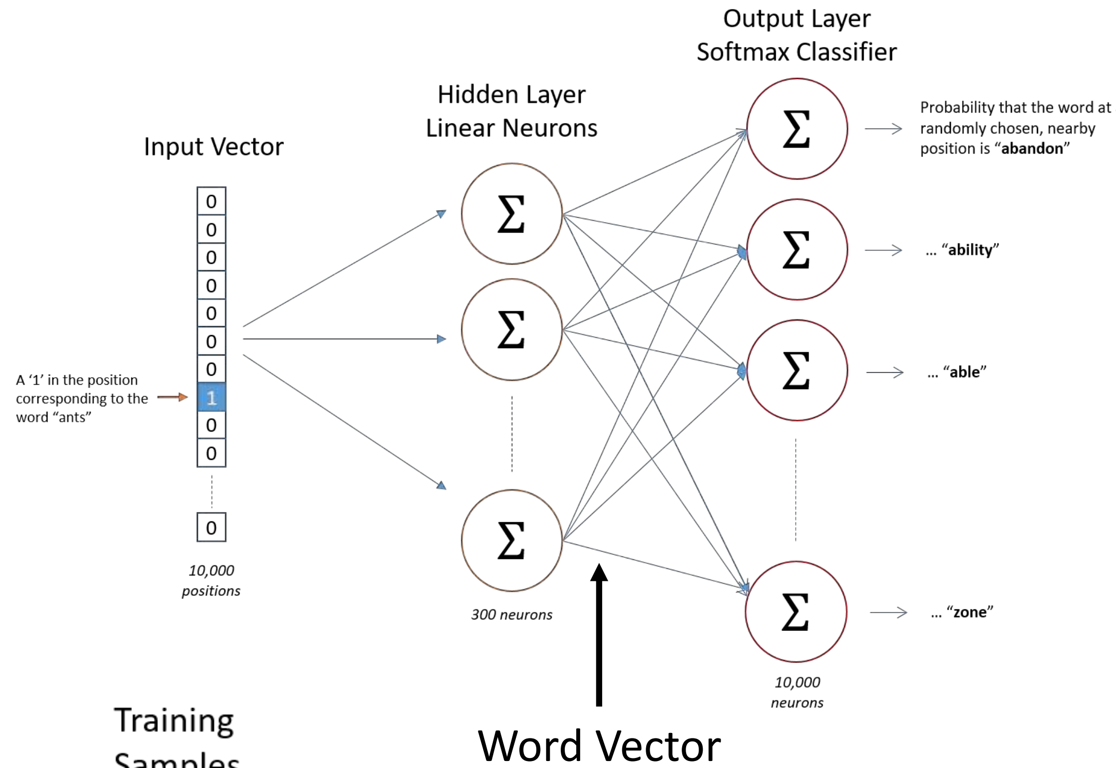
More details: http://jalammar.github.io/illustrated-transformer/

Transformer

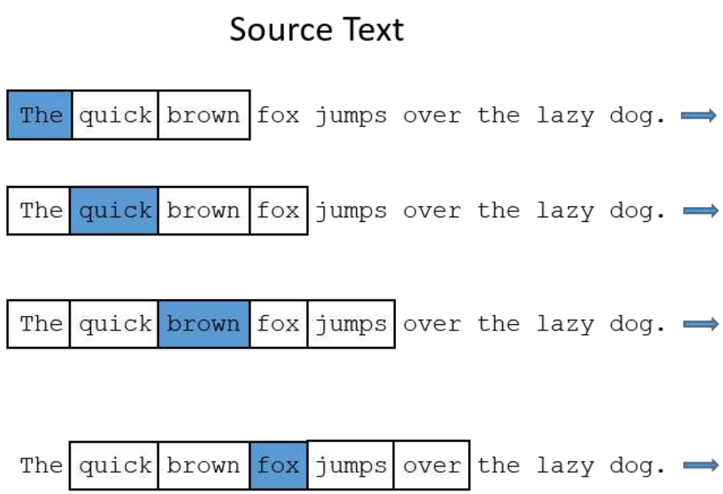


Vaswani, Ashish, et al. "Attention is all you need." *Advances in Neural Information Processing Systems*. 2017.

Word Embeddings (Word2Vec)



Skip Gram Model:



Training Samples

(the, quick)
(the, brown)

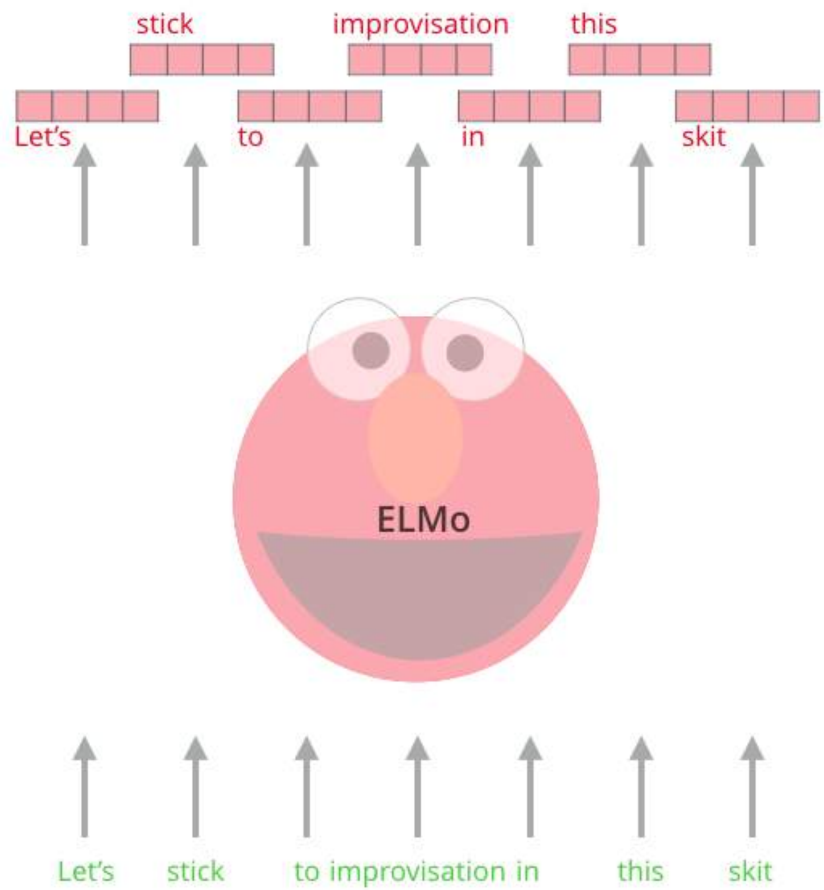
(quick, the)
(quick, brown)
(quick, fox)

(brown, the)
(brown, quick)
(brown, fox)
(brown, jumps)

(fox, quick)
(fox, brown)
(fox, jumps)
(fox, over)

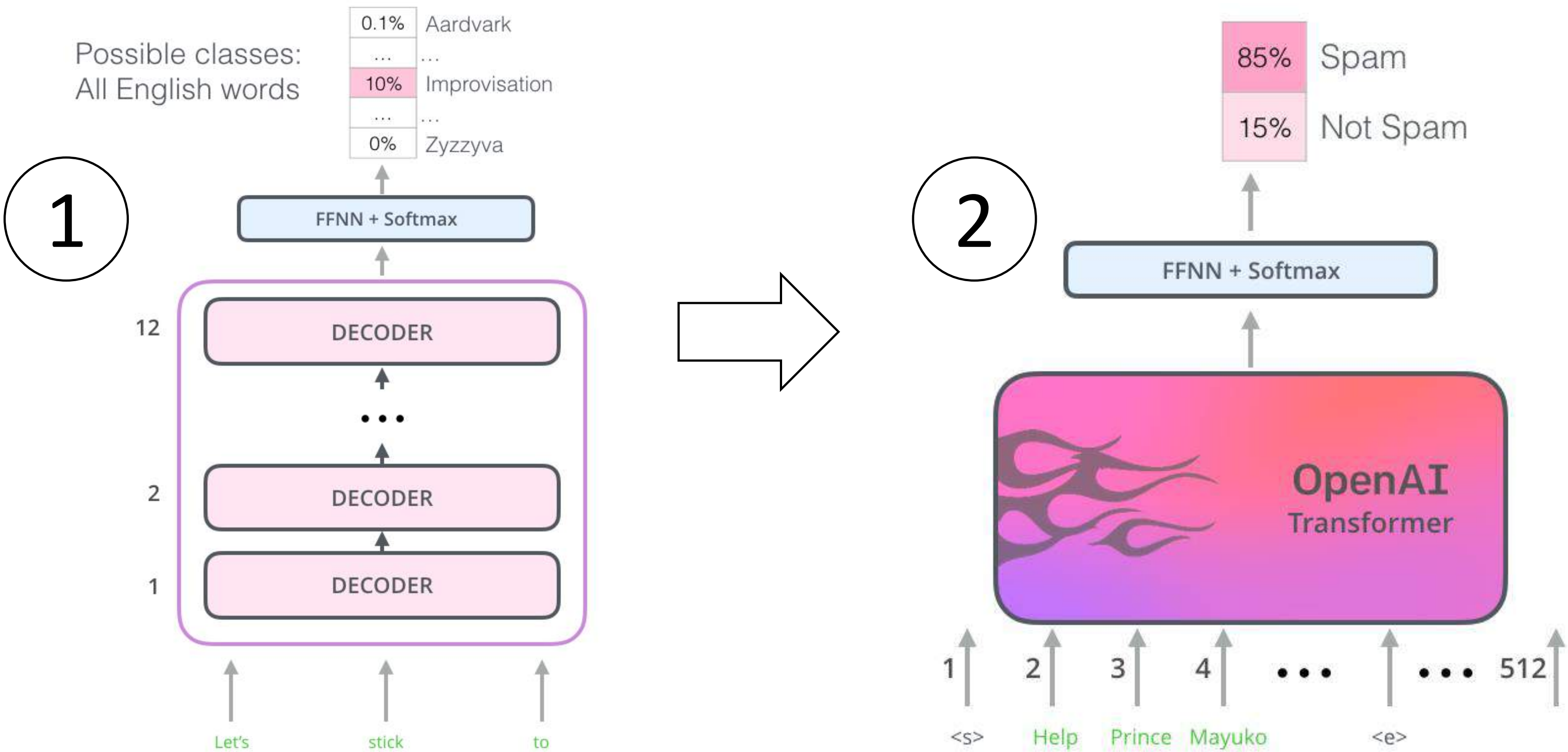
Context-Aware Embeddings

ELMo
Embeddings



Words to embed

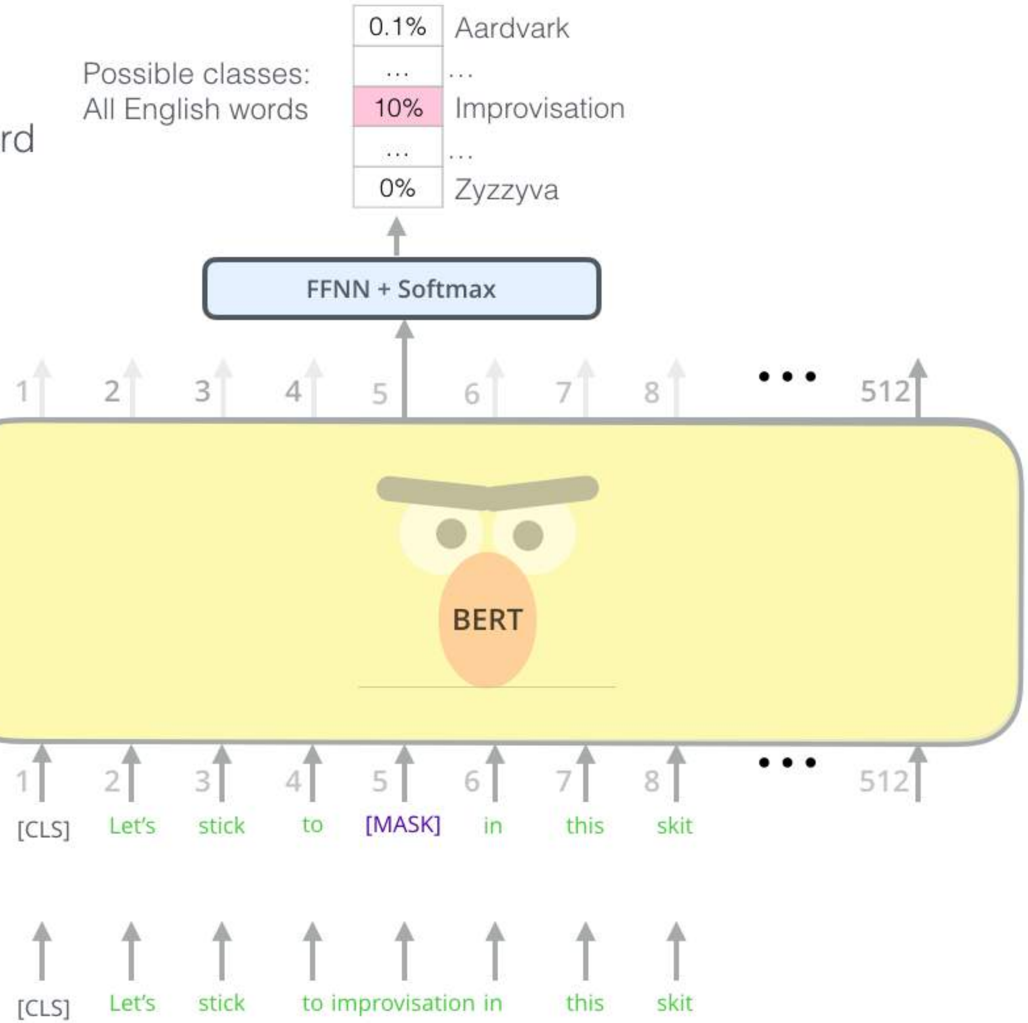
OpenAI Transformer



1. Pre-train a Transformer's decoder for language modeling
2. Train it on, for example, a sentence classification task

BERT

Use the output of the masked word's position to predict the masked word

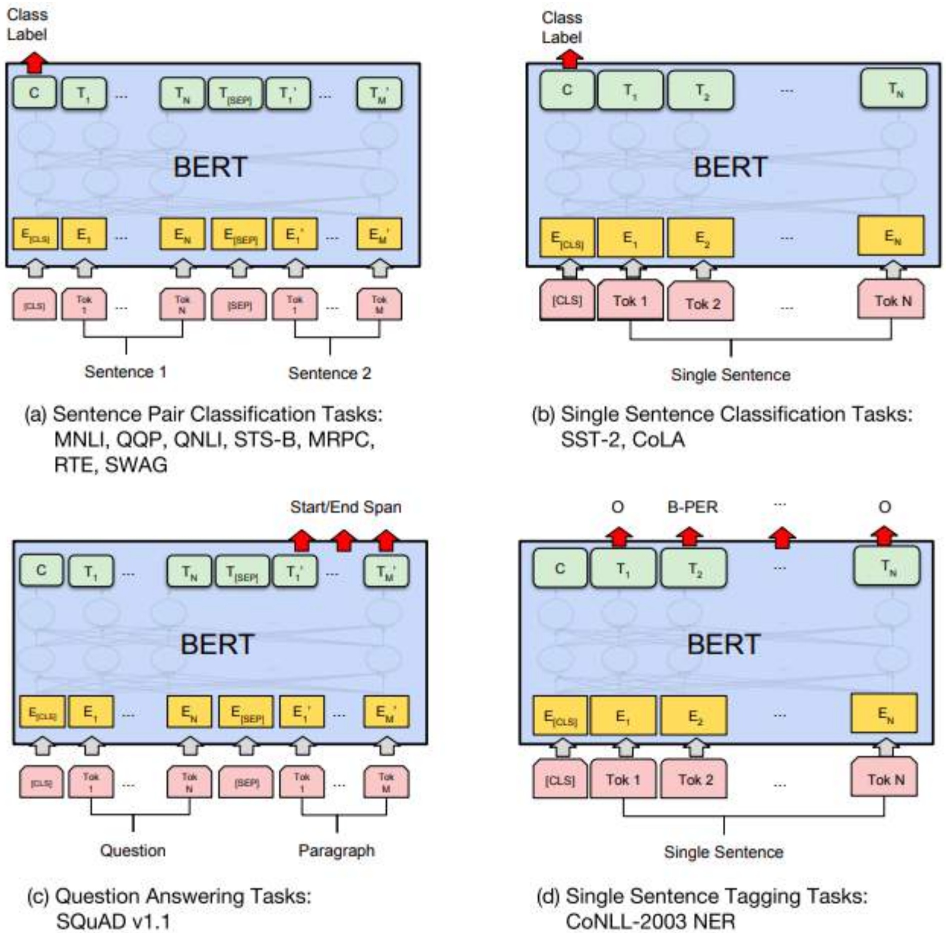


Randomly mask 15% of tokens

Input

Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." (2018).

