Empowering Citizens. Smarter Societies.

Insight
Centre for Data Analytics

Successes and Frontiers of Deep Learning

Sebastian Ruder
Insight @ NUIG
Aylien

A World Leading SFI Research Centre

DCU NUI Galway OÉ Gaillimh UCC University College Cork, Ireland Coláiste na hOllscoile Corcaigh UCD Dublin Science Foundation Ireland For what's next

Agenda

1. Deep Learning fundamentals
2. Deep Learning successes
3. Frontiers
    - Unsupervised learning and transfer learning

Empowering Citizens. Smarter Societies.

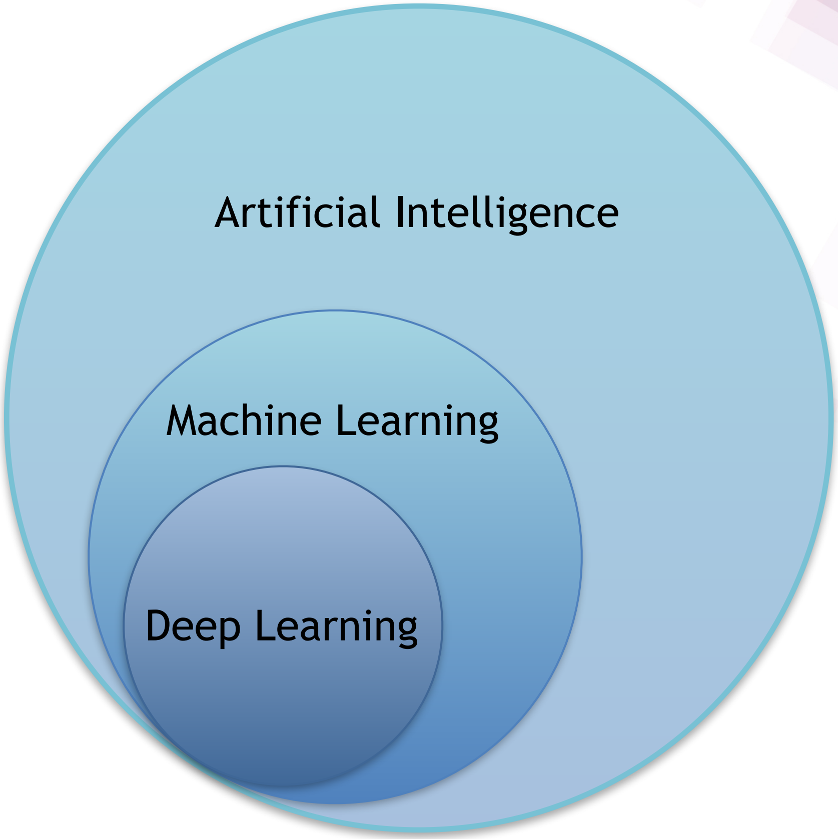
Insight
Centre for Data Analytics

Deep Learning fundamentals

A World Leading SFI Research Centre

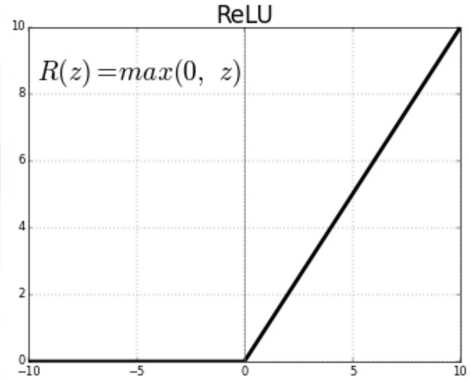
DCU NUI Galway OÉ Gaillimh UCC University College Cork, Ireland Coláiste na hOllscoile Corcaigh UCD Dublin Science Foundation Ireland sfi For what's next

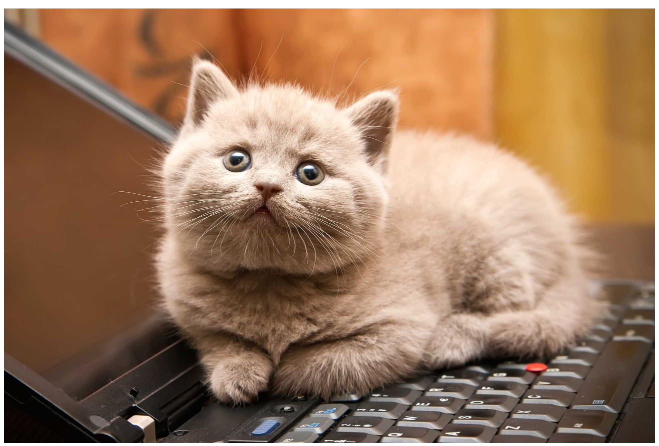
AI and Deep Learning

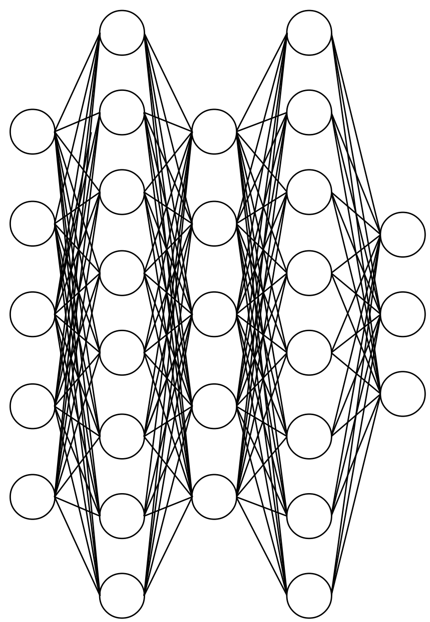


A Deep Neural Network

- A sequence of linear transformations (matrix multiplications) with non-linear activation functions in between
- Maps from an input to (typically) output probabilities





