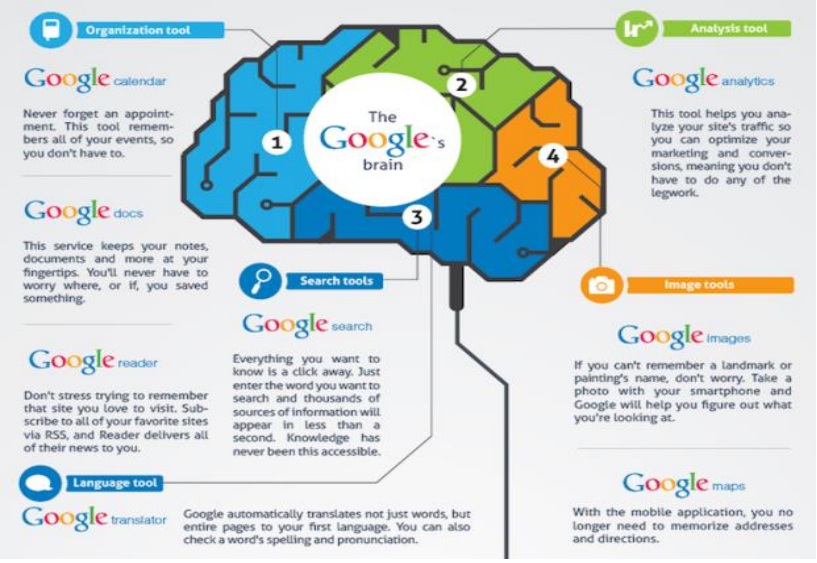
A Seminar Report
On
"GOOGLE BRAIN- AI Perspective"



Under the Supervision of:
Dr. Rachna Asthana, Professor, HOD,
Electronics Engineering Department,
HBTU, Kanpur

Submitted By:
SHIVAM GUPTA (S.R. No. 209/14)
SHASHWAT GAUTAM (S.R. No. 208/14)
Final B. Tech. Electronics Engineering,

HBTU Kanpur

CERTIFICATE

It is certified that Mr. SHIVAM GUPTA and Mr. SHASHWAT GAUTAM, student of Final B. Tech. Electronics Engineering H. B. T. U., Kanpur have been working on the seminar titled "**GOOGLE BRAIN- AI Perspective**" under my guidance and supervision. They have shown sincere efforts and keen interest during the preparation of this seminar report, and this work has not been submitted elsewhere for the award of any degree.

Dr. RACHNA ASTHANA
Professor
Electronics Engg. Department
HOD,HBTU, Kanpur
Seminar Supervisor

DECLARATION

We, **Shivam Gupta, Shashwat Gautam** studying in the final semester(8th) of Bachelor of Technology in Electronics Engineering at Harcourt Butler Technical University, Kanpur, hereby declare that this Seminar work entitled **"GOOGLE BRAIN- AI Perspective"** which is being submitted by us in the partial fulfillment for the award of the degree of Bachelor of Technology in Electronics Engineering at Harcourt Butler Technical University, Kanpur is an authentic record of our carried out during the academic year 2017-2018, under the guidance of **Dr. Rachna Asthana,** Professor, Department of Electronics Engineering at Harcourt Butler Technical University, Kanpur.

We further undertake that the matter embodied in the dissertation has not been submitted previously for the award of any degree or diploma by us to any other university or institution.

Place: Kanpur

Shivam Gupta (209/14)

Shashwat Gautam(208/14)

Acknowledgement

I would like to take the opportunity to express my gratitude to all those people who have helped in various ways in successful my dissertation on "**GOOGLE BRAIN- AI Perspective**", I would specially like to thanks my thesis adviser, Dr. RACHNA ASTHANA, H.O.D., Electronics & Communication Engineering Department at H.B.T.U., Kanpur, for his valuable guidance, advice and positive gesture during the preparation of the thesis. I convey my sincere thanks to all the faculty members and my classmates for their valuable support**.**

Shivam Gupta

Shashwat Gautam

Date: 4th April 2018
Place: KANPUR

B. Tech (Final Year)
Electronics Engineering

ABSTRACT

Google Brain is the biggest Artificial Intelligence research-Based Project going at Google Inc. since 2011. It includes the research in the sub-fields of Artificial Intelligence like Machine Learning, Deep Learning, Computer-Vison, Natural Language processing just to enhance the research going at Google. Google Brain has created such great stuffs which made the life of the human beings so easy. The Google Brain Team has recently reached significant breakthroughs for Google Translate, which is part of the Google Brain Project. Google brain has helped in the sentences prediction like Convert every sentence in a document to a though vector, in a way that similar thoughts are nearby. We can do basic natural reasoning by learning to predict next thought vector based on a sequence of previous thought vectors. Thereby, by reading every document on the web, computers might be able to reason like humans do by mimicking the thoughts expressed in content.

Google Brain as recently launched Google assistant in which Users primarily interact with the Google Assistant [10] through natural voice, though keyboard input is also supported. In the same nature and manner as Google Now, the Assistant is able to search the Internet, schedule events and alarms, adjust hardware settings on the user's device, and show information from the user's Google account with highest accuracy as compared to the other technologies like Cortana (Microsoft), Siri (Apple). Several Google services use TensorFlow in production, They have released it as an open-source project, and it has become widely used for machine learning and deep learning research.

TABLE OF CONTENTS

1. INTRODUCTION

Google Brain is a deep learning-Artificial Intelligence research Project going at Google since 2011. It combines open-ended machine learning research with system engineering and Google-scale computing resources. Google Brain's mission is to improve people's lives by making machines smarter. To do this, the team focuses on constructing models with high degrees of flexibility that are capable of learning their own features, and use data and computation efficiently [10].

As the Google Brain Team describes "This approach fits into the broader Deep Learning subfield of ML (Machine Learning) and ensures our work will ultimately make a difference for problems of practical importance. Furthermore, our expertise in systems complements this approach by allowing us to build tools to accelerate ML research and unlock its practical value for the world.

Google Brain team members set their own research agenda, with the team as a whole maintaining a portfolio of projects across different time horizons and levels of risk. As part of Google and Alphabet, the team has resources and access to projects impossible to find elsewhere. Our broad and fundamental research goals allow us to actively collaborate with, and contribute uniquely to, many other teams across Alphabet who deploy our cutting edge technology into products. The google brain team believes that openly disseminating research is critical to a healthy exchange of ideas, leading to rapid progress in the field. As such, they publish their research regularly at top academic conferences and release our tools, such as TensorFlow [2], as open source projects. In October 2016, Google Brain Team successfully developed an AI-devised encryption system. In February 2017, they announced a neural network image enhancement system that would fill in details in low resolution images — it could transform 8*8 resolution into 32*32.

2. HISTORY

The so-called "Google Brain" project began in 2011 as a part-time research collaboration between Google Fellow Jeff Dean, Google Researcher Greg Corrado, and Stanford University professor Andrew Ng. Ng had been interested in using deep learning techniques to crack the problem of artificial intelligence since 2006, and in 2011 began collaborating with Dean and Corrado to build a large-scale deep learning software system, DistBelief, on top of Google's cloud computing infrastructure. Google Brain started as a Google X project and became so successful that it was graduated back to Google: Astro Teller has said that Google Brain paid for the entire cost of Google X.

In June 2012, the New York Times reported that a cluster of 16,000 computers dedicated to mimicking some aspects of human brain activity had successfully trained itself to recognize a cat based on 10 million digital images taken from YouTube videos. The story was also covered by National Public Radio and SmartPlanet.

In March 2013, Google hired Geoffrey Hinton, a leading researcher in the deep learning field, and acquired the company DNN Research Inc. headed by Hinton. Hinton said that he would be dividing his future time between his university research and his work at Google.

Google Brain was initially established by Google Fellow Jeff Dean and visiting Stanford professor Andrew Ng (Ng later left to lead the artificial intelligence group at Baidu). In 2014, the team includes Jeff Dean, Geoffrey Shields, Greco Rado, Quoc Le, Ilya Sutskever, Alex Kelly Forth Alex Krizhevsky, Samy Bengio and Vincent Vanhoucke. In 2017, team members include Anelia Angelova, Samy Bengio, Greg Corrado, George Dahl, Michael Isard, Anjuli Kannan, Hugo Larochelle, Quoc Le, Chris Olah, Vincent Vanhoucke, Vijay Vasudevan and Fernanda Viegas. Chris Lattner, who created Apple's new programming language Swift and then ran Tesla's autonomy team for six months joined Google Brain's team in August 2017.