BERT Applications



Now you can use BERT:

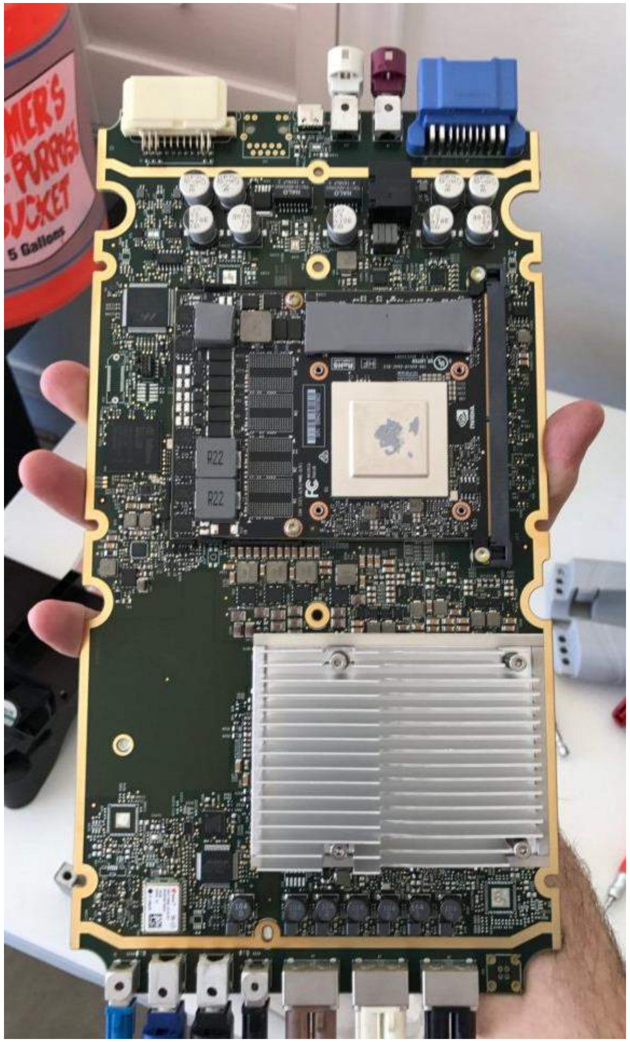
- Create contextualized word embeddings (like ELMo)
- Sentence classification
- Sentence pair classification
- Sentence pair similarity
- Multiple choice
- Sentence tagging
- Question answering

Deep Learning: State of the Art*
(Breakthrough Developments in 2017 & 2018)

- BERT and Natural Language Processing
- **Tesla Autopilot Hardware v2+: Neural Networks at Scale**
- AdaNet: AutoML with Ensembles
- AutoAugment: Deep RL Data Augmentation
- Training Deep Networks with Synthetic Data
- Segmentation Annotation with Polygon-RNN++
- DAWNBench: Training Fast and Cheap
- BigGAN: State of the Art in Image Synthesis
- Video-to-Video Synthesis
- Semantic Segmentation
- AlphaZero & OpenAI Five
- Deep Learning Frameworks

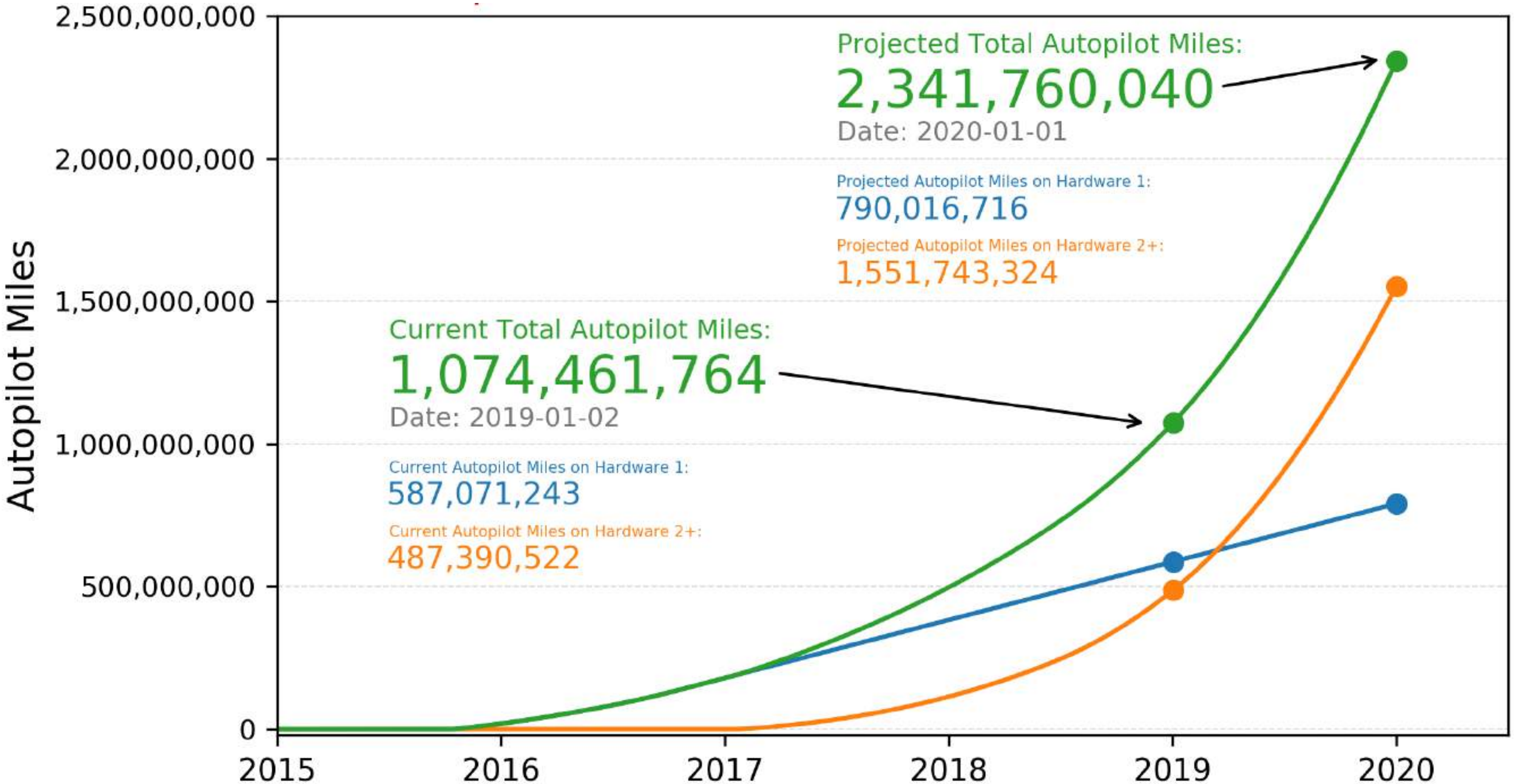
* This is not a list of state-of-the-art results on main machine learning benchmark datasets. It's an overview of exciting recent developments.

Tesla Autopilot Hardware v2+



- Specialized NVIDIA Drive PX 2 hardware
- Neural network takes all 8 cameras as input
- Based on Inception v1 architecture

Autopilot Reaches 1 Billion Miles
(~0.5 Billion on Hardware v2+)

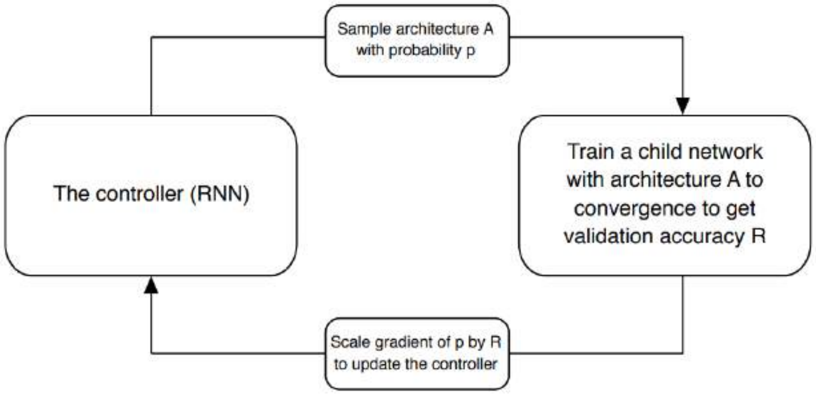


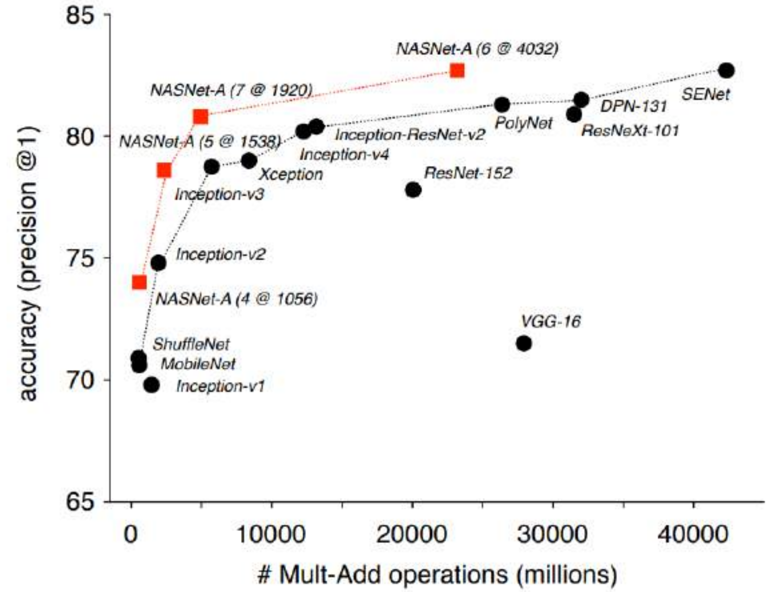
Deep Learning: State of the Art*
(Breakthrough Developments in 2017 & 2018)

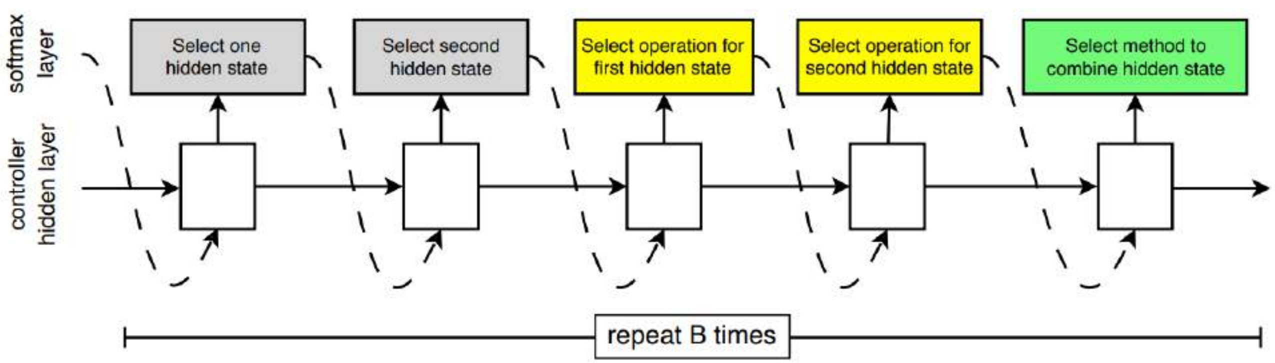
- BERT and Natural Language Processing
- Tesla Autopilot Hardware v2+: Neural Networks at Scale
- **AdaNet: AutoML with Ensembles**
- AutoAugment: Deep RL Data Augmentation
- Training Deep Networks with Synthetic Data
- Segmentation Annotation with Polygon-RNN++
- DAWNBench: Training Fast and Cheap
- BigGAN: State of the Art in Image Synthesis
- Video-to-Video Synthesis
- Semantic Segmentation
- AlphaZero & OpenAI Five
- Deep Learning Frameworks

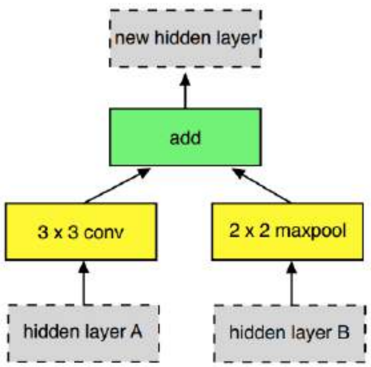
* This is not a list of state-of-the-art results on main machine learning benchmark datasets. It's an overview of exciting recent developments.