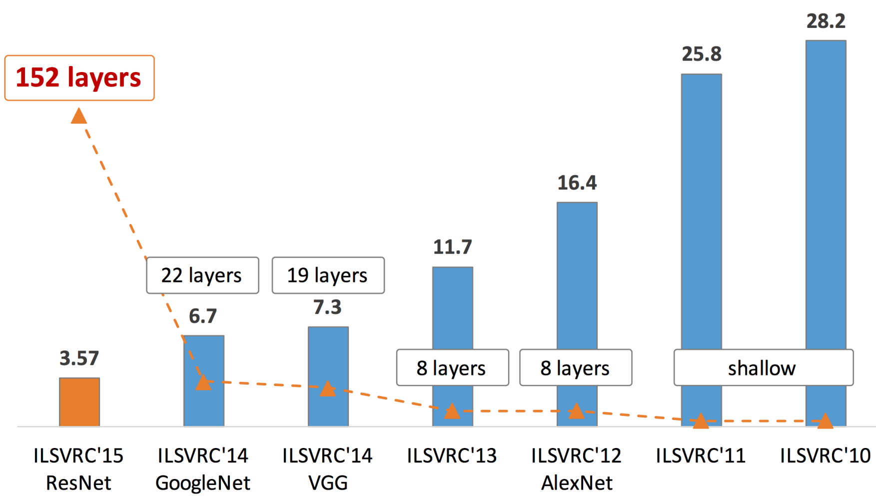


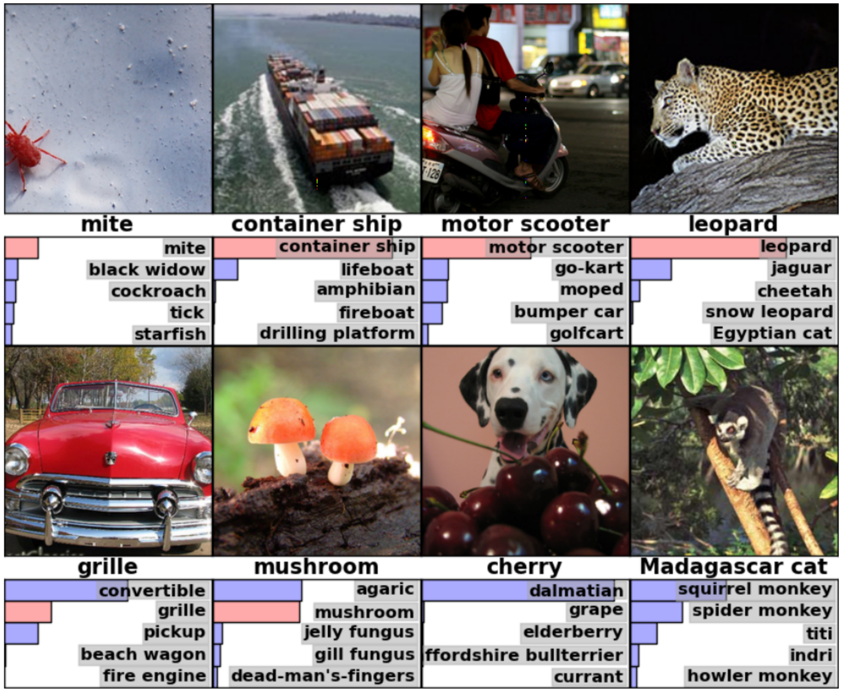
PREDICTED CONCEPT	PROBABILITY
cat	0.985
cute	0.976
little	0.949
mammal	0.945
no person	0.907
pet	0.906
kitten	0.890
baby	0.888
portrait	0.882
one	0.874
funny	0.857

Deep Learning successes

Image recognition

- AlexNet in 2012
- Deeper architectures, more connections: Inception, ResNet, DenseNet, ResNeXt
- SotA (May ’18): 2.4% Top-5, 14.6% Top-1

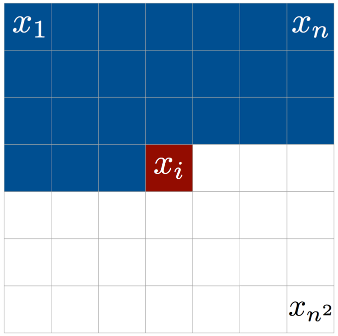


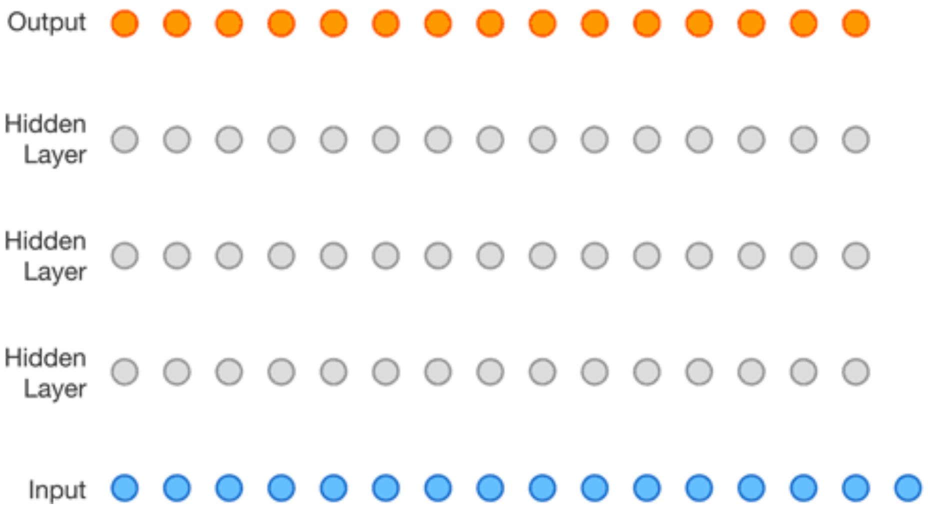


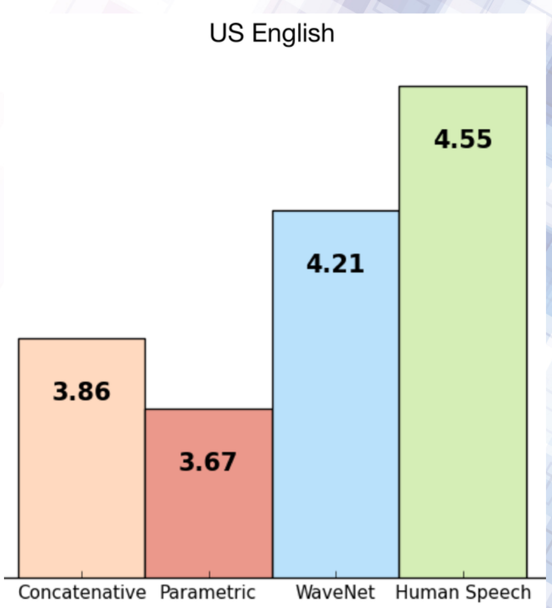
Sources: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf, http://kaiminghe.com/cvpr16resnet/cvpr2016_deep_residual_learning_kaiminghe.pdf, https://research.fb.com/publications/exploring-the-limits-of-weakly-supervised-pretraining/

