Read time: 10 minutes

The major advancements in Deep Learning in 2018

Javier Wed, Dec 19, 2018 in **#MACHINE LEARNING**

MACHINE LEARNING DEEP LEARNING TRANSFER LEARNING

Deep learning has changed the entire landscape over the past few years. Every day, there are more applications that rely on deep learning techniques in fields as diverse as healthcare, finance, human resources, retail, earthquake detection, and self-driving cars. As for existing applications, the results have been steadily improving.

At the academic level, the field of **machine learning** has become so important that **a new scientific article is born every 20 minutes**.

In this article, I will present some of the main advances in deep learning for 2018. As with the **2017 version on deep learning advancements**, an exhaustive review is impossible. I'd simply like to share some of the accomplishments in the field that have most impressed me.

Language models: Google's BERT representation

In **Natural Language Processing** (NLP), a **language model** is a model that can estimate the probability distribution of a set of linguistic units, typically a sequence of words. These are interesting models since they can be built at little cost and have significantly improved several NLP tasks such as **machine translation**, **speech recognition**, and **parsing**.

Historically, one of the best-known approaches is based on **Markov models** and **n-grams**. With the emergence of deep learning, more powerful models generally based on **long short-term memory networks** (LSTM) appeared. Although highly effective, existing models are usually unidirectional, meaning that only the left (or right) context of a word ends up being considered.

Last October, the Google AI Language team published a paper that caused a stir in the community. **BERT (Bidirectional Encoder Representations from Transformers)** is a new bidirectional language model that has achieved state of the art results for 11 complex NLP tasks, including **sentiment analysis**, **question answering**, and **paraphrase detection**.

System	MNLI-(m/mm) 392k	QQP 363k	QNLI 108k	SST-2 67k	CoLA 8.5k	STS-B 5.7k	MRPC 3.5k	RTE 2.5k	**Average** -
Pre-OpenAI SOTA	80.6/80.1	66.1	82.3	93.2	35.0	81.0	86.0	61.7	74.0
BiLSTM+ELMo+Attn	76.4/76.1	64.8	79.9	90.4	36.0	73.3	84.9	56.8	71.0
OpenAI GPT	82.1/81.4	70.3	88.1	91.3	45.4	80.0	82.3	56.0	75.2
$BERT_{BASE}$	84.6/83.4	71.2	90.1	93.5	52.1	85.8	88.9	66.4	79.6
$BERT_{LARGE}$	**86.7/85.9**	**72.1**	**91.1**	**94.9**	**60.5**	**86.5**	**89.3**	**70.1**	**81.9**

Comparative results for the GLUE Benchmark.

Image source

The strategy for pre-training BERT differs from the traditional left-to-right or right-to-left options. The novelty consists of:

- masking some percentage of the input tokens at random, then predicting only those masked tokens; this keeps, in a multi-layered context, the words from indirectly "seeing themselves".
- building a binary classification task to predict if sentence B follows immediately after sentence A, which allows the model to determine the relationship between sentences, a phenomenon not directly captured by classical language modeling.

As for the implementation, Google AI open-sourced **the code for their paper**, which is based on TensorFlow. Some PyTorch implementations also exist, such as those by **Thomas Wolf** and **Junseong Kim**.