AutoML and Neural Architecture Search (NASNet)









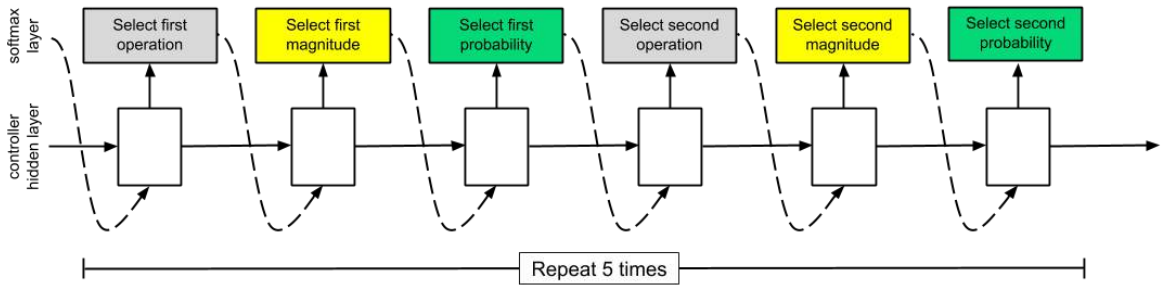
AdaNet

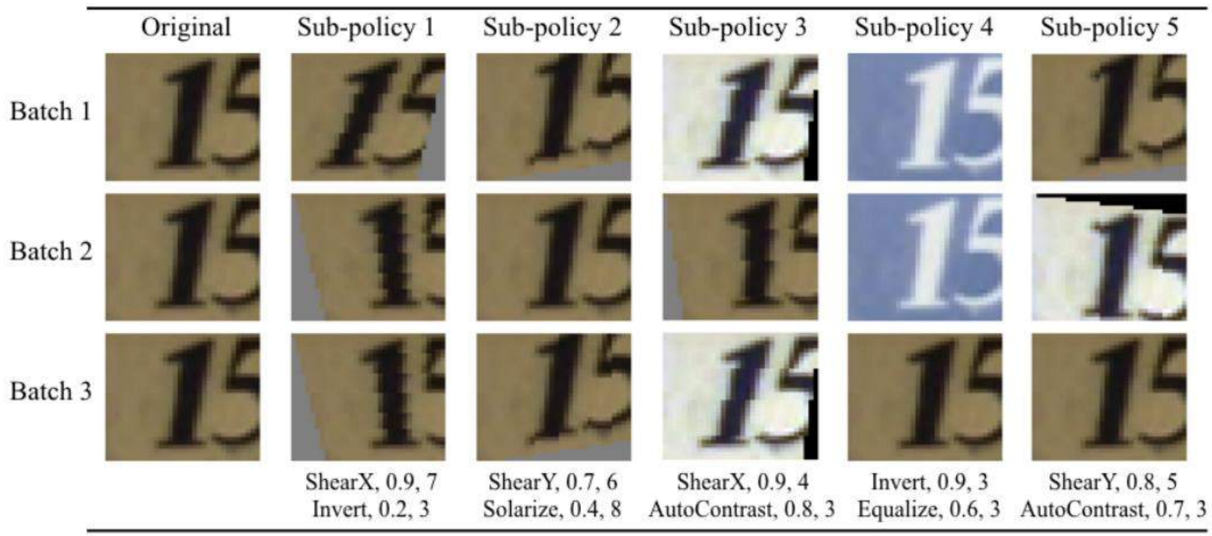
Deep Learning: State of the Art*
(Breakthrough Developments in 2017 & 2018)

- BERT and Natural Language Processing
- Tesla Autopilot Hardware v2+: Neural Networks at Scale
- AdaNet: AutoML with Ensembles
- **AutoAugment: Deep RL Data Augmentation**
- Training Deep Networks with Synthetic Data
- Segmentation Annotation with Polygon-RNN++
- DAWNBench: Training Fast and Cheap
- BigGAN: State of the Art in Image Synthesis
- Video-to-Video Synthesis
- Semantic Segmentation
- AlphaZero & OpenAI Five
- Deep Learning Frameworks

* This is not a list of state-of-the-art results on main machine learning benchmark datasets. It's an overview of exciting recent developments.

AutoAugment: RL for Data Augmentation





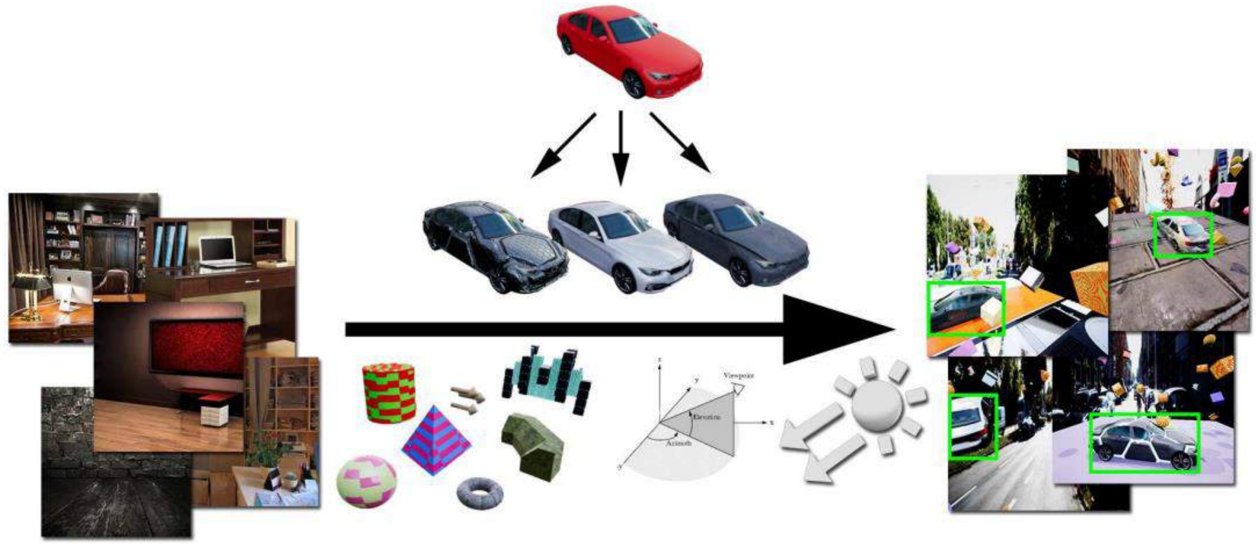
- Show that transfer learning can **also** be done augmentation policies instead of weights (or with).

Deep Learning: State of the Art*
(Breakthrough Developments in 2017 & 2018)

- BERT and Natural Language Processing
- Tesla Autopilot Hardware v2+: Neural Networks at Scale
- AdaNet: AutoML with Ensembles
- AutoAugment: Deep RL Data Augmentation
- **Training Deep Networks with Synthetic Data**
- Segmentation Annotation with Polygon-RNN++
- DAWNBench: Training Fast and Cheap
- BigGAN: State of the Art in Image Synthesis
- Video-to-Video Synthesis
- Semantic Segmentation
- AlphaZero & OpenAI Five
- Deep Learning Frameworks

* This is not a list of state-of-the-art results on main machine learning benchmark datasets. It's an overview of exciting recent developments.

Training on Randomized Synthetic Data



- number and types of objects
- number, types, colors, and scales of distractors
- texture on the object of interest, and background photograph
- location of the virtual camera with respect to the scene
- angle of the camera with respect to the scene
- number and locations of point lights

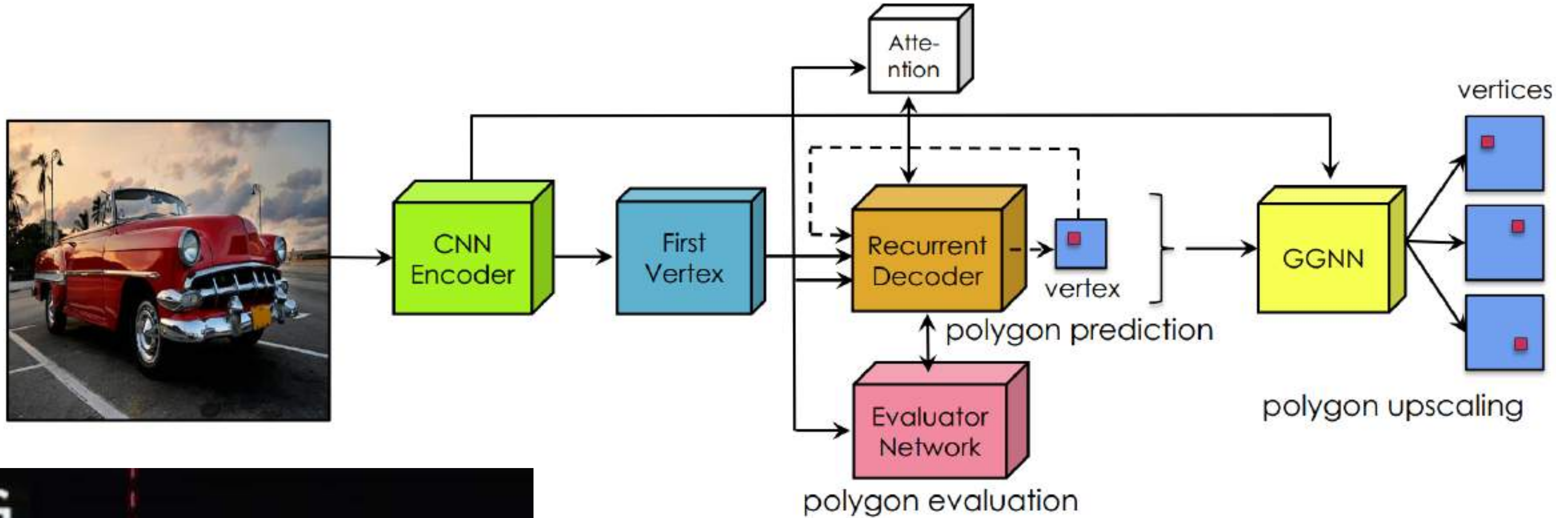
Tremblay, Jonathan, et al. "Training deep networks with synthetic data: Bridging the reality gap by domain randomization." (2018).

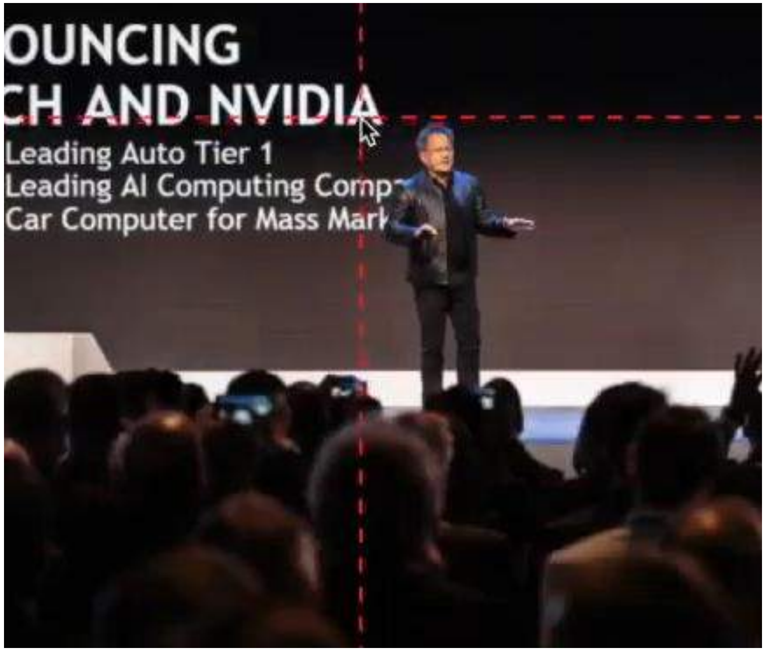
Deep Learning: State of the Art*
(Breakthrough Developments in 2017 & 2018)

- BERT and Natural Language Processing
- Tesla Autopilot Hardware v2+: Neural Networks at Scale
- AdaNet: AutoML with Ensembles
- AutoAugment: Deep RL Data Augmentation
- Training Deep Networks with Synthetic Data
- **Segmentation Annotation with Polygon-RNN++**
- DAWNBench: Training Fast and Cheap
- BigGAN: State of the Art in Image Synthesis
- Video-to-Video Synthesis
- Semantic Segmentation
- AlphaZero & OpenAI Five
- Deep Learning Frameworks

* This is not a list of state-of-the-art results on main machine learning benchmark datasets. It's an overview of exciting recent developments.

Segmentation Annotation with Polygon-RNN++





- High-resolution polygon segmentation generation with human-in-the-loop

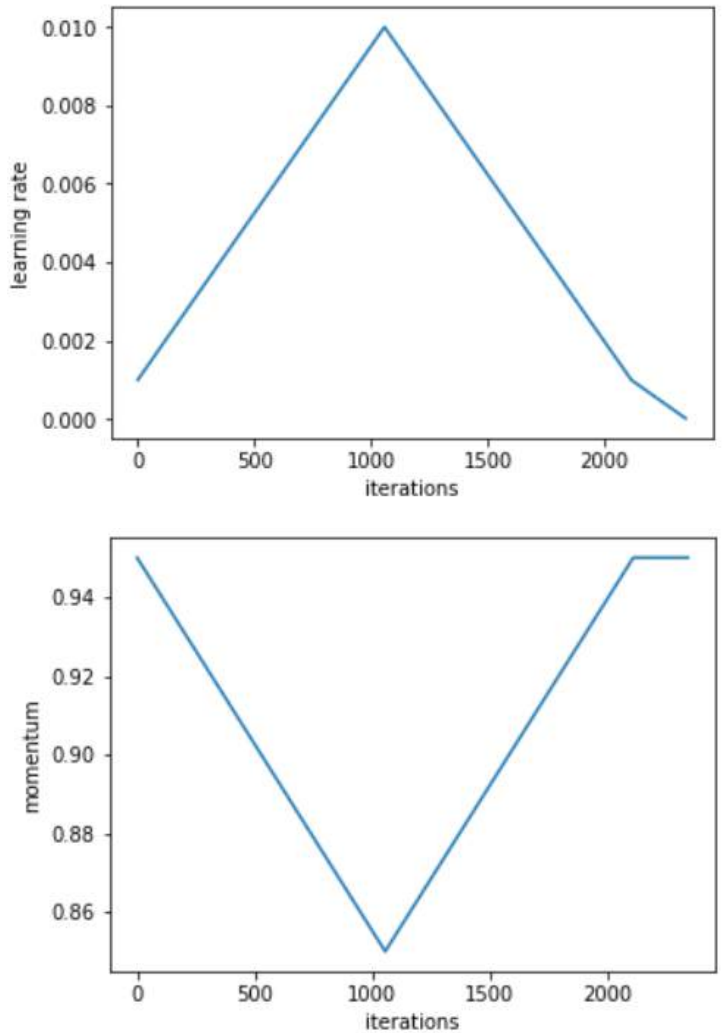
Acuna, David, et al. "Efficient Interactive Annotation of Segmentation Datasets With Polygon-RNN++." CVPR 2018.

Deep Learning: State of the Art*
(Breakthrough Developments in 2017 & 2018)

- BERT and Natural Language Processing
- Tesla Autopilot Hardware v2+: Neural Networks at Scale
- AdaNet: AutoML with Ensembles
- AutoAugment: Deep RL Data Augmentation
- Training Deep Networks with Synthetic Data
- Segmentation Annotation with Polygon-RNN++
- **DAWNBench: Training Fast and Cheap**
- BigGAN: State of the Art in Image Synthesis
- Video-to-Video Synthesis
- Semantic Segmentation
- AlphaZero & OpenAI Five
- Deep Learning Frameworks

* This is not a list of state-of-the-art results on main machine learning benchmark datasets. It's an overview of exciting recent developments.