Speech synthesis

- WaveNet, Parallel WaveNet enable human-like speech synthesis
- Adapted from PixelRNNs, PixelCNNs
- Based on fully convolutional nets with different dilation factors
- Vast number of applications, e.g. personal assistants



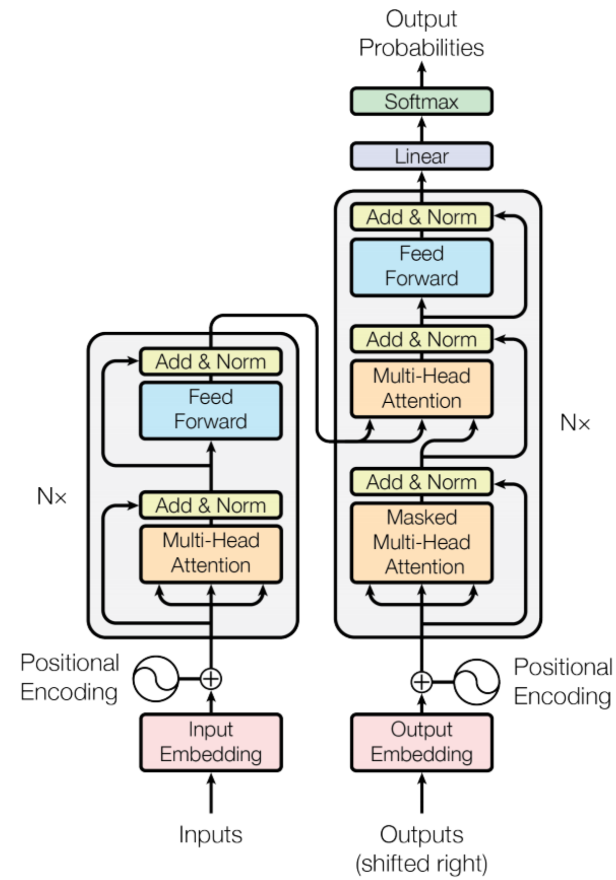




Sources: https://deepmind.com/blog/wavenet-generative-model-raw-audio/, https://deepmind.com/blog/wavenet-launches-google-assistant/, https://arxiv.org/abs/1711.10433, https://arxiv.org/pdf/1601.06759.pdf

Machine translation

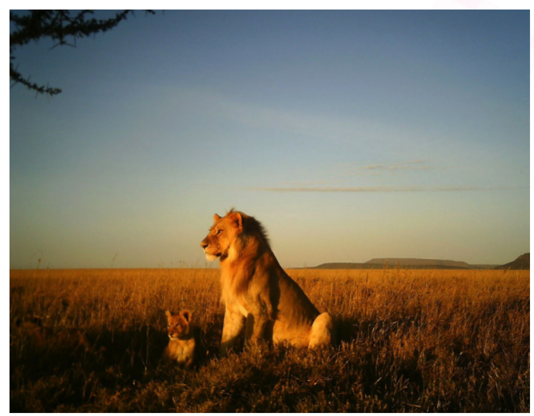
- State-of-the-art: Neural machine translation
- Transformer, based on self-attention; RNMT+
- Combined with other improvements: human parity on English-Chinese (Microsoft)
- Babel-fish earbuds (Google)
- Phrase-based still better on low-resource languages
- Exciting future directions:
  - Unsupervised MT
- But: Only superficial understanding; not close to a 'real' understanding of language yet

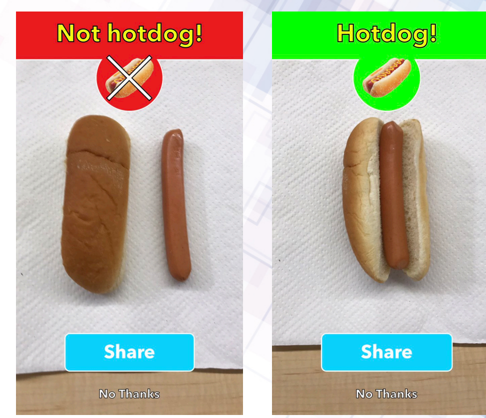


Sources: https://arxiv.org/abs/1706.03762, https://arxiv.org/abs/1804.09849, https://arxiv.org/abs/1803.05567, https://www.citi.io/2018/03/09/top-breakthrough-technologies-for-2018-babel-fish-earbuds/, https://arxiv.org/abs/1711.00043, https://arxiv.org/abs/1710.11041, https://www.theatlantic.com/technology/archive/2018/01/the-shallowness-of-google-translate/551570/

Making ML more accessible

- Accessible libraries such as TensorFlow, PyTorch, Keras, etc.
- Mobile, on-device ML
- Opens up many applications:
  - Predicting disease in Cassava plants
  - Sorting cucumbers
  - Identifying wild animals in trap images
  - Identifying hot dogs





Sources: https://medium.com/tensorflow/highlights-from-tensorflow-developer-summit-2018-cd86615714b2, https://cloud.google.com/blog/big-data/2016/08/how-a-japanese-cucumber-farmer-is-using-deep-learning-and-tensorflow, https://news.developer.nvidia.com/automatically-identify-wild-animals-in-camera-trap-images/, https://medium.com/@timanglade/how-hbos-silicon-valley-built-not-hotdog-with-mobile-tensorflow-keras-react-native-ef03260747f3

Reading comprehension

Microsoft | Office | Windows | Surface | Xbox | Deals | Support | More

The AI Blog | The Official Microsoft Blog | Microsoft On the Issues | Transform

Microsoft creates AI that can read a document and answer questions about it as well as a person