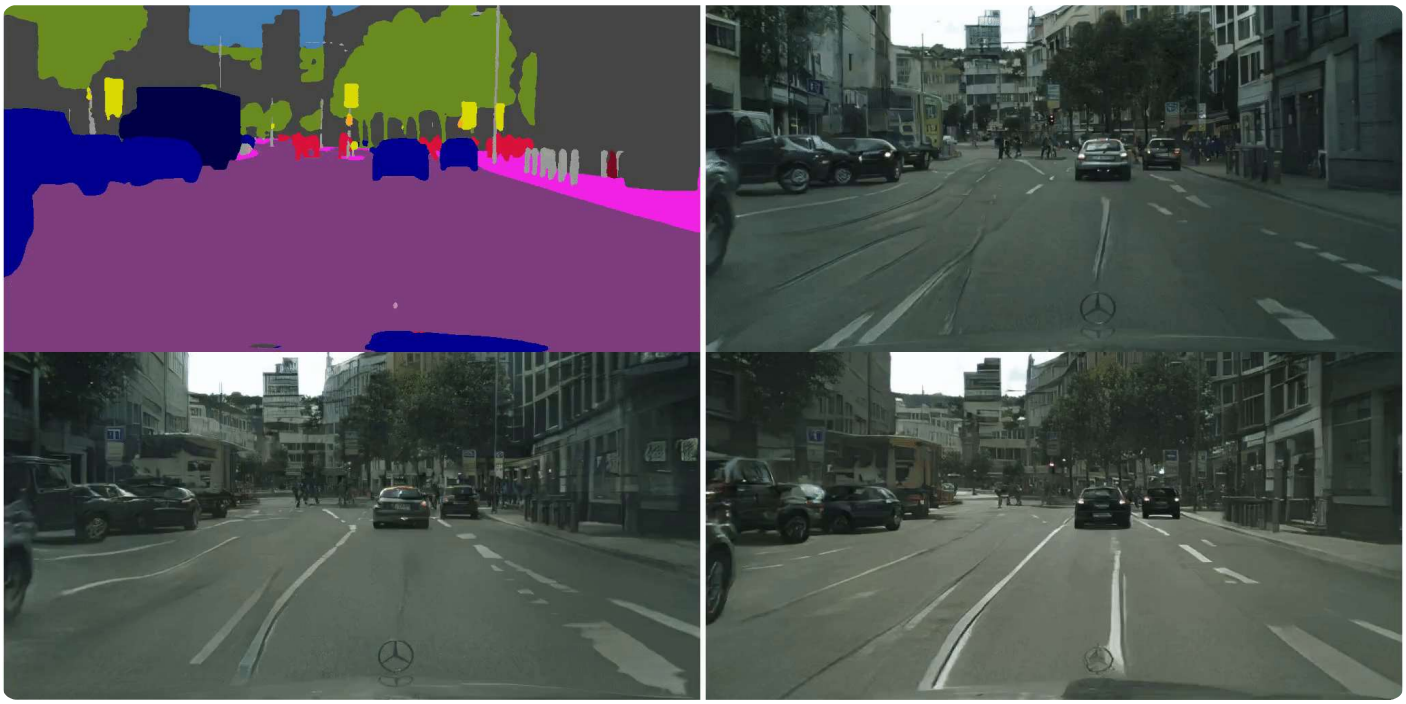
The impact on business applications is huge since this improvement affects various areas of NLP. This could lead to more accurate results in machine translation, chatbot behavior, automated email responses, and customer review analysis.

Video-to-video Synthesis

We are quite used to the interactive environments of simulators and video games typically created by **graphics engines**. While impressive, the classic approaches are costly in that the scene geometry, materials, lighting, and other parameters must be meticulously specified. A very good question is; whether it is possible to automatically build these environments using, for example, deep learning techniques.

In their **video-to-video synthesis paper**, researchers from NVIDIA address this problem. Basically, their goal is to come up with a mapping function between a source video and a photorealistic output video that precisely depicts the input content. The authors model it as a distribution matching problem, where the goal is to get the conditional distribution of the automatically created videos as close as possible to that of the actual videos. To achieve this, they build a model based on **generative adversarial networks** (GAN). The key idea, within the GAN framework, is that the *generator* tries to produce realistic synthetic data such that the *discriminator* cannot differentiate between real and synthesized data. They define a spatio-temporal learning objective, with the aim of achieving temporarily coherent videos.

The results are absolutely amazing, as can be seen in the video below.



Video-to-video synthesis.

Video source

The input video is in the top left quadrant. It is a segmentation map of a video of a street scene from the **Cityscapes dataset**. The authors compare their results (bottom right) with two baselines: pix2pixHD (top right) and COVST (bottom left).

This approach can be applied to many other tasks, like a sketch-to-video synthesis for face swapping. In the filmstrip linked to below, for each person we have an original video (left), an extracted sketch (bottom-middle), and a synthesized video.

Video synthesis for face swapping.

Image source

This approach can even be used to perform future video prediction; that is predicting the future video given a few observed frames with, again, very impressive results.

Since NVIDIA open-sourced **the vid2vid code** (based on PyTorch), you might enjoy experimenting with it.