DAWNBench: Training Fast and Cheap



DAWNBench:

- Competition on speed and cost of training and inference that achieves 93% for ImageNet and 94% for CIFAR 10

fast.ai – Training:

- ImageNet in 3 hours for $25
- CIFAR10 for $0.26
- **Key idea:** During training, if you very slowly increase **learning rate** while decreasing **momentum**, you can train at extremely high learning rates, thus avoiding over-fitting, and training in far fewer epochs.
- Details: http://bit.ly/2H6yv6H

Deep Learning: State of the Art*
(Breakthrough Developments in 2017 & 2018)

- BERT and Natural Language Processing
- Tesla Autopilot Hardware v2+: Neural Networks at Scale
- AdaNet: AutoML with Ensembles
- AutoAugment: Deep RL Data Augmentation
- Training Deep Networks with Synthetic Data
- Segmentation Annotation with Polygon-RNN++
- DAWNBench: Training Fast and Cheap
- **BigGAN: State of the Art in Image Synthesis**
- Video-to-Video Synthesis
- Semantic Segmentation
- AlphaZero & OpenAI Five
- Deep Learning Frameworks

* This is not a list of state-of-the-art results on main machine learning benchmark datasets. It's an overview of exciting recent developments.

BigGAN: State of the Art in Image Synthesis

- Same GAN techniques, much larger scale
- Increase model capacity + increase batch size

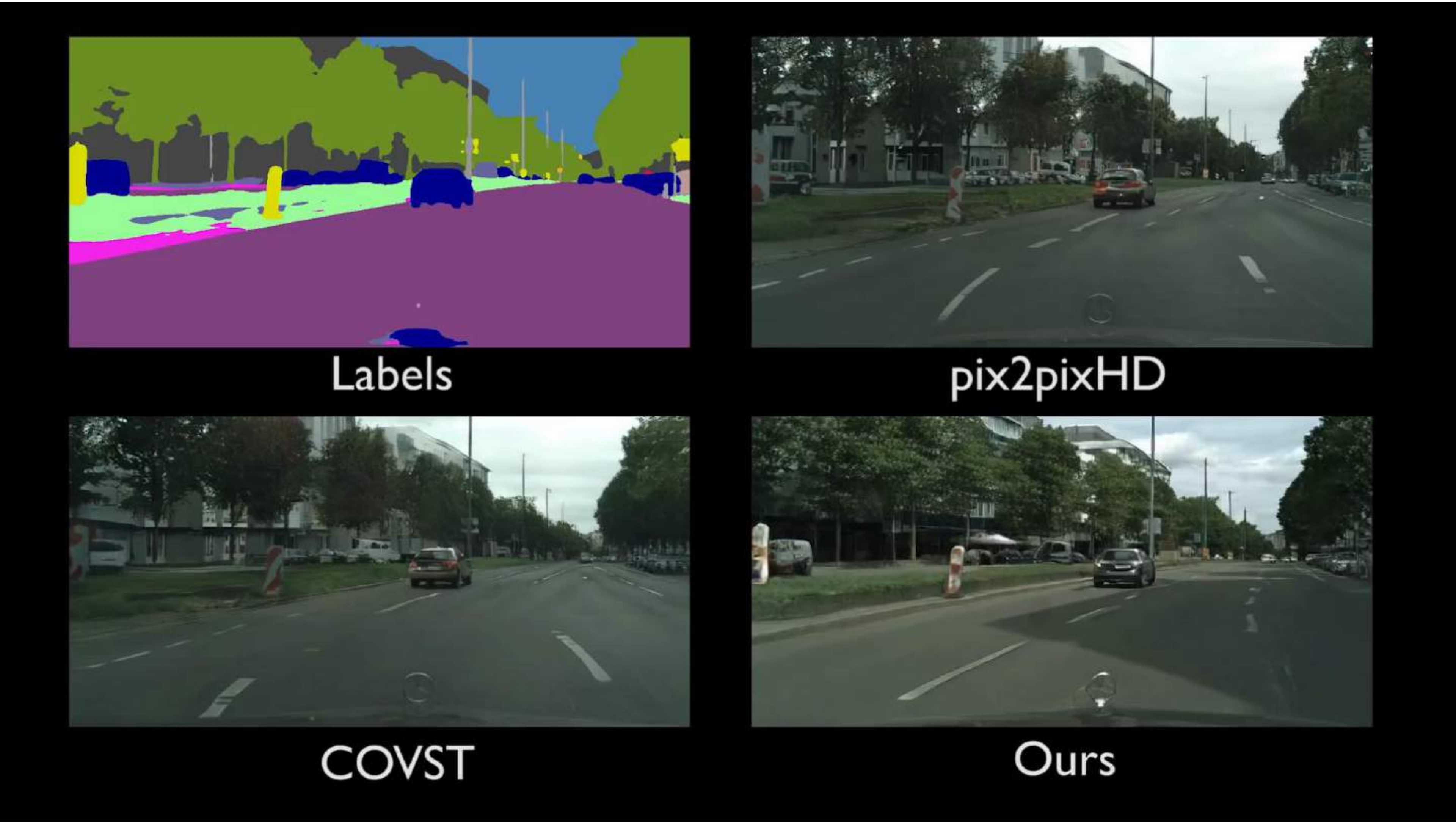
Brock, Andrew, Jeff Donahue, and Karen Simonyan. "Large scale gan training for high fidelity natural image synthesis." (2018).

Deep Learning: State of the Art*
(Breakthrough Developments in 2017 & 2018)

- BERT and Natural Language Processing
- Tesla Autopilot Hardware v2+: Neural Networks at Scale
- AdaNet: AutoML with Ensembles
- AutoAugment: Deep RL Data Augmentation
- Training Deep Networks with Synthetic Data
- Segmentation Annotation with Polygon-RNN++
- DAWNBench: Training Fast and Cheap
- BigGAN: State of the Art in Image Synthesis
- **Video-to-Video Synthesis**
- Semantic Segmentation
- AlphaZero & OpenAI Five
- Deep Learning Frameworks

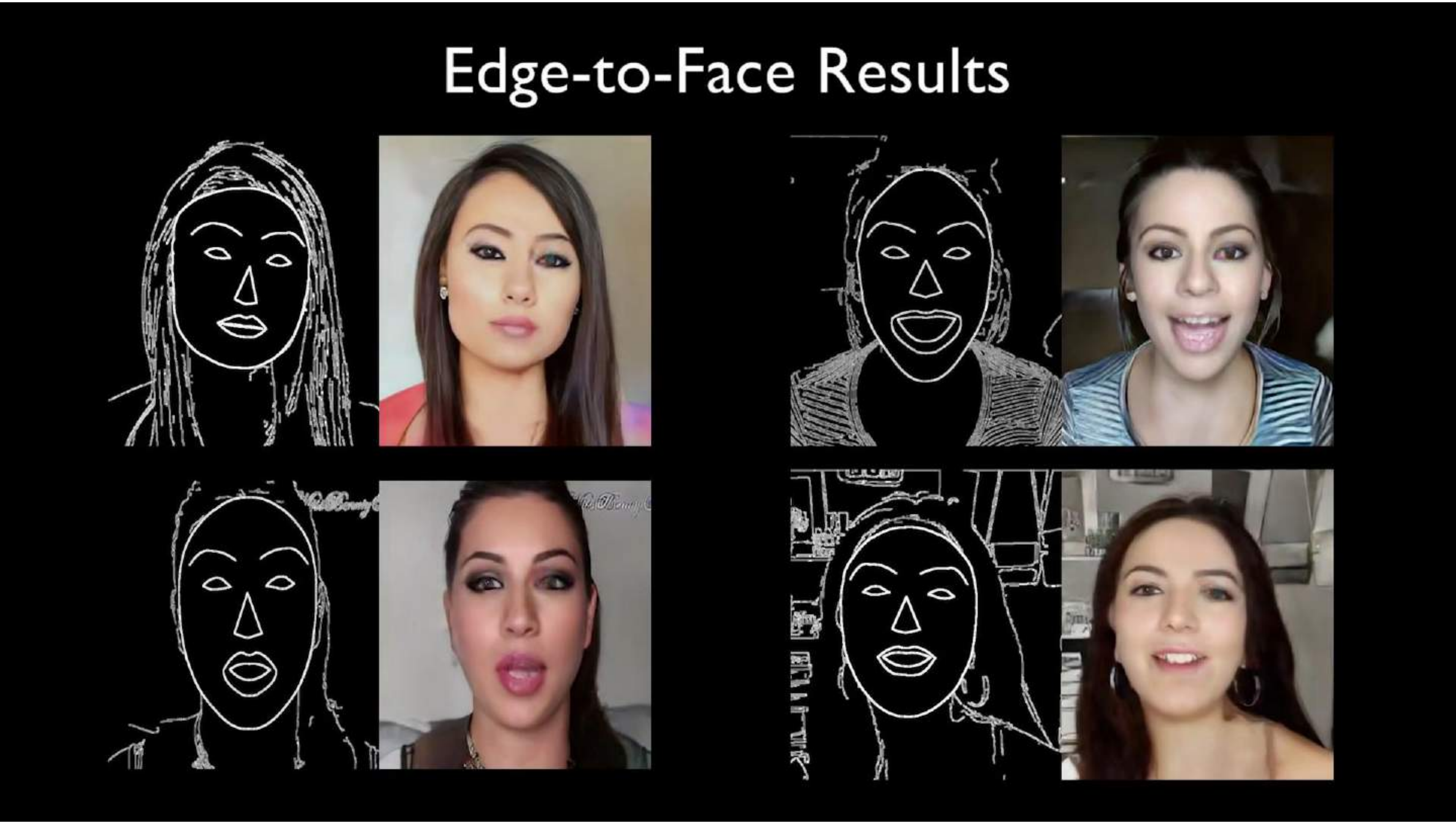
* This is not a list of state-of-the-art results on main machine learning benchmark datasets. It's an overview of exciting recent developments.

Video-to-Video Synthesis



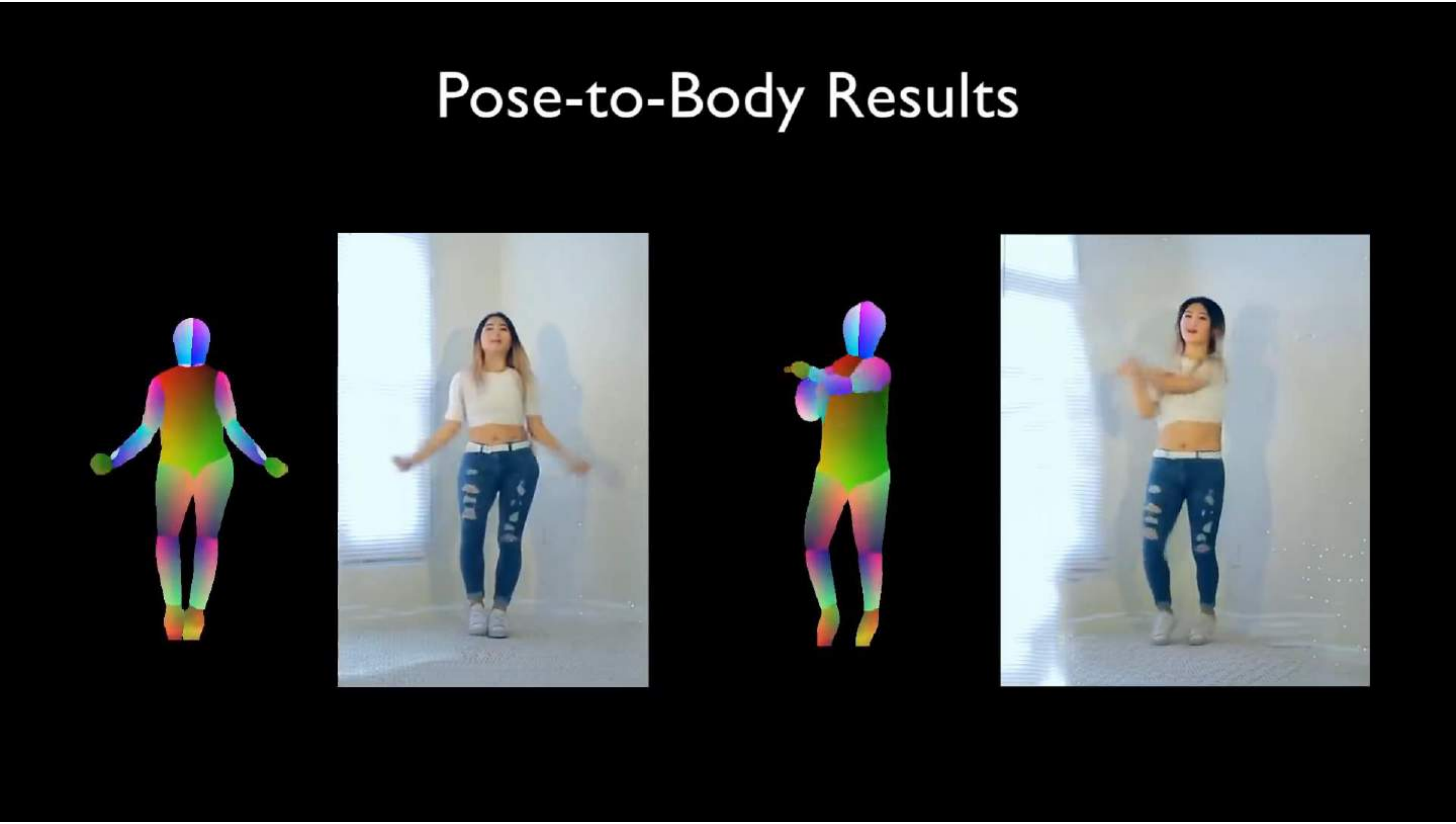
Wang, Ting-Chun, et al. "Video-to-video synthesis." (2018).

Video-to-Video Synthesis



Wang, Ting-Chun, et al. "Video-to-video synthesis." (2018).

Video-to-Video Synthesis



Wang, Ting-Chun, et al. "Video-to-video synthesis." (2018).