January 15, 2018 | Allison Linn

ALIZILA NEWS FROM ALIBABA GROUP | GATEWAY TO CHINA | ECOSYSTEM IN ACTION | CHINA INSIGHTS | FILTER BY BUSINESS

Never Miss a Story | Enter Your E-mail Address

ALIBABA AI MODEL TOPS HUMANS IN READING COMPREHENSION

ADAM NAJBERG | JANUARY 15, 2018

U.S. EDITION | Thu, May 03, 2018 | Newsweek

U.S. | World | Business | Tech & Science | Culture | Sports | Health | Opi

ROBOTS CAN NOW READ BETTER THAN HUMANS, PUTTING MILLIONS OF JOBS AT RISK

BY ANTHONY CUTHBERTSON ON 1/15/18 AT 8:00 AM

CNN tech | BUSINESS | CULTURE | GADGETS | FUTURE | STARTUPS

Computers are getting better than humans at reading

Sources: https://blogs.microsoft.com/ai/microsoft-creates-ai-can-read-document-answer-questions-well-person/, http://www.alizila.com/alibaba-ai-model-tops-humans-in-reading-comprehension/, http://www.newsweek.com/robots-can-now-read-better-humans-putting-millions-jobs-risk-781393, http://money.cnn.com/2018/01/15/technology/reading-robot-alibaba-microsoft-stanford/index.html

Reading comprehension — success?

- Constrained task
- Models exploit superficial pattern-matching
- Can be easily fooled
  —> adversarial examples

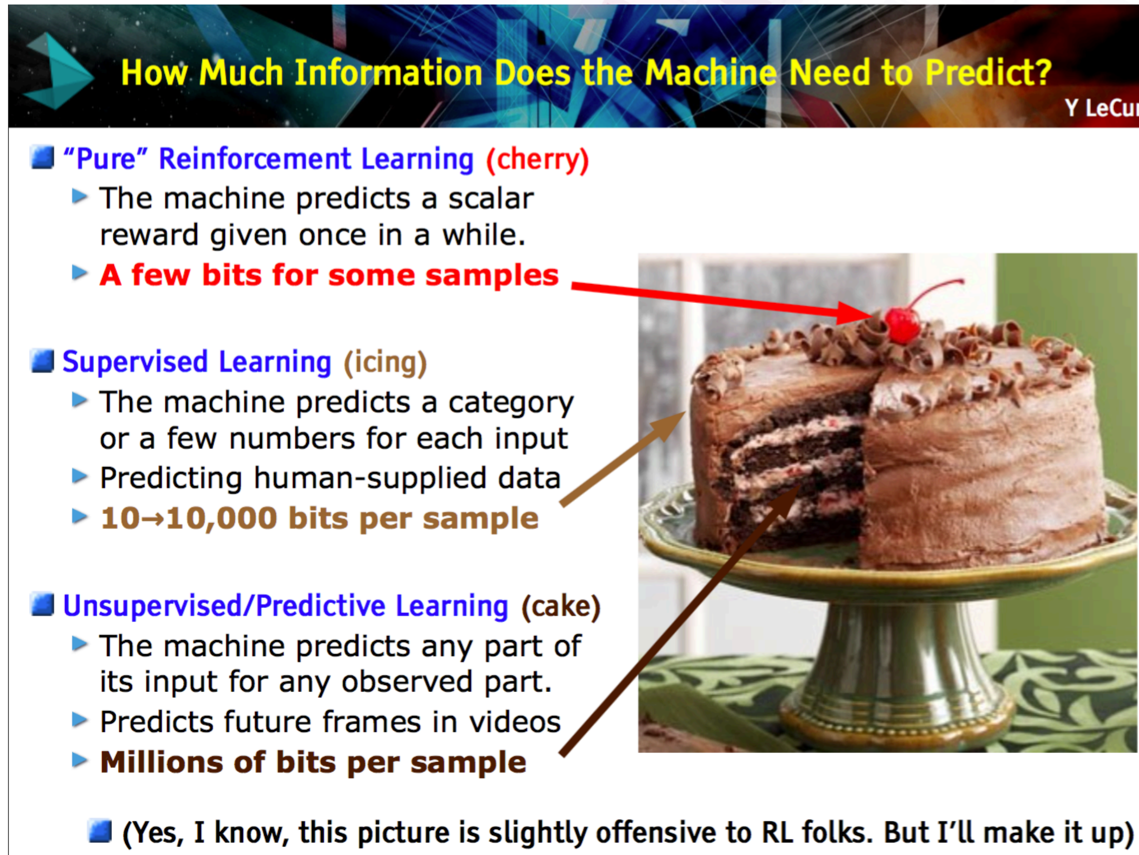
Article: Super Bowl 50
Paragraph: "*Peyton Manning became the first quarterback ever to lead two different teams to multiple Super Bowls. He is also the oldest quarterback ever to play in a Super Bowl at age 39. The past record was held by John Elway, who led the Broncos to victory in Super Bowl XXXIII at age 38 and is currently Denver's Executive Vice President of Football Operations and General Manager. Quarterback Jeff Dean had jersey number 37 in Champ Bowl XXXIV.*"
Question: "*What is the name of the quarterback who was 38 in Super Bowl XXXIII?*"
Original Prediction: John Elway
Prediction under adversary: Jeff Dean

Sources: http://u.cs.biu.ac.il/~yogo/squad-vs-human.pdf, http://www.aclweb.org/anthology/D17-1215

Frontiers

Unsupervised learning

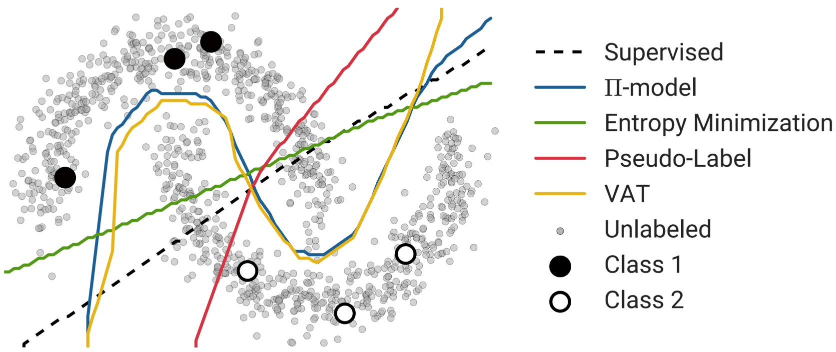
- Unsupervised learning is important
- Still elusive
- Most successful recent direction: Generative Adversarial Networks (GANs)
- Still hard to train
- Mainly useful so far for generative tasks in computer vision