Google Brain is based in Mountain View, California and has satellite groups in Cambridge, Massachusetts, London, Montreal, New York City, San Francisco, Toronto, and Zurich [10].

3. Recent Breakthroughs in Google Brain

3.1 Artificial-Intelligence (AI)-devised encryption system

In October 2016, the Google Brain ran an experiment concerning the encrypting of communications. In it, two sets of AI's devised their own cryptographic algorithms to protect their communications from another AI, which at the same time aimed at evolving its own system to crack the AI-generated encryption. The study proved to be successful, with the two initial AIs being able to learn and further develop their communications from scratch [10].

In this experiment, three AIs were created: Alice, Bob and Eve. The goal of the experiment was for Alice to send a message to Bob, which would decrypt it, while in the meantime Eve would try to intercept the message. In it, the AIs were not given specific instructions on how to encrypt their messages, they were solely given a loss function. The consequence was that during the experiment, if communications between Alice and Bob were not successful, with Bob misinterpreting Alice's message or Eve intercepting the communications, the following rounds would show an evolution in the cryptography so that Alice and Bob could communicate safely. Indeed, this study allowed for concluding that it is possible for AIs to devise their own encryption system without having any cryptographic algorithms prescribed beforehand, which would reveal a breakthrough for message encryption in the future.

3.2 Image Enhancement

In February 2017, Google Brain announced an image enhancement system using neural networks to fill in details in very low resolution pictures. The examples provided would transform pictures with an 8x8 resolution into 32x32 ones.

The software utilizes two different neural networks to generate the images. The first, called a “conditioning network,” maps the pixels of the low-resolution picture to a similar high-resolution one, lowering the resolution of the latter to 8×8 and trying to make a match. The second is a “prior network”, which analyses the pixelated image and tries to add details based on a large number of high resolution pictures. Then, upon upscaling of the original 8×8 picture, the system adds pixels based on its knowledge of what the picture should be. Lastly, the outputs from the two networks are combined to create the final image.

This represents a breakthrough in the enhancement of low resolution pictures. Despite the fact that the added details are not part of the real image, but only best guesses, the technology has

shown impressive results when facing real-world testing. Upon being shown the enhanced picture and the real one, humans were fooled 10% of the time in case of celebrity faces, and 28% in case of bedroom pictures. This compares to previous disappointing results from normal bicubic scaling, which did not fool any human.

3.3 Google Translate

The Google Brain Team has recently reached significant breakthroughs for Google Translate, which is part of the Google Brain Project. In September 2016, the team launched the new system, Google Neural Machine Translation (GNMT) [11], which is an end-to-end learning framework, able to learn from a large amount of examples. While its introduction has greatly increased the quality of Google Translate's translations for the pilot languages, it was very difficult to create such improvements for all of its 103 languages. Addressing this problem, the Google Brain Team was able to develop a Multilingual GNMT system, which extended the previous one by enabling translations between multiple languages. Furthermore, it allows for Zero-Shot Translations, which are translations between two languages that the system has never explicitly seen before. Recently, Google announced that Google Translate can now also translate without transcribing, using neural networks. This means that it is possible to translate speech in one language directly into text in another language, without first transcribing it to text. According to the Researchers at Google Brain, this intermediate step can be avoided using neural networks. In order for the system to learn this, they exposed it to many hours of Spanish audio together with the corresponding English text. The different layers of neural networks, replicating the human brain, were able to link the corresponding parts and subsequently manipulate the audio waveform until it was transformed to English text [11].

3.4 Robotics

Different from the traditional robotics, robotics searched by the Google Brain Team could automatically learn to acquire new skills by machine learning. In 2016, the Google Brain Team collaborated with researchers at Google X to demonstrate how robotics could use their experiences to teach themselves more efficiently. Robots made about 800,000 grasping attempts during research. Later in 2017, the team explored three approaches for learning new skills, i.e., through reinforcement learning, through their own interaction with objects, and

through human demonstration. To build on the goal of the Google Brain Team, they would continue making robots that are able to learn new tasks through learning and practice, as well as deal with complex tasks.

4. Research Areas of Google Brain

4.1 Computer Systems for Machine Learning

Key to the success of deep learning in the past few years is that we finally reached a point where we had interesting real-world datasets and enough computational resources to actually train large, powerful models on these datasets. The needs of new applications, such as training and inference for deep neural network models, often require interesting innovations in computer systems, at many levels of the stack. At the same time, the appearance of new, powerful hardware platforms is a great stimulus and enabler for computer systems research.

One key way to accelerate machine learning research is to have rapid turnaround time on machine learning experiments, and we have strived to build systems that enable this. Our group has built multiple generations of machine learning software platforms to enable research and production uses of our research, with a focus on the following characteristics:

- Flexibility: it should be easy to express state-of-the-art machine learning models, such as the ones that our colleagues are developing (e.g. RNNs, attention-based models, Neural Turing Machines, reinforcement learning models, etc.) [1].
- Scalability: turnaround time for research experiments on real-world, large-scale datasets should be measured in hours not weeks
- Portability: models expressed in the system should run on phones, desktops, and datacenters, using GPUs, CPUs, and even custom accelerator hardware
- Production readiness: it should be easy to move new research from idea to experiment to production
- Reproducibility: it should be easy to share and reproduce research results.

Their first system, DistBelief, described in a NIPS 2012 paper, did well on most of these, except for flexibility and external reproducibility, and was used by hundreds of teams within Google to deploy real-world deep neural networks systems across dozens of products. Our more recent system, TensorFlow, was designed based on our experience with DistBelief and improved its flexibility by generalizing the programming model to arbitrary dataflow graphs; it is now the basis of hundreds of research projects and production systems at Google. In November, 2015, we open-sourced TensorFlow (blog post) and there’s now a vibrant and growing set of Google

and non-Google contributors improving the core TensorFlow system on the TensorFlow GitHub repository. As we had hoped when we open-sourced TensorFlow, there's also a thriving community of TensorFlow users across the world, using it for research and for real-world deployments, suggesting new directions, and improving and extending it. TensorFlow was the most forked new repository on GitHub in 2015 (source: Donne Martin), despite only launching in November of that year.

We also have a close working relationship with Google's datacenter and hardware platforms teams, which has allowed us to have significant input on the design and deployment of machine configurations that work well for machine learning (e.g., clusters of machines that have many GPUs and significant cross-machine bandwidth), as well as the requirements for Google's Tensor Processing Unit (TPU), a custom ASIC designed explicitly with neural network computations in mind and offering an order of magnitude performance and performance-per-watt improvement over other solutions. The Tensor Processing Unit is used in production for many kinds of models, including those used in ranking documents for every search query, and the use of many TPUs was also a key aspect of the recent AlphaGo victory over Lee Sedol in Seoul, Korea, in March 2016 (see Google Cloud Platform blog: Google supercharges machine learning tasks with TPU custom chip, by Norm Jouppi, May, 2016).

This close collaboration ensures that we design, build and deploy the right computational platforms for machine learning, allowing us to make our researchers more productive, as well as to enable product teams within Google to use machine learning in ambitious ways.

4.1.1 TensorFlow: A System for Large-Scale Machine Learning

TensorFlow [2] is a machine learning system that operates at large scale and in heterogeneous environments. TensorFlow uses dataflow graphs to represent computation, shared state, and the operations that mutate that state. It maps the nodes of a dataflow graph across many machines in a cluster, and within a machine across multiple computational devices, including multicore CPUs, generalpurpose GPUs, and custom-designed ASICs known as Tensor Processing Units (TPUs). This architecture gives flexibility to the application developer: whereas in previous "parameter server" designs the management of shared state is built into the system, TensorFlow enables developers to experiment with novel optimizations and training algorithms.