Deep Learning: State of the Art*
(Breakthrough Developments in 2017 & 2018)

- BERT and Natural Language Processing
- Tesla Autopilot Hardware v2+: Neural Networks at Scale
- AdaNet: AutoML with Ensembles
- AutoAugment: Deep RL Data Augmentation
- Training Deep Networks with Synthetic Data
- Segmentation Annotation with Polygon-RNN++
- DAWNBench: Training Fast and Cheap
- BigGAN: State of the Art in Image Synthesis
- Video-to-Video Synthesis
- **Semantic Segmentation**
- AlphaZero & OpenAI Five
- Deep Learning Frameworks

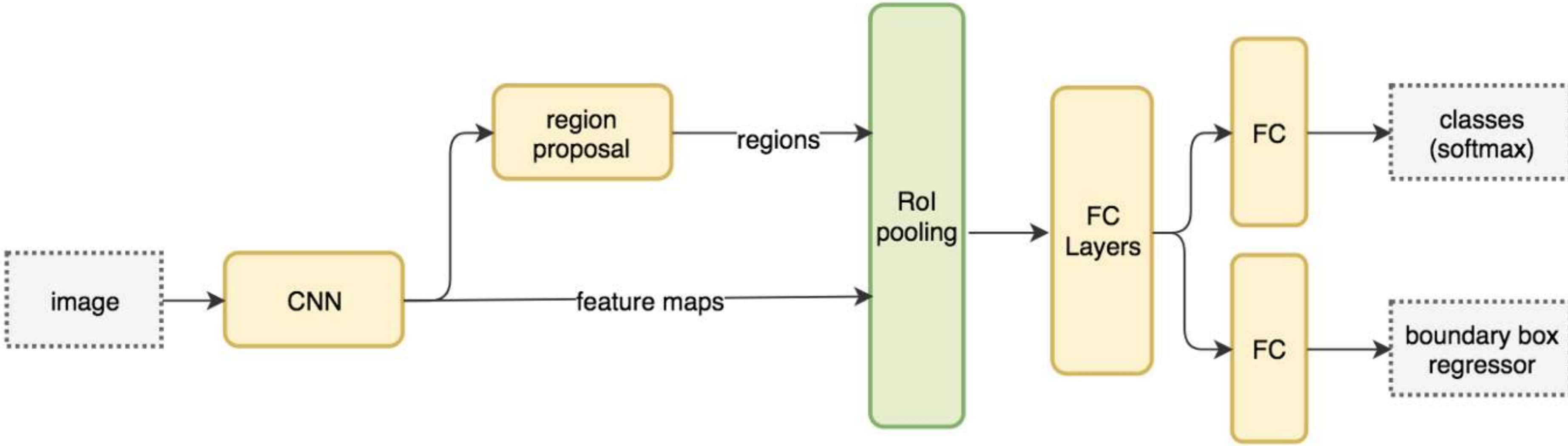
* This is not a list of state-of-the-art results on main machine learning benchmark datasets. It's an overview of exciting recent developments.

Human error (5.1%) surpassed in 2015

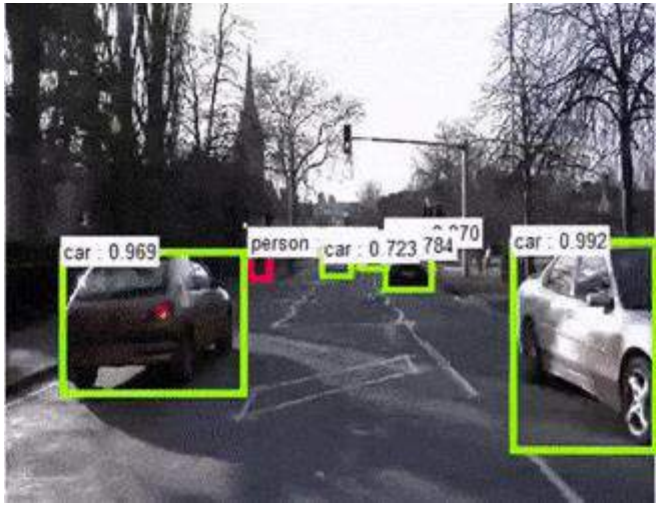
- **AlexNet (2012): First CNN (15.4%)**
  - 8 layers
  - 61 million parameters
- **ZFNet (2013): 15.4% to 11.2%**
  - 8 layers
  - More filters. Denser stride.
- **VGGNet (2014): 11.2% to 7.3%**
  - Beautifully uniform:
    3x3 conv, stride 1, pad 1, 2x2 max pool
  - 16 layers
  - 138 million parameters
- **GoogLeNet (2014): 11.2% to 6.7%**
  - Inception modules
  - 22 layers
  - 5 million parameters
    (throw away fully connected layers)
- **ResNet (2015): 6.7% to 3.57%**
  - More layers = better performance
  - 152 layers
- **CUImage (2016): 3.57% to 2.99%**
  - Ensemble of 6 models
- **SENet (2017): 2.99% to 2.251%**
  - Squeeze and excitation block: network is allowed to adaptively adjust the weighting of each feature map in the convolutional block.

Object Detection / Localization

Region-Based Methods | Shown: Faster R-CNN

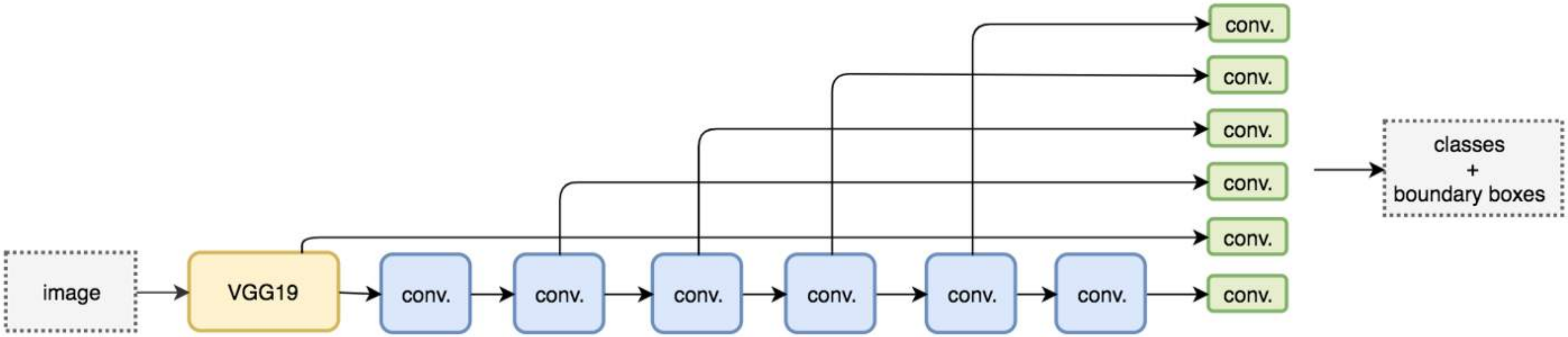


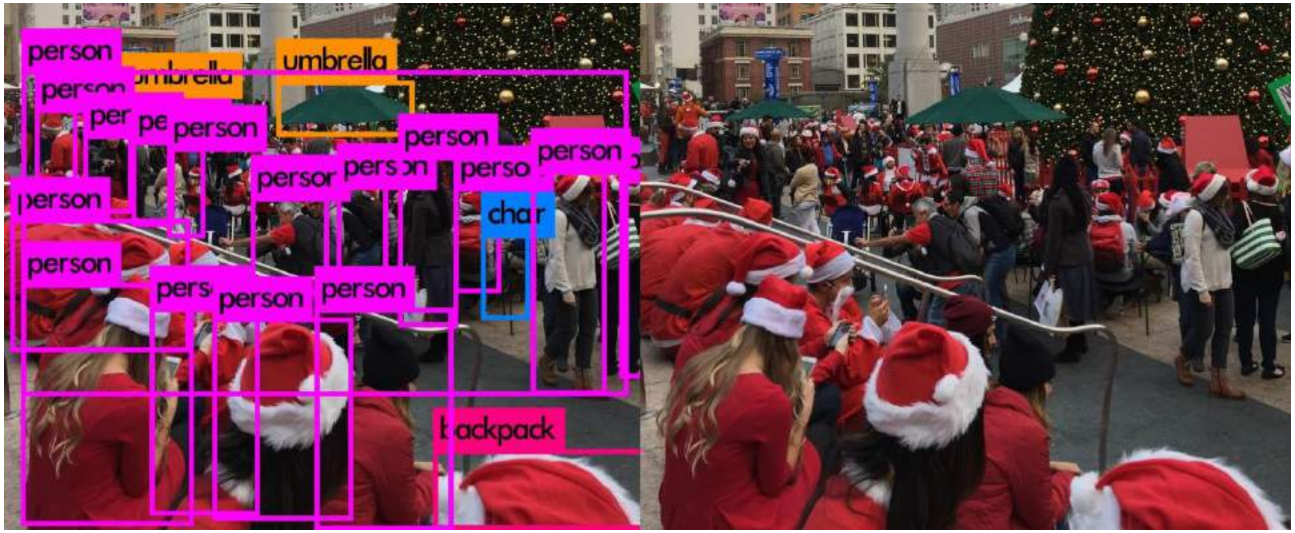

```
ROIs = region_proposal(image)
for ROI in ROIs
    patch = get_patch(image, ROI)
    results = detector(patch)
```



Object Detection / Localization

Single-Shot Methods | Shown: SSD





Object Detection: State of the Art Progress

R-CNN (2013.11) → **OverFeat** (ICLR' 14) → MultiBox (CVPR' 14) → SPP-Net (ECCV' 14) → MR-CNN (ICCV' 15) → DeepBox (ICCV' 15) → AttentionNet (ICCV' 15) →

Fast R-CNN (ICCV' 15) → DeepProposal (ICCV' 15) → **Faster R-CNN** (NIPS' 15) → **OHEM** (CVPR' 16) → **YOLO v1** (CVPR' 16) → G-CNN (CVPR' 16) → AZNet (CVPR' 16) →

Inside-OutsideNet(ION) (CVPR' 16) → HyperNet (CVPR' 16) → CRAFT (CVPR' 16) → MultiPathNet(MPN) (BMVC' 16) → **SSD** (ECCV' 16) → GBDNet (ECCV' 16) →

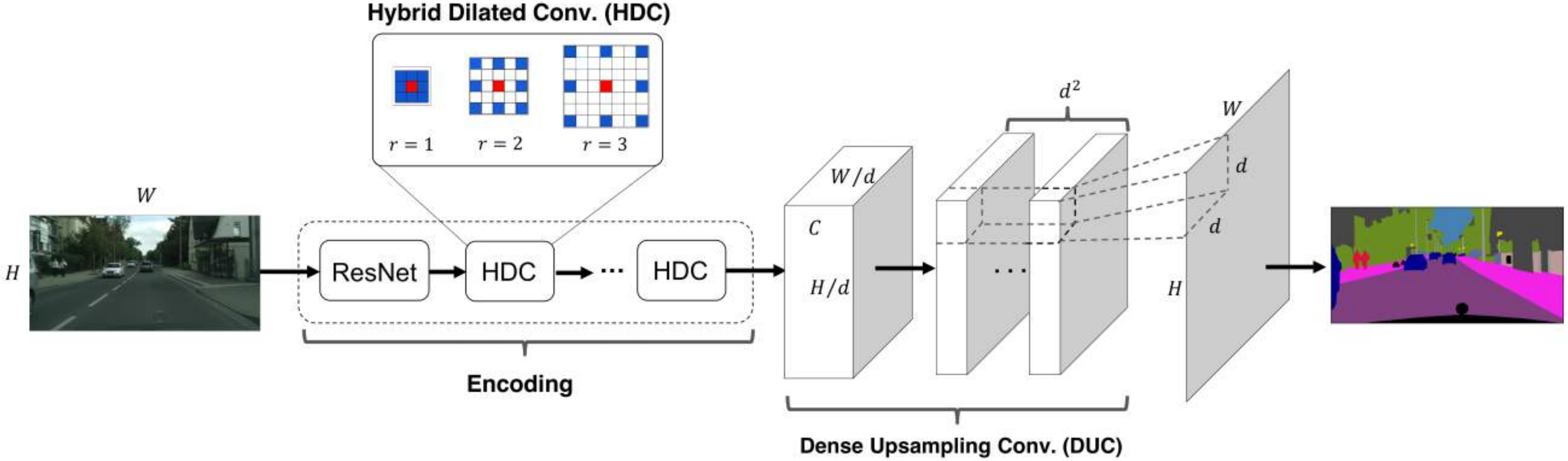
CPF (ECCV' 16) → MS-CNN (ECCV' 16) → **R-FCN** (NIPS' 16) → PVANET (NIPSW' 16) → DeepID-Net (PAMI' 16) → NoC (TPAMI' 16) → DSSD (arXiv' 17) → TDM (CVPR' 17) → **YOLO v2** (CVPR' 17) →

Feature Pyramid Net(**FPN**) (CVPR' 17) → RON (CVPR' 17) → DCN (ICCV' 17) → DeNet (ICCV' 17) → CoupleNet (ICCV' 17) → **RetinaNet** (ICCV' 17) → DSOD (ICCV' 17) →

Mask R-CNN (ICCV' 17) → SMN (ICCV' 17) → **YOLO v3** (arXiv' 18) → SIN (CVPR' 18) → STDN (CVPR' 18) → **RefineDet** (CVPR' 18) → MLKP (CVPR' 18) → Relation-Net (CVPR' 18) →

Cascade R-CNN (CVPR' 18) → RFBNet (ECCV' 18) → CornetNet (ECCV' 18) → Pelee (NIPS' 18) → MethAnchor (NIPS' 18) → SNIPER (NIPS' 18) → **M2Det** (AAAI' 19) ⋯

Semantic Segmentation



State-of-the-Art: DeepLab v3

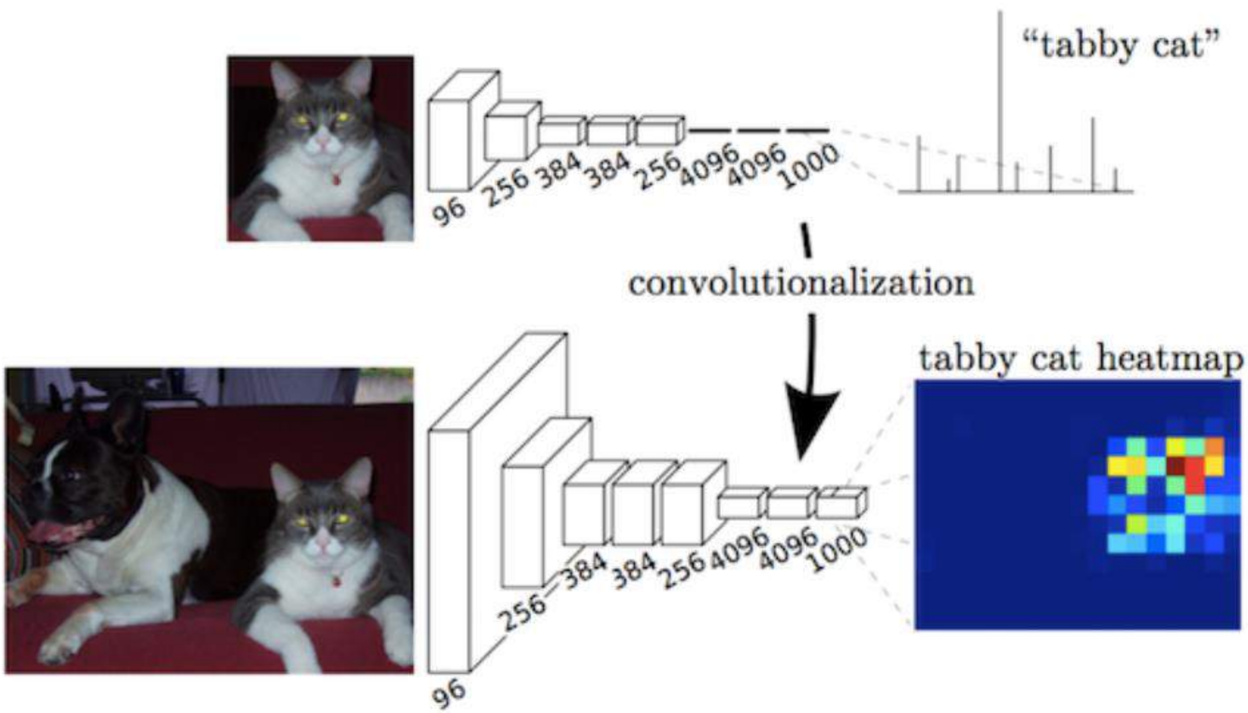
- PASCAL VOC Challenge: http://bit.ly/2HdzTEu