Improving word embeddings

Last year, I wrote about **the importance of word embeddings in NLP** and the conviction that it was a research topic that was going to get more attention in the near future. Anyone who has utilized word embeddings knows that once the initial excitement of checking via compositionality (i.e. King - Man + Woman = Queen) has passed, there are several limitations in practice. Perhaps the most important ones are insensitivity to polysemy and inability to characterize the exact established relationship between words. Synonyms? Hyponyms? Hyperonyms? Another limitation concerns morphological relationships: word embeddings are commonly not able to determine that words such as driver and driving are morphologically related.

In the paper titled, **Deep contextualized word representations** (recognized as an **Outstanding paper at NAACL 2018**), researchers from the Allen Institute for Artificial Intelligence and the Paul G. Allen School of Computer Science & Engineering propose a new kind of deep contextualized word representation that simultaneously models complex characteristics of word use (e.g. syntax and semantics) as well as how these uses vary across linguistic contexts (i.e. polysemy).

The central theme of their proposal, called Embeddings from Language Models (ELMo), is to vectorize each word using the entire context in which it is used, or the entire sentence. To achieve this, the authors rely on a deep bidirectional language model (biLM), which is pre-trained on a very large body of text. Additionally, since representation is based on characters, the morphosyntactic relationships between words are captured. Consequently, the model behaves quite well when dealing with words that were not seen in training (i.e. out-of-vocabulary words).

TASK	PREVIOUS SOTA		OUR BASELINE	ELMO + BASELINE	INCREASE (ABSOLUTE/ RELATIVE)
SQuAD	Liu et al. (2017)	84.4	81.1	85.8	4.7 / 24.9%
SNLI	Chen et al. (2017)	88.6	88.0	88.7 ± 0.17	0.7 / 5.8%
SRL	He et al. (2017)	81.7	81.4	84.6	3.2 / 17.2%
Coref	Lee et al. (2017)	67.2	67.2	70.4	3.2 / 9.8%
NER	Peters et al. (2017)	91.93 ± 0.19	90.15	92.22 ± 0.10	2.06 / 21%
SST-5	McCann et al. (2017)	53.7	51.4	54.7 ± 0.5	3.3 / 6.8%

Comparative results of state-of-the-art models across six benchmark NLP tasks.

Image source

The authors show that by simply adding ELMo to existing state-of-the-art solutions, the outcomes improve considerably for difficult NLK tasks such as **textual entailment**, **coreference resolution**, and **question answering**. As in the case of Google's BERT representation, ELMo is a significant contribution to the field, and therefore promises to have a significant impact on business applications.

