Inteligencia Artificial

Una Revolución

Pablo Zegers, Ph.D.
pablozegers@gmail.com

2019

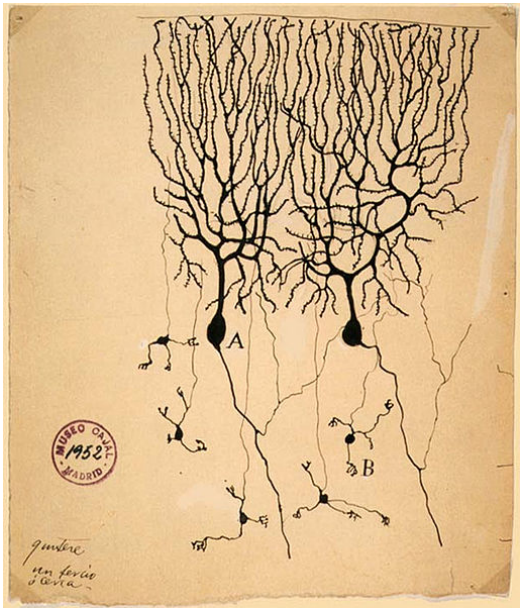
Presentación

Científico y Emprendedor, Enfocado en Inteligencia Artificial

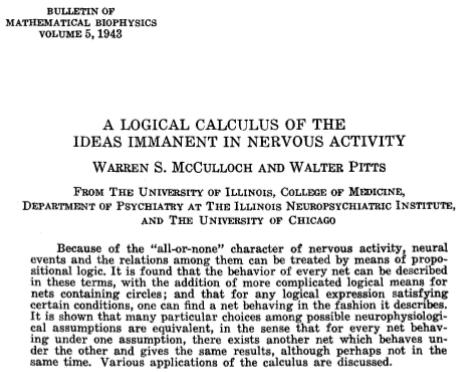
- No es una ciencia más, es una **meta ciencia**.
- **Inteligencia Artificial**
  - Aprendizaje de Máquinas
    - Aprendizaje Profundo (*Deep Learning*)
- **Robótica**, IA en un sistema físico capaz de interactuar con el mundo físico.

Inteligencia Artificial

Una Historia que Partió al **Inicio del Siglo XX**



Ramón y Cajal, *Revista Trimestral de Histología Normal y Patológica*, Laboratorio de Histología de la Facultad de Medicina de Barcelona, **1888**. Dibujo del cerebelo de un pichón, hecho por Santiago Ramón y Cajal (Wikipedia).



BULLETIN OF
MATHEMATICAL BIOPHYSICS
VOLUME 5, 1943

A LOGICAL CALCULUS OF THE
IDEAS IMMANENT IN NERVOUS ACTIVITY

WARREN S. McCULLOCH AND WALTER PITTS

FROM THE UNIVERSITY OF ILLINOIS, COLLEGE OF MEDICINE,
DEPARTMENT OF PSYCHIATRY AT THE ILLINOIS NEUROPSYCHIATRIC INSTITUTE,
AND THE UNIVERSITY OF CHICAGO

Because of the "all-or-none" character of nervous activity, neural events and the relations among them can be treated by means of propositional logic. It is found that the behavior of every net can be described in these terms, with the addition of more complicated logical means for nets containing circles; and that for any logical expression satisfying certain conditions, one can find a net behaving in the fashion it describes. It is shown that many particular choices among possible neurophysiological assumptions are equivalent, in the sense that for every net behaving under one assumption, there exists another net which behaves under the other and gives the same results, although perhaps not in the same time. Various applications of the calculus are discussed.

McCullock and Pitts, Journal of Mathematical Biophysics, **1943**.

$$y = g\left(\sum_{i=1}^{n} w_i \cdot x_i - b\right)$$

$$g(s) = \frac{2}{1 + e^{-a \cdot s}}$$

Rosenblatt, *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*, Spartan Books, **1962**.

- A partir del 2011, sorpresivamente y en contra del sentido común, las **redes neuronales profundas resolvieron problemas difíciles con facilidad**.
- Schwartz-Ziv and Tishby, el 29 of April of 2017, **demuestran en "Opening the black box of Deep Neural Networks via Information", ArXiv, que se cruzó un umbral matemático, hasta el momento desconocido, que permitió diseñar con facilidad IAs complejas**.

Full-Resolution Residual Networks for Semantic Segmentation in Street Scenes

Tobias Pohlen Alexander Hermans Markus Mathias Bastian Leibe
Visual Computing Institute
RWTH Aachen University
tobias.pohlen@rwth-aachen.de, {hermans, mathias, leibe}@vision.rwth-aachen.de

Abstract

Semantic image segmentation is an essential component of modern autonomous driving systems, as an accurate understanding of the surrounding scene is crucial to navigation and action planning. Current state-of-the-art approaches in semantic image segmentation rely on pre-trained networks that were initially developed for classifying images as a whole. While these networks exhibit outstanding recognition performance (i.e., what is visible?), they lack localization accuracy (i.e., where precisely is something located?). Therefore, additional processing steps have to be performed in order to obtain pixel-accurate segmentation masks at the full image resolution. To alleviate this problem we propose a novel ResNet-like architecture that exhibits strong localization and recognition performance. We combine multi-scale context with pixel-level accuracy by using two processing streams within our network: One stream carries information at the full image resolution, enabling precise adherence to segment boundaries. The other stream undergoes a sequence of pooling operations to obtain robust features for recognition. The two streams are coupled at the full image resolution using residuals. Without additional processing steps and without pre-training, our approach achieves an intersection-over-union score of 71.8% on the Cityscapes dataset.

1. Introduction

Recent years have seen an increasing interest in self driving cars and in driver assistance systems. A crucial aspect of autonomous driving is to acquire a comprehensive understanding of the surroundings in which a car is moving. Semantic image segmentation [49, 38, 21, 53, 33], the task of assigning a set of predefined class labels to image pixels, is an important tool for modeling the complex relationships of the semantic entities usually found in street scenes, such as cars, pedestrians, road, or sidewalks. In automotive scenarios it is used in various ways, *e.g.* as a pre-processing step to discard image regions that are unlikely to contain objects of interest [42, 15], to improve object detection [4, 23, 24, 58],

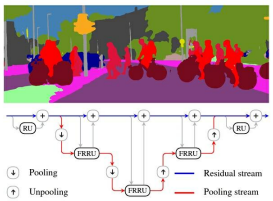


Figure 1. Example output and the abstract structure of our full-resolution residual network. The network has two processing streams. The residual stream (blue) stays at the full image resolution, the pooling stream (red) undergoes a sequence of pooling and unpooling operations. The two processing streams are coupled using full-resolution residual units (FRRUs).

or in combination with 3D scene geometry [32, 17, 35]. Many of those applications require precise region boundaries [20]. In this work, we therefore pursue the goal of achieving high-quality semantic segmentation with precise boundary adherence.

Current state-of-the-art approaches for image segmentation all employ some form of *fully convolutional network (FCNs)* [38] that takes the image as input and outputs a probability map for each class. Many papers rely on network architectures that have already been proven successful for image classification such as variants of the ResNet [25] or the VGG architecture [50]. Starting from pre-trained nets, where a large number of weights for the target task can be pre-set by an auxiliary classification task, reduces training time and often yields superior performance compared to training a network from scratch using the (possibly limited amount of) data of the target application. However, a main limitation of using such pre-trained networks is that they severely restrict the design space of novel approaches, since new network elements such as batch normalization [27] or

2016-12-6

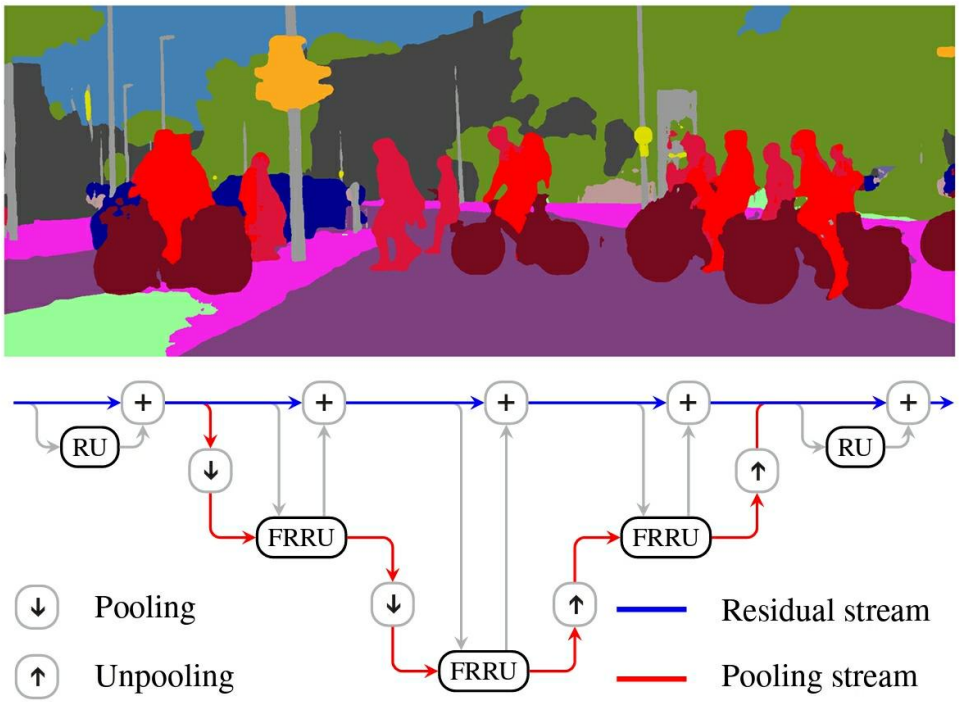


Figure 1. Example output and the abstract structure of our full-resolution residual network. The network has two processing streams. The residual stream (blue) stays at the full image resolution, the pooling stream (red) undergoes a sequence of pooling and unpooling operations. The two processing streams are coupled using full-resolution residual units (FRRUs).

Deep Video Portraits

HYEONGWOO KIM, Max Planck Institute for Informatics, Germany
PABLO GARRIDO, Technicolor, France
AYUSH TEWARI and WEIPENG XU, Max Planck Institute for Informatics, Germany
JUSTUS THIES and MATTHIAS NIESSNER, Technical University of Munich, Germany
PATRICK PÉREZ, Technicolor, France
CHRISTIAN RICHARDT, University of Bath, United Kingdom
MICHAEL ZOLLHÖFER, Stanford University, United States of America
CHRISTIAN THEOBALT, Max Planck Institute for Informatics, Germany

Fig. 1. Unlike current face reenactment approaches that only modify the expression of a target actor in a video, our novel deep video portrait approach enables full control over the target by transferring the rigid head pose, facial expression and eye motion with a high level of photorealism.