▶	DeepLabv3+_JFT [?]	**89.0**
▷	SRC-B-MachineLearningLab [?]	88.5
▷	DeepLabv3+_AASPP [?]	88.5
▷	MSCI [?]	88.0
▷	ExFuse [?]	87.9
▷	DeepLabv3+ [?]	87.8
▷	DeepLabv3-JFT [?]	86.9

FCN (Nov 2014)

Paper: "Fully Convolutional Networks for Semantic Segmentation"

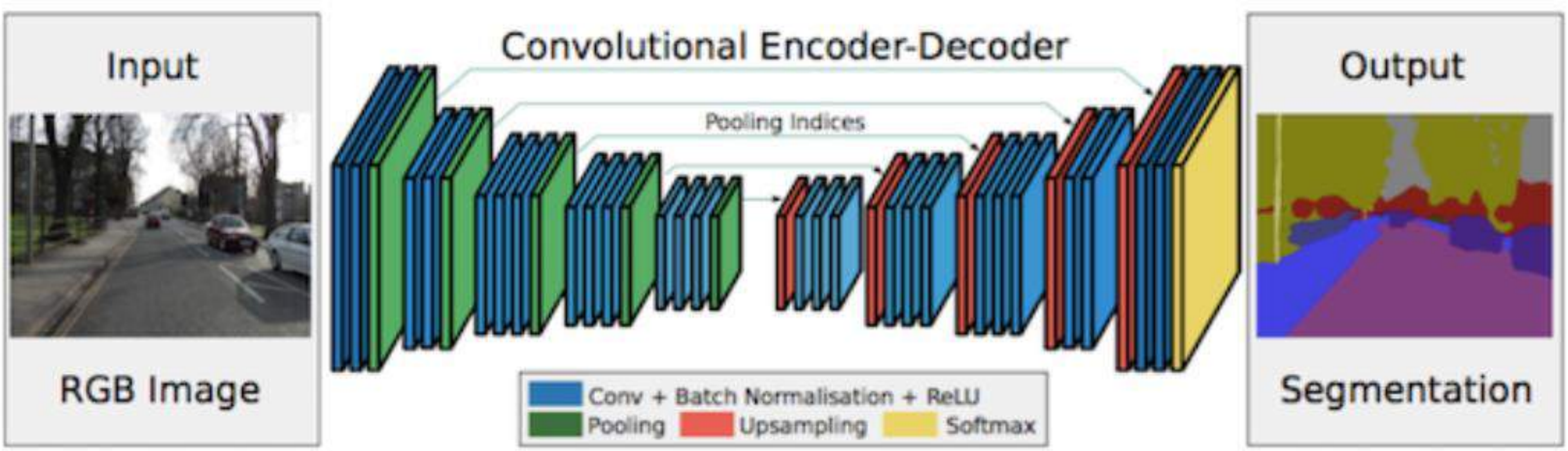
- Repurpose Imagenet pretrained nets
- Upsample using deconvolution
- Skip connections to improve coarseness of upsampling



SegNet (Nov 2015)

Paper: "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation"

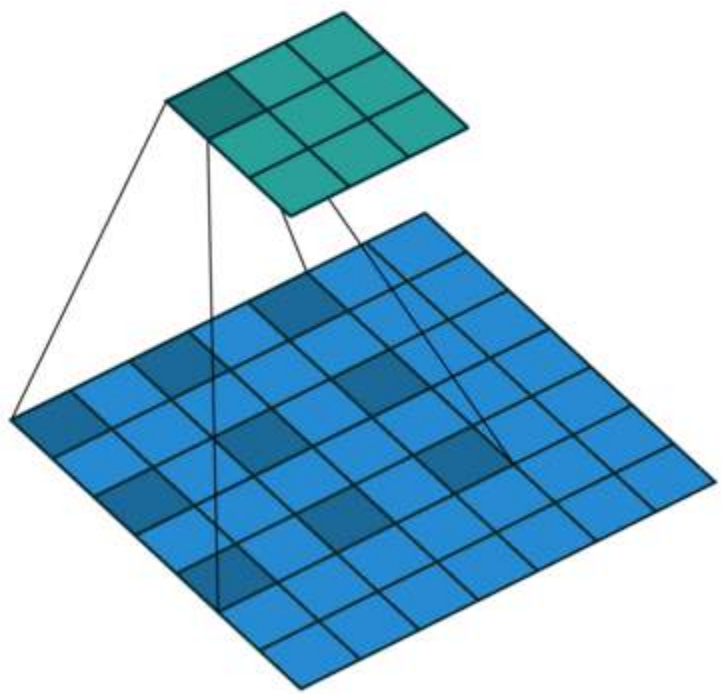
- Maxpooling indices transferred to decoder to improve the segmentation resolution.



Dilated Convolutions (Nov 2015)

Paper: “Multi-Scale Context Aggregation by Dilated Convolutions”

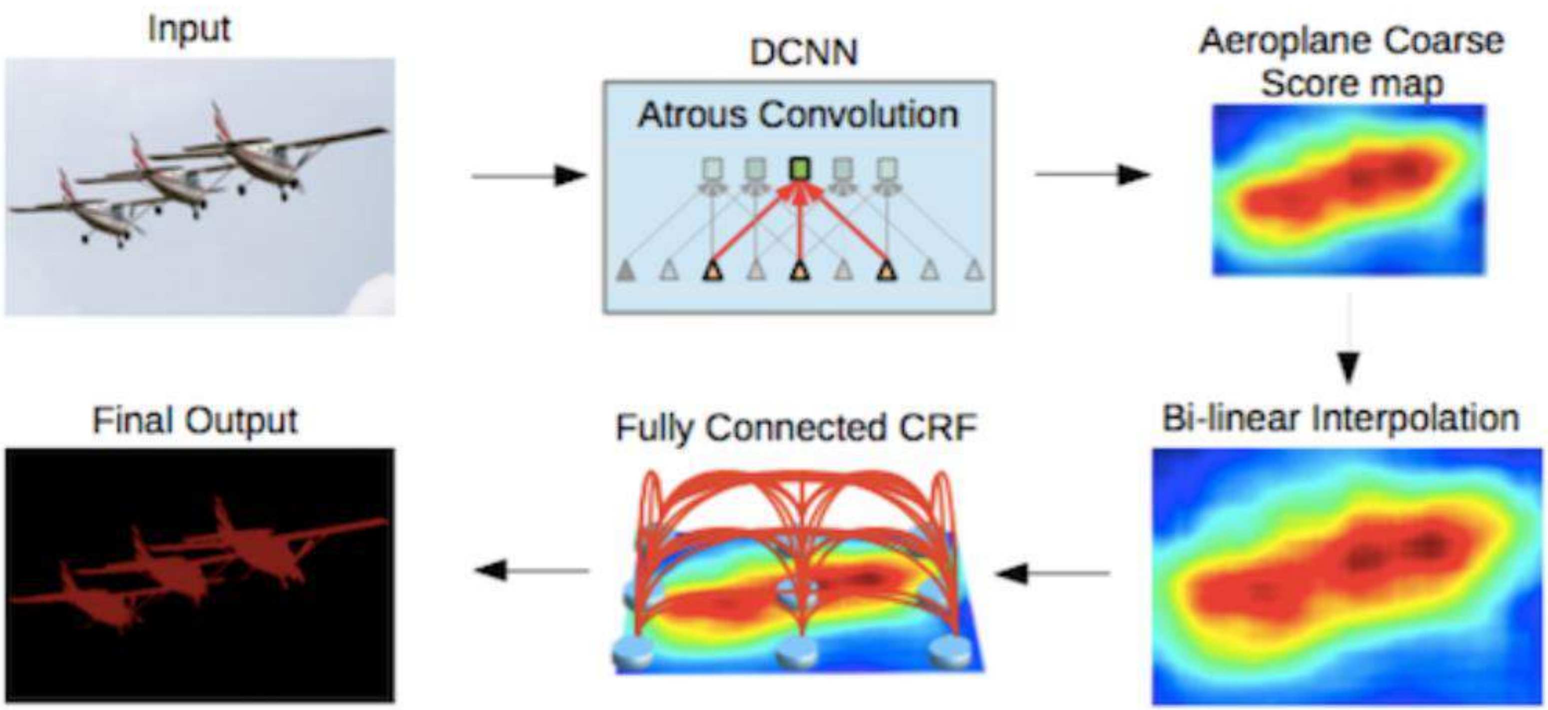
- Since pooling decreases resolution:
  - Added “dilated convolution layer”
- Still interpolate up from 1/8 of original image size



DeepLab v1, v2 (Jun 2016)

Paper: "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs"

- Added fully-connected Conditional Random Fields (CRFs) – as a post-processing step
  - Smooth segmentation based on the underlying image intensities



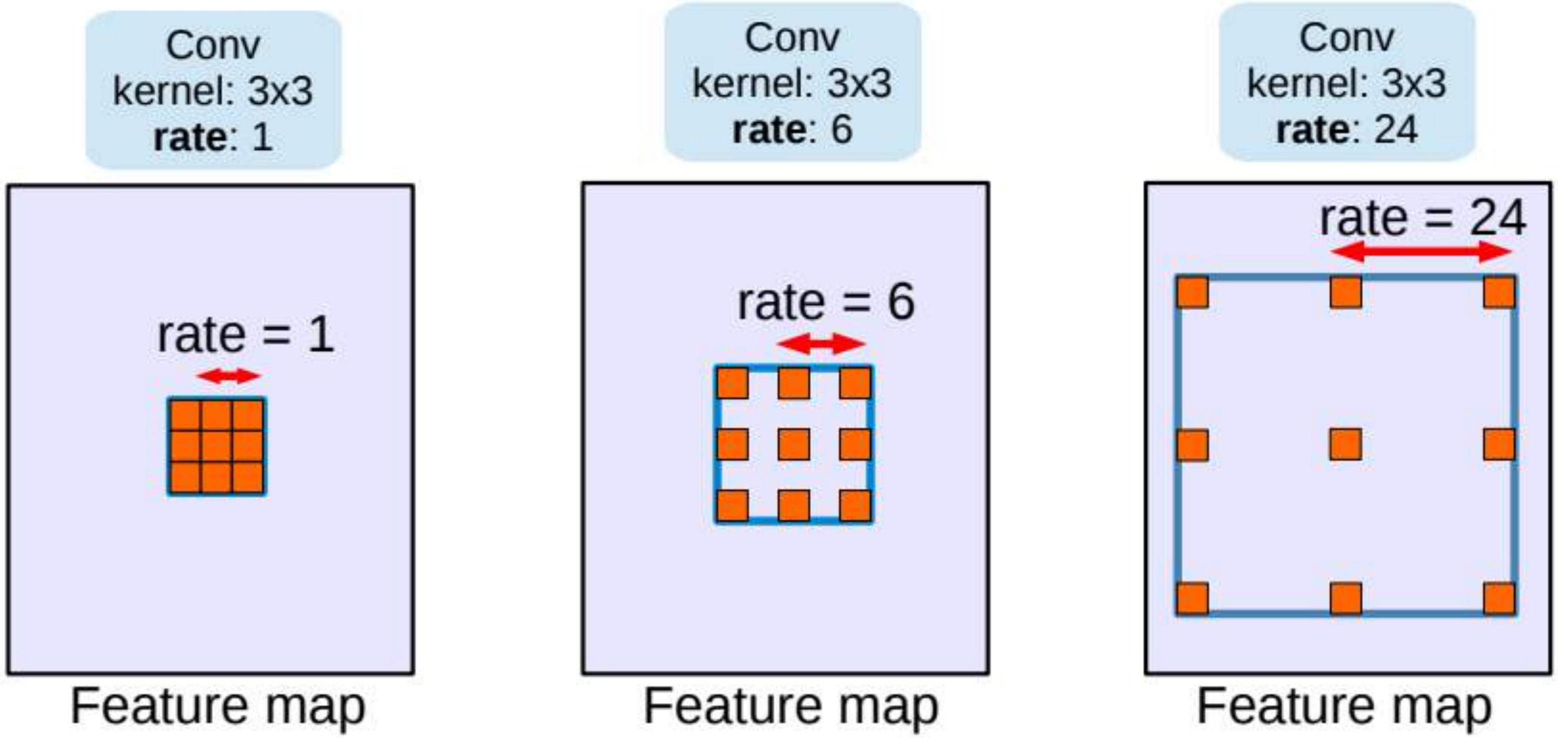
Key Aspects of Segmentation

- **Fully convolutional networks (FCNs)** - replace fully-connected layers with convolutional layers. Classification network is where the biggest gains come from.
  - Deeper, updated models (ResNet, etc) consistent with ImageNet Challenge object classification tasks.
- **Conditional Random Fields (CRFs)** to capture both local and long-range dependencies within an image to refine the prediction map.
- **Dilated convolution** (aka Atrous convolution) – maintain computational cost, increase resolution of intermediate feature maps
- Process at multiple scales and combine the information together

DeepLab v3

Paper: "Rethinking Atrous Convolution for Semantic Image Segmentation"

- Multi-scale processing, **without** increasing parameters.
- Increasing "atrous rate" enlarges the model's "field-of-view"



