December 19, 2018 | Daniel DeMillard

NeurIPS 2018: AI Advancements, Insights, and 2019 Predictions

Estimated Reading Time: 12 minutes

What are the most recent advances in artificial intelligence? What new technologies can we expect to see in 2019? Will there be new regulations on AI? These are the questions that can be answered by keeping a pulse on the Neural Information Processing Conference (NeurIPS previously NIPS).

NeurIPS brings together some of the greatest minds in artificial intelligence and deep learning and with its popularity in recent years has become more difficult to get into than a Taylor Swift concert. The conference is primarily focused on deep learning. Deep learning is the process of using many layers of connected artificial neural networks to model highly dimensional data. It is responsible for breakthroughs in image classification, voice recognition, self-driving cars, facial recognition, and even your phone's autocorrect features. Since deep learning has become such a central part of recent AI advances, NeurIPS can be seen as a proxy for the larger AI ecosystem.

The conference is sponsored by over 100 companies and big names in AI like Google, Nvidia, Microsoft, Facebook, IBM, Amazon, and many more—all with booths and demonstrations showcasing recent advances. NeurIPS spans six days and varies in its location, with this year’s in Montreal. The content and sessions are mostly technical. Researchers spend the days deep diving into emerging topics, presenting academic papers, and demonstrating new technologies. Given there are three parallel tracks (neuroscience, machine learning theory, and applied machine learning), it’s impossible for a single researcher (human, that is) to see and experience everything. But here are the insights and perspective I gathered at the 2018 conference—as well as some exciting advancements to look for in 2019.

Table of Contents

AI Topics & Takeaways

This year saw many breakthroughs in AI including improvements in generative networks, unsupervised learning and few shot learning, meta-learning and automatic machine learning, reinforcement learning, and theoretical explanations of some of the deep learning wizardry.

Generative Models for Image, Text, & Speech

Generative models estimate an entire probability distribution so that new content can be generated such as images, text, or speech. Predictive models learn some classifier to predict the probability of an output label given inputs. For example the probability that an image is of Donald Trump given the raw pixels, pr(Trump|pixels).

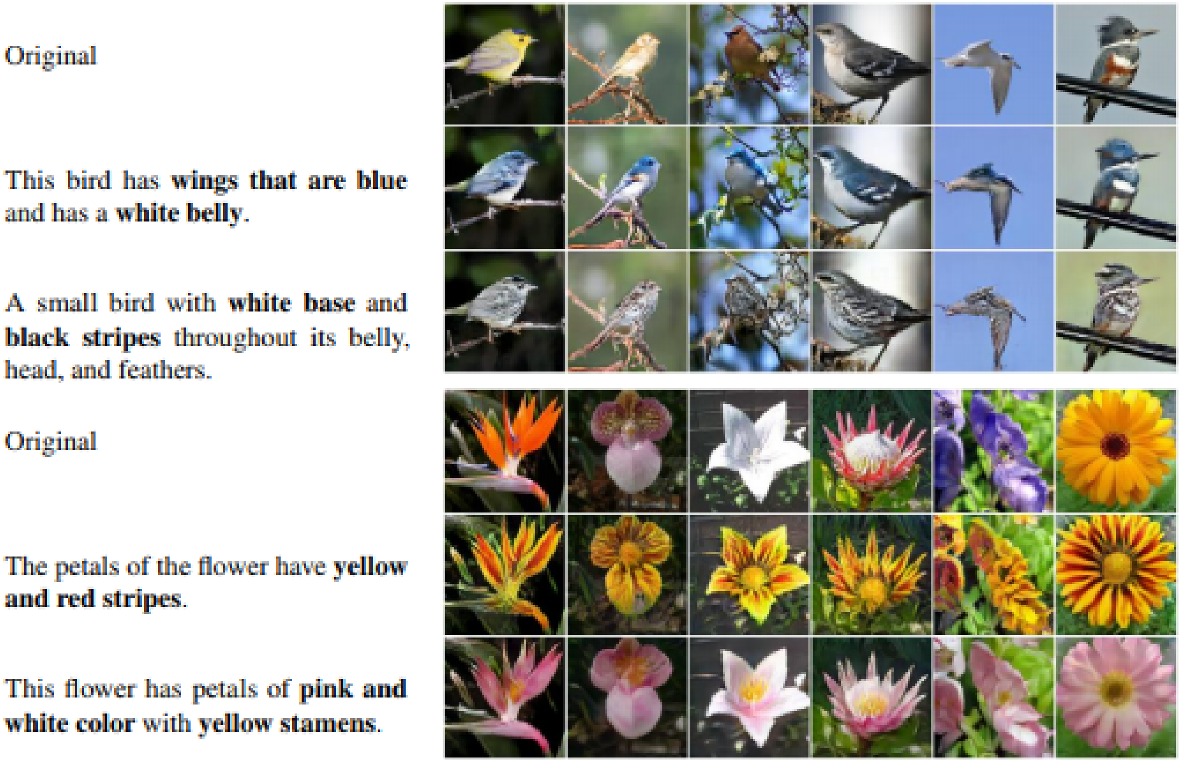
A generative model instead learns the entire probability distribution over the inputs for some desired output pr(pixels|Trump), allowing entirely new Donald Trump images to be generated upon request. Recurrent neural networks have been generating text for some time but only recently have advances in deep learning improved enough for images to be generated.