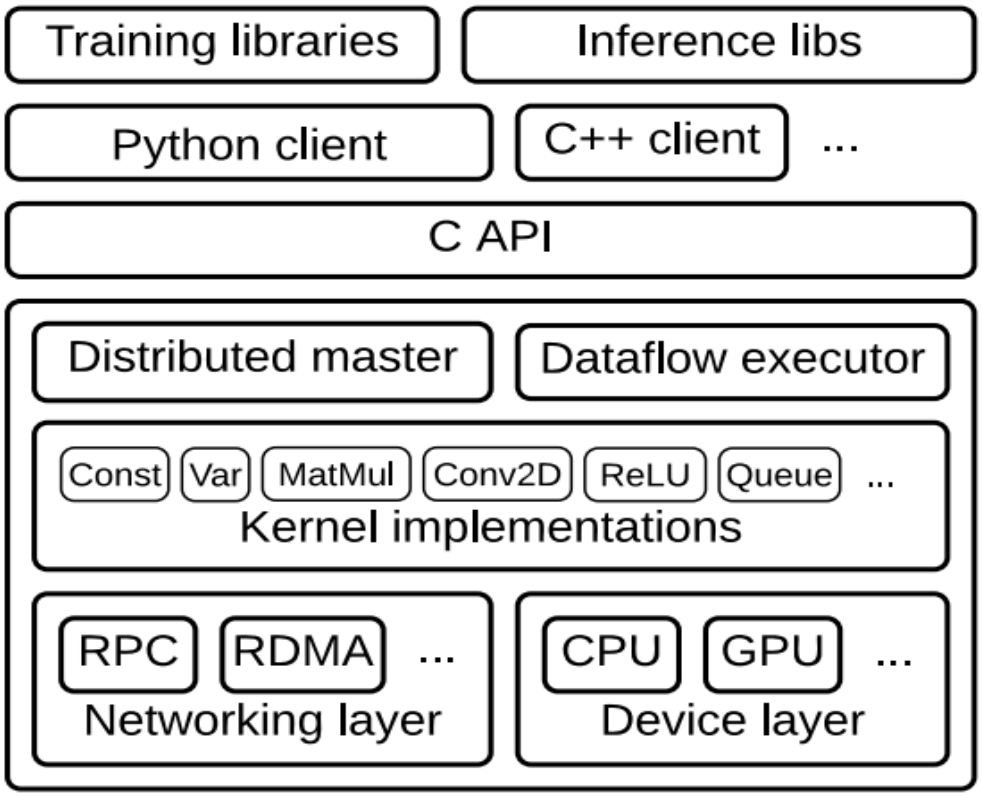


Figure 1: The layered TensorFlow architecture [2].

4.2 Music and Art Generation

DeepDream [1] is a method we developed to visualize what deep neural net models were learning. The resulting images were visually fascinating, and suggested neural nets could be used for artistic purposes. Inspired by DeepDream, the Google Brain team launched the Magenta project, which directly addresses the question, “Can machines be creative?"
Magenta encompasses two goals. It’s first a research project to advance the state-of-the art in music, video, image and text generation. So much has been done with machine learning to understand content— for example speech recognition and translation; Magenta explores content generation and creativity. Second, Magenta is an attempt to build a community of artists, coders and machine learning researchers around machine learning tools for creation, via open source. Research areas in Brain surrounding creativity include generative models for media, learning in domains where we lack a clear measure of success (related to unsupervised learning), and reinforcement learning for real-time interaction with artists and musicians.

Magenta [2] is a research project exploring the role of machine learning in the process of creating art and music. Primarily this involves developing new deep learning and reinforcement learning algorithms for generating songs, images, drawings, and other materials. But it's also an exploration in building smart tools and interfaces that allow artists and musicians to extend (not replace!) their processes using these models. Magenta was started by some researchers and engineers from the Google Brain team but many others have contributed significantly to the project. We use TensorFlow and release our models and tools in open source on our GitHub.

Google Creative Lab just released A.I. Duet, an interactive experiment which lets you play a music duet with the computer. You no longer need code or special equipment to play along with a Magenta music generation model. Just point your browser at A.I. Duetand use your laptop keyboard or a MIDI keyboard to make some music. You can learn more by reading Alex Chen's Google Blog post. A.I. Duet is a really fun way to interact with a Magenta music model. As AI Duet is open source, it can also grow into a powerful tool for machine learning research.

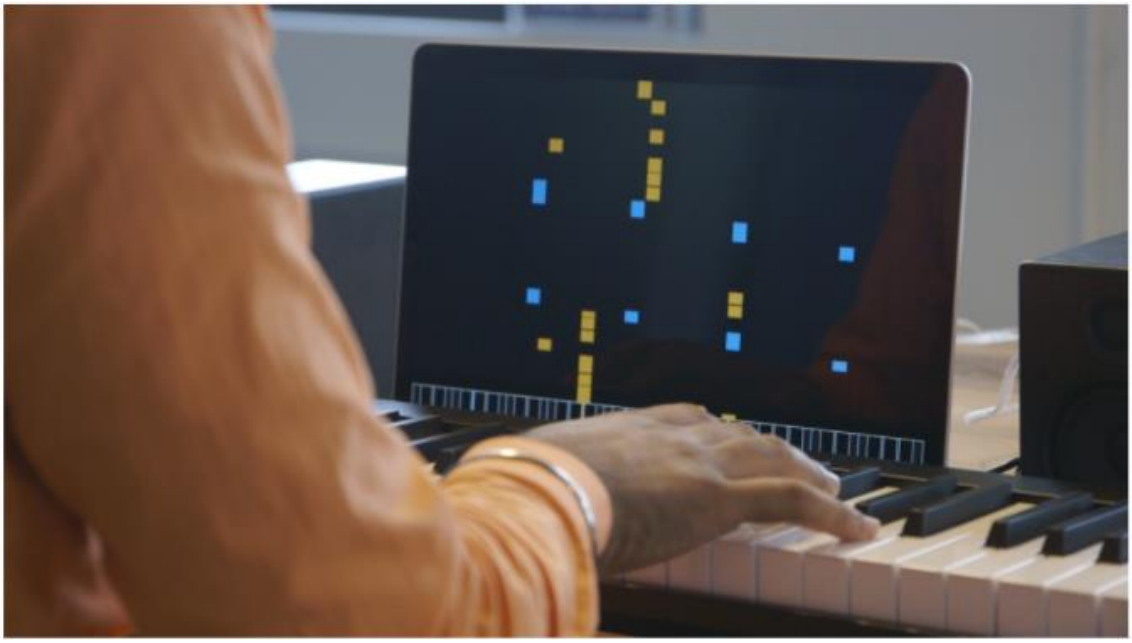


Figure 2: AI Duet [2]

4.3 Genomics

The genomics team in Google Brain focuses on ways that deep learning can transform genome sciences, with a goal of enabling, creating and validating new capabilities and tools that will

empower researchers, accelerate discoveries, and ultimately improve people's lives. Our efforts fall into three broad areas:

(1) Extending TensorFlow to better support genomics data;

(2) Developing deep learning models for genomics problems;

(3) Releasing new tools and capabilities as open source software.

Our first major project is DeepVariant [2], a universal SNP and small indel variant caller created using deep neural networks. DeepVariant is a collaboration between Google and Verily Life Sciences, and is available as open source.

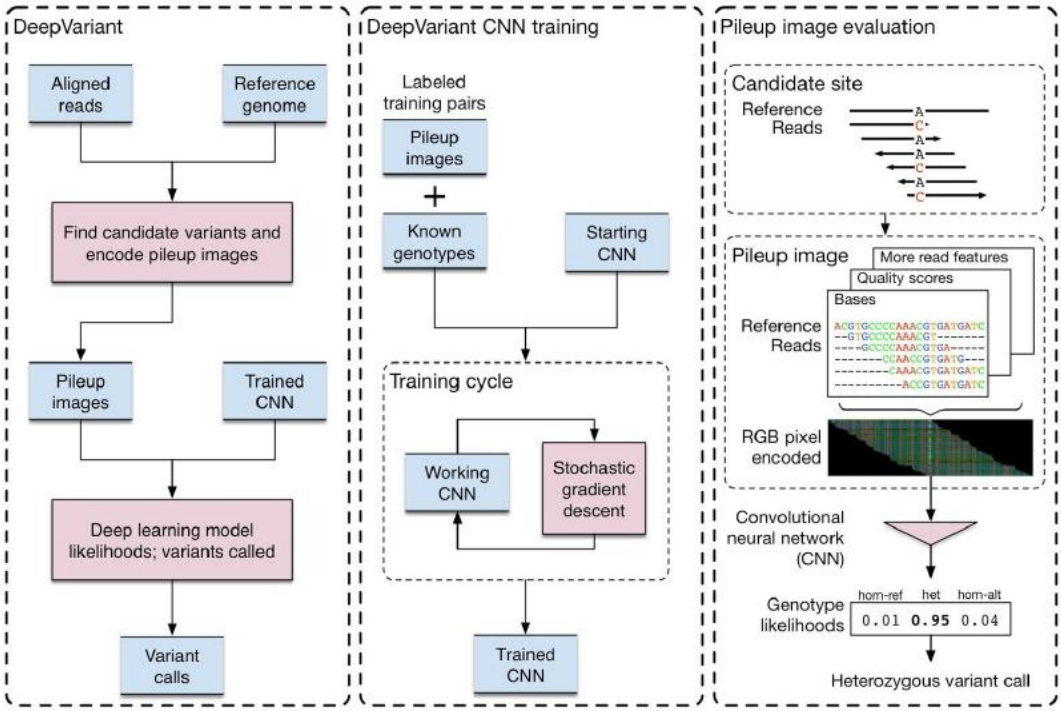


Figure 3: DeepVariant workflow overview [4].