DeepLab v3 trained on CityScapes

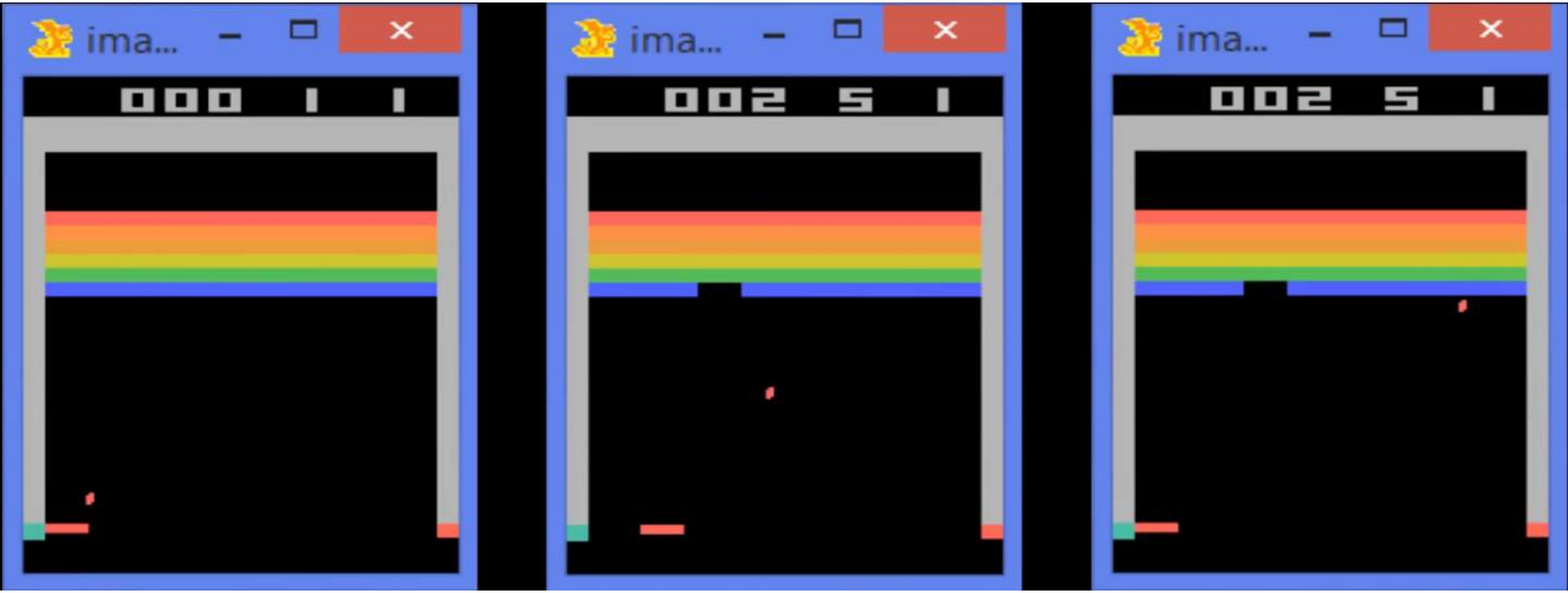
Tutorial: https://github.com/lexfridman/mit-deep-learning

Deep Learning: State of the Art*
(Breakthrough Developments in 2017 & 2018)

- BERT and Natural Language Processing
- Tesla Autopilot Hardware v2+: Neural Networks at Scale
- AdaNet: AutoML with Ensembles
- AutoAugment: Deep RL Data Augmentation
- Training Deep Networks with Synthetic Data
- Segmentation Annotation with Polygon-RNN++
- DAWNBench: Training Fast and Cheap
- BigGAN: State of the Art in Image Synthesis
- Video-to-Video Synthesis
- Semantic Segmentation
- **AlphaZero & OpenAI Five**
- Deep Learning Frameworks

* This is not a list of state-of-the-art results on main machine learning benchmark datasets. It's an overview of exciting recent developments.

Atari Breakout

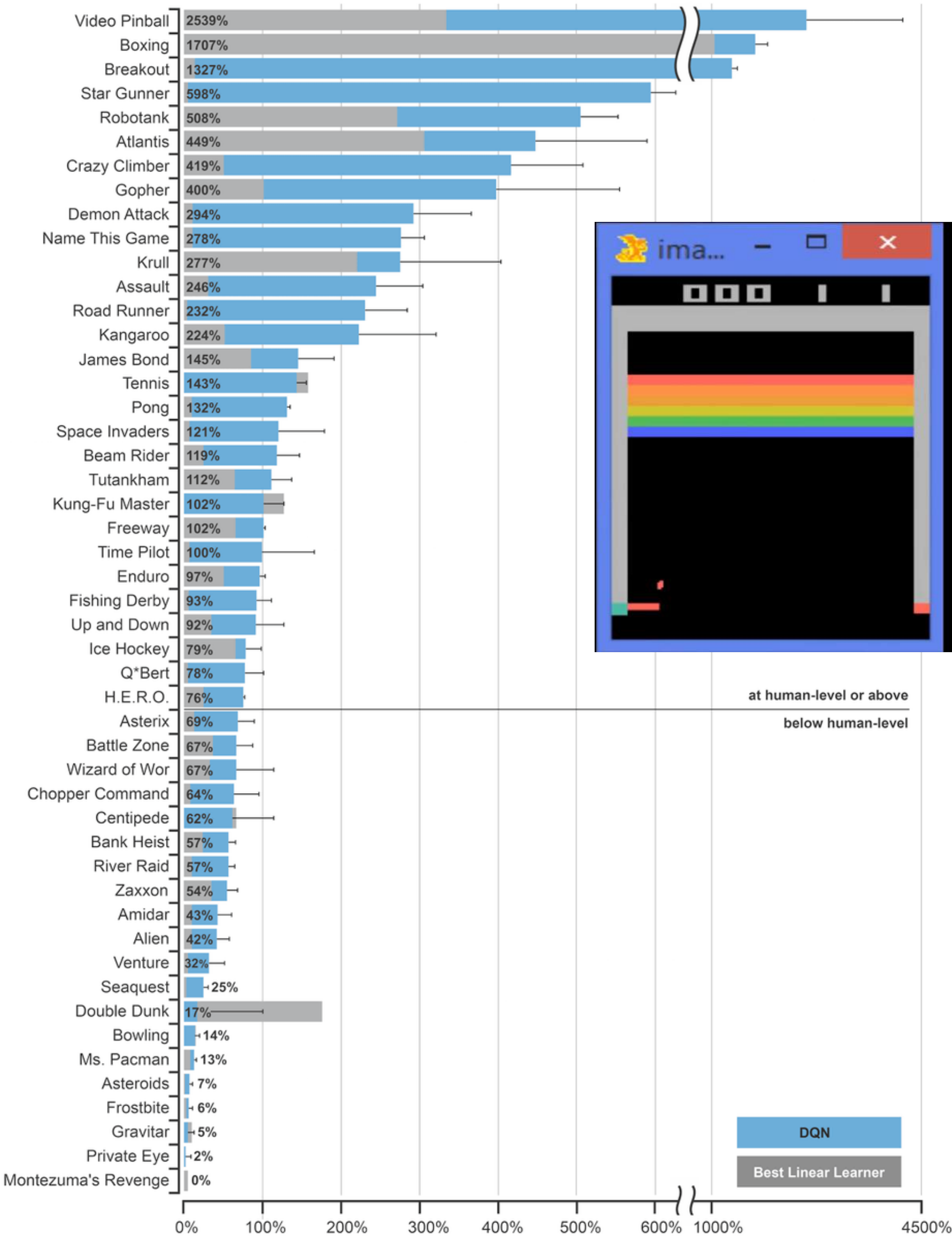


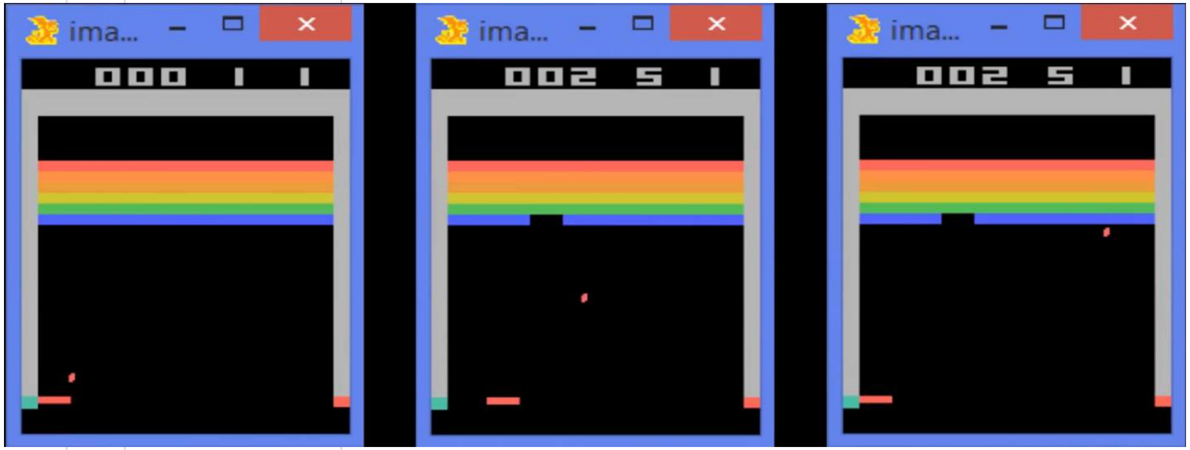
After **10 Minutes** of Training

After **120 Minutes** of Training

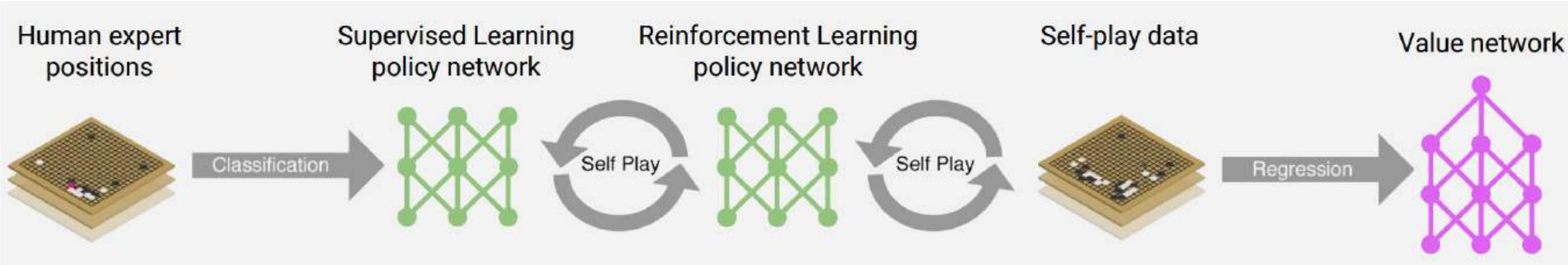
After **240 Minutes** of Training

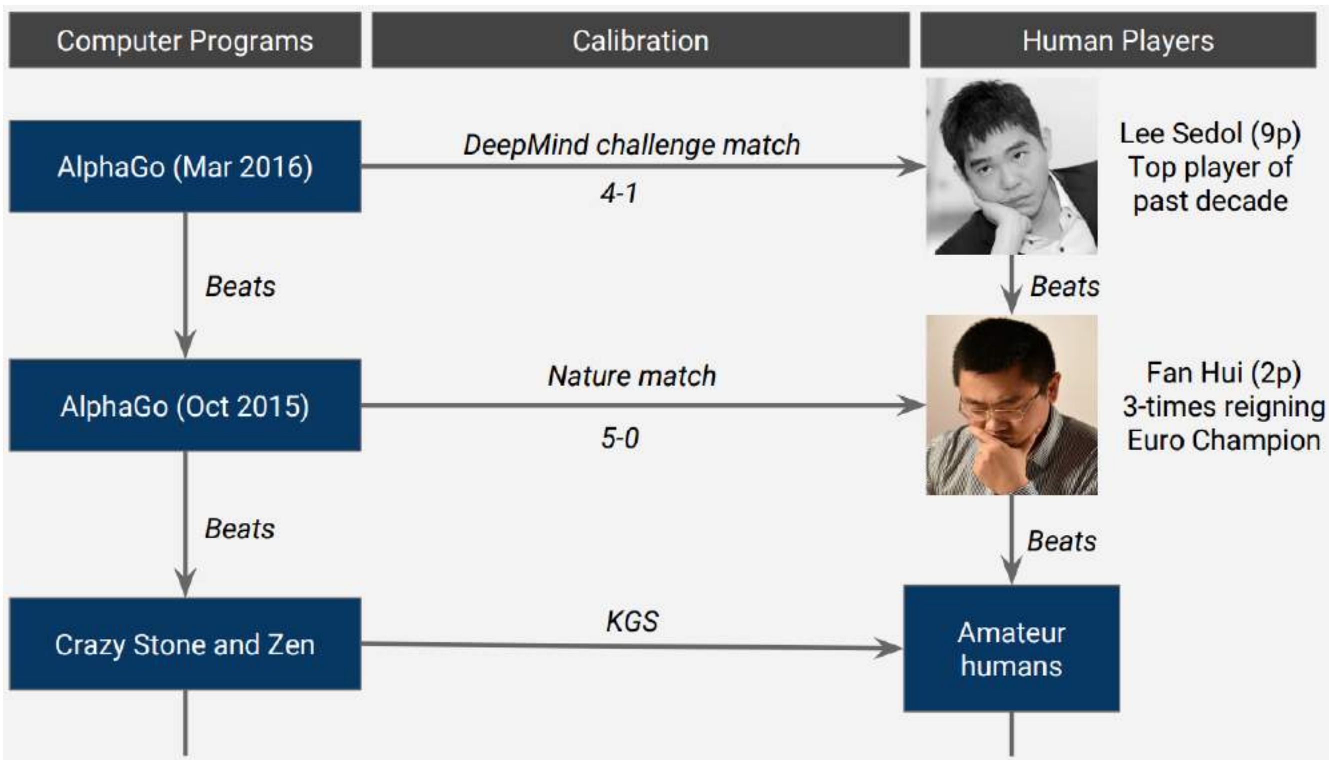
DQN: Atari Games (2015)