Models that can generate photorealistic images have evolved very quickly in the last few years. They can now generate some truly impressive results. Look at the pictures below, can you tell which ones are real and which ones were generated by a neural network?

Can you guess which of these images were generated by a neural network?

That was a trick question, because they were all images that have been algorithmically generated. These people do not actually exist.

There were many talks dedicated to advances in this space including being able to modify images with text-based descriptions alone. Photos of birds and flowers can now be instantly edited with a simple request.


Original
This bird has wings that are blue and has a white belly.
A small bird with white base and black stripes throughout its belly, head, and feathers.
Original
The petals of the flower have yellow and red stripes.
This flower has petals of pink and white color with yellow stamens.

Generative modeling allows us to modify Images with text-based inputs

This technology is cool and really exciting. However, generative models haven't really had their day in the sun because no one really knows what to use them for... yet. Nvidia may have began to bridge this gap in their demo at the conference. They showcased a virtual simulation of a user navigating a street scene that wasn't generated by a graphic designer or a video game designer but was entirely generated with images taken from an environment and then stitched together into a coherent 3D environment.

Prediction for 2019

A company will use generative models to automatically create digital walkthroughs for house buyers, museum tours, or even generate content for video games.

Learning with Fewer Labels (and None at All)

Unsupervised learning is the long-term dream of AI. Quite often, we have plenty of data but the data is unlabeled. This means that it is largely unusable beyond basic data exploration and anomaly detection. Unfortunately, this problem remains unsolved in 2018 and won't be solved any time soon. However, there have been many advances in at least being able to leverage your unlabeled data to improve your supervised learning process (labeled data).

A recent paper entitled, "Unsupervised Neural Machine Translation" showed that unsupervised learning can help in machine translation. There is an abundance of text in an given language. Think about all of the books, blogs, and correspondence that exists in any language you like. The problem is that you need parallel sentences to show a deep learning algorithm how to directly map words from a source text (say French) into a target language (say English). These data are far less abundant and hard to come by for specific language pairs, for instance, Swahili to Lao. This paper showed that by training a joint embedding space between an encoder and decoder on monolingual corpora—a weak form of unsupervised translation could be learned. Although, direct unsupervised translation isn't very good, the joint embeddings that were learned can greatly accelerate training of a supervised model, increasing accuracy with fewer mapped sentences and less training.

There were many talks and papers at the conference that continued this vein of obtaining state-of-the-art accuracy with fewer number of labelled samples. A few examples include applying transfer learning to clone voices using small amounts of audio recordings and object detection with far fewer examples for each class.

Prediction for 2019

There will be no fundamental breakthroughs in using unsupervised learning to directly learn novel information. However, unsupervised pre-training, transfer learning, and few shot learning will accelerate training in new domains and help bootstrap training time and the number of examples required to learn accurate classifiers.

Automatic Machine Learning and Meta-Learning

There is a trend away from “fiddling-by-hand” with all of the details involved in deep learning such as neural network architectures and hyperparameter tuning. Meta-learning attempts to “learn how to learn” and then transfer that learning ability to new domains that it hasn’t been exposed to. Although, meta-learning is still a work in progress, automatic machine learning is being applied effectively to quickly do hyperparameter search more efficiently than brute-force grid search and to automatically learn exotic connectionist architectures. There is even a handy plugin to the popular, scikit-learn, that has been recently released for automatic machine learning.

Prediction for 2019

Machine learning researchers will spend less of their time rerunning experiments with different hyperparameters and will use tools provided by automatic machine learning to run single experiments instead. This will speed up research time and enable researchers to explore more ideas.