We present a novel approach that enables photo-realistic re-animation of portrait videos using only an input video. In contrast to existing approaches that are restricted to manipulations of facial expressions only, we are the first to transfer the full 3D head position, head rotation, face expression, eye gaze, and eye blinking from a source actor to a portrait video of a target actor. The core of our approach is a generative neural network with a novel space-time architecture. The network takes as input synthetic renderings of a parametric face model, based on which it predicts photo-realistic video frames for a given target actor. The realism in this rendering-to-video transfer is achieved by careful adversarial training, and as a result, we can create modified target videos that mimic the behavior of the synthetically-created input. In order to enable source-to-target video re-animation, we render a synthetic target video with the reconstructed head animation parameters from a source video, and feed it into the trained network – thus taking full control of the target. With the ability to freely recombine source and target parameters, we are able to demonstrate a large variety of video rewrite applications without explicitly modeling hair, body or background. For instance, we can reenact the full head using interactive user-controlled editing, and realize high-fidelity visual dubbing. To demonstrate the high quality of our output, we conduct an extensive series of experiments and evaluations, where for instance a user study shows that our video edits are hard to detect.

CCS Concepts: • **Computing methodologies** → **Computer graphics**; **Neural networks**; *Appearance and texture representations*; *Animation*; *Rendering*;

Additional Key Words and Phrases: Facial Reenactment, Video Portraits, Dubbing, Deep Learning, Conditional GAN, Rendering-to-Video Translation

ACM Reference Format:
Hyeongwoo Kim, Pablo Garrido, Ayush Tewari, Weipeng Xu, Justus Thies, Matthias Nießner, Patrick Pérez, Christian Richardt, Michael Zollhöfer, and Christian Theobalt. 2018. Deep Video Portraits. *ACM Trans. Graph.* 37, 4, Article 163 (August 2018), 14 pages. https://doi.org/10.1145/3197517.3201283

1 INTRODUCTION

Synthesizing and editing video portraits, i.e., videos framed to show a person's head and upper body, is an important problem in computer graphics, with applications in video editing and movie post-production, visual effects, visual dubbing, virtual reality, and telepresence, among others. In this paper, we address the problem of synthesizing a photo-realistic video portrait of a *target* actor that mimics the actions of a *source* actor, where source and target can be different subjects. More specifically, our approach enables a source actor to take full control of the rigid head pose, face expressions and

Authors' addresses: Hyeongwoo Kim, Max Planck Institute for Informatics, Campus E1.4, Saarbrücken, 66123, Germany, hyeongwoo.kim@mpi-inf.mpg.de; Pablo Garrido, Technicolor, 975 Avenue des Champs Blancs, Cesson-Sévigné, 35576, France, pablo.garrido.adrian@gmail.com; Ayush Tewari, atewari@mpi-inf.mpg.de; Weipeng Xu, wxu@mpi-inf.mpg.de, Max Planck Institute for Informatics, Campus E1.4, Saarbrücken, 66123, Germany; Justus Thies, justus.thies@tum.de; Matthias Nießner, niessner@tum.de, Technical University of Munich, Boltzmannstraße 3, Garching, 85748, Germany; Patrick Pérez, Technicolor, 975 Avenue des Champs Blancs, Cesson-Sévigné, 35576, France, Patrick.Perez@technicolor.com; Christian Richardt, University of Bath, Claverton Down, Bath, BA2 7AY, United Kingdom, christian@richardt.name; Michael Zollhöfer, Stanford University, 353 Serra Mall, Stanford, CA, 94305, United States of America, zollhoefer@cs.stanford.edu; Christian Theobalt, Max Planck Institute for Informatics, Campus E1.4, Saarbrücken, 66123, Germany, theobalt@mpi-inf.mpg.de.

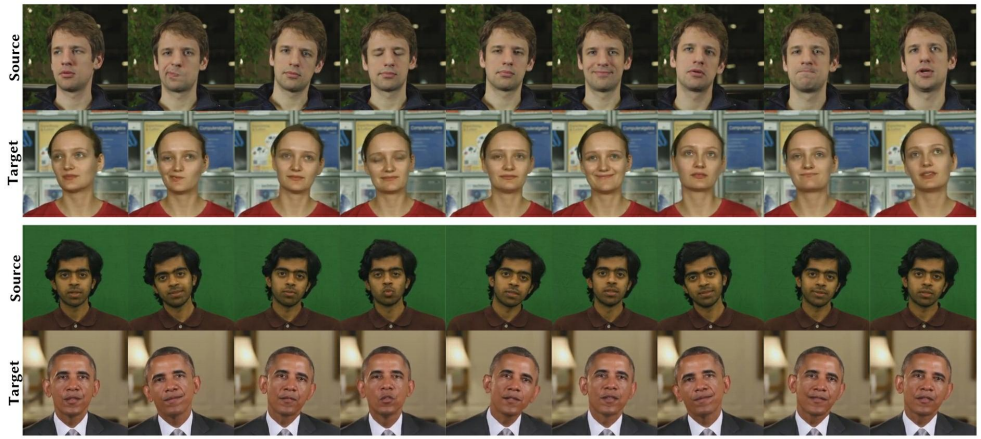


Fig. 5. Qualitative results of full-head reenactment: our approach enables full-frame target video portrait synthesis under full 3D head pose control. The output video portraits are photo-realistic and hard to distinguish from real videos. Note that even the shadow in the background of the second row moves consistently with the modified foreground head motion. In the sequence at the top, we only transfer the translation in the camera plane, while we transfer the full 3D translation for the sequence at the bottom. For full sequences, please refer to our video. *Obama* video courtesy of the White House (public domain).

2018-5-29

Learning Dexterous In-Hand Manipulation

OpenAI*

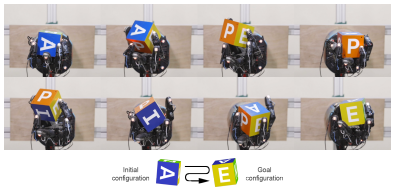


Figure 1: A five-fingered humanoid hand trained with reinforcement learning manipulating a block from an initial configuration to a goal configuration using vision for sensing.

Abstract

We use reinforcement learning (RL) to learn dexterous in-hand manipulation policies which can perform vision-based object reorientation on a physical Shadow Dexterous Hand. The training is performed in a simulated environment in which we randomize many of the physical properties of the system like friction coefficients and an object's appearance. Our policies transfer to the physical robot despite being trained entirely in simulation. Our method does not rely on any human demonstrations, but many behaviors found in human manipulation emerge naturally, including finger gaiting, multi-finger coordination, and the controlled use of gravity. Our results were obtained using the same distributed RL system that was used to train OpenAI Five [43]. We also include a video of our results: https://youtu.be/jwSbzNHGf1M.

1 Introduction

While dexterous manipulation of objects is a fundamental everyday task for humans, it is still challenging for autonomous robots. Modern-day robots are typically designed for specific tasks in constrained settings and are largely unable to utilize complex end-effectors. In contrast, people are able to perform a wide range of dexterous manipulation tasks in a diverse set of environments, making the human hand a grounded source of inspiration for research into robotic manipulation.

The Shadow Dexterous Hand [58] is an example of a robotic hand designed for human-level dexterity; it has five fingers with a total of 24 degrees of freedom. The hand has been commercially available

*Built by a team of researchers and engineers at OpenAI (in alphabetical order).

Marcin Andrychowicz Bowen Baker Maciek Chociej Rafal Józefowicz Bob McGrew Jakub Pachocki Arthur Petron Matthias Plappert Glenn Powell Alex Ray Jonas Schneider Szymon Sidor Josh Tobin Peter Welinder Lilian Weng Wojciech Zaremba

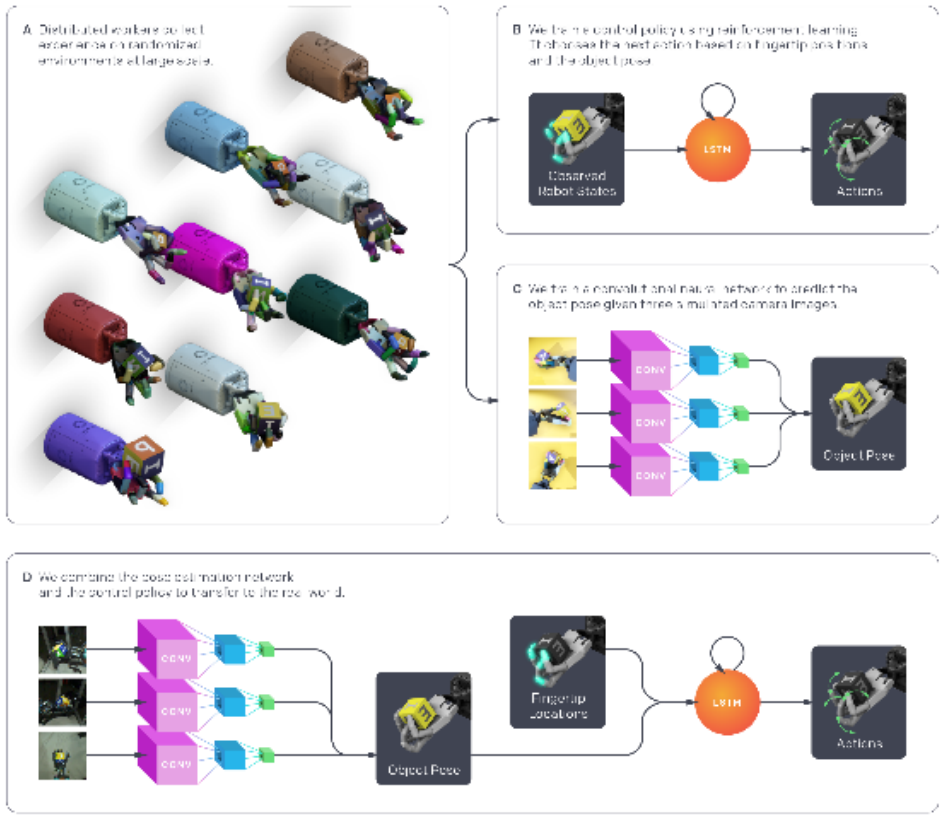


Figure 2: System Overview. (a) We use a large distribution of simulations with randomized parameters and appearances to collect data for both the control policy and vision-based pose estimator. (b) The control policy receives observed robot states and rewards from the distributed simulations and learns to map observations to actions using a recurrent neural network and reinforcement learning. (c) The vision based pose estimator renders scenes collected from the distributed simulations and learns to predict the pose of the object from images using a convolutional neural network (CNN), trained separately from the control policy. (d) To transfer to the real world, we predict the object pose from 3 real camera feeds with the CNN, measure the robot fingertip locations using a 3D motion capture system, and give both of these to the control policy to produce an action for the robot.

2018-7-30

Linguistic Regularities in Continuous Space Word Representations

Tomas Mikolov[*], **Wen-tau Yih, Geoffrey Zweig**
Microsoft Research
Redmond, WA 98052

Abstract

Continuous space language models have recently demonstrated outstanding results across a variety of tasks. In this paper, we examine the vector-space word representations that are implicitly learned by the input-layer weights. We find that these representations are surprisingly good at capturing syntactic and semantic regularities in language, and that each relationship is characterized by a relation-specific vector offset. This allows vector-oriented reasoning based on the offsets between words. For example, the male/female relationship is automatically learned, and with the induced vector representations, "King - Man + Woman" results in a vector very close to "Queen." We demonstrate that the word vectors capture syntactic regularities by means of syntactic analogy questions (provided with this paper), and are able to correctly answer almost 40% of the questions. We demonstrate that the word vectors capture semantic regularities by using the vector offset method to answer SemEval-2012 Task 2 questions. Remarkably, this method outperforms the best previous systems.