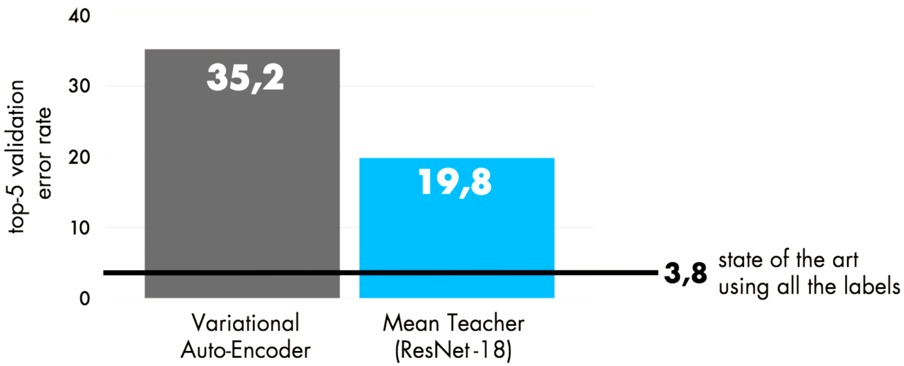


Sources: http://ruder.io/highlights-nips-2016/

Semi-supervised learning

- Advances in semi-supervised learning (SSL)
- 19.8% Top-5 error on ImageNet with only 10% of the labels
- However:
  - Simple baselines—when properly tuned—or heavy regularised models are often competitive
  - Transfer learning often works better
  - Performance can quickly degrade with a domain shift

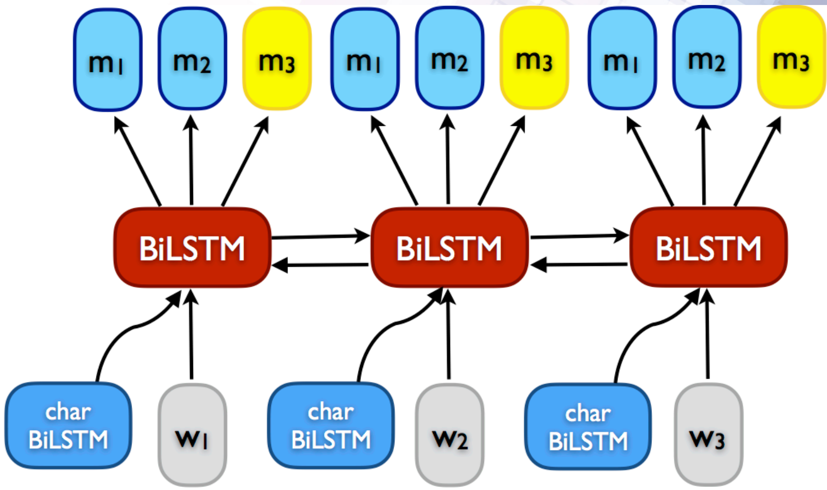




Sources: https://github.com/CuriousAI/mean-teacher/blob/master/nips_2017_slides.pdf, https://arxiv.org/abs/1804.09170

Semi-supervised learning

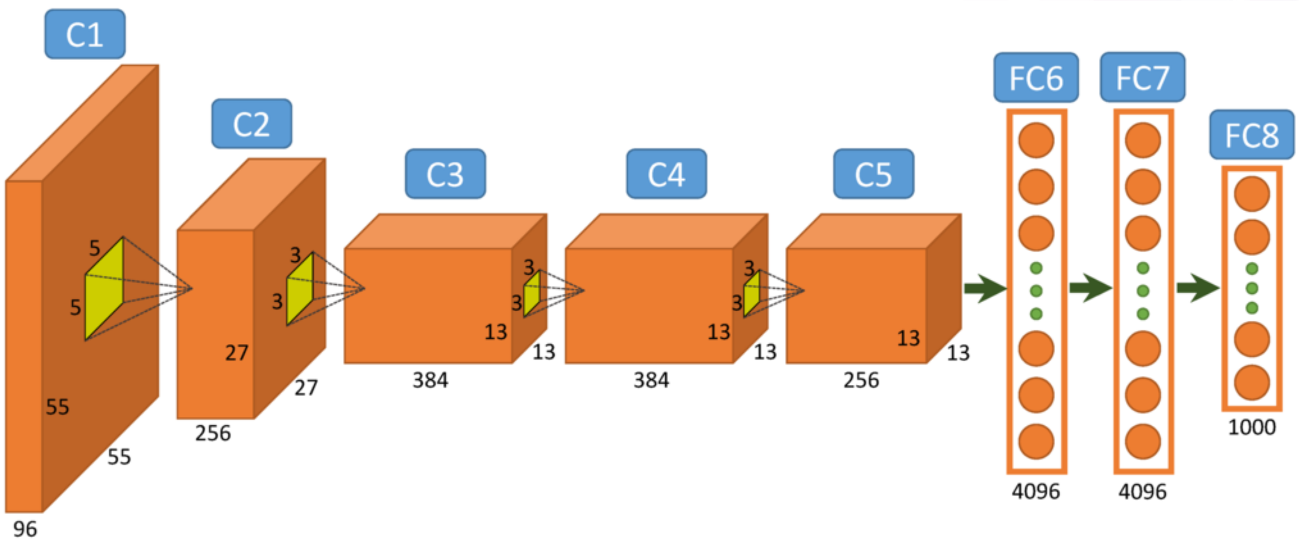
- New methods for SSL and domain adaptation rarely compare against classic SSL algorithms
- Re-evaluate general-purpose bootstrapping approaches with NNs vs. state-of-the-art
- Classic tri-training (with some modifications) performs best
- Tri-training with neural networks is expensive
- Propose a more time- and space-efficient version, Multi-task tri-training