The Theory of Deep Learning

Deep learning models have often been referred to as a “black-box”. This is because a deep learning model now contains millions and millions of parameters connected in complicated ways. It can be difficult even for the researchers who created the models to really understand every aspect of what is going on. Part of the allure of deep learning is that a researcher doesn’t have to build features by hand or even fully understand the domain that they are applying an algorithm to. In a very real way, the deep learning model knows more about the problem being solved than the researcher does.

This presents concerns for mission critical systems like self-driving cars, healthcare medical imaging, and financial transactions. If we don't know how the machinery is working, can we really call it safe? That is why progress in the theory behind deep learning is so important. This conference saw a number of talks and papers focussed on just that.

"On neuronal capacity" was a very interesting talk that mathematically showed how many bits could be stored in feedforward neural networks for varying sizes. This knowledge can be applied to know the degree of abstraction vs. memorization that is occuring in neural nets. For example, if you have 20,000 images of dimensions 256×256 color pixels, then you know that you have 20000X256X256X3 ~= 1 billion pieces of information. If this is comparable to the neural capacity of your network, then your model is simply too big and will overfit your data by simply memorizing it. It is unlikely that this model will generalize well to out-of-sample test sets. This information can be used by researchers when selecting the size of their networks.

Additionally, we saw "On the dimensionality of word embeddings" which gave a theoretical explanation of embedding sizes, a largely magic number in the deep learning community. These explanations can help in correctly selecting hyperparameters such as the size of the embedding without costly empirical searches.

READ: Data Piracy Challenges to DaaS Business Models

Finally, we even have a neuroscience explanation of how backpropagation may occur in the brain. It has long been criticized that deep learning neural networks don't really mimic connections in the brain due to the use of backpropagation, the method of communicating errors in a classifier back through the network's weights. However, a recent paper in neuroscience outlines a model that can explain how backpropagation-like-processes can occur in the brain. This may provide some liberation for AI researchers who believe the intelligence of their algorithms is akin to human intelligence.

Reinforcement Learning

Reinforcement learning (RL) is the method of solving sequential decision making problems that are common in game playing, financial markets, and robotics. Many of the talks regarding reinforcement learning this year focussed on one thing, data efficiency. There have been amazing breakthroughs applying what are called “policy networks” to play chess, Go, atari games, and more recently Dota. However, these methods require vast resources to simulate millions and millions of iterations of a game. Many of the papers that have been published by the likes of Google’s DeepMind, a leader in reinforcement learning, cannot be reproduced by general AI researchers because we don’t have access to the hundreds and hundreds of cloud GPU’s that DeepMind does. This is why there has been a focus to create RL algorithms that not only play games well, but are also reproducible, reusable, and robust. To do this, models have to be simplified and become more “data efficient”, meaning that fewer iterations of the simulation (less data) are required to end up with an accurate model.

Prediction for 2019

Alpha Go Zero, the DeepMind algorithm that taught itself to play Go, Chess, and Shogi at super-human level purely through self-play (no real-world examples), will be replicated by the research community and you will be able to train a version on a single consumer GPU. RL algorithms will start beating humans at more complex modern competitive gaming competitions like Starcraft and Dota.

Conversational Chatbots vs. Goal-Oriented Virtual Assistants

I was able to attend “The 2nd Conversational AI Workshop: today’s practice and tomorrow’s potential”. This workshop put a clear delineation on “conversation chatbots” vs. “goal-oriented virtual assistants”. Conversational chatbots involve social bots that can talk about a variety of domains and generally don’t have a goal beyond keeping a user engaged. An example of conversational chatbots are Microsoft’s twitterbot Tay and its successor Zo. Goal-oriented virtual assistants; however, focus on completing some pre-defined task. Examples of these assistants include Amazon’s Alexa, Apple’s Siri, Microsoft’s Cortana, as well as automated help-center bots.

Conversational chatbots have historically been much more difficult to build because the types of responses are varied and depend on the context or even the user they are communicating with. This workshop showcased recent advances in both of these categories.

Advancements follow a similar vein to some of the technologies above. Improvements in unsupervised learning, reinforcement learning, and memory networks have led to less repetition, better information retrieval, and better understanding of context.

Prediction for 2019

We will still not see conversational chatbots in 2019 (the type of chatbot that you can talk about anything with and sounds like a human). However, we will see more sophisticated virtual assistants that help you while shopping online, ordering at drive throughs, interacting with mall directories, and engaging for support needs with call centers. We will see more appliances outfitted with voice controls. You may be able to control your oven or garage door with your voice alone (without hooking it up to Alexa).