1 Introduction

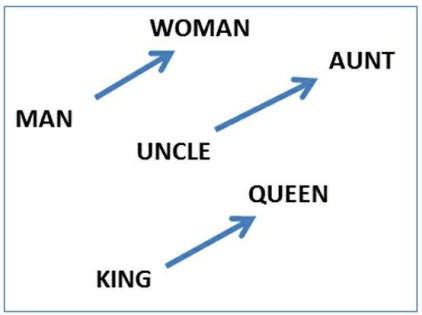
A defining feature of neural network language models is their representation of words as high dimensional real valued vectors. In these models (Bengio et al., 2003; Schwenk, 2007; Mikolov et al., 2010), words are converted via a learned lookup-table into real valued vectors which are used as the inputs to a neural network. As pointed out by the original proposers, one of the main advantages of these models is that the distributed representation achieves a level of generalization that is not possible with classical n-gram language models; whereas a n-gram model works in terms of discrete units that have no inherent relationship to one another, a continuous space model works in terms of word vectors where similar words are likely to have similar vectors. Thus, when the model parameters are adjusted in response to a particular word or word-sequence, the improvements will carry over to occurrences of similar words and sequences.

By training a neural network language model, one obtains not just the model itself, but also the learned word representations, which may be used for other, potentially unrelated, tasks. This has been used to good effect, for example in (Collobert and Weston, 2008; Turian et al., 2010) where induced word representations are used with sophisticated classifiers to improve performance in many NLP tasks.

In this work, we find that the learned word representations in fact capture meaningful syntactic and semantic regularities in a very simple way. Specifically, the regularities are observed as constant vector offsets between pairs of words sharing a particular relationship. For example, if we denote the vector for word i as x_i, and focus on the singular/plural relation, we observe that $x_{apple} - x_{apples} \approx x_{car} - x_{cars}, x_{family} - x_{families} \approx x_{car} - x_{cars}$, and so on. Perhaps more surprisingly, we find that this is also the case for a variety of semantic relations, as measured by the SemEval 2012 task of measuring relation similarity.

[*] Currently at Google, Inc.

Proceedings of NAACL-HLT 2013, pages 746–751,
Atlanta, Georgia, 9–14 June 2013.



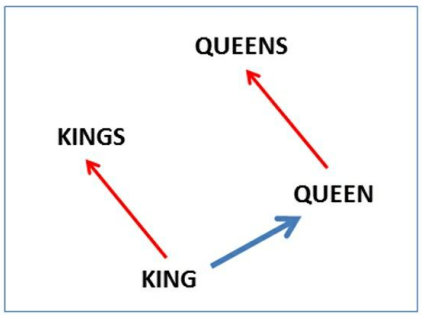


Figure 2: Left panel shows vector offsets for three word pairs illustrating the gender relation. Right panel shows a different projection, and the singular/plural relation for two words. In high-dimensional space, multiple relations can be embedded for a single word.

2013-6-10

Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation

Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi
yonghui,schuster,zhifengc,qvl,mnorouzi@google.com

Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Łukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, Jeffrey Dean

Abstract

Neural Machine Translation (NMT) is an end-to-end learning approach for automated translation, with the potential to overcome many of the weaknesses of conventional phrase-based translation systems. Unfortunately, NMT systems are known to be computationally expensive both in training and in translation inference – sometimes prohibitively so in the case of very large data sets and large models. Several authors have also charged that NMT systems lack robustness, particularly when input sentences contain rare words. These issues have hindered NMT's use in practical deployments and services, where both accuracy and speed are essential. In this work, we present GNMT, Google's Neural Machine Translation system, which attempts to address many of these issues. Our model consists of a deep LSTM network with 8 encoder and 8 decoder layers using residual connections as well as attention connections from the decoder network to the encoder. To improve parallelism and therefore decrease training time, our attention mechanism connects the bottom layer of the decoder to the top layer of the encoder. To accelerate the final translation speed, we employ low-precision arithmetic during inference computations. To improve handling of rare words, we divide words into a limited set of common sub-word units ("wordpieces") for both input and output. This method provides a good balance between the flexibility of "character"-delimited models and the efficiency of "word"-delimited models, naturally handles translation of rare words, and ultimately improves the overall accuracy of the system. Our beam search technique employs a length-normalization procedure and uses a coverage penalty, which encourages generation of an output sentence that is most likely to cover all the words in the source sentence. To directly optimize the translation BLEU scores, we consider refining the models by using reinforcement learning, but we found that the improvement in the BLEU scores did not reflect in the human evaluation. On the WMT'14 English-to-French and English-to-German benchmarks, GNMT achieves competitive results to state-of-the-art. Using a human side-by-side evaluation on a set of isolated simple sentences, it reduces translation errors by an average of 60% compared to Google's phrase-based production system.

1 Introduction

Neural Machine Translation (NMT) [41, 2] has recently been introduced as a promising approach with the potential of addressing many shortcomings of traditional machine translation systems. The strength of NMT lies in its ability to learn directly, in an end-to-end fashion, the mapping from input text to associated output text. Its architecture typically consists of two recurrent neural networks (RNNs), one to consume the input text sequence and one to generate translated output text. NMT is often accompanied by an attention mechanism [2] which helps it cope effectively with long input sequences.

An advantage of Neural Machine Translation is that it sidesteps many brittle design choices in traditional phrase-based machine translation [26]. In practice, however, NMT systems used to be worse in accuracy than phrase-based translation systems, especially when training on very large-scale datasets as used for the very best publicly available translation systems. Three inherent weaknesses of Neural Machine Translation are

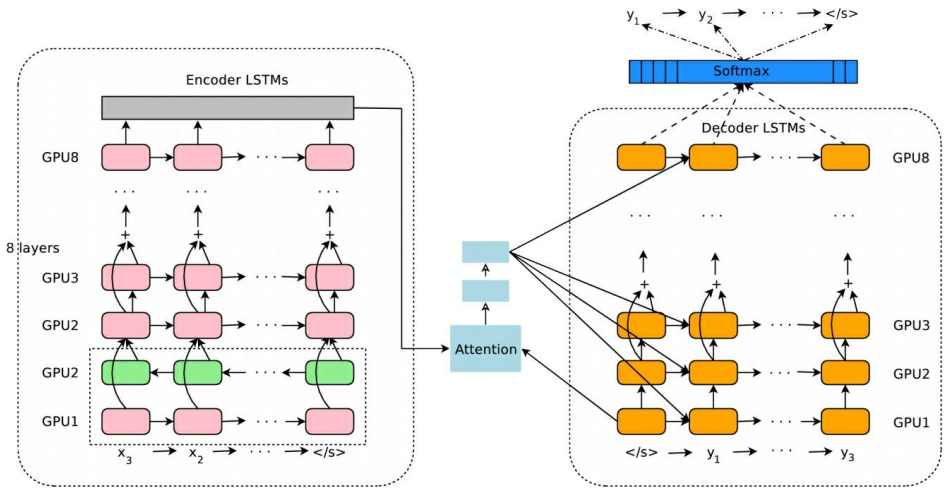


Figure 1: The model architecture of GNMT, Google's Neural Machine Translation system. On the left is the encoder network, on the right is the decoder network, in the middle is the attention module. The bottom encoder layer is bi-directional: the pink nodes gather information from left to right while the green nodes gather information from right to left. The other layers of the encoder are uni-directional. Residual connections start from the layer third from the bottom in the encoder and decoder. The model is partitioned into multiple GPUs to speed up training. In our setup, we have 8 encoder LSTM layers (1 bi-directional layer and 7 uni-directional layers), and 8 decoder layers. With this setting, one model replica is partitioned 8-ways and is placed on 8 different GPUs typically belonging to one host machine. During training, the bottom bi-directional encoder layers compute in parallel first. Once both finish, the uni-directional encoder layers can start computing, each on a separate GPU. To retain as much parallelism as possible during running the decoder layers, we use the bottom decoder layer output only for obtaining recurrent attention context, which is sent directly to all the remaining decoder layers. The softmax layer is also partitioned and placed on multiple GPUs. Depending on the output vocabulary size we either have them run on the same GPUs as the encoder and decoder networks, or have them run on a separate set of dedicated GPUs.

2016-10-8

A Deep Reinforced Model for Abstractive Summarization

Romain Paulus, Caiming Xiong and **Richard Socher**
{rpaulus,cxiong,rsocher}@salesforce.com

Abstract

Attentional, RNN-based encoder-decoder models for abstractive summarization have achieved good performance on short input and output sequences. However, for longer documents and summaries, these models often include repetitive and incoherent phrases. We introduce a neural network model with intra-attention and a new training method. This method combines standard supervised word prediction and reinforcement learning (RL). Models trained only with the former often exhibit "exposure bias" – they assume ground truth is provided at each step during training. However, when standard word prediction is combined with the global sequence prediction training of RL the resulting summaries become more readable. We evaluate this model on the CNN/Daily Mail and New York Times datasets. Our model obtains a 41.16 ROUGE-1 score on the CNN/Daily Mail dataset, a 5.7 absolute points improvement over previous state-of-the-art models. It also performs well as the first abstractive model on the New York Times corpus. Human evaluation also shows that our model produces higher quality summaries.

1 Introduction

Text summarization is the process of automatically generating natural language summaries from an input document while retaining the important points.

By condensing large quantities of information into short, informative summaries, summarization can aid many downstream applications such as creating news digests, search, and report generation.

There are two prominent types of summarization algorithms. First, extractive summarization systems form summaries by copying parts of the input (Neto et al., 2002; Dorr et al., 2003; Nallapati et al., 2017). Second, abstractive summarization systems generate new phrases, possibly rephrasing or using words that were not in the original text (Chopra et al., 2016; Nallapati et al., 2016; Zeng et al., 2016).

Recently, neural network models (Nallapati et al., 2016; Zeng et al., 2016), based on the attentional encoder-decoder model for machine translation (Bahdanau et al., 2014), were able to generate abstractive summaries with high ROUGE scores. However, these systems have typically focused on summarizing short input sequences (one or two sentences) to generate even shorter summaries. For example, the summaries on the DUC-2004 dataset generated by the state-of-the-art system by Zeng et al. (2016) are limited to 75 characters.

Nallapati et al. (2016) also applied their abstractive summarization model on the CNN/Daily Mail dataset (Hermann et al., 2015), which contains input sequences of up to 800 tokens and multi-sentence summaries of up to 100 tokens. The analysis by Nallapati et al. (2016) illustrate a key problem with attentional encoder-decoder models: they often generate unnatural summaries consisting of repeated phrases.

