TTIC 31230, Fundamentals of Deep Learning

David McAllester, Autumn 2020

The History of Deep Learning

and Moore's Law of AI

Early History

1943: McCullock and Pitts introduced the linear threshold "neuron".

1962: Rosenblatt applies a "Hebbian" learning rule. Novikoff proved the perceptron convergence theorem.

1969: Minsky and Papert publish the book *Perceptrons*.

The Perceptrons book greatly discourages work in artificial neural networks. Symbolic methods dominate AI research through the 1970s.

80s Renaissance

1980: Fukushima introduces the neocognitron (a form of CNN)

1984: Valiant defines PAC learnability and stimulates learning theory. Wins the Turing Award in 2010.

1985: Hinton and Sejnowski introduce the Boltzman machine

1986: Rummelhart, Hinton and Williams demonstrate empirical success with backpropagation (itself dating back to 1961).

90s and 00s: Research In the Shadows

1997: Schmidhuber et al. introduce LSTMs

1998: LeCunn introduces convolutional neural networks (CNNs) (LeNet).

2003: Bengio introduces neural language modeling.

Current Era

2012: Alexnet dominates the Imagenet computer vision challenge.

Google speech recognition converts to deep learning.

Both developments come out of Hinton's group in Toronto.

2013: Refinement of AlexNet continues to dramatically improve computer vision.

Current Era

2014: Neural machine translation appears (Seq2Seq models).

Variational auto-encoders (VAEs) appear.

Generative Adversarial Networks (GANs) appear.

Graph neural networks appear (GNNs) revolutionizing the prediction of molecular properties.

Dramatic improvement in computer vision and speech recognition continues.

Current Era

2015: Google converts to neural machine translation leading to dramatic improvements.

ResNet (residual connections) appear. This makes yet another dramatic improvement in computer vision.

2016: Alphago defeats Lee Sedol.

Current Era

2017: AlphaZero learns both go and chess at super-human levels in a mater of hours entirely form self-play and advances computer go far beyond human abilities.

Unsupervised machine translation is demonstrated.

Progressive GANs demonstrate high resolution realistic face generation.

Current Era

2018: Unsupervised pre-training significantly improves a broad range of NLP tasks including question answering (but dialogue remains unsolved).

AlphaFold revolutionizes protein structure prediction.

Style GANs lead to yet another dramatic imrpovement in generation of face images.

2019: Vector quantized VAEs (VQ-VAE) demstrate that VAEs can be competative with GANs for high-resolution image generation.

Super-human performance is achieved on the GLUE natural langauge understanding benchmark.

Current Era

2020: GPT-3 writes a fake blog post that landed in the No. 1 spot on Hacker News.

Natural Language Understanding

GLUE: General Language Understanding Evaluation

ArXiv 1804.07461

Corpus	|Train|	|Test|	Task	Metrics	Domain
			Single-Sentence Tasks		
CoLA	8.5k	**1k**	acceptability	Matthews corr.	misc.
SST-2	67k	1.8k	sentiment	acc.	movie reviews
			Similarity and Paraphrase Tasks		
MRPC	3.7k	1.7k	paraphrase	acc./F1	news
STS-B	7k	1.4k	sentence similarity	Pearson/Spearman corr.	misc.
QQP	364k	**391k**	paraphrase	acc./F1	social QA questions
			Inference Tasks		
MNLI	393k	**20k**	NLI	matched acc./mismatched acc.	misc.
QNLI	105k	5.4k	QA/NLI	acc.	Wikipedia
RTE	2.5k	3k	NLI	acc.	news, Wikipedia
WNLI	634	**146**	coreference/NLI	acc.	fiction books

Table 1: Task descriptions and statistics. All tasks are single sentence or sentence pair classification, except STS-B, which is a regression task. MNLI has three classes; all other classification tasks have two. Test sets shown in bold use labels that have never been made public in any form.

BERT and GLUE

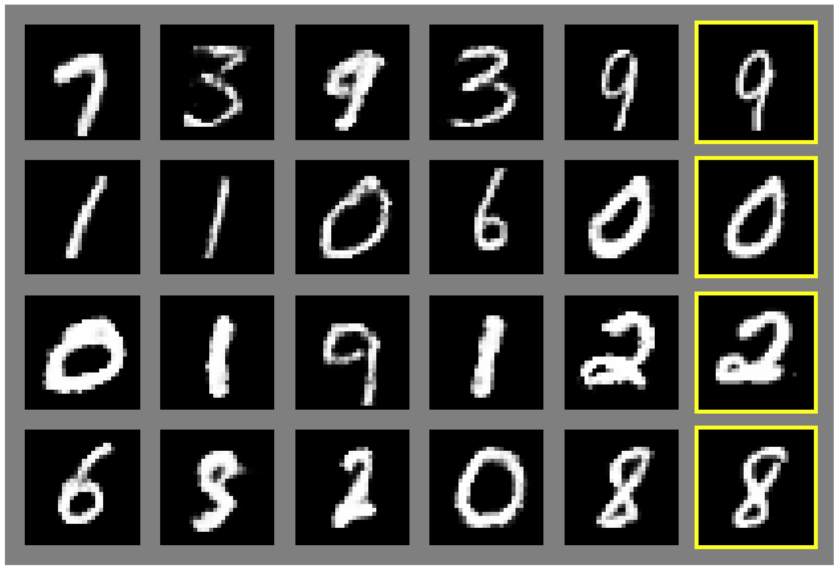
Rank	Name	Model	URL	Score
1	T5 Team - Google	T5		90.3
2	ERNIE Team - Baidu	ERNIE		90.1
3	Microsoft D365 AI & MSR AI & GATECH	MT-DNN-SMART		89.9
4	王玮	ALICE v2 large ensemble (Alibaba DAMO NLP)		89.7
5	Microsoft D365 AI & UMD	FreeLB-RoBERTa (ensemble)		88.4
6	Junjie Yang	HIRE-RoBERTa		88.3
7	Facebook AI	RoBERTa		88.1
8	Microsoft D365 AI & MSR AI	MT-DNN-ensemble		87.6
9	GLUE Human Baselines	GLUE Human Baselines		87.1