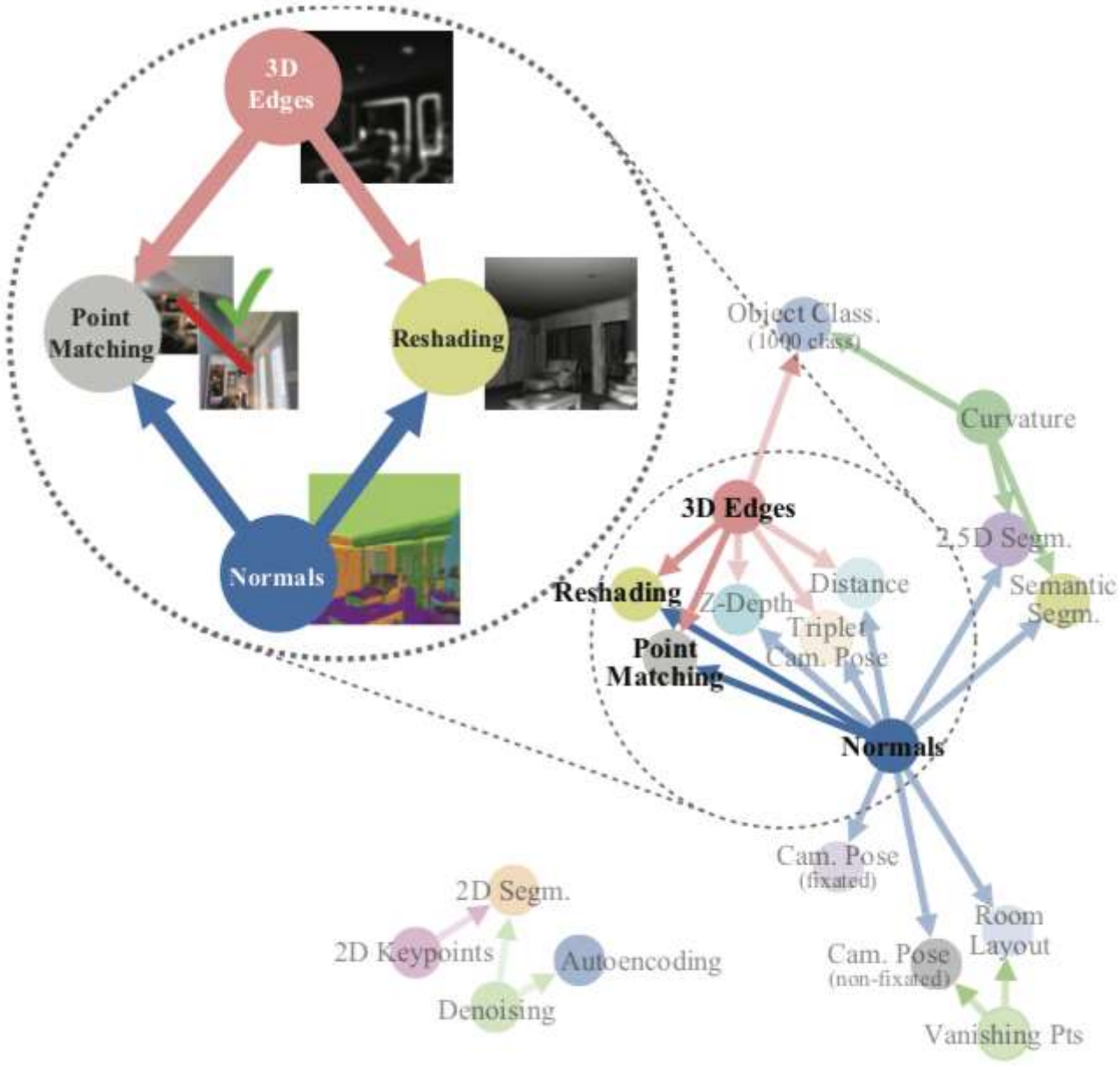
Modeling the structure of space of visual tasks

Are visual tasks related or not? This is the question addressed by researchers at Stanford and UC Berkeley in the paper titled, **Taskonomy: Disentangling Task Transfer Learning**, which won the **Best Paper Award** at **CVPR 2018**.

It can reasonably be argued that some kind of connection exists between certain visual tasks. For example, knowing surface normals can help in estimating the depth of an image. In such a scenario, **transfer learning** techniques – or the possibility to reuse supervised learning results – are very useful.

The authors propose a computational approach to modeling this structure by finding transfer-learning dependencies across 26 common visual tasks, including

object recognition, edge detection, and depth estimation. The output is a computational taxonomy map for task transfer learning.



A sample task structure discovered by the computational task taxonomy.

Image source

The figure above shows a sample task structure discovered by the computational taxonomy task. In this example, the approach informs us that if the learned features of a surface normal estimator and occlusion edge detector are combined, then models for reshading and point matching can be rapidly trained with little labeled data.

Reducing the demand for labeled data is one of the main concerns of this work. The authors demonstrate that the total number of labeled data points required for solving a set of 10 tasks can be reduced by roughly $^2/_3$ (compared with independent

training) while maintaining near identical performance. This is an important finding for real use cases, and therefore promises to have a significant impact on business applications.

Fine-tuning the universal language model for text classification

Deep learning models have contributed significantly to the field of NLP, yielding state-of-the-art results for some common tasks. However, models are usually trained from scratch, which requires large amounts of data and takes considerable time.

In their work, **Howard and Ruder** propose an inductive **transfer learning** approach dubbed Universal Language Model Fine-tuning (ULMFiT). The main idea is to fine tune pre-trained language models, in order to adapt them to specific NLP tasks. This is an astute approach that enables us to tackle specific tasks for which we do not have large amounts of 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**

Test error rates (%) on two text classification datasets (lower is better).

Image source

Their method outperforms state-of-the-art results for six **text classification** tasks, reducing the error rate by 18-24%. Regarding the volume of training data, the results are also pretty astounding: with only 100 labeled and 50K unlabeled samples, the approach achieves the same performance as models trained from scratch on 10K labeled samples.