We present a new abstractive summarization model that achieves state-of-the-art results on the CNN/Daily Mail and similarly good results on the New York Times dataset (NYT) (Sandhaus, 2008). To our knowledge, this is the first model for abstractive summarization on the NYT dataset. We introduce a key attention mechanism and a new learning objective to address the repeating phrase

Source document
Jenson Button was denied his 100th race for McLaren after an ERS prevented him from making it to the start-line. It capped a miserable weekend for the Briton; his time in Bahrain plagued by reliability issues. Button spent much of the race on Twitter delivering his verdict as the action unfolded. 'Kimi is the man to watch,' and 'loving the sparks', were among his pearls of wisdom, but the tweet which courted the most attention was a rather mischievous one: 'Ooh is Lewis backing his team mate into Vettel?' he quizzed after Rosberg accused Hamilton of pulling off such a manoeuvre in China. Jenson Button waves to the crowd ahead of the Bahrain Grand Prix which he failed to start Perhaps a career in the media beckons Lewis Hamilton has out-qualified and finished ahead of Nico Rosberg at every race this season. Indeed Rosberg has now beaten his Mercedes team-mate only once in the 11 races since the pair infamously collided in Belgium last year. Hamilton secured the 36th win of his career in Bahrain and his 21st from pole position. Only Michael Schumacher (40), Ayrton Senna (29) and Sebastian Vettel (27) have more. He also became only the sixth F1 driver to lead 2,000 laps. Nico Rosberg has been left in the shade by Lewis Hamilton who celebrates winning his third race of the year Kimi Raikkonen secured a record seventh podium finish in Bahrain following his superb late salvo, although the Ferrari driver has never won in the Gulf Kingdom. It was the Finn's first trip to the rostrum since the 2013 Korean Grand Prix, but his triumph brought a typically deadpan response: 'You're never happy when you finish second... I'm a bit pleased to get a result.' Sparks fly off the back of Kimi Raikkonen's Ferrari en route to finishing second in Bahrain Bernie Ecclestone was in the Bahrain paddock this weekend. He denied trying to engineer a deal for Hamilton, out of contract at the end of the season, to join Ferrari despite earlier insisting that such a move would be 'great' for the sport. The 84-year-old also confirmed that F1 would be in Azerbaijan for the first time next year, even with concerns surrounding the countrys human rights record. 'I think everybody seems to be happy,' Ecclestone said. 'There doesn't seem to be any big problem there. There's no question of it not being on the calendar. It's going to be another good race. Formula One supremo Bernie Ecclestone speaks to Nico Rosberg ahead of the Bahrain Grand Prix
Ground truth summary
Button denied 100th race start for McLaren after ERS failure. Button then spent much of the Bahrain Grand Prix on Twitter delivering his verdict on the action as it unfolded. Lewis Hamilton has out-qualified and finished ahead of Mercedes team-mate Nico Rosberg at every race this season. Bernie Ecclestone confirms F1 will make its bow in Azerbaijan next season.
ML, with intra-attention (ROUGE-1 41.58)
Button was denied his 100th race for McLaren. ERS prevented him from making it to the start-line. The Briton. He quizzed after Nico Rosberg accused Lewis Hamilton of pulling off such a manoeuvre in China. Button has been in Azerbaijan for the first time since 2013.
RL, with intra-attention (ROUGE-1 50.00)
Button was denied his 100th race for McLaren after an ERS prevented him from making it to the start-line. It capped a miserable weekend for the Briton. Button has out-qualified. Finished ahead of Nico Rosberg at Bahrain. Lewis Hamilton has. In 11 races. . The race. To lead 2,000 laps. . In. . . And. .
ML+RL, with intra-attention (ROUGE-1 44.00)
Button was denied his 100th race for McLaren. The ERS prevented him from making it to the start-line. Button was his team mate in the 11 races in Bahrain. He quizzed after Nico Rosberg accused Lewis Hamilton of pulling off such a manoeuvre in China.

Table 3: Example from the CNN/Daily Mail test dataset showing the outputs of our three best models after de-tokenization, re-capitalization, replacing anonymized entities, and replacing numbers. The ROUGE score corresponds to the specific example.

2017-5-19

Deep Visual-Semantic Alignments for Generating Image Descriptions

Andrej Karpathy Li Fei-Fei
Department of Computer Science, Stanford University
{karpathy,feifeili}@cs.stanford.edu

Abstract

We present a model that generates natural language descriptions of images and their regions. Our approach leverages datasets of images and their sentence descriptions to learn about the inter-modal correspondences between language and visual data. Our alignment model is based on a novel combination of Convolutional Neural Networks over image regions, bidirectional Recurrent Neural Networks over sentences, and a structured objective that aligns the two modalities through a multimodal embedding. We then describe a Multimodal Recurrent Neural Network architecture that uses the inferred alignments to learn to generate novel descriptions of image regions. We demonstrate that our alignment model produces state of the art results in retrieval experiments on Flickr8K, Flickr30K and MSCOCO datasets. We then show that the generated descriptions significantly outperform retrieval baselines on both full images and on a new dataset of region-level annotations.

1. Introduction

A quick glance at an image is sufficient for a human to point out and describe an immense amount of details about the visual scene [14]. However, this remarkable ability has proven to be an elusive task for our visual recognition models. The majority of previous work in visual recognition has focused on labeling images with a fixed set of visual categories and great progress has been achieved in these endeavors [45, 11]. However, while closed vocabularies of visual concepts constitute a convenient modeling assumption, they are vastly restrictive when compared to the enormous amount of rich descriptions that a human can compose.

Some pioneering approaches that address the challenge of generating image descriptions have been developed [29, 13]. However, these models often rely on hard-coded visual concepts and sentence templates, which imposes limits on their variety. Moreover, the focus of these works has been on reducing complex visual scenes into a single sentence, which we consider to be an unnecessary restriction.

In this work, we strive to take a step towards the goal of generating dense descriptions of images (Figure 1). The primary challenge towards this goal is in the design of a model that is rich enough to simultaneously reason about contents of images and their representation in the domain of natural language. Additionally, the model should be free of assumptions about specific hard-coded templates, rules or categories and instead rely on learning from the training data. The second, practical challenge is that datasets of image captions are available in large quantities on the internet [21, 58, 37], but these descriptions multiplex mentions of several entities whose locations in the images are unknown.

Our core insight is that we can leverage these large image-sentence datasets by treating the sentences as weak labels, in which contiguous segments of words correspond to some particular, but unknown location in the image. Our approach is to infer these alignments and use them to learn a generative model of descriptions. Concretely, our contributions are twofold:

- We develop a deep neural network model that infers the latent alignment between segments of sentences and the region of the image that they describe.

Figure 1. Motivation/Concept Figure: Our model treats language as a rich label space and generates descriptions of image regions.

2015-4-14

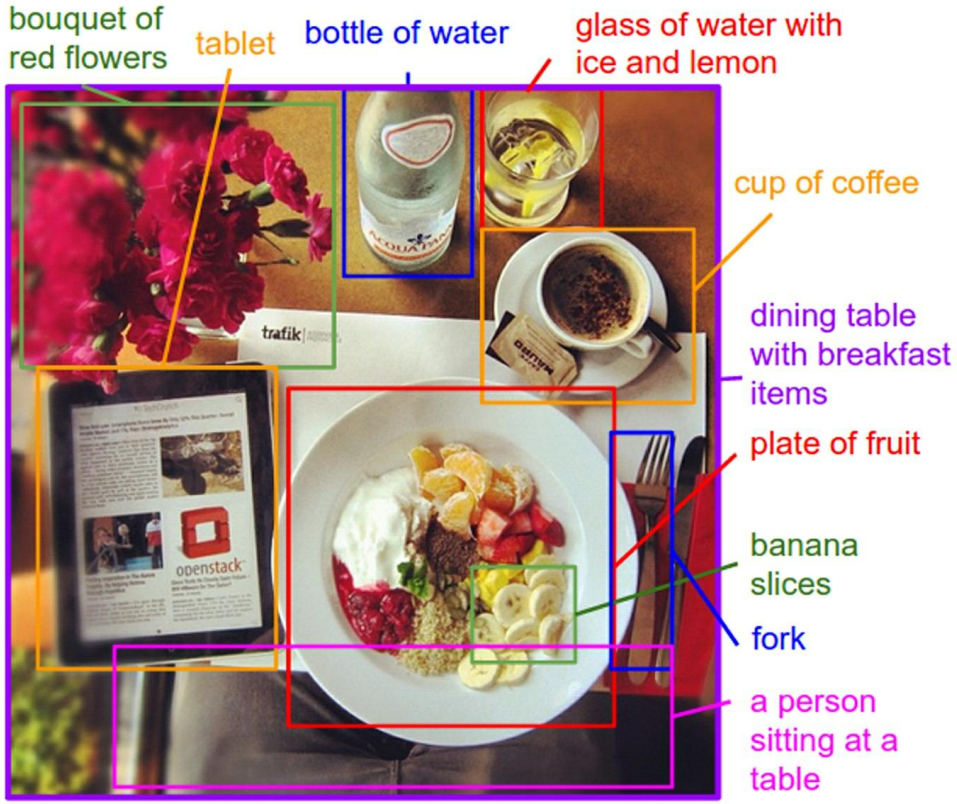


Figure 1. Motivation/Concept Figure: Our model treats language as a rich label space and generates descriptions of image regions.

StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks

Han Zhang*[1], Tao Xu[2], Hongsheng Li[3], Shaoting Zhang[4], Xiaolei Huang[2], Xiaogang Wang[3], Dimitris Metaxas[1]

[1]Department of Computer Science, Rutgers University
[2]Department of Computer Science and Engineering, Lehigh University
[3]Department of Electronic Engineering, The Chinese University of Hong Kong
[4]Department of Computer Science, University of North Carolina at Charlotte

Abstract

Synthesizing photo-realistic images from text descriptions is a challenging problem in computer vision and has many practical applications. Samples generated by existing text-to-image approaches can roughly reflect the meaning of the given descriptions, but they fail to contain necessary details and vivid object parts. In this paper, we propose stacked Generative Adversarial Networks (StackGAN) to generate photo-realistic images conditioned on text descriptions. The Stage-I GAN sketches the primitive shape and basic colors of the object based on the given text description, yielding Stage-I low resolution images. The Stage-II GAN takes Stage-I results and text descriptions as inputs, and generates high resolution images with photo-realistic details. The Stage-II GAN is able to rectify defects and add compelling details with the refinement process. Samples generated by StackGAN are more plausible than those generated by existing approaches. Importantly, our StackGAN for the first time generates realistic 256 × 256 images conditioned on only text descriptions, while state-of-the-art methods can generate at most 128 × 128 images. To demonstrate the effectiveness of the proposed StackGAN, extensive experiments are conducted on CUB and Oxford-102 datasets, which contain enough object appearance variations and are widely-used for text-to-image generation analysis.

1. Introduction

Generating photo-realistic images from text descriptions is a challenging problem. It has tremendous applications including photo editing and computer-aided design when fully automatic synthesis systems are available. However, even the most advanced methods failed in generating high resolution images with photo-realistic details using text descriptions. The main challenge of this problem is that the space of plausible images given text descriptions is multimodal. There are a large number of images that correctly fit the given text description.

Recently, Generative Adversarial Networks (GAN) [7, 3, 19] have shown promising results in modeling complex multimodal data and synthesizing real-world images. Reed *et al*. [22, 20] demonstrated that GAN can effectively gen-

*han.zhang@cs.rutgers.edu

Figure 1. Photo-realistic images generated by our StackGAN from unseen text descriptions. Descriptions for birds and flowers are from CUB [32] and Oxford-102 [18] datasets, respectively. (a) Given text descriptions, Stage-I of StackGAN sketches rough shapes and basic colors of objects, yielding low resolution images. (b) Stage-II of StackGAN takes Stage-I results and text descriptions as inputs, and generates high resolution images with photo-realistic details.