4.4 Natural Language Processing

For Machine Intelligence to truly be useful, it should excel at tasks that humans are good at, such as natural language understanding. The Google Brain team's language understanding research focuses on developing learning algorithms that are capable of understanding language to enable machines to translate text, answer questions, summarize documents, or conversationally interact with humans.

Our research in this area started with neural language models and word vectors. Our work on word vectors, word2vec - which learns to map words to vectors, was opensourced in 2013 and has since gained widespread adoption in the research community and industry. Our work on language models has also made great strides (see this and this) in improving state-of-art prediction accuracies.

We also conduct fundamental research that leads to a series of advances using neural networks for end-to-end language (or language-related) tasks such as translation, parsing, speech recognition, image captioning and conversation modeling. The underlying technology is the seq2seq framework, which is also now used in SmartReply (and other products) at Google and is opensourced in TensorFlow [2].

Our recent research highlights are also in the areas of semi/unsupervised learning, multitask learning, learning to manipulate symbols and learning with augmented logic and arithmetic.

4.4.1 Listen, Attend and Spell

Listen, Attend and Spell (LAS) [11], a neural network that learns to transcribe speech utterances to characters. Unlike traditional DNN-HMM models [11], this model learns all the components of a speech recognizer jointly. Our system has two components: a listener and a speller. The listener is a pyramidal recurrent network encoder that accepts filter bank spectra as inputs. The speller is an attention-based recurrent network decoder that emits characters as outputs. The network produces character sequences without making any independence assumptions between the characters. This is the key improvement of LAS over previous end-to-end CTC models. On a subset of the Google voice search task, LAS achieves a word error rate (WER) of 14.1% without a dictionary or a language model, and 10.3% with language model rescoring over the top 32 beams. By comparison, the state-of-the-art CLDNN-HMM model achieves a WER of 8.0%.

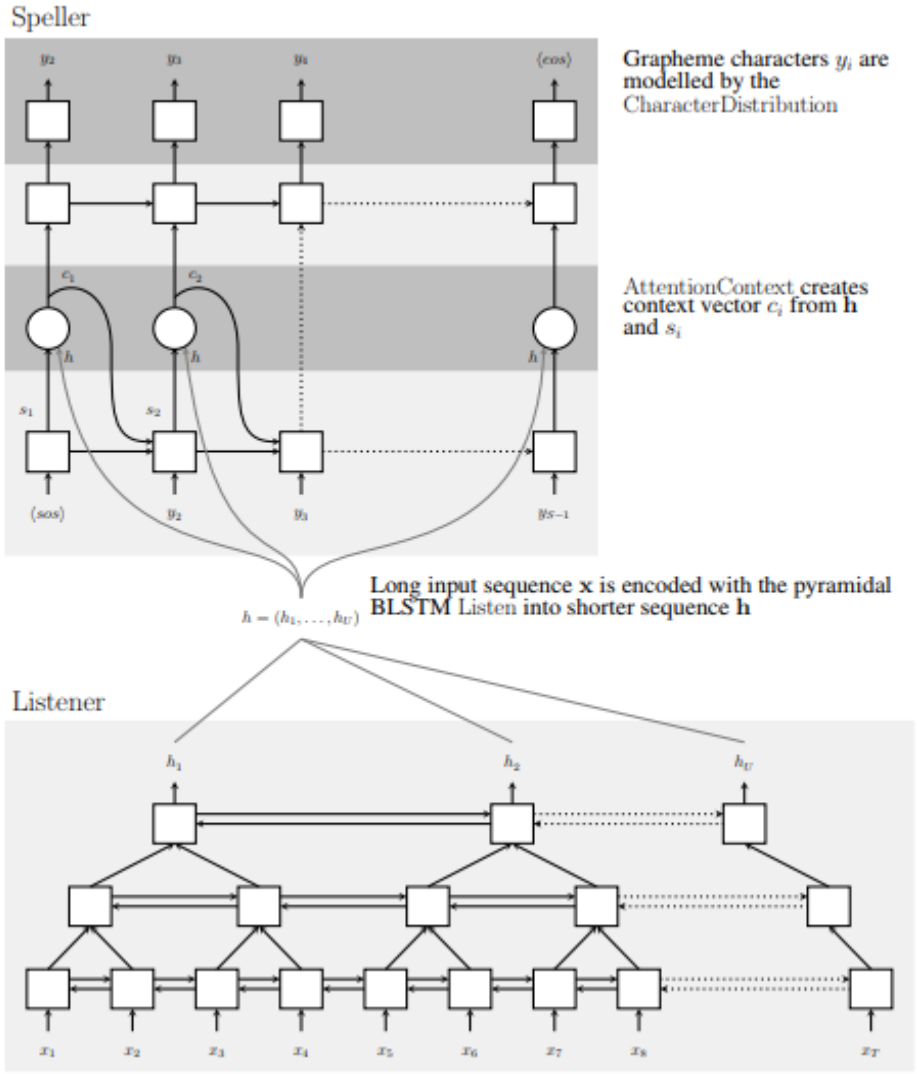


Figure 4: Listen, Attend and Spell (LAS) model: the listener is a pyramidal BLSTM encoding our input sequence x into high level features h, the speller is an attention-based decoder generating the y characters from h [11].

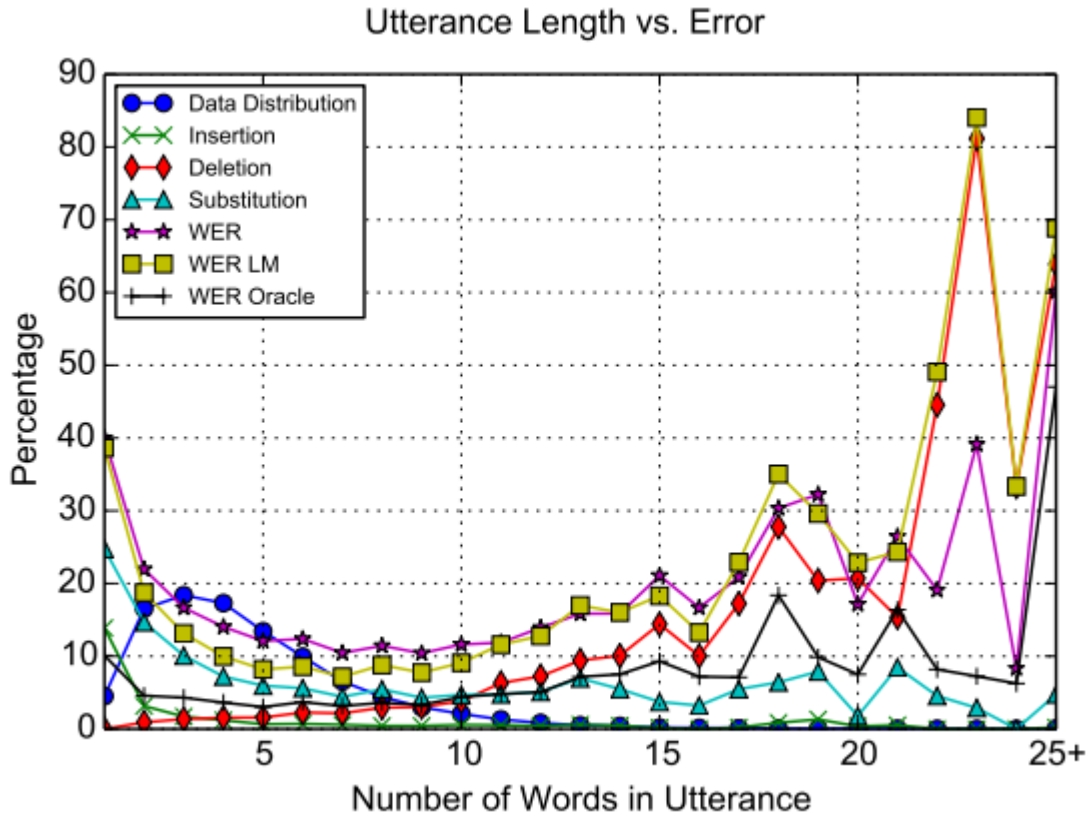


Figure 5: The correlation between error rates (insertion, deletion, substitution and WER) and the number of words in an utterance. The WER is reported without a dictionary or language model, with language model rescoring and the oracle WER for the clean Google voice search task. The data distribution with respect to the number of words in an utterance is overlaid in the figure. LAS performs poorly with short utterances despite an abundance of data. LAS also fails to generalize well on longer utterances when trained on a distribution of shorter utterances. Insertions and substitutions are the main sources of errors for short utterances, while deletions dominate the error for long utterances [11].