S. Ruder, B. Plank. Strong Baselines for Neural Semi-supervised Learning under Domain Shift. *ACL 2018.*

Transfer learning

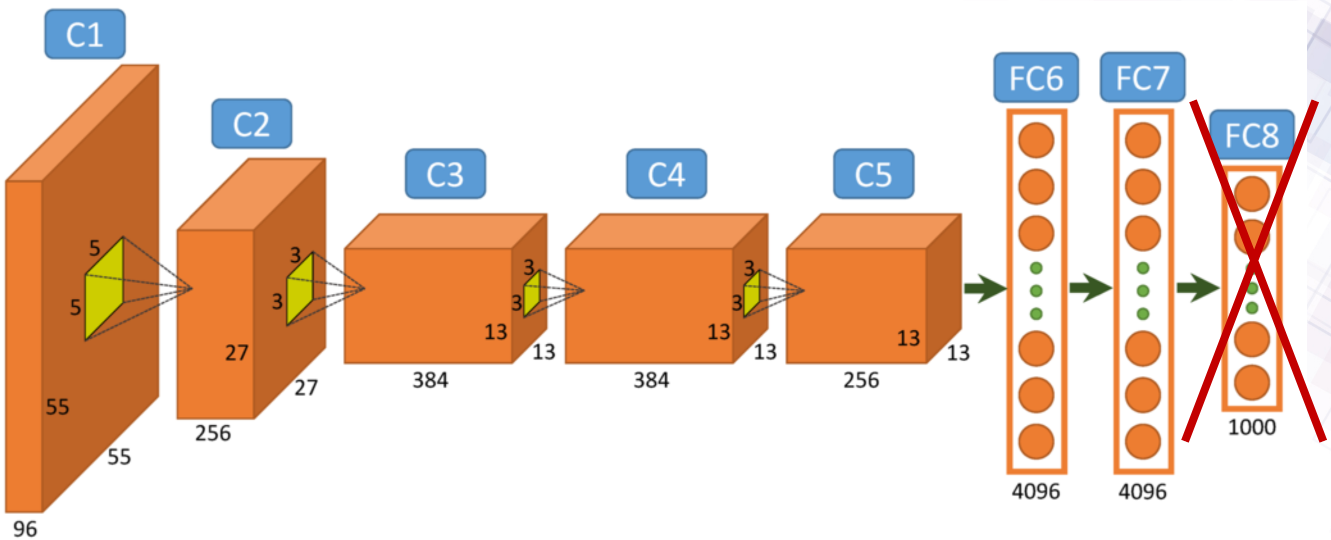
- Solve a problem by leveraging knowledge from a related problem
- Most useful when few labeled examples are available
- Most common use-case: Fine-tune a pre-trained ImageNet model