2016-12-10

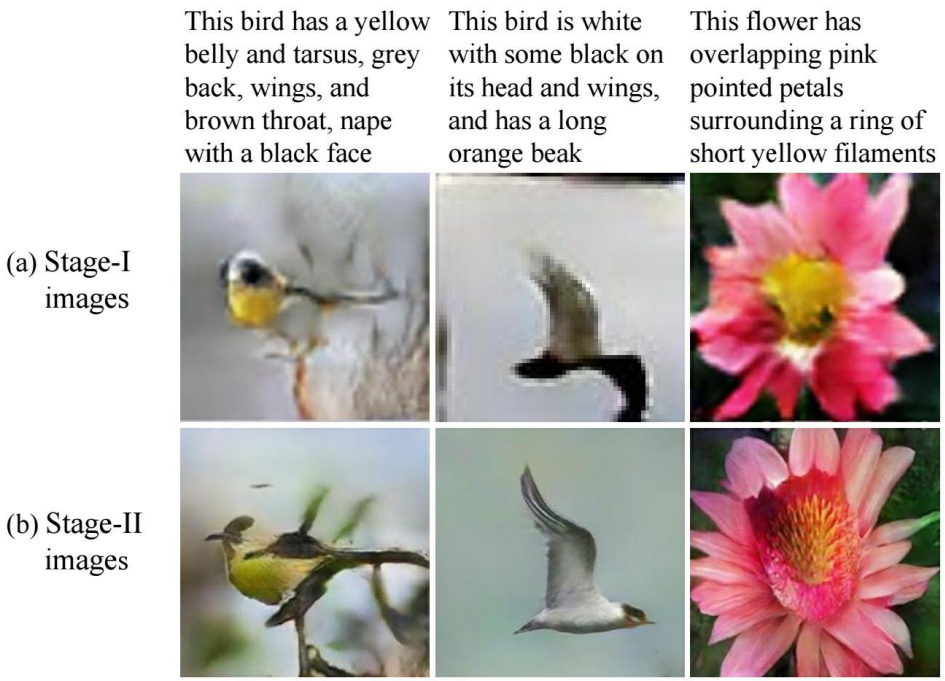


Figure 1. Photo-realistic images generated by our StackGAN from unseen text descriptions. Descriptions for birds and flowers are from CUB [32] and Oxford-102 [18] datasets, respectively. (a) Given text descriptions, Stage-I of StackGAN sketches rough shapes and basic colors of objects, yielding low resolution images. (b) Stage-II of StackGAN takes Stage-I results and text descriptions as inputs, and generates high resolution images with photo-realistic details.

ARTICLE

doi:10.1038/nature16961

Mastering the game of Go with deep neural networks and tree search

David Silver[1]*, Aja Huang[1]*, Chris J. Maddison[1], Arthur Guez[1], Laurent Sifre[1], George van den Driessche[1], Julian Schrittwieser[1], Ioannis Antonoglou[1], Veda Panneershelvam[1], Marc Lanctot[1], Sander Dieleman[1], Dominik Grewe[1], John Nham[2], Nal Kalchbrenner[1], Ilya Sutskever[2], Timothy Lillicrap[1], Madeleine Leach[1], Koray Kavukcuoglu[1], Thore Graepel[1] & Demis Hassabis[1]

The game of Go has long been viewed as the most challenging of classic games for artificial intelligence owing to its enormous search space and the difficulty of evaluating board positions and moves. Here we introduce a new approach to computer Go that uses 'value networks' to evaluate board positions and 'policy networks' to select moves. These deep neural networks are trained by a novel combination of supervised learning from human expert games, and reinforcement learning from games of self-play. Without any lookahead search, the neural networks play Go at the level of state-of-the-art Monte Carlo tree search programs that simulate thousands of random games of self-play. We also introduce a new search algorithm that combines Monte Carlo simulation with value and policy networks. Using this search algorithm, our program AlphaGo achieved a 99.8% winning rate against other Go programs, and defeated the human European Go champion by 5 games to 0. This is the first time that a computer program has defeated a human professional player in the full-sized game of Go, a feat previously thought to be at least a decade away.

All games of perfect information have an optimal value function, $v^*(s)$, which determines the outcome of the game, from every board position or state s, under perfect play by all players. These games may be solved by recursively computing the optimal value function in a search tree containing approximately b^d possible sequences of moves, where b is the game's breadth (number of legal moves per position) and d is its depth (game length). In large games, such as chess ($b \approx 35$, $d \approx 80$)[1] and especially Go ($b \approx 250$, $d \approx 150$)[1], exhaustive search is infeasible[2,3], but the effective search space can be reduced by two general principles. First, the depth of the search may be reduced by position evaluation: truncating the search tree at state s and replacing the subtree below s by an approximate value function $v(s) \approx v^*(s)$ that predicts the outcome from state s. This approach has led to superhuman performance in chess[4], checkers[5] and othello[6], but it was believed to be intractable in Go due to the complexity of the game[7]. Second, the breadth of the search may be reduced by sampling actions from a policy $p(a|s)$ that is a probability distribution over possible moves a in position s. For example, Monte Carlo rollouts[8] search to maximum depth without branching at all, by sampling long sequences of actions for both players from a policy p. Averaging over such rollouts can provide an effective position evaluation, achieving superhuman performance in backgammon[8] and Scrabble[9], and weak amateur level play in Go[10].

Monte Carlo tree search (MCTS)[11,12] uses Monte Carlo rollouts to estimate the value of each state in a search tree. As more simulations are executed, the search tree grows larger and the relevant values become more accurate. The policy used to select actions during search is also improved over time, by selecting children with higher values. Asymptotically, this policy converges to optimal play, and the evaluations converge to the optimal value function[12]. The strongest current Go programs are based on MCTS, enhanced by policies that are trained to predict human expert moves[13]. These policies are used to narrow the search to a beam of high-probability actions, and to sample actions during rollouts. This approach has achieved strong amateur play[13–15]. However, prior work has been limited to shallow policies[13–15] or value functions[16] based on a linear combination of input features.

Recently, deep convolutional neural networks have achieved unprecedented performance in visual domains: for example, image classification[17], face recognition[18], and playing Atari games[19]. They use many layers of neurons, each arranged in overlapping tiles, to construct increasingly abstract, localized representations of an image[20]. We employ a similar architecture for the game of Go. We pass in the board position as a 19×19 image and use convolutional layers to construct a representation of the position. We use these neural networks to reduce the effective depth and breadth of the search tree: evaluating positions using a value network, and sampling actions using a policy network.

We train the neural networks using a pipeline consisting of several stages of machine learning (Fig. 1). We begin by training a supervised learning (SL) policy network p_σ directly from expert human moves. This provides fast, efficient learning updates with immediate feedback and high-quality gradients. Similar to prior work[13,15], we also train a fast policy p_π that can rapidly sample actions during rollouts. Next, we train a reinforcement learning (RL) policy network p_ρ that improves the SL policy network by optimizing the final outcome of games of self-play. This adjusts the policy towards the correct goal of winning games, rather than maximizing predictive accuracy. Finally, we train a value network v_θ that predicts the winner of games played by the RL policy network against itself. Our program AlphaGo efficiently combines the policy and value networks with MCTS.

Supervised learning of policy networks

For the first stage of the training pipeline, we build on prior work on predicting expert moves in the game of Go using supervised learning[13,21–23]. The SL policy network $p_\sigma(a|s)$ alternates between convolutional layers with weights σ, and rectifier nonlinearities. A final softmax layer outputs a probability distribution over all legal moves a. The input s to the policy network is a simple representation of the board state (see Extended Data Table 2). The policy network is trained on randomly

[1]Google DeepMind, 5 New Street Square, London EC4A 3TW, UK. [2]Google, 1600 Amphitheatre Parkway, Mountain View, California 94043, USA.
*These authors contributed equally to this work.

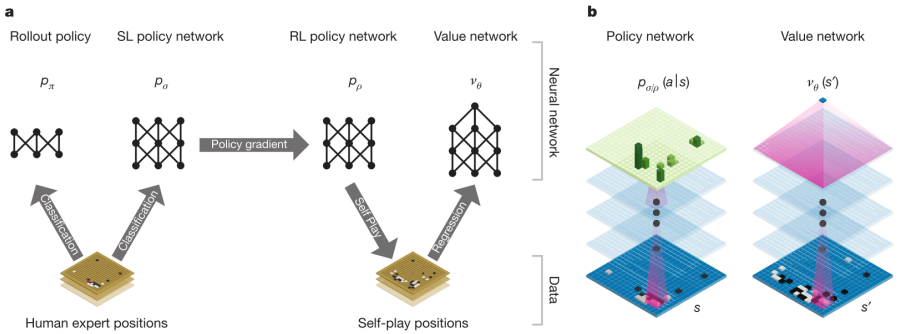
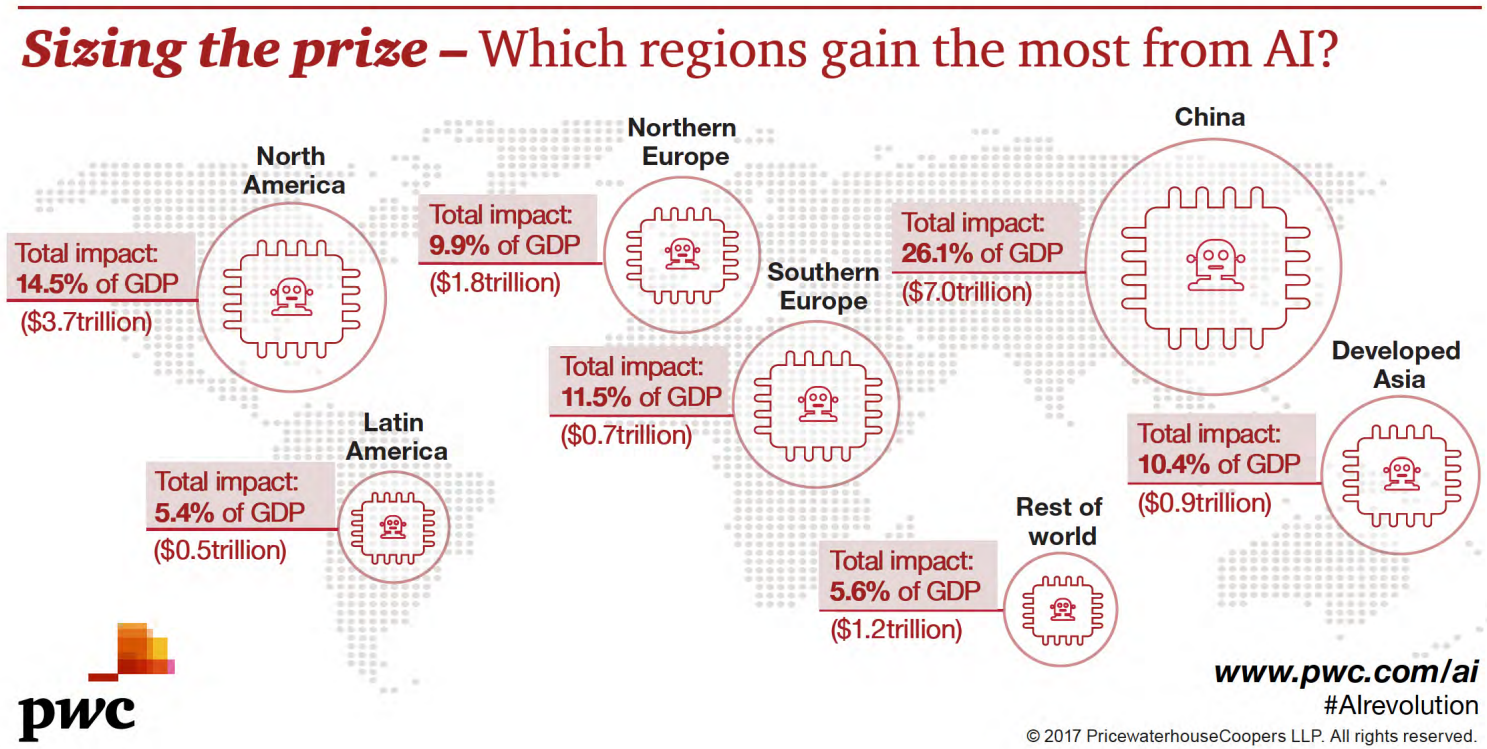