4.5 Healthcare

We think that AI is poised to transform medicine, delivering new, assistive technologies that will empower doctors to better serve their patients. Machine learning has dozens of possible application areas, but healthcare stands out as a remarkable opportunity to benefit people — and working closely with clinicians and medical providers, we're developing tools that we hope will dramatically improve the availability and accuracy of medical services [1].

Deep learning has already revolutionized the field of computer vision, making practical, in-your-pocket technologies out of what seemed like science fiction just a few years ago. If these

new computer vision systems can reach human-level accuracy in identifying dog breeds or cars, we asked ourselves, might those same systems be capable of learning to identify disease in medical images? Over the last few years, we've been working with doctors and clinicians to explore this question, and our research has shown that this is indeed possible — and not just in some far off future, but today. Two of the areas we're most excited about and where we've made the most progress in research to date are ophthalmology and digital pathology.

In the area of ophthalmology, we began exploring computer-aided diagnostic screening for a disease of the eye called diabetic retinopathy. Diabetic retinopathy is the fastest growing cause of preventable blindness globally. The condition is normally diagnosed by a highly trained doctor examining a retinal scan of the eye. If caught early, effective treatments are available, but if undetected, the disease progresses into irreversible blindness, and in much of the world, there simply are not enough doctors available to support the volume of screening required to protect the population.

Collaborating closely with doctors and international healthcare systems, we developed a state-of-the-art computer vision system for reading retinal fundus images for diabetic retinopathy and determined our algorithm's performance is on par with U.S. board-certified ophthalmologists. We've recently published some of our research in the Journal of the American Medical Association and summarized the highlights in a blog post. It's early days, and there's still work to do to bring the benefits of this research to patients, but ultimately, we hope to help real doctors and clinics expand global screening capacity to cover all at-risk individuals in the world.

In the field of digital pathology, we've focused our initial research on algorithms that might assist pathologists in detecting breast cancer in lymph node biopsies. Reviewing pathology slides is a complex task that requires years of training, expertise, and experience. Even with this extensive training, there can be substantial variability in the diagnoses given by different pathologists for the same patient — which isn't surprising given the massive amount of information that pathologists must review in order to make an accurate diagnosis.

To address these issues of limited time and diagnostic variability, we built an automated detection algorithm that can naturally complement pathologists' workflow. Our algorithm was designed to be highly sensitive to make it easier for pathologists to find even small instances of breast cancer metastasis in lymph node biopsies. We're encouraged by these results, but we think they're just the beginning. There's significant opportunity for AI to improve the accuracy

and availability of healthcare, and we hope that our research will serve as one of many demonstrations of that potential.

As part of this ongoing exploration, we're also partnering with healthcare providers to see how machine learning might predict healthcare events.

Metastasis detection [5] is currently performed by pathologists reviewing large expanses of biological tissues. This process is labor intensive and error-prone. We present a framework to automatically detect and localize tumors as small as 100 ×100 pixels in gigapixel microscopy images sized 100 , 000 ×100 , 000 pixels. Our method leverages a convolutional neural network (CNN) architecture and obtains state-of-the-art results on the Camelyon16 dataset in the challenging lesion-level tumor detection task. At 8 false positives per image, we detect 92 .4% of the tumors, relative to 82 .7% by the previous best automated approach. For comparison, a human pathologist attempting exhaustive search achieved 73.2% sensitivity. We achieve image-level AUC scores above 97% on both the Camelyon16 test set and an independent set of 110 slides. In addition, we discover that two slides in the Camelyon16 training set were erroneously labeled normal. Our approach could considerably reduce false negative rates in metastasis detection.

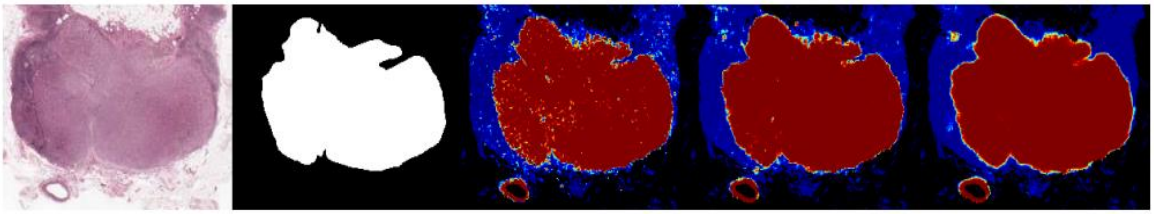


Figure 6: Left to right: sample image, ground truth (tumor in white), and heatmap outputs (40X-ensemble-of-3, 40X+20X, and 40X+10X). Heatmaps of 40X and 40X- ensemble-of-3 look identical. The red circular regions at the bottom left quadrant of the heatmaps are unannotated tumor. Some of the speckles are either out of focus patches on the image or non-tumor patches within a large tumor [6].

4.6 Perception

The goal of the Google Brain team's machine perception efforts is to improve a machine's ability to hear and see so that machines may naturally interact with humans. Historically, computers have been poor at perceiving visual and audio information that humans are able to process with ease. In the last few years, advances in deep learning have changed this equation

substantially and visual and audio recognition systems continue to approach human-level performance [1].

Our team within Google Brain has focused on building deep learning systems to advance the state of the art in these domains and apply these ideas to real products that affect the quality of user experience. Several notable advances which have stemmed from researchers within our team and the wider Google research community include:

- Advancing the state-of-the-art for image recognition through steady progress in designing and scaling convolutional neural network architectures [Krizhevsky et al, 2012, Szegedy et al, 2014]. This work has been recognized as the winner of the ImageNet ILSVRC Challenge in 2012 and 2014.

- Replacing highly handcrafted and hand-tuned speech systems with carefully built component models with deep, recurrent and convolutional neural network architectures that are increasingly being trained end-to-end [Jaitly et al, 2011; Sainath et al 2015, Chan et al, 2016]. Our contribution to the area of end-to-end models for speech recognition has been recognized with the ICASSP 2016 Speech and Language Processing Student Paper Award.

- Combining machine learning systems with different perceptual modalities to perform unique machine perception tasks, e.g. zero-shot learning or neural image captioning [Frome et al, 2012; Vinyals et al 2015]. The latter work has the distinction of winning the first CoCo Image Captioning Challenge in 2015.

Our long term goal is to make human perception a seamless component of future software systems including mobile devices, robotics and healthcare. While we have made great strides in the last few years, much work is yet to be done and we are excited about future directions.

4.6.1 Rethinking the Inception Architecture for Computer Vision

Convolutional networks are at the core of most state of-the-art computer vision solutions [8] for a wide variety of tasks. Since 2014 very deep convolutional networks started to become mainstream, yielding substantial gains in various benchmarks. Although increased model size and computational cost tend to translate to immediate quality gains for most tasks (as long as

enough labeled data is provided for training), computational efficiency and low parameter count are still enabling factors for various use cases such as mobile vision and big-data scenarios. Here we are exploring ways to scale up networks in ways that aim at utilizing the added computation as efficiently as possible by suitably factorized convolutions and aggressive regularization. We benchmark our methods on the ILSVRC 2012 classification challenge validation set demonstrate substantial gains over the state of the art: 21.2% top-1 and 5.6% top-5 error for single frame evaluation using a network with a computational cost of 5 billion multiply-adds per inference and with using less than 25 million parameters. With an ensemble of 4 models and multi-crop evaluation, we report 3.5% top-5 error and 17.3% top-1 error.