Sources: https://www.saagie.com/blog/object-detection-part1

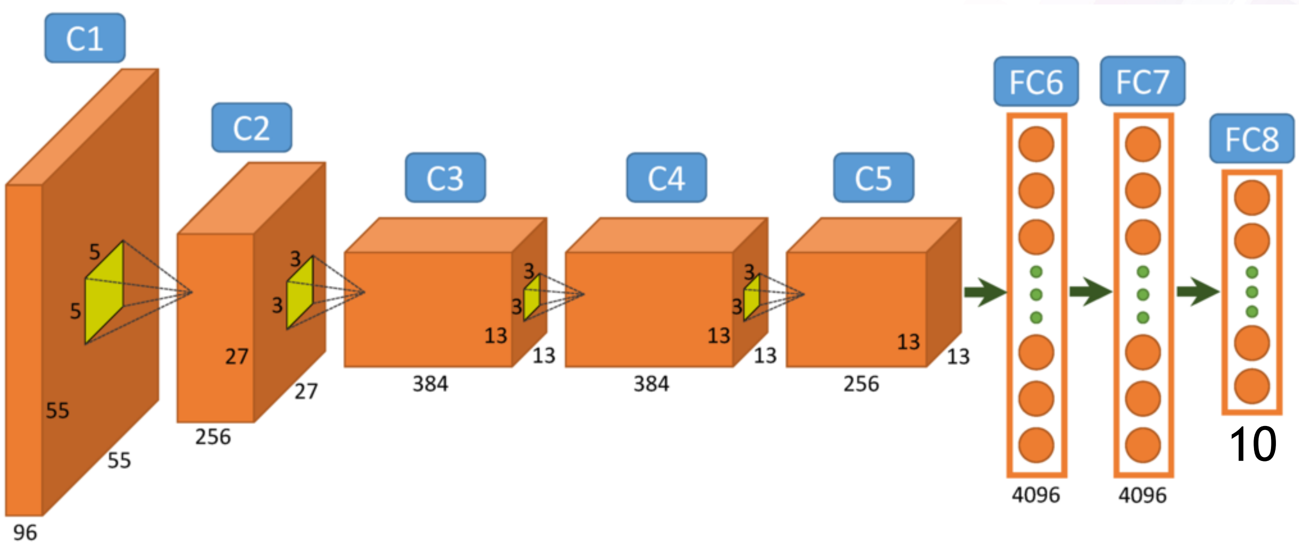
Transfer learning

1. Remove final classification layer



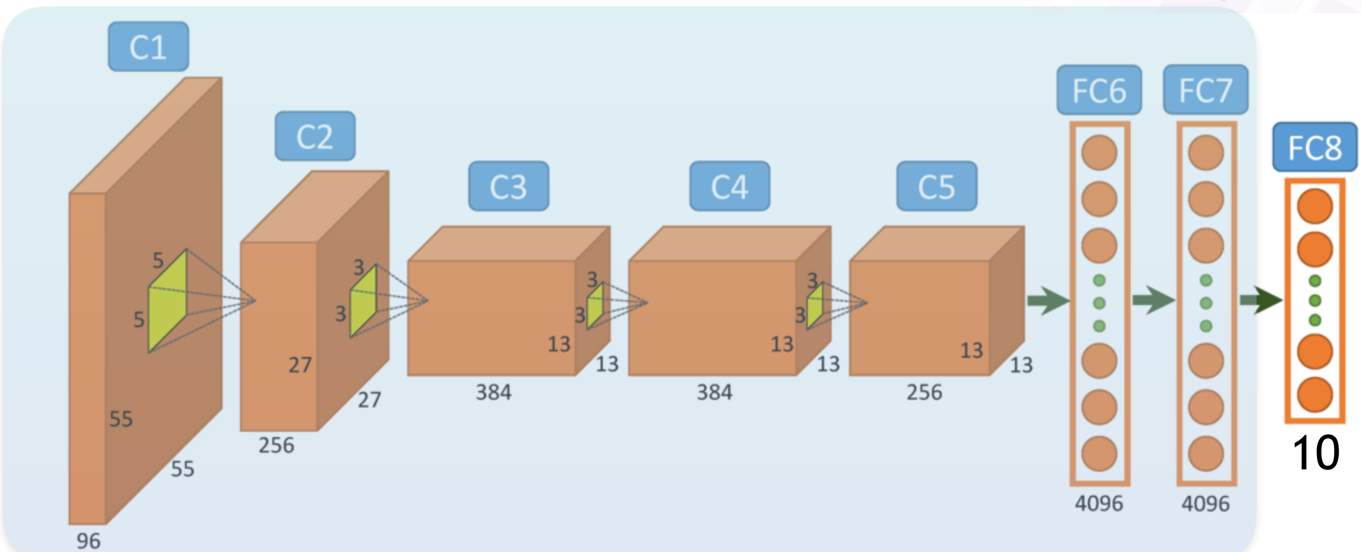
Transfer learning

1. Remove final classification layer
2. Replace with own classification layer



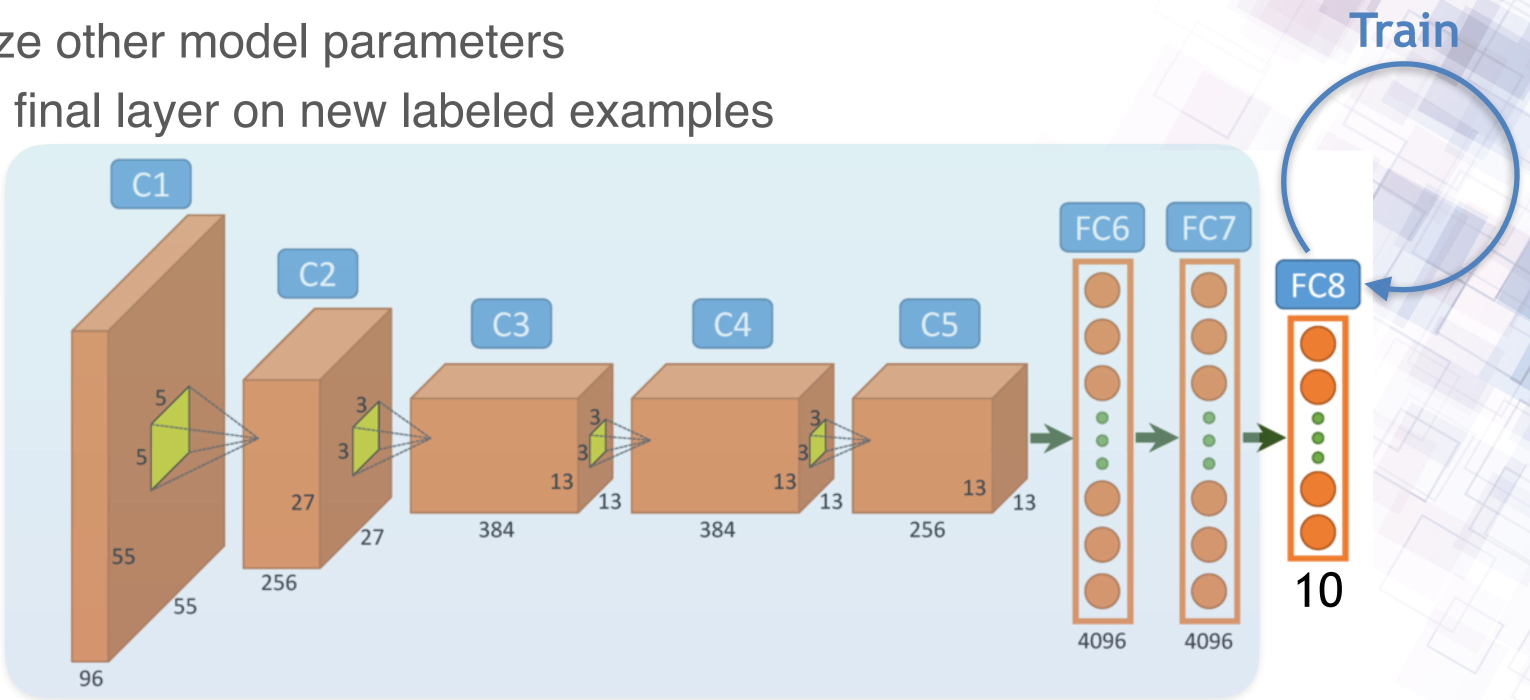
Transfer learning

1. Remove final classification layer
2. Replace with own classification layer
3. Freeze other model parameters



Transfer learning

1. Remove final classification layer
2. Replace with own classification layer
3. Freeze other model parameters
4. Train final layer on new labeled examples



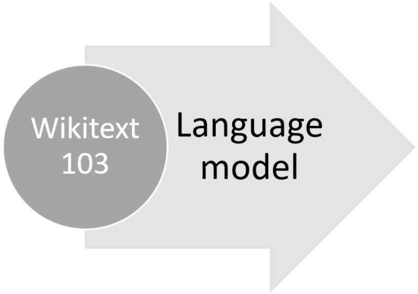
Transfer learning for NLP

- Transfer via pre-trained word embeddings is common
- Only affects first layer
- Recent methods still initialize most parameters randomly
- Universal Language Model Fine-tuning (ULMFit)