Figure 1 | Neural network training pipeline and architecture. a, A fast rollout policy p_π and supervised learning (SL) policy network p_σ are trained to predict human expert moves in a data set of positions. A reinforcement learning (RL) policy network p_ρ is initialized to the SL policy network, and is then improved by policy gradient learning to maximize the outcome (that is, winning more games) against previous versions of the policy network. A new data set is generated by playing games of self-play with the RL policy network. Finally, a value network v_θ is trained by regression to predict the expected outcome (that is, whether the current player wins) in positions from the self-play data set. **b**, Schematic representation of the neural network architecture used in AlphaGo. The policy network takes a representation of the board position s as its input, passes it through many convolutional layers with parameters σ (SL policy network) or ρ (RL policy network), and outputs a probability distribution $p_\sigma(a|s)$ or $p_\rho(a|s)$ over legal moves a, represented by a probability map over the board. The value network similarly uses many convolutional layers with parameters θ, but outputs a scalar value $v_\theta(s')$ that predicts the expected outcome in position s'.

2016-01-18

Inteligencia Artificial

Creación de US $15,7 Millones de Millones de Valor Económico

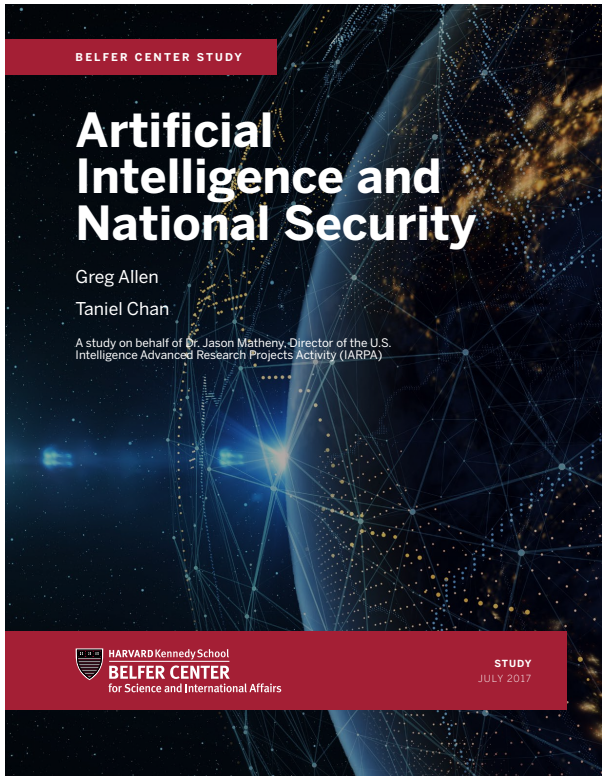


Personas caminando por el Sendero Chilkoot durante la Fiebre del Oro en Yukon, Alaska, a fines del siglo XIX (Wikipedia).

- Incide en **todas las actividades humanas**, no es sobre un producto o servicio.
- Afecta a **donde sea que haya personas**, no está localizada geográficamente.
- Produce una **reacción de empresas y gobiernos** que tiene como fin asegurar posiciones en este nuevo escenario.

Reacciones de los Gobiernos

Visión



- Executive summary
  - Researchers in the field of Artificial Intelligence (AI) have demonstrated significant technical progress over the past five years, **much faster** than was previously anticipated.
  - Most AI research advances are occurring in the **private sector and academia**.
  - **Existing capabilities in AI have significant potential for national security.**
  - Future progress in AI has the potential to be a transformative national security technology, **on a par with nuclear weapons, aircraft, computers, and biotech**.
  - Advances in AI will affect national security by driving change in three areas: **military superiority, information superiority, and economic superiority**.
  - We analyzed four prior cases of transformative military technologies—nuclear, aerospace, cyber, and biotech—and generated **"lessons learned" for AI**.
  - Taking a "whole of government" frame, **we provide three goals for U.S. national security policy toward AI technology and provide 11 recommendations**.

Reacciones de los Gobiernos

Fuerzas Armadas

Reacciones de los Gobiernos

Capacidad de Manipulación

PNAS PNAS PNAS PNAS

Psychological targeting as an effective approach to digital mass persuasion

S. C. Matz[a,1], M. Kosinski[b,2], G. Nave[c], and D. J. Stillwell[d,2]

[a]Columbia Business School, Columbia University, New York City, NY 10027; [b]Graduate School of Business, Stanford University, Stanford, CA 94305; [c]Wharton School of Business, University of Pennsylvania, Philadelphia, PA 19104; and [d]Cambridge Judge Business School, University of Cambridge, Cambridge, CB2 3EB, United Kingdom

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved October 17, 2017 (received for review June 17, 2017)

People are exposed to persuasive communication across many different contexts: Governments, companies, and political parties use persuasive appeals to encourage people to eat healthier, purchase a particular product, or vote for a specific candidate. Laboratory studies show that such persuasive appeals are more effective in influencing behavior when they are tailored to individuals' unique psychological characteristics. However, the investigation of large-scale psychological persuasion in the real world has been hindered by the questionnaire-based nature of psychological assessment. Recent research, however, shows that people's psychological characteristics can be accurately predicted from their digital footprints, such as their Facebook Likes or Tweets. Capitalizing on this form of psychological assessment from digital footprints, we test the effects of psychological persuasion on people's actual behavior in an ecologically valid setting. In three field experiments that reached over 3.5 million individuals with psychologically tailored advertising, we find that matching the content of persuasive appeals to individuals' psychological characteristics significantly altered their behavior as measured by clicks and purchases. Persuasive appeals that were matched to people's extraversion or openness-to-experience level resulted in up to 40% more clicks and up to 50% more purchases than their mismatching or unpersonalized counterparts. Our findings suggest that the application of psychological targeting makes it possible to influence the behavior of large groups of people by tailoring persuasive appeals to the psychological needs of the target audiences. We discuss both the potential benefits of this method for helping individuals make better decisions and the potential pitfalls related to manipulation and privacy.

persuasion | digital mass communication | psychological targeting | personality | targeted marketing

Persuasive mass communication is aimed at encouraging large groups of people to believe and act on the communicator's viewpoint. It is used by governments to encourage healthy behaviors, by marketers to acquire and retain consumers, and by political parties to mobilize the voting population. Research suggests that persuasive communication is particularly effective when tailored to people's unique psychological characteristics and motivations (1–5), an approach that we refer to as *psychological persuasion*. The proposition of this research is simple yet powerful: What convinces one person to behave in a desired way might not do so for another. For example, matching computer-generated advice to participants' dominance level elicited higher ratings of source credibility and increased the likelihood of participants changing their initial opinions in response to the advice (2). Similarly, participants' positive attitudes and purchase intentions were stronger when the marketing message for a mobile phone was tailored to their personality profile (4). While these studies provide promising evidence for the effectiveness of psychological persuasion, their validity is limited by the fact that they were mainly conducted in small-scale, controlled laboratory settings using self-report questionnaires. Self-reports are known to be affected by a whole range of response biases (6), and there are numerous reasons why people's natural behavior might differ from that displayed in the laboratory (7). Consequently, it is questionable whether—and to what extent—these findings can be generalized to the application of psychological persuasion in real-world mass persuasion (see ref. 8 for initial evidence).

A likely explanation for the lack of ecologically valid research in the context of psychological persuasion is the questionnaire-based nature of psychological assessment. Whereas researchers can ask participants to complete a psychological questionnaire in the laboratory, it is unrealistic to expect millions of people to do so before sending them persuasive messages online. Recent research in the field of computational social sciences (9), however, suggests that people's psychological profiles can be accurately predicted from the digital footprints they leave with every step they take online (10). For example, people's personality profiles have been predicted from personal websites (11), blogs (12), Twitter messages (13), Facebook profiles (10, 14–16), and Instagram pictures (17). This form of *psychological assessment from digital footprints* makes it paramount to establish the extent to which behaviors of large groups of people can be influenced through the application of psychological mass persuasion—both in their own interest (e.g., by persuading them to eat healthier) and against their best interest (e.g., by persuading them to gamble). We begin this endeavor in a domain that is relatively uncontroversial from an ethical point of view: consumer products.

Significance

Building on recent advancements in the assessment of psychological traits from digital footprints, this paper demonstrates the effectiveness of psychological mass persuasion—that is, the adaptation of persuasive appeals to the psychological characteristics of large groups of individuals with the goal of influencing their behavior. On the one hand, this form of psychological mass persuasion could be used to help people make better decisions and lead healthier and happier lives. On the other hand, it could be used to covertly exploit weaknesses in their character and persuade them to take action against their own best interest, highlighting the potential need for policy interventions.

Author contributions: S.C.M. and M.K. designed research; S.C.M., M.K., and D.J.S. performed research; S.C.M. analyzed data; and S.C.M., M.K., G.N., and D.J.S. wrote the paper.

Conflict of interest statement: D.J.S. received revenue as the owner of the myPersonality Facebook application until it was discontinued in 2012. Revenue was received from displaying ads within the application and charging for a premium personality test. The revenue received by D.J.S. was unrelated to the studies presented in this paper, and there will be no future revenue generated from the application. None of the authors received any compensation for working on the marketing campaigns used to collect data for the studies presented in this manuscript.

This article is a PNAS Direct Submission.

Data deposition: The data reported in this paper have been deposited in the Open Science Framework (OSF), https://osf.io/ey7r7/view_only=f0310b0c2e0642bbbc3a1cd857a9e3c7.

[1]To whom correspondence should be addressed. Email: sm4409@gsb.columbia.edu.

[2]M.K. and D.J.S. contributed equally to this work.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1710966114-/DCSupplemental.

2017-11-28

A

High Extraversion

Dance like no one's watching (but they totally are)

Low Extraversion

Beauty doesn't have to shout

B

High Openness

Aristoteles? The Seychelles? Unleash your creativity and challenge your imagination with an ulimited number of crossword puzzles!

Low Openness

Settle in with an all-time favorite! The crossword puzzle that has challenged players for generations.