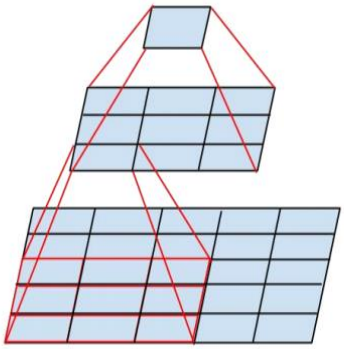


Figure 7: Mini-network replacing the 5 × 5 convolutions [8]

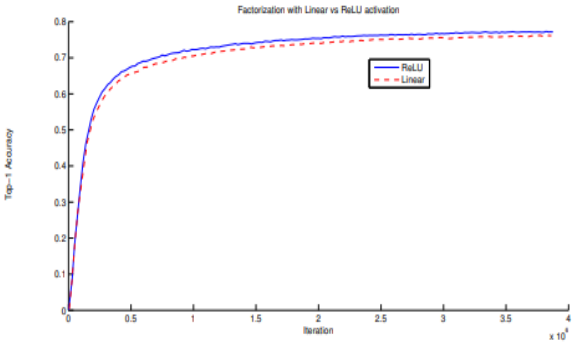


Figure 8: One of several control experiments between two Inception models, one of them uses factorization into linear + ReLU layers, the other uses two ReLU layers. After 3.86 million operations, the former settles at 76.2%, while the latter reaches 77.2% top-1 Accuracy on the validation set [8].

4.7 Machine Learning Algorithms and Techniques

The Google Brain team’s mission is "Make machines intelligent. Improve people’s lives." We combine open-ended machine learning research with world-class system engineering and Google-scale computing resources to realize this mission.

Our research started with the development of DistBelief, as a common platform to experiment with various unsupervisedand supervised learning algorithms for computer vision, speech recognition and other areas. In computer vision, our team members have played key roles in developing award-winning AlexNet and InceptionNet [8] models, and DeepDream. In Speech Recognition, our team members have pioneered the use of deep Learning for acoustic modeling. In natural language understanding, our team members have advanced word vectors, neural language modeling and pioneered sequence to sequence learning.

We currently conduct fundamental research to further advance key areas in machine intelligence and to create a better theoretical understanding of deep learning such as in Toward Deeper Understanding of Neural Networks: The Power of Initialization and a Dual View on Expressivity. Our recent research achievements also include: unsupervised learning, adversarial training, structured learning, long-term dependencies, knowledge distillation, general learning algorithms, understanding of learning algorithms, reinforcement learning, AI safety and TensorFlow.

4.7.1 Inceptionism: Going Deeper into Neural Networks

Artificial Neural Networks have surred remarkable recent progress in image classification and speech recognition. But even though these are very useful tools based on well-known mathematical methods, we actually understand surprisingly little of why certain models work and others don’t. So let’s take a look at some simple techniques for peeking inside these networks.

We train an artificial neural network by showing it millions of training examples and gradually adjusting the network parameters until it gives the classifications we want. The network typically consists of 10-30 stacked layers of artificial neurons [8]. Each image is fed into the input layer, which then talks to the next layer, until eventually the “output” layer is reached. The network’s “answer” comes from this final output layer.

One of the challenges of neural networks is understanding what exactly goes on at each layer. We know that after training, each layer progressively extracts higher and higher-level features of the image, until the final layer essentially makes a decision on what the image shows. For

example, the first layer maybe looks for edges or corners. Intermediate layers interpret the basic features to look for overall shapes or components, like a door or a leaf. The final few layers assemble those into complete interpretations—these neurons activate in response to very complex things such as entire buildings or trees.

One way to visualize what goes on is to turn the network upside down and ask it to enhance an input image in such a way as to elicit a particular interpretation. Say you want to know what sort of image would result in "Banana." Start with an image full of random noise, then gradually tweak the image towards what the neural net considers a banana itself, that doesn't work very well, but it does if we impose a prior constraint that the image should have similar statistics to natural images, such as neighboring pixels needing to be correlated.

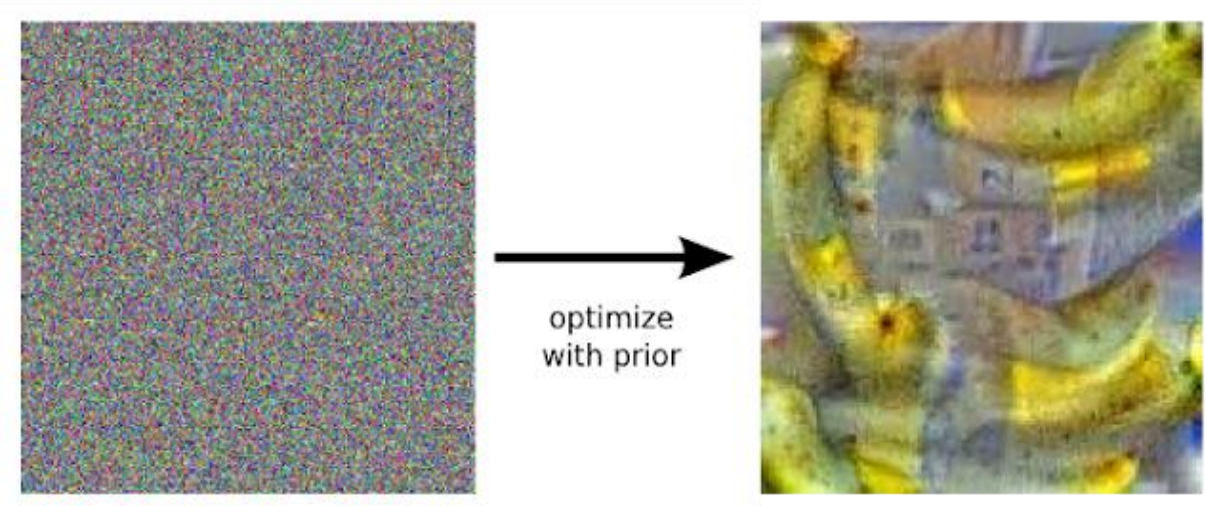


Figure 9: Classification of Banana from the Imagenets [8].

So here's one surprise: neural networks that were trained to discriminate between different kinds of images have quite a bit of the information needed to generate images too. Check out some more examples across different classes:

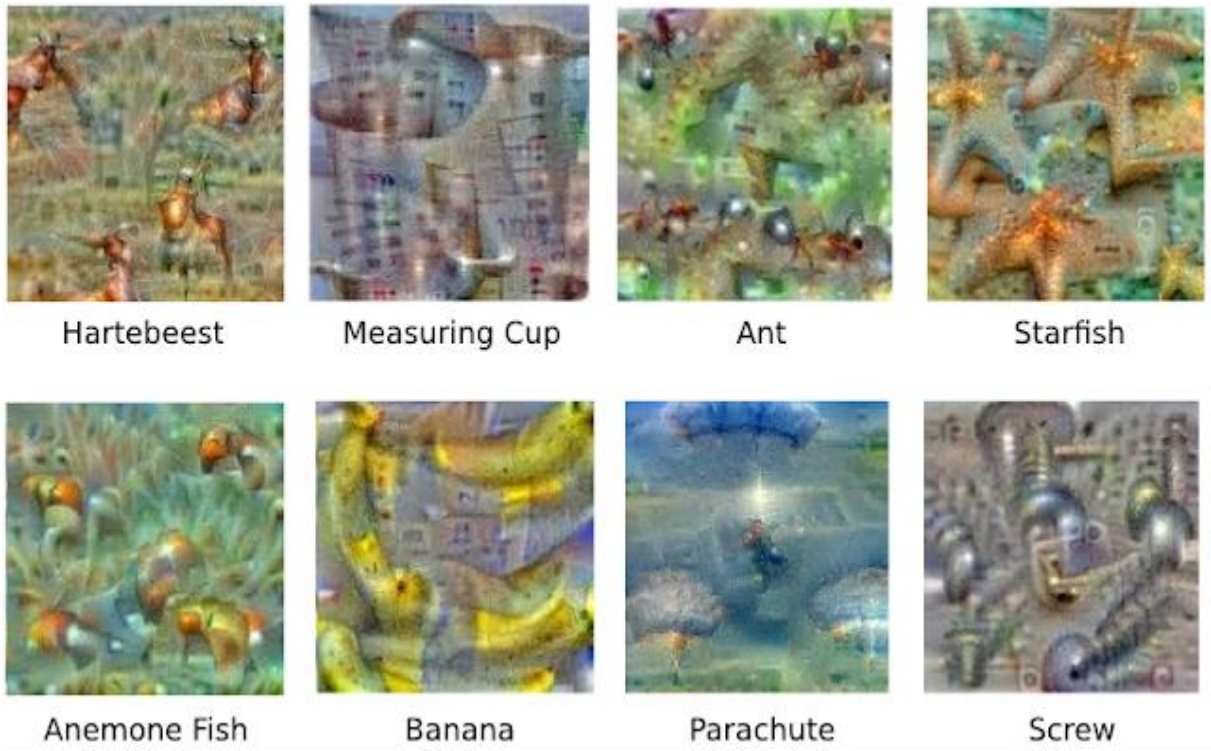


Figure 10: Some more Classification examples across different classes [8].