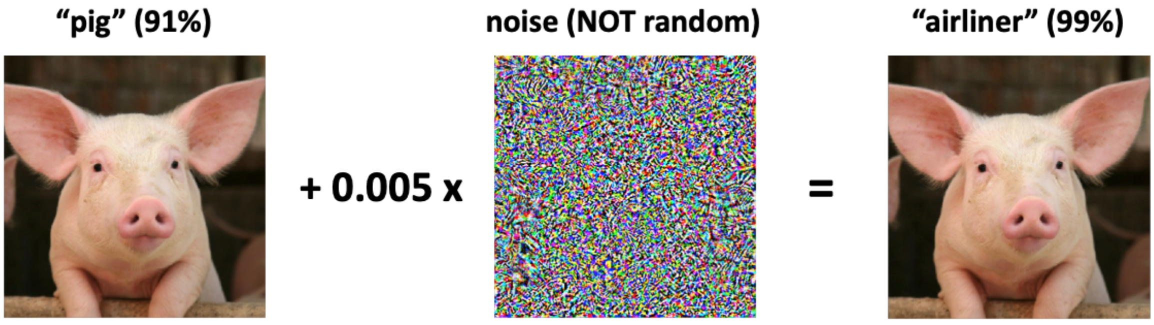
Concerns

This year also saw many concerns regarding AI, including security, safety, discriminatory bias, and lack of inclusion and diversity amongst AI researchers. These concerns may result in a regulatory backlash if companies are unable to self-regulate.

Adversarial Attacks on Image Classifiers

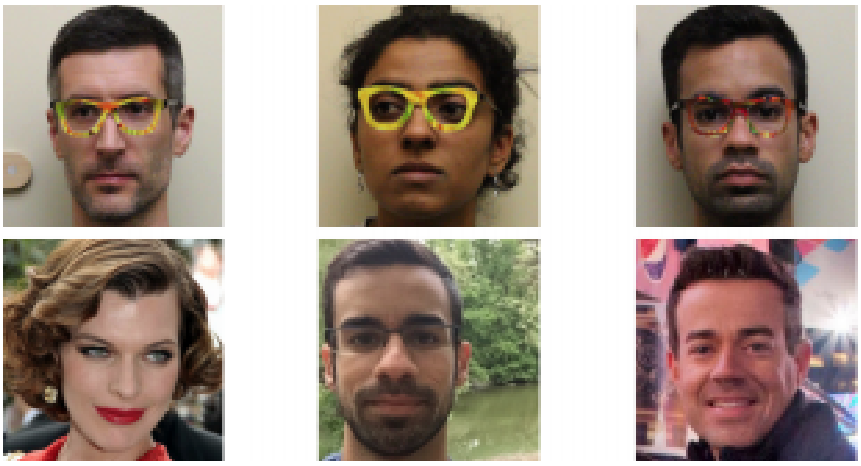
I counted nearly a dozen papers regarding "adversarial attacks" to deep learning models. For those unfamiliar, these attacks tend to focus on image classification algorithms and attempt to fool the classifier by adding small perturbations to the pixel values of the image.

By adding just the right amount of noise, an image of say, a "pig", can be transformed to fool the deep learning model to classify it as an "airliner".



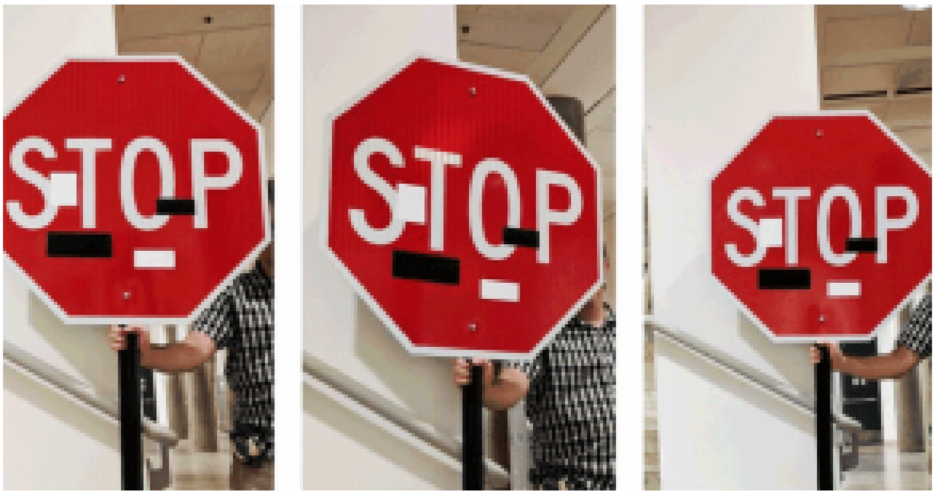
(Szegedy Zaremba Sutskever Bruna Erhan Goodfellow Fergus 2013)
(Biggio Corona Maiorca Nelson Srndic Laskov Giacinto Roli 2013)