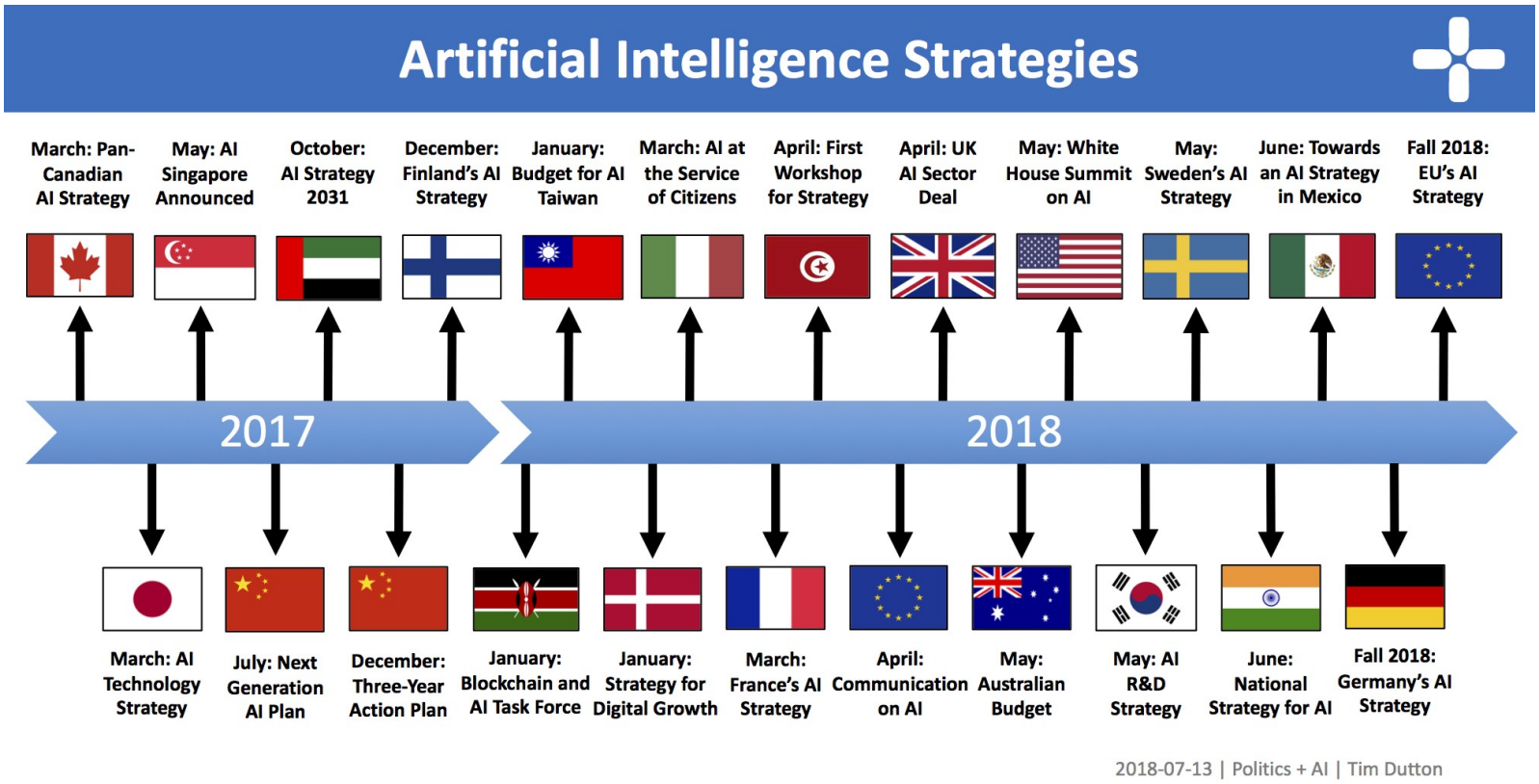
Reacciones de los Gobiernos

Riesgos

- **"Some countries are already moving in this direction. China has begun to construct a digital authoritarian state by using surveillance and machine learning tools to control restive populations, and by creating what it calls a "social credit system." Several like-minded countries have begun to buy or emulate Chinese systems. Just as competition between liberal democratic, fascist, and communist social systems defined much of the twentieth century, so the struggle between liberal democracy and digital authoritarianism is set to define the twenty-first."** (*How Artificial Intelligence Will Reshape the Global Order*, Foreign Affairs, 10 de julio de 2018)

Reacciones de los Gobiernos

Planes Gubernamentales



Reacciones de los Gobiernos

Objetivos Generales

- Aprovechar la **oportunidad**.
- Crear **capacidades humanas**.
- Ayudar a la **transición**.

La Sociedad Humana

Falsas Expectativas sobre la IA Gracias al "Efecto Hollywood"

The Terminator (Wikipedia).

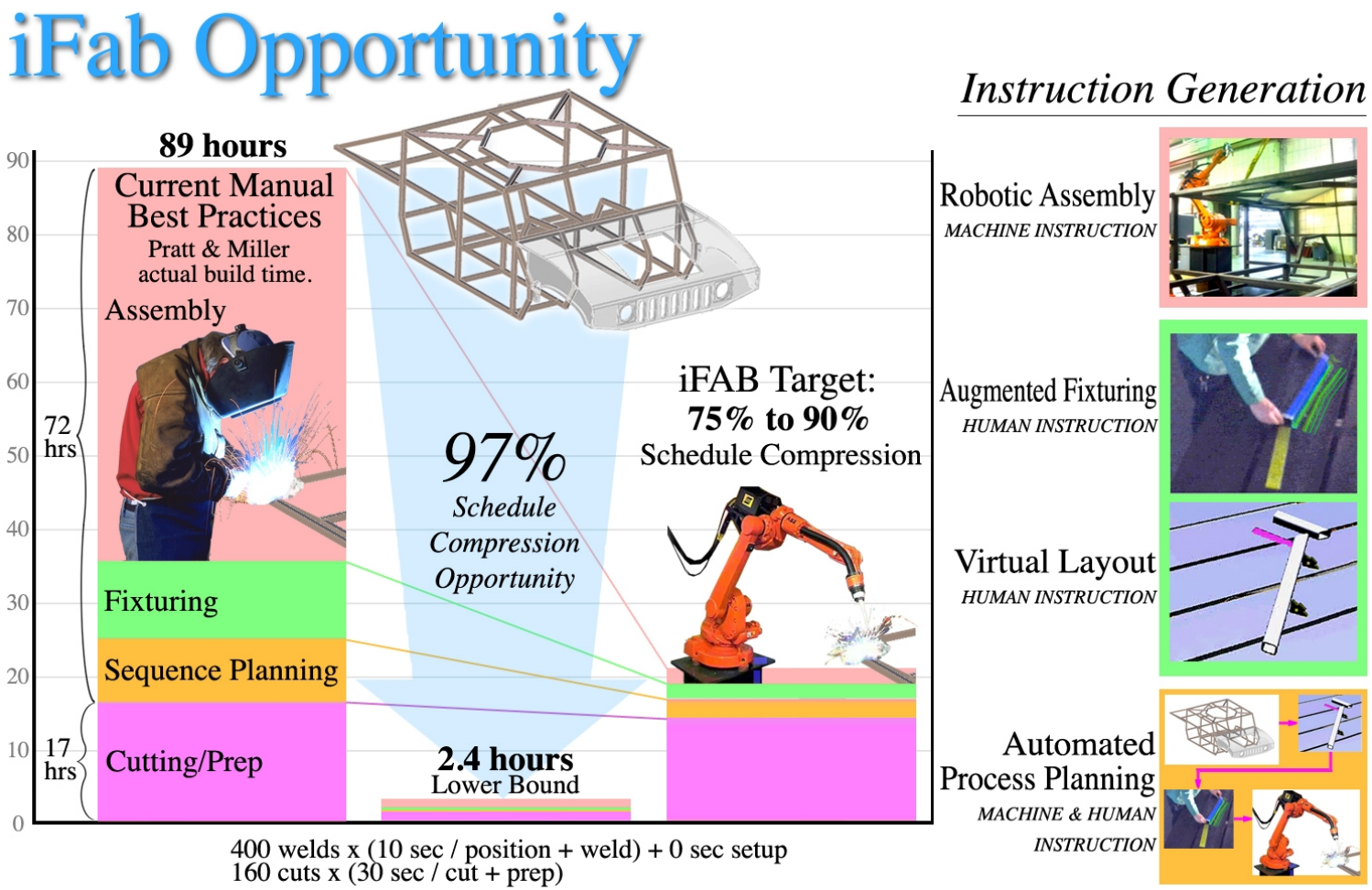
- Sí, **la IA está cambiando las cosas**.
- Pero hasta el momento sólo hay **IAs angostas**: se necesitan a las personas antes, durante y después de la introducción de estas tecnologías.
- La tecnología actual está **muy lejos de ser capaz de duplicar la enorme capacidad humana**.
- Queda por verse **hasta dónde y cuán rápido** va a ser el cambio.

La Sociedad Humana

Predicciones sobre la Automatización del Trabajo

When	Where	Jobs Destroyed	Jobs Created	Predictor
2016	worldwide		900,000 to 1,500,000	Metra Martech
2018	US jobs	13,852,530*	3,078,340*	Forrester
2020	worldwide		1,000,000-2,000,000	Metra Martech
2020	worldwide	1,800,000	2,300,000	Gartner
2020	sampling of 15 countries	7,100,000	2,000,000	WEF
2021	worldwide		1,900,000-3,500,000	IFR
2021	US jobs	9,108,900*		Forrester
2022	worldwide	1,000,000,000		Thomas Frey
2025	US jobs	24,186,240*	13,604,760*	Forrester
2025	US jobs	3,400,000		ScienceAlert
2027	US jobs	24,700,000	14,900,000	Forrester
2030	worldwide	2,000,000,000		Thomas Frey
2030	worldwide	400,000,000-800,000,000	555,000,000-890,000,000	McKinsey
2030	US jobs	58,164,320*		PWC
2035	US jobs	80,000,000		Bank of England
2035	UK jobs	15,000,000		Bank of England
No Date	US jobs	13,594,320*		OECD
No Date	UK jobs	13,700,000		IPPR

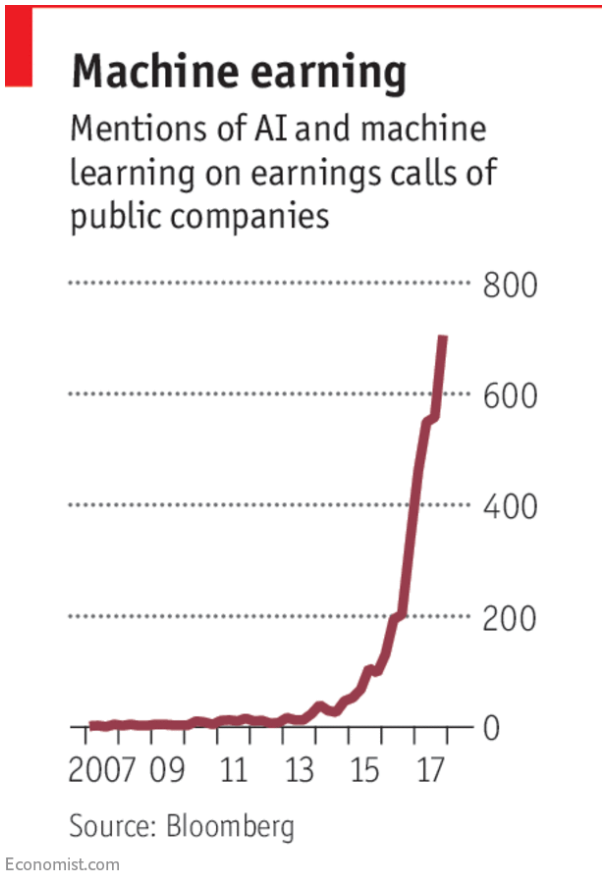
Every study we could find on what automation will do to jobs, in one chart, Erin Winick, Technology Review, 25 de enero de 2018.