Why is this important? Well, we train networks by simply showing them many examples of what we want them to learn, hoping they extract the essence of the matter at hand (e.g., a fork needs a handle and 2-4 tines), and learn to ignore what doesn't matter (a fork can be any shape, size, color or orientation). But how do you check that the network has correctly learned the right features? It can help to visualize the network's representation of a fork.

Indeed, in some cases, this reveals that the neural net isn't quite looking for the thing we thought it was. For example, here's what one neural net we designed thought dumbbells looked like:

Figure 11: One neural net they designed thought dumbbells looked like [8].

There are dumbbells in there alright, but it seems no picture of a dumbbell is complete without a muscular weightlifter there to lift them. In this case, the network failed to completely distill the essence of a dumbbell. Maybe it's never been shown a dumbbell without an arm holding it. Visualization can help us correct these kinds of training mishaps.

Instead of exactly prescribing which feature we want the network to amplify, we can also let the network make that decision. In this case we simply feed the network an arbitrary image or photo and let the network analyze the picture. We then pick a layer and ask the network to enhance whatever it detected. Each layer of the network deals with features at a different level of abstraction, so the complexity of features we generate depends on which layer we choose to enhance. For example, lower layers tend to produce strokes or simple ornament-like patterns, because those layers are sensitive to basic features such as edges and their orientations.

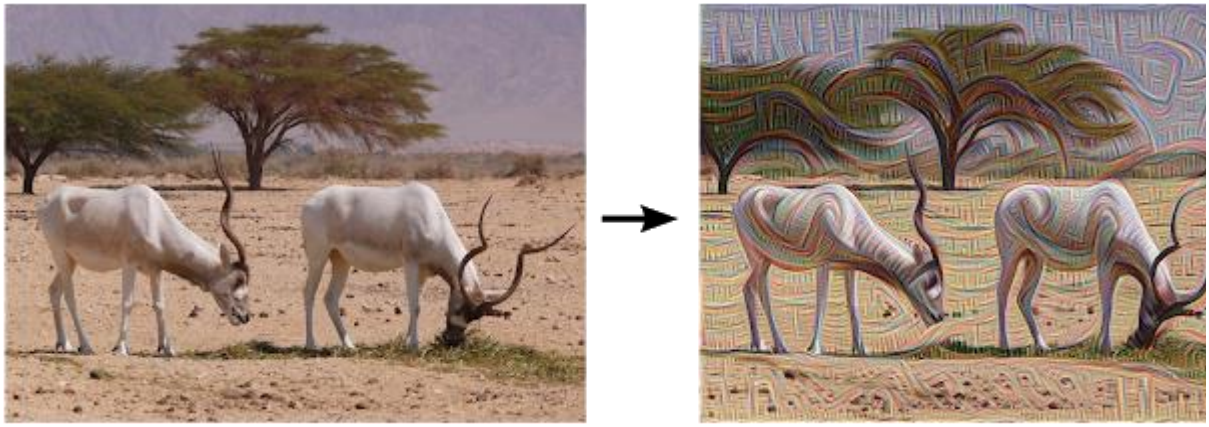


Figure 12: Left: Original photo by Zachi Evenor. Right: processed by Günther Noack, Software Engineer [8].

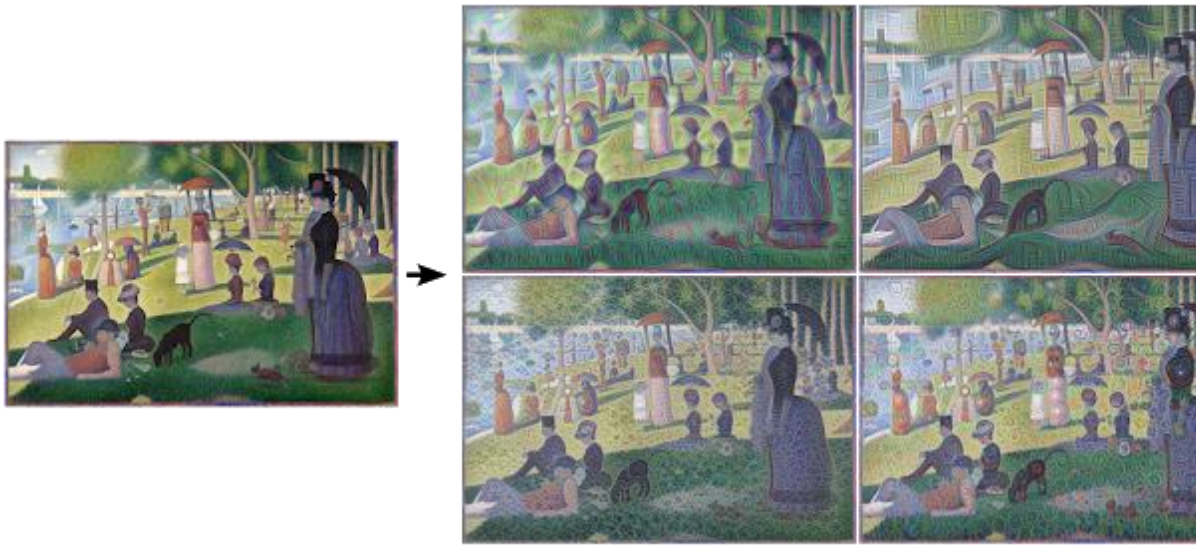


Figure 13: Left: Original painting by Georges Seurat. Right: processed images by Matthew McNaughton, Software Engineer [8].

If we choose higher-level layers, which identify more sophisticated features in images, complex features or even whole objects tend to emerge. Again, we just start with an existing image and give it to our neural net. We ask the network: "Whatever you see there, I want more of it!" This creates a feedback loop: if a cloud looks a little bit like a bird, the network will make it look more like a bird. This in turn will make the network recognize the bird even more strongly on the next pass and so forth, until a highly detailed bird appears, seemingly out of nowhere.

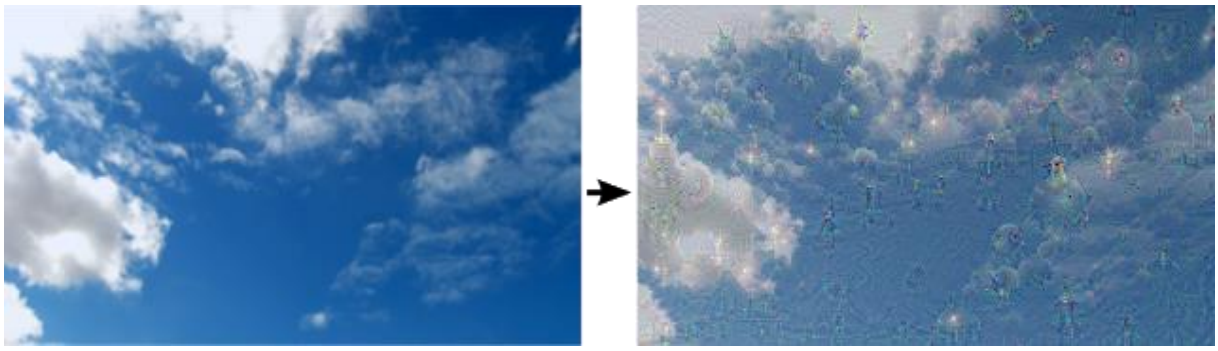


Figure 14: Simple neural network can be used to over-interpret an image [8].