At first glance, this may seem like an innocuous scientific oddity. However, consider that facial recognition is becoming a bigger part of security authorization e.g. smartphone logins. Researchers have shown that you can become someone else in the eyes of a classifier by simply modifying the pixels around an image. Of course a human can instantly tell that something is off, but for automated systems, fraud is as easy as putting on some funky digital glasses.



(Sharif Bhagavatula Bauer Reiter 2016)

Or consider what this means for self-driving cars. Many of the breakthroughs that are making self-driving cars feasible are deep learning image classifiers that can accurately read street signs, detect people, and interpret signals. It has been shown that these classifiers can be "thrown off" to miscategorize by simply applying black and white strips of tape strategically on the street sign.



(Eykholt Evtimov Fernandes Li Rahmati Xiao Prakash Kohno Song 2017)

Creating models that are impervious to these types of attacks is essential to ensuring people's safety and security. Fortunately, there were many approaches to build these self-defensive models such as regularization, data augmentation, vigilance about where your data comes from (see data poisoning), adding interpretability to your models, collecting more data, and building new types of detectors. The upshot of building models that cannot be attacked is that generalization is also improved. This means that the models will work on data that varies from your initial training set and overall accuracy is enhanced. Unfortunately, even though there has been a lot of headway in making models that cannot be attacked, not everyone will follow best practices when building AI models.

Prediction for 2019

Facial recognition will become ever more important for biometric security. For example you may log into your bank account using your face. A company this year will fail to apply the proper precautions to making their networks robust to adversarial attacks and someone will be the victim of fraudulent activity.

Inherent Biases and Discrimination in Machine Learning

Another backlash that AI has experienced this year is in models that exhibit racial and sexual discrimination. Models that have been trained on biased datasets, for example where minority populations are underrepresented or where the data provides an inaccurate snapshot, have been shown to make glaring mistakes and negatively target one group. This problem is twofold, the first problem is that interpretability and transparency are necessary in general when algorithms are making critical decisions such as who should be approved for a loan, who should be considered for a job interview, or even in determining criminal activity. The second problem is with a lack of fairness and accountability, imbalanced and incomplete datasets can end up singling out a group.

With a large database of faces and an imperfect classifier, it can be easy to find a match for a mugshot of someone that was completely unrelated to a crime. Facial recognition in America performs worse on African Americans and Asians than Caucasians due to heavily biased datasets that feature primarily white subjects. Likewise, this bias is reversed in Asian classifiers which have the opposite bias in their datasets.

AI researchers are not ethicists and should not be the end-all-be-all decision makers about what is fair or right or wrong. What AI researchers should do is to strive to add accountability, transparency and interpretability to their models. Unfortunately, this is one of the weakest points regarding black-box deep learning models. They are extremely powerful algorithms that can make truly amazing classifications but remain largely inscrutable.

There are some approaches to fixing this issue such as adding specific features for factor importance analysis, performing ablation studies, and explicitly testing for discriminatory bias e.g. A/B test on test sets of black and white faces. This remains a challenge however, and there is not an easy solution to this problem yet. If AI companies fail to self-police this issue, there is likely to be a regulatory backlash as individuals civil rights are violated.

Prediction for 2019

There will be an increase in the number of lawsuits that claim that machine learning models were discriminatory and violated anti-discrimination laws. Deep learning models will remain largely inscrutable but regulatory pressures, security concerns, and advances in theory will see much more time and funding for research to improve interpretability and transparency in deep learning models.

Wrap-Up & Conclusions

We saw many amazing breakthrough in 2018—and 2019 holds opportunities as well as a number of challenges. There will undoubtedly be many more surprises along the way but I hope that we can continue progressing technology in a socially responsible and conscientious way.

Happy Holidays and I look forward to bringing you more in the new year! See everyone at NeurIPS 2019!