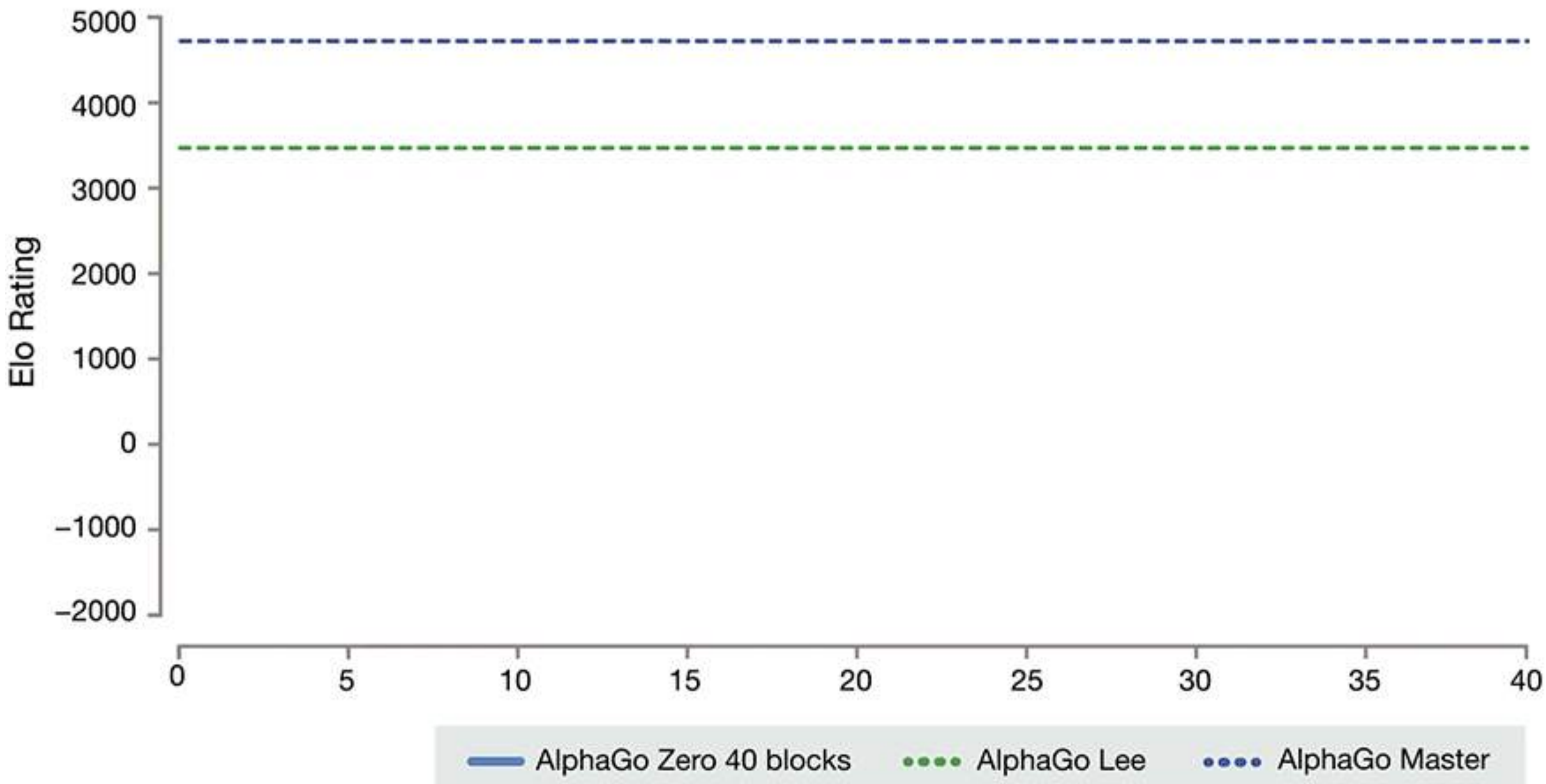


AlphaGo (2016): Beat Top Human at Go

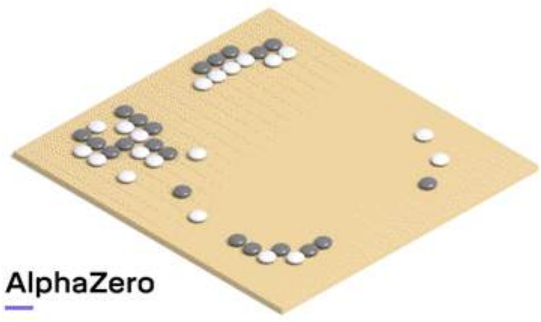


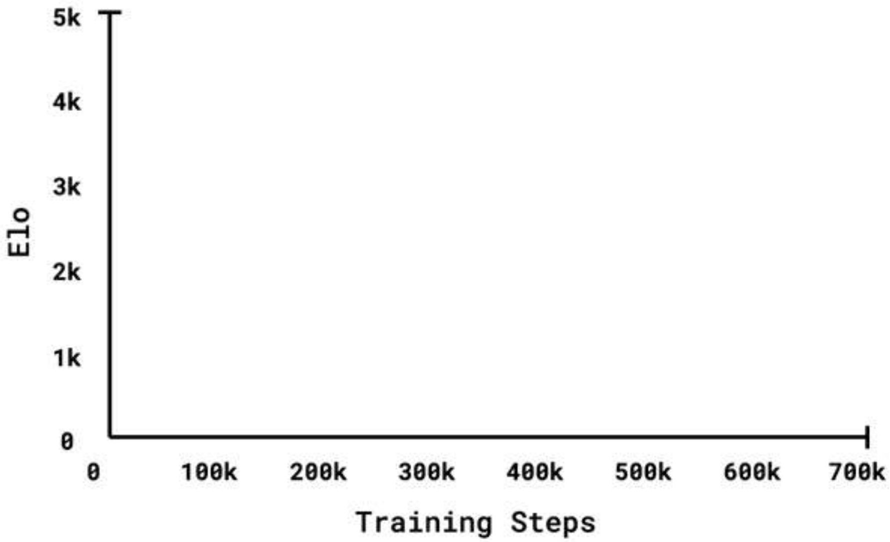


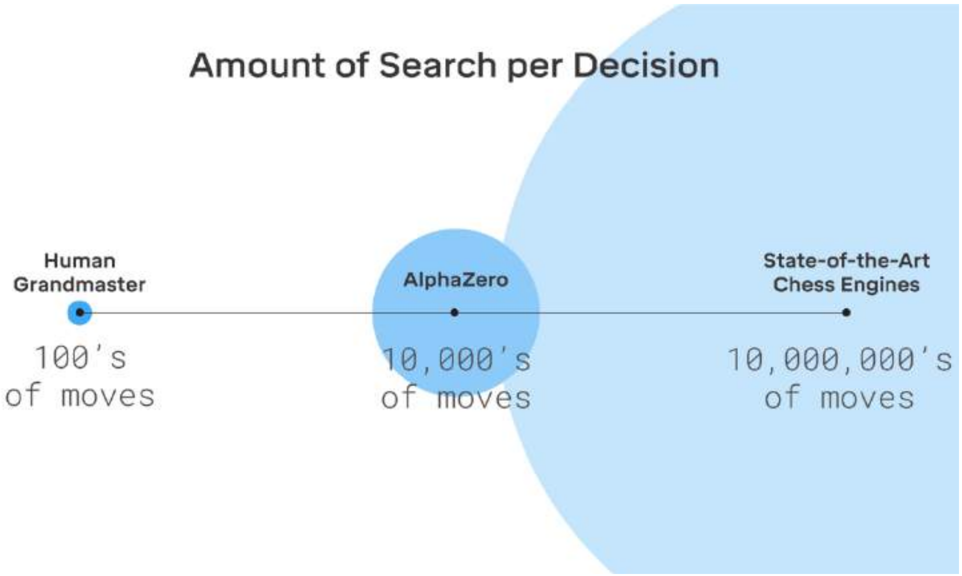
AlphaGo Zero (2017): Beats AlphaGo



AlphaZero (Dec 2017) *vs StockFish (Chess) & Elmo (Shogi)*







OpenAI & Dota 2

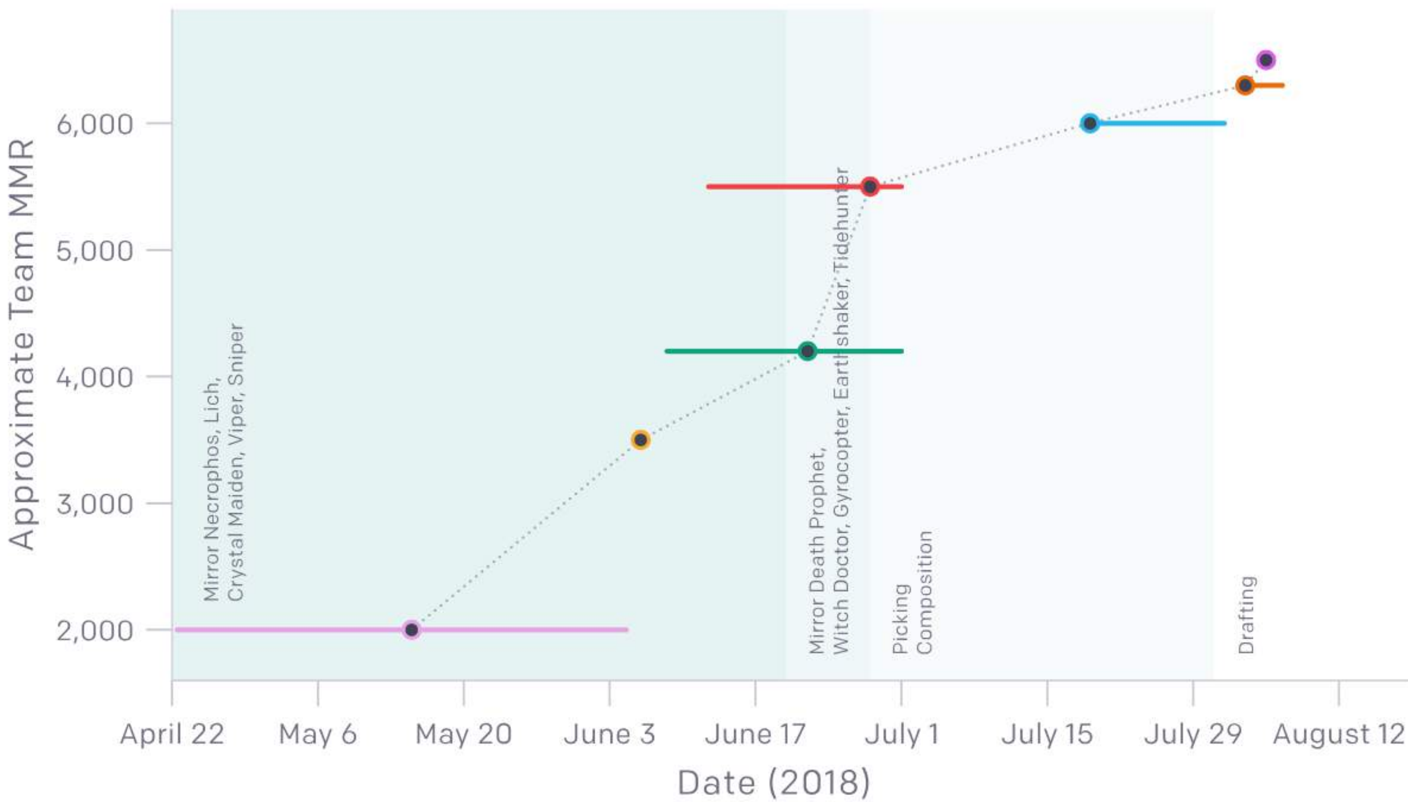
- Dota 2 as a testbed for the messiness and continuous nature of the real world: teamwork, long time horizons, and hidden information.

The International
DOTA 2 CHAMPIONSHIPS

Place	Team	Prize money
1st	OG	$11,190,158
2nd	PSG.LGD	$4,069,148
3rd	Evil Geniuses	$2,670,379
4th	Team Liquid	$1,780,252
5th/6th	Team Secret	$1,144,448
	Virtus.pro	
7th/8th	OpTic Gaming	$635,804
	VGJ.Storm	
9th–12th	Mineski	$381,483
	Team Serenity	
	Vici Gaming	
	Winstrike Team	
13th–16th	Fnatic	$127,161
	Newbee	
	TNC Predator	
	VGJ.Thunder	
17th–18th	Invictus Gaming	$63,580
	paiN Gaming	

OpenAI & Dota 2 Progress

- Aug, 2017: 1v1 bot beats top professional Dota 2 players.
- Aug, 2018: OpenAI Five lost two games against top Dota 2 players at The International. "We are looking forward to pushing Five to the next level."

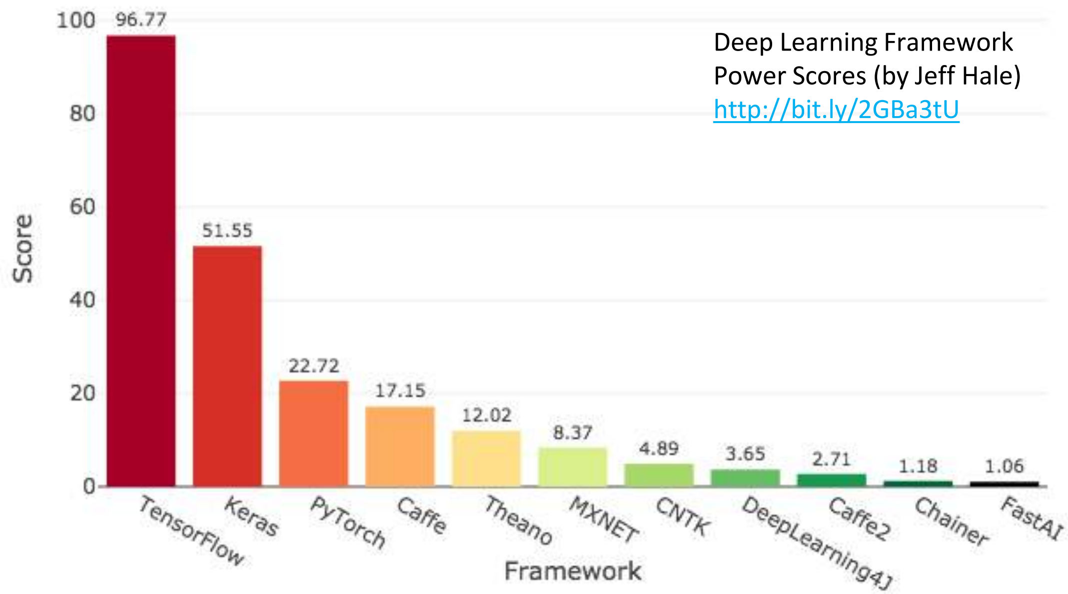


Deep Learning: State of the Art*
(Breakthrough Developments in 2017 & 2018)

- BERT and Natural Language Processing
- Tesla Autopilot Hardware v2+: Neural Networks at Scale
- AdaNet: AutoML with Ensembles
- AutoAugment: Deep RL Data Augmentation
- Training Deep Networks with Synthetic Data
- Segmentation Annotation with Polygon-RNN++
- DAWNBench: Training Fast and Cheap
- BigGAN: State of the Art in Image Synthesis
- Video-to-Video Synthesis
- Semantic Segmentation
- AlphaZero & OpenAI Five
- **Deep Learning Frameworks**

* This is not a list of state-of-the-art results on main machine learning benchmark datasets. It's an overview of exciting recent developments.

Deep Learning Frameworks



Factors to consider:

- Learning curve
- Speed of development
- Size and passion of community
- Number of papers implemented in framework
- Likelihood of long-term growth and stability
- Ecosystem of tooling

1. TensorFlow
2. Keras
3. PyTorch
4. Caffe
5. theano
6. Apache mxnet
7. Microsoft CNTK
8. DL4J
9. Caffe2
10. Chainer
11. fast.ai

Deep Learning: 2019 and Beyond

- On backpropagation: "My view is throw it all away and start again."
- "The future depends on some graduate student who is deeply suspicious of everything I have said."

\- Geoffrey Hinton
"Godfather of Deep Learning"

Thank You

Website:

deeplearning.mit.edu

- Videos and slides posted on the website
- Code posted on GitHub:
  https://github.com/lexfridman/mit-deep-learning