BERT and SuperGLUE

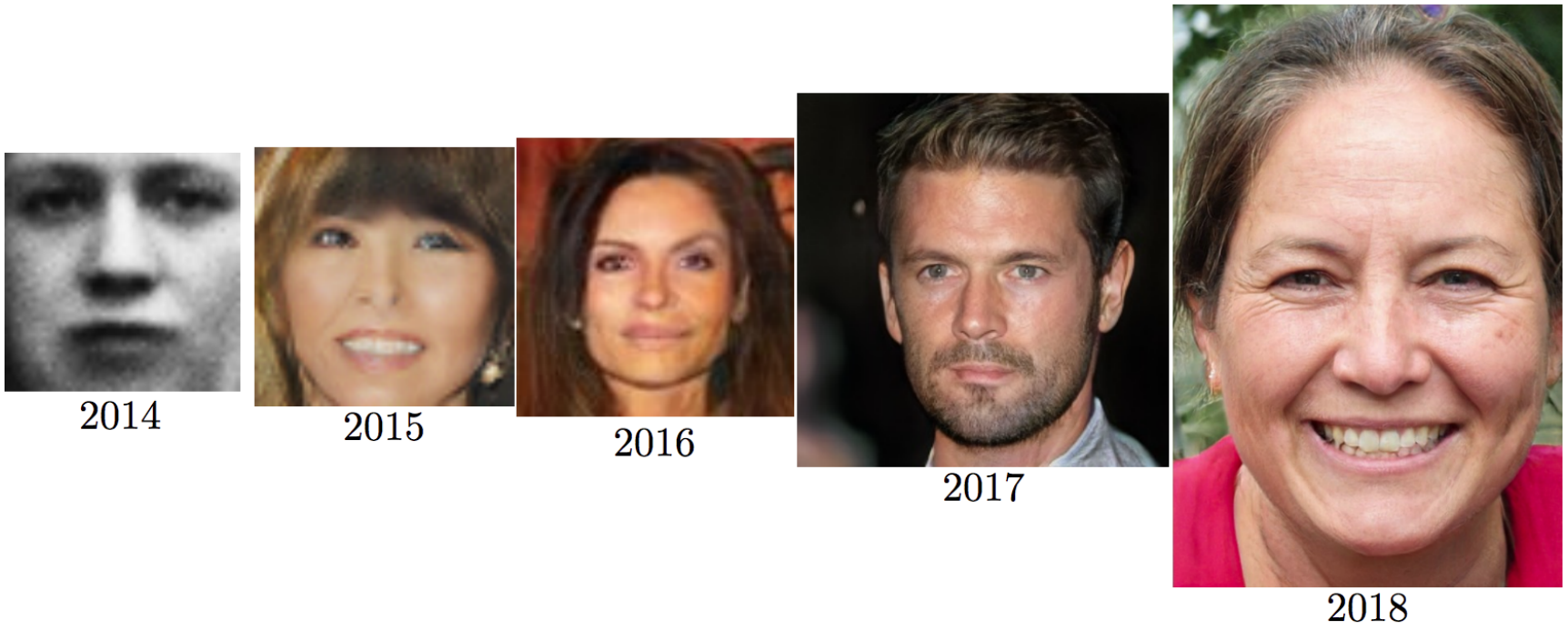
	Rank	Name	Model	URL	Score
	1	SuperGLUE Human Baselines	SuperGLUE Human Baselines		89.8
+	2	T5 Team - Google	T5		89.3
	3	Zhuiyi Technology	RoBERTa-mtl-adv		85.7
	4	Facebook AI	RoBERTa		84.6
	5	IBM Research AI	BERT-mtl		73.5

Generative Adversarial Nets (GANs)

Goodfellow et al., 2014



Moore's Law of AI
4.5 years of progress on faces



(Goodfellow 2019)

ArXiv 1406.2661, 1511.06434, 1607.07536, 1710.10196, 1812.04948
Goodfellow, ICLR 2019 Invited Talk

GANs for Imagenet

(Odena 2018)

BigGANs, Brock et al., 2018

Figure 1: Class-conditional samples generated by our model.

Variational Auto Encoders (VAEs, 2015)

[Alec Radford, 2015]

VAEs in 2019

VQ-VAE-2, Razavi et al. June, 2019

VAEs in 2019

Figure 1: Class-conditional 256x256 image samples from a two-level model trained on ImageNet.

VQ-VAE-2, Razavi et al. June, 2019

END