The results are intriguing—even a relatively simple neural network can be used to over-interpret an image, just like as children we enjoyed watching clouds and interpreting the random shapes. This network was trained mostly on images of animals, so naturally it tends to interpret shapes as animals. But because the data is stored at such a high abstraction, the results are an interesting remix of these learned features.

Figure 15: We can do more than cloud watching with this technique [8].

Of course, we can do more than cloud watching with this technique. We can apply it to any kind of image. The results vary quite a bit with the kind of image, because the features that are entered bias the network towards certain interpretations. For example, horizon lines tend to get filled with towers and pagodas. Rocks and trees turn into buildings. Birds and insects appear in images of leaves.

Figure 16: This technique (Inceptionism) gives us a qualitative sense of the level of abstraction that a particular layer has achieved in its understanding of images [8].

The original image influences what kind of objects form in the processed image.

This technique gives us a qualitative sense of the level of abstraction that a particular layer has achieved in its understanding of images. We call this technique “Inceptionism” in reference to the neural net architecture used. See our Inceptionism gallery for more pairs of images and their processed results, plus some cool video animations.

If we apply the algorithm iteratively on its own outputs and apply some zooming after each iteration, we get an endless stream of new impressions, exploring the set of things the network knows about. We can even start this process from a random-noise image, so that the result becomes purely the result of the neural network, as seen in the following images:

Figure 17: Neural net “dreams”— generated purely from random noise, using a network trained on places by MIT Computer Science and AI Laboratory. See our Inceptionism gallery for hi-res versions of the images above and more (Images marked “Places205-GoogLeNet” were made using this network) [8].

4.8 Robotics

Having a machine learning agent interact with its environment requires true unsupervised learning, skill acquisition, active learning, exploration and reinforcement, all ingredients of human learning that are still not well understood or exploited through the supervised approaches that dominate deep learning today.

Our goal is to improve robotics via machine learning, and improve machine learning via robotics. We foster close collaborations between machine learning researchers and roboticists to enable learning at scale on real and simulated robotic systems [6].

We're exploring how to teach robots transferrable skills, by learning in parallel across many manipulation arms in our one-of-a-kind lab purpose-built for machine learning research:

We're teaching robots to predict what happens when they move objects around, in order to learn about the world around them and make better, safer decisions without supervision, and we are sharing our training data publicly to help advance the state of the art in this field. We're also bringing advances in deep learning to the exciting and demanding world of self-driving cars to improve their safety and reliability.

4.8.1 Deep Learning for Robotics: Learning from Large-Scale Interaction

We describe a learning-based approach to hand-eye coordination for robotic grasping from monocular images. To learn hand-eye coordination for grasping, we trained a large convolutional neural network to predict the probability that task- space motion of the gripper will result in successful grasps, using only monocular camera images independent of camera calibration or the current robot pose. This requires the network to observe the spatial relationship between the gripper and objects in the scene, thus learning hand-eye coordination. We then use this network to servo the gripper in real time to achieve successful grasps. We describe two large-scale experiments that we conducted on two separate robotic platforms. In the first experiment, about 800,000 grasp attempts were collected over the course of two months, using between 6 and 14 robotic manipulators at any given time, with differences in camera placement and gripper wear and tear. In the second experiment, we used a different robotic platform and 8 robots to collect a dataset consisting of over 900,000 grasp attempts. The second robotic platform was used to test transfer between robots, and the degree to which data from a different set of robots can be used to aid learning. Our experimental results

demonstrate that our approach achieves effective real-time control, can successfully grasp novel objects, and corrects mistakes by continuous servoing. Our transfer experiment also illustrates that data from different robots can be combined to learn more reliable and effective grasping.

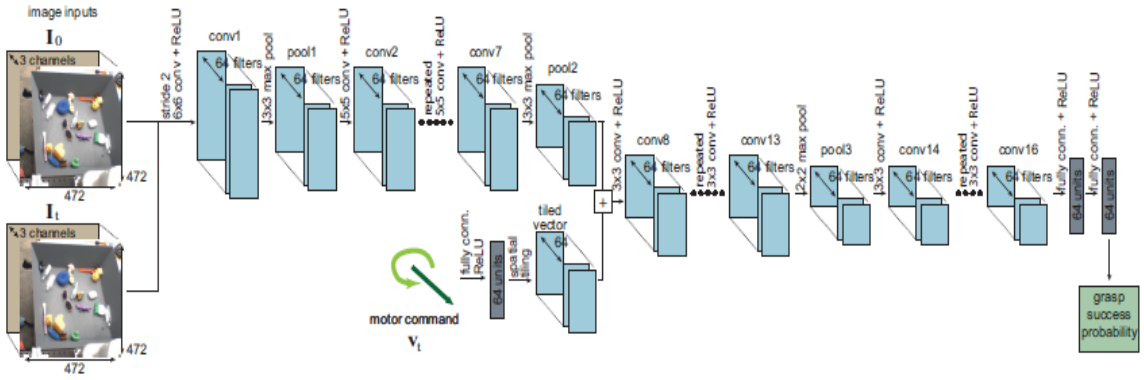


Figure 18: The architecture of our CNN grasp predictor. The input image It , as well as the pregrasp image I0, are fed into a 6×6 convolution with stride 2, followed by 3 × 3 max-pooling and six 5 × 5 convolutions. This is followed by a 3 × 3 max-pooling layer. The motor command v_t is processed by one fully connected layer, which is then pointwise added to each point in the response map of pool2 by tiling the output over the spatial dimensions. The result is then processed by 6 3 × 3 convolutions, 2 × 2 max-pooling, 3 more 3 × 3 convolutions, and two fully connected layers with 64 units, after which the network outputs the probability of a successful grasp through a sigmoid. Each convolution is followed by batch normalization [6].

5. CONCLUSION AND FUTURE WORK

In this report we would like to conclude that Google Brain has having a great impact on the research in Artificial Intelligence which is making the life so easy for humans, just one touch away from your destination. Google assistant which is having better accuracy than Siri (Apple), Cortana (Microsoft).It is allo helping in the Image captioning, Recommendation Replies in G-Mail and also in YouTube.

Google brain is trying it's best to just act like the human Brain or even Better making the things fully automated. It may just a matter of time, getting the accuracy as close as possible to 100%. It has also helped a lot in the Medical-fields like detection of tumours, Cancers,etc. These methods could improve accuracy and consistency of evaluating breast cancer cases, and potentially improve patient outcomes. Future work will focus on improvements utilizing larger datasets. In future work, we plan to further explore the relationship between our self-supervised continuous grasping approach and reinforcement learning, to allow the methods, to learn a wider variety of grasp strategies from large datasets of robotic experience.

6. REFERENCES

[1] https://research.google.com/teams/brain/

[2] Mart´ın Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, TensorFlow: A system for large-scale machine learning, ICML-2018

[3] Jeffrey Dean, Greg S. Corrado, Rajat Monga, Kai Chen, Matthieu Devin, Quoc V. Le, Mark Z. Mao, Marc'Aurelio Ranzato, Andrew Senior, Paul Tucker, Ke Yang, Andrew Y. Ng, Large Scale Distributed Deep networks,NIPS-2017

[4] Ryan Poplin, Dan Newburger, Jojo Dijamco, Nam Nguyen, Dion Loy, Sam S. Gross, Cory Y. McLean, Mark A. DePristo, Creating a universal SNP and small indel variant caller with deep neural networks, ICML-2017

[5] Yun Liu1, Krishna Gadepalli, Mohammad Norouzi, George E. Dahl,Timo Kohlberger, Aleksey Boyko, Subhashini Venugopalan, Aleksei Timofeev, Philip Q. Nelson2, Detecting Cancer Metastases on Gigapixel Pathology Images, CVPR-2017

[6] Sergey Levine, Peter Pastor, Alex Krizhevsky, Julian Ibarz and Deirdre Quillen, Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection, The International Journal of Robotics Research, IJRR-2017

[7] William Chan, Navdeep Jaitly, Quoc V. Le, Oriol Vinyals, ICASSP 2016, Listen, Attend and Spell: A Neural Network for Large Vocabulary Conversational Speech Recognition, ICASSP 2016

[8] https://research.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html

[9] magenta.tensorflow.org

[10] https://en.wikipedia.org/wiki/Google_Brain

[11] William Chan Navdeep Jaitly, Quoc V. Le, Oriol Vinyals, Listen, Attend and Spell, CVPR-2015