Again, these results are evidence that transfer learning is a key concept in the field. You can take a look at their code and pretrained models **here**.

Final Thoughts

As was the case last year, 2018 saw a sustained increase in the use of deep learning techniques. In particular, this year was marked by a growing interest in transfer

learning techniques. From a strategic point of view, this is probably the best outcome of the year in my opinion, and I hope this trend continues in the near future.

Some other advances I do not explore in this post are equally remarkable. For instance, advancements in reinforcement learning such as the amazing **OpenAI Five bots**, capable of defeating professional players of **Dota 2**, deserve mention. So do **spherical CNN**, particularly efficient at analyzing spherical images, as well as **PatternNet and PatternAttribution**, two techniques that confront a major shortcoming of neural networks: the ability to explain deep networks.

The impact on business applications of all the above is massive, since they affect so many different areas of NLP and computer vision. We may observe improved results in the areas of machine translation, healthcare diagnostics, chatbot behavior, warehouse inventory management, automated email responses, facial recognition, and customer review analysis, just to name a few.

From a scientific point of view, I loved the **review on deep learning written by Gary Marcus**. He lucidly points out the limitations of current deep learning approaches and suggests that the field of AI would gain a considerable amount if deep learning methods were supplemented by insights from other disciplines and techniques, such as cognitive and developmental psychology, and symbol manipulation and hybrid modeling. Whether or not you agree with him, I think it's worth reading his paper.

I hope you enjoyed this year-in-review. Please feel free to comment on how these advancements struck you. Are there any additional ones from this year that I didn't mention here? Let us know!

> *"If you're interested in discussing how these advancements could impact your industry, we'd love to **chat with you**.*

Bibliography

- **Spherical CNNs**. Taco S. Cohen, Mario Geiger, Jonas Koehler, and Max Welling.
- **BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding**. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova.
- **Universal Language Model Fine-tuning for Text Classification**. Jeremy Howard and Sebastian Ruder.

- **Learning how to explain neural networks: PatternNet and PatternAttribution**. Pieter-Jan Kindermans, Kristof T. Schütt, Maximilian Alber, Klaus-Robert Müller, Dumitru Erhan, Been Kim, and Sven Dähne.
- **Deep Learning: A Critical Appraisal**. Gary Marcus.
- **Deep contextualized word representations**. Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer.
- **GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding**. Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R. Bowman.
- **Video-to-Video Synthesis**. Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Guilin Liu, Andrew Tao, Jan Kautz, and Bryan Catanzaro.
- **Taskonomy: Disentangling Task Transfer Learning**. Amir Zamir, Alexander Sax, William Shen, Leonidas Guibas, Jitendra Malik, and Silvio Savarese.

What do you think?

23 Responses

Upvote | Love | Surprised

2 Comments | Tryolabs Blog | Login

Recommend | Tweet | Share | Sort by Best

Join the discussion…

LOG IN WITH

OR SIGN UP WITH DISQUS

Name

Sharply Unclear • a month ago

A very enjoyable and informative article, thank you.

However, BERT is so... ...October 2018!
I see that Microsoft submitted something last week that significantly improves upon it (look at the https://gluebenchmark.com/l.... The table says they have taken a "Multi-task joint learning" approach, rather than BERT's "Bidirectional pre-trained deep transformer"

1 • Reply • Share ›

Amit Jindia • 8 days ago

Thanks for sharing.

• Reply • Share ›

Subscribe | Add Disqus to your siteAdd DisqusAdd | Disqus' Privacy PolicyPrivacy PolicyPrivacy

What to read next

How we built a stand-in robot for remote workers using IoT and computer vision

How Machine Learning is reshaping Price Optimization

My PyCon APAC 2018 experience in Singapore

Get in touch

Do you have a project in mind?

We'd love to e-meet you!

YOUR NAME

YOU@EMAIL.COM

COMPANY

TELL US ABOUT YOUR PROJECT

☐ *Subscribe to receive news and blog updates*

CONTACT US

ABOUT

About us
What we do
ML Consulting

OUR WORK

Clients
Products
Brochure

COMMUNITY

Blog
Open Source
Careers

CONTACT

UY Phone: (598) 2716 8997
hello@tryolabs.com

OFFICES

US: 156 2nd Street, SF.
UY: Rambla Gandhi 655/701, MVD.

SOCIAL