Robots and People Can Work Faster Together, David Bourne, Director del Rapid Manufacturing Lab, Robotics Institute, Carnegie Mellon University, 25 de julio de 2013. Leer *Collaborative Intelligence: Humans and AI Are Joining Forces*, Harvard Business Review, Julio-Agosto, 2018.

Adaptaciones de las Empresas

Reacciones



Non-tech businesses are beginning to use artificial intelligence at scale, The Economist, 31 de marzo de 2018.

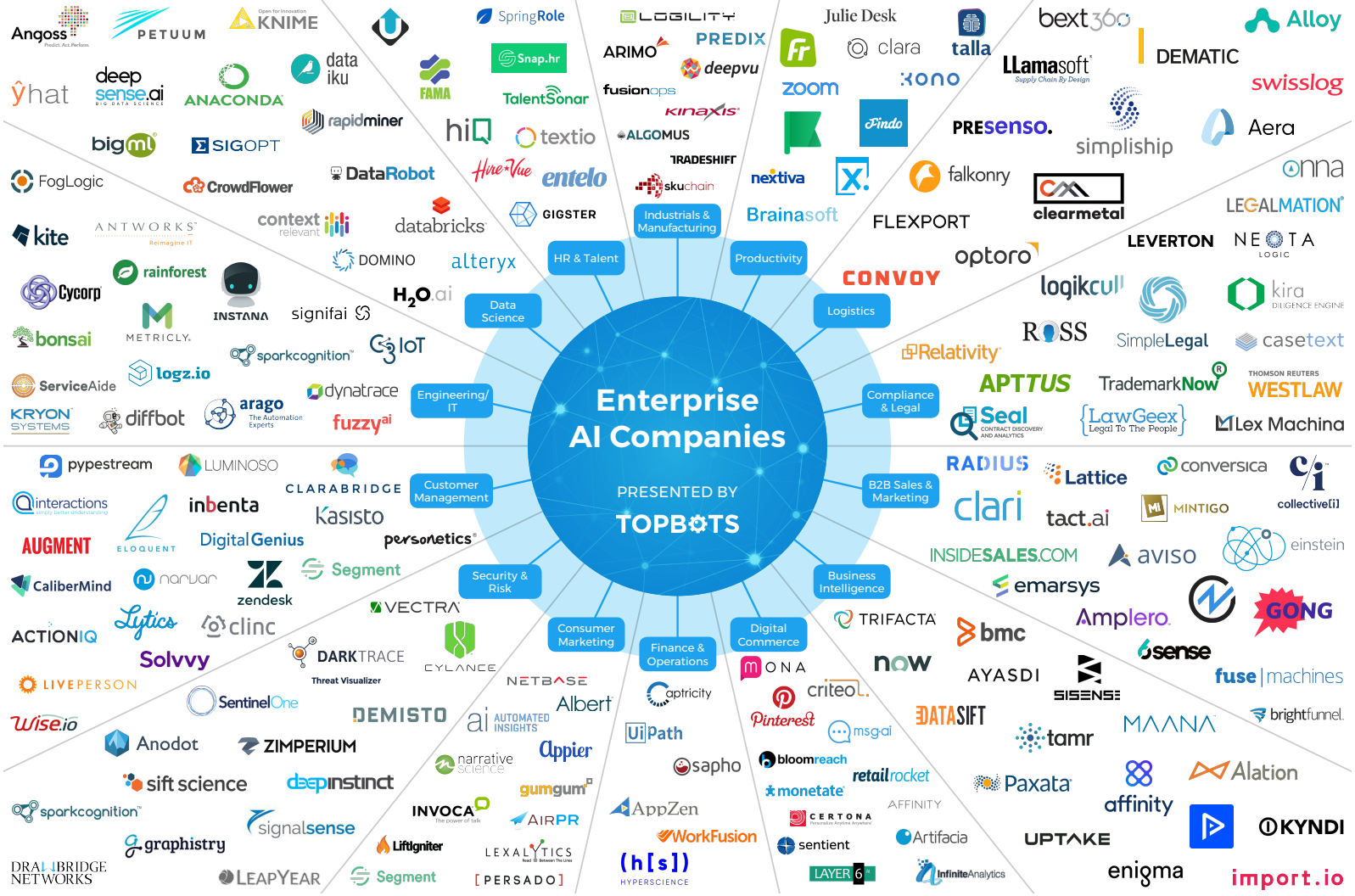
- Mercados horizontales compuestos por consumidores generales están dominados por **Alibaba, Amazon, Apple, Baidu, Facebook, Google, IBM, Microsoft y Tencent**.
- El resto son **silos aislados**.
- **Quién tiene la gente y los datos domina.**
- Los algoritmos se están transformando en un **commodity**.

Adaptaciones de las Empresas

IA en la **Internet**

Adaptaciones de las Empresas

IA en la **Empresa**



AI Lanscape 2018, Topbots, septiembre de 2018.

Adaptaciones de las Empresas

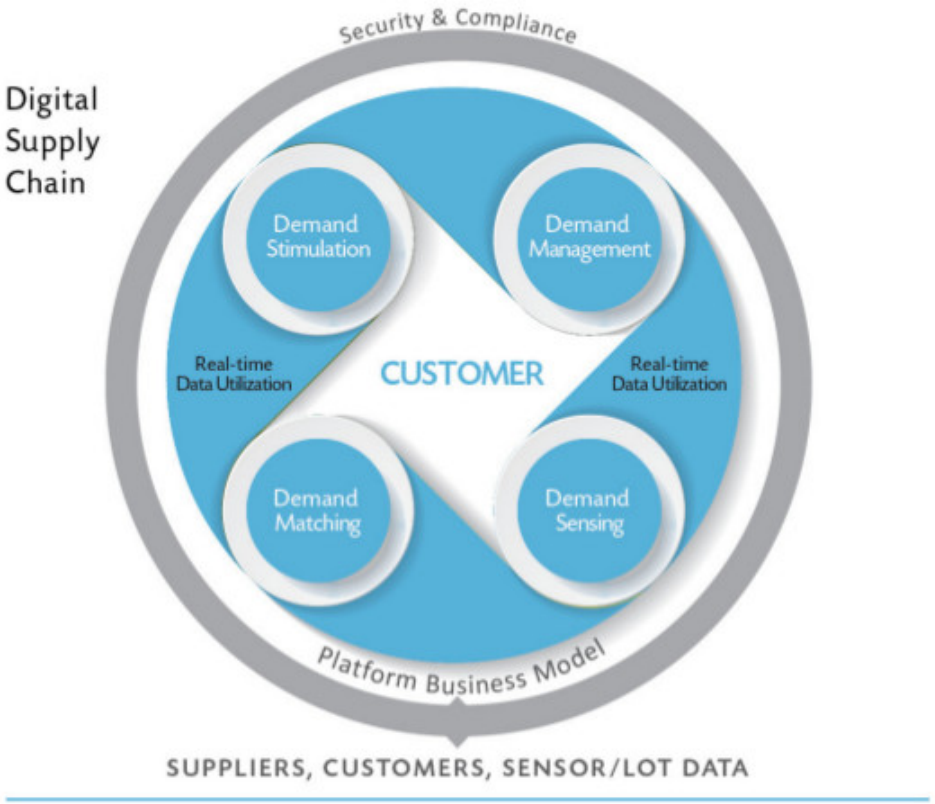
IA en el **Mundo Físico**

YuMi de ABB.

- **El último paso.**

Adaptaciones de las Empresas

Ciclo de **Percepción y Acción**



DRIVING DEMAND IN THE DIGITAL SUPPLY CHAIN: Algorithms and the Untapped Power of Applying Real-Time Big Data and AI/ML, DSCI Institute, The Center for Global Enterprise, enero de 2018.

Adaptaciones de las Empresas

Data Driven

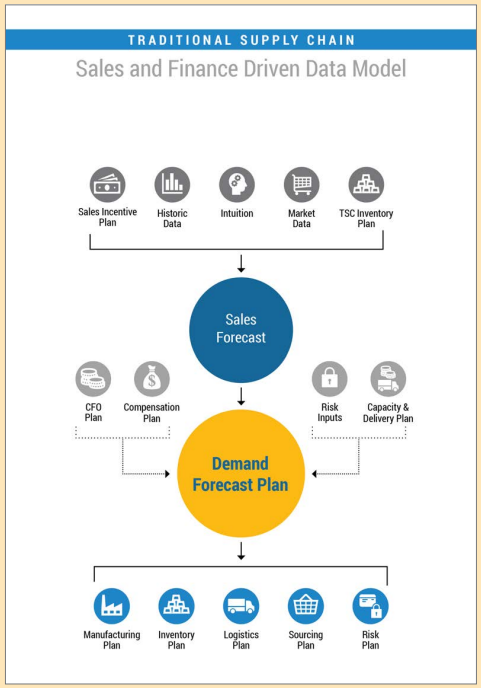


EXHIBIT 1

Driving Demand in the Digital Supply Chain 38

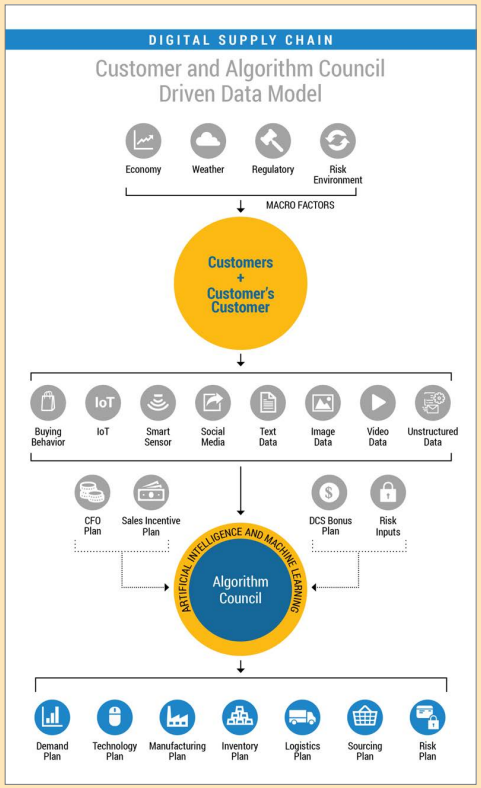


EXHIBIT 2

Driving Demand in the Digital Supply Chain 39

DRIVING DEMAND IN THE DIGITAL SUPPLY CHAIN: Algorithms and the Untapped Power of Applying Real-Time Big Data and AI/ML, DSCI Institute, The Center for Global Enterprise, enero de 2018.

Adaptaciones de las Empresas

Desafíos

- Las empresas en el mundo están **acelerando** la introducción de la inteligencia artificial.
- Es necesario educar a las personas, **partiendo por los líderes**.
- Fuerte cambio de cultura interna que **rompe los silos** típicos.
- Es necesaria la ingeniería **antes, durante y después**.

Conclusión

¿Preguntas?