J. Howard*, S. Ruder*. Universal Language Model Fine-tuning for Text Classification. *ACL 2018.*

Transfer learning for NLP

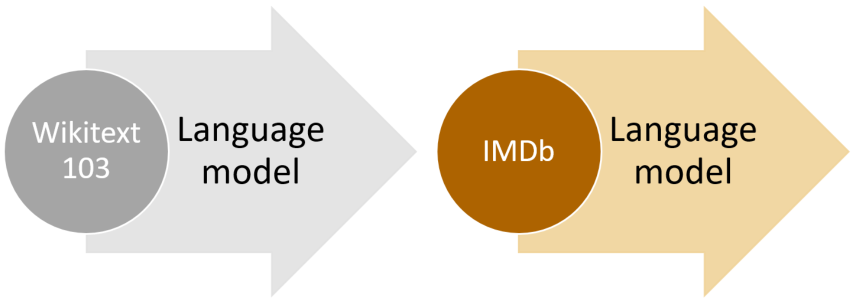
- Transfer via pre-trained word embeddings is common
- Only affects first layer
- Recent methods still initialize most parameters randomly
- Universal Language Model Fine-tuning (ULMFit)
    1. Pre-train a language model (LM) on a large corpus, e.g. Wikipedia



J. Howard*, S. Ruder*. [Universal Language Model Fine-tuning for Text Classification](). *ACL 2018.*

Transfer learning for NLP

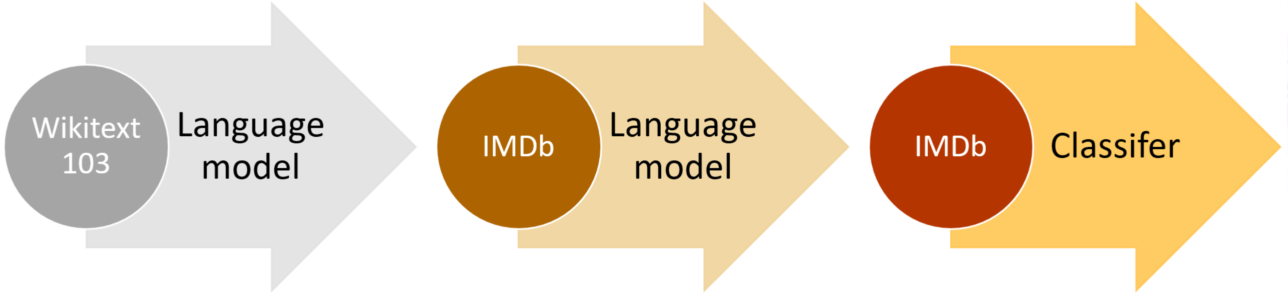
- Transfer via pre-trained word embeddings is common
- Only affects first layer
- Recent methods still initialize most parameters randomly
- Universal Language Model Fine-tuning (ULMFit)
    1. Pre-train a language model (LM) on a large corpus, e.g. Wikipedia
    2. Fine-tune the LM on unlabelled data with new techniques



J. Howard*, S. Ruder*. Universal Language Model Fine-tuning for Text Classification. *ACL 2018.*

Transfer learning for NLP

- Transfer via pre-trained word embeddings is common
- Only affects first layer
- Recent methods still initialize most parameters randomly
- Universal Language Model Fine-tuning (ULMFit)
    1. Pre-train a language model (LM) on a large corpus, e.g. Wikipedia
    2. Fine-tune the LM on unlabelled data with new techniques
    3. Fine-tune classifier on labeled examples



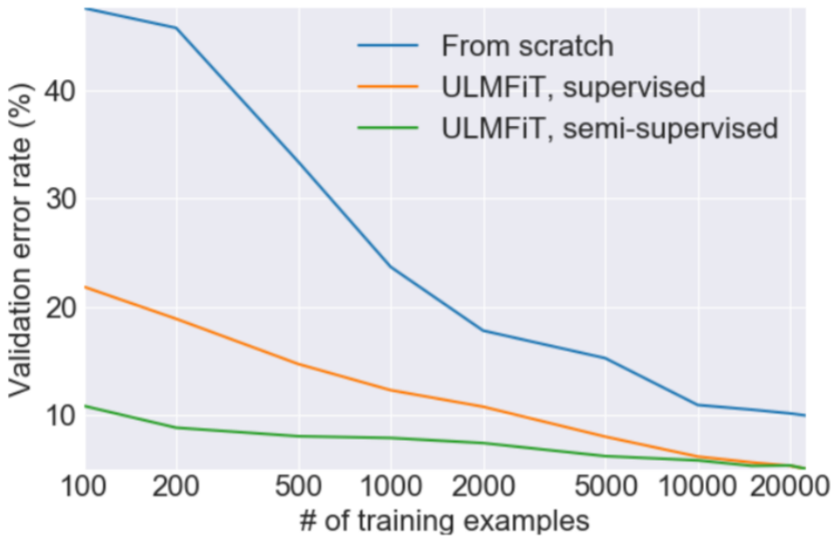
J. Howard*, S. Ruder*. [Universal Language Model Fine-tuning for Text Classification](). *ACL 2018.*

Transfer learning for NLP

- State-of-the-art on six widely used text classification datasets
- Enables training with orders of magnitude less data

	Model	Test		Model	Test
IMDb	CoVe (McCann et al., 2017)	8.2	TREC-6	CoVe (McCann et al., 2017)	4.2
	oh-LSTM (Johnson and Zhang, 2016)	5.9		TBCNN (Mou et al., 2015)	4.0
	Virtual (Miyato et al., 2016)	5.9		LSTM-CNN (Zhou et al., 2016)	3.9
	ULMFiT (ours)	**4.6**		ULMFiT (ours)	**3.6**

	AG	DBpedia	Yelp-bi	Yelp-full
Char-level CNN (Zhang et al., 2015)	9.51	1.55	4.88	37.95
CNN (Johnson and Zhang, 2016)	6.57	0.84	2.90	32.39
DPCNN (Johnson and Zhang, 2017)	6.87	0.88	2.64	30.58
ULMFiT (ours)	**5.01**	**0.80**	**2.16**	**29.98**



J. Howard*, S. Ruder*. Universal Language Model Fine-tuning for Text Classification. *ACL 2018.*

Many more frontiers

- Reasoning
- Bias
- Interpretability
- Adversarial examples
- Generalization

Thanks for your attention!

More info: http://ruder.io/

Twitter: @seb_ruder

Aylien: jobs@aylien.com — We are hiring!