

Inteligencia Artificial

Una Revolución

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

2019

Presentación

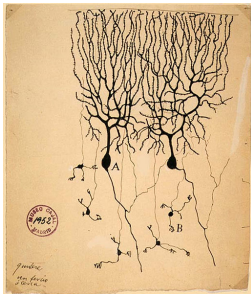
Científico y Emprendedor, Enfocado en Inteligencia Artificial



- 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 1, 1943

A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN NERVOUS ACTIVITY

WARREN S. MCCULLOCK 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 some specialized logical means for nets containing circles; and that for any logical expression qualifying certain conditions, one can find a net behaving in the fashion it describes. It is shown that many particular circuits among possible neurophysiological arrangements are equivalent, in the sense that, for every, and 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.**

Inteligencia Artificial

Estado del Arte

arXiv:1611.08235v2 [cs.CV] 6 Dec 2016

Full-Resolution Residual Networks for Semantic Segmentation in Street Scenes

Tobias Pohlen, Alexander Hermans, Markus Matthias, Bastian Leibe

Visual Computing Institute

RWTH Aachen University

tobias.pohlen@vis-vc.informatik.rwth-aachen.de, {hermans, matthias, leibe}@informatik.rwth-aachen.de

Abstract

Semantic image representation 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 to 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 results at the full image resolution. To achieve this problem we propose a novel ResNet-like architecture that exhibits strong localization and recognition power. We combine multi-scale context with pixel-level accuracy by using two processing streams: while our network does semantic segmentation at the full image resolution, enabling precise adherence to object boundaries. The other stream undergoes a sequence of pooling operations to obtain coarse features for recognition. The two streams are coupled at the full image resolution using residual units. Without additional pre-training and without pre-training, our approach achieves an intersection-over-union score of 77.3% 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 [1, 2, 3, 4, 5, 6], 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 content usually found in street scenes, such as cars, pedestrians, road, or sidewalks. In autonomous navigation it is used to estimate motion, e.g., in lane-keeping systems to detect image regions that are unlikely to contain objects of interest [7, 8] to improve object detection [9, 10, 11, 12].



(a) Pooling (b) Residual stream

(c) Pooling (d) Pooling stream

(e) Pooling (f) Pooling stream

(g) Pooling (h) Pooling stream

(i) Pooling (j) Pooling stream

(k) Pooling (l) Pooling stream

(m) Pooling (n) Pooling stream

(o) Pooling (p) Pooling stream

(q) Pooling (r) Pooling stream

(s) Pooling (t) Pooling stream

(u) Pooling (v) Pooling stream

(w) Pooling (x) Pooling stream

(y) Pooling (z) Pooling stream

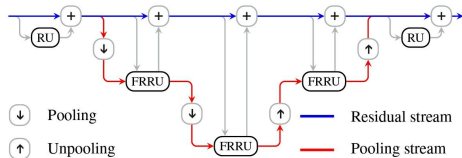


Figure 1. Example output and the abstract structure of our full-resolution residual network. The network has two processing streams. The residual stream (blue) stays at the full image resolution, the pooling stream (red) undergoes a sequence of pooling and unpooling operations. The two processing streams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

2016-12-6

Inteligencia Artificial

Estado del Arte

arXiv:1805.11714v1 [cs.CV] 29 May 2018

Deep Video Portraits

HYEONWOO KIM, Max Planck Institute for Informatics, Germany
 PABLO CARRERO, Technical Fellow
 AJUSH TAWARI and WEIPENG XU, Max Planck Institute for Informatics, Germany
 JUSTUS THIES and MATTHIAS NIESSNER, Technical University of Munich, Germany
 PATRICK PÉREZ, Technical Fellow
 CHRISTIAN RICHARDT, University of Bath, United Kingdom
 MICHAEL ZOLLHOFFER, Stanford University, United States of America
 CHRISTIAN THEOBALT, Max Planck Institute for Informatics, Germany

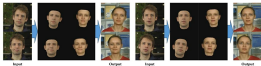


Fig. 1. Unlike current face reenactment approaches that only modify the expression of a target actor in a video, our novel deep-video-portrait approach enables full-frame-to-full-frame transfer of the target’s head/pose and facial expression and eye motion with a high level of photo-realism.

We present a novel approach that enables photo-realistic reenactment of portrait videos using only one target video. In contrast to existing approaches that require simultaneous recordings of both expression and eye motion, we transfer the full facial expression and motion, not just eye motion, from the source actor to a general video of a target actor. The core of our approach is a generative neural network with several sparse non-linearities. The network takes as input spatial coordinates of pixels in the target video, head pose and identity photo-realistic video frames for a given target actor. The module in this rendering is video transfer learned from spatial adversarial training, which is a novel cross-modal modified GAN module that makes the behavior of the spatially-coordinated input to video transfer appear as target video appearance, to render a realistic target video. We also introduce full-identity generation, which is a neural network that takes the trained network—thus taking full control of the

target. With the ability to freely combine source and target parameters, we are able to demonstrate a large variety of video transfer applications without explicitly modifying face, body or background. The network is more powerful than the full-face translation one in terms of editing and transfer high fidelity video editing. To demonstrate the high quality of our output, we conduct an extensive series of experiments and evaluation, which for instance a recent study shows that our video-clip are hard to detect.

CC BY 4.0 International license. arXiv:1805.11714v1 [cs.CV] 29 May 2018

Additional Key Words and Phrases: Facial Reenactment, Video Forensics, Editing, Deep Learning, Conditional GAN, Clustering in Video Transitions

ACM Reference Format:

Hyewoo Kim, Pablo Carrero, Ajush Tawari, Weipeng Xu, Justus Thies, Patrick Pérez, Michael Zollhofer, Christian Richardt, Matthias Nießner, and Christian Theobalt. 2018. Deep Video Portraits: Full-Frame-to-Full-Frame Transfer of Facial Expression and Motion. In Proceedings of the ACM Conference on Computer Graphics and Interactive Techniques (SIGGRAPH ’18). ACM, New York, NY, USA, 14 pages. <https://doi.org/10.1145/3196489>

Authors’ address: Hyewoo Kim, Max Planck Institute for Informatics, Informatik 3, Ludwig-Maximilians-Universität München, Amalienstr. 69, D-80539 München, Germany; Pablo Carrero, Max Planck Institute for Informatics, Informatik 3, Ludwig-Maximilians-Universität München, Amalienstr. 69, D-80539 München, Germany; Ajush Tawari, Max Planck Institute for Informatics, Informatik 3, Ludwig-Maximilians-Universität München, Amalienstr. 69, D-80539 München, Germany; Weipeng Xu, Max Planck Institute for Informatics, Informatik 3, Ludwig-Maximilians-Universität München, Amalienstr. 69, D-80539 München, Germany; Justus Thies, Technical University of Munich, School of Computer Science, Technical University of Munich, Garching, Germany; Patrick Pérez, Technical University of Munich, School of Computer Science, Technical University of Munich, Garching, Germany; Michael Zollhofer, Stanford University, Stanford, CA, USA; Christian Richardt, Technical University of Munich, School of Computer Science, Technical University of Munich, Garching, Germany; Matthias Nießner, Technical University of Munich, School of Computer Science, Technical University of Munich, Garching, Germany; Christian Theobalt, Max Planck Institute for Informatics, Informatik 3, Ludwig-Maximilians-Universität München, Amalienstr. 69, D-80539 München, Germany.

© 2018 ACM. This is not an authorized version of the manuscript. This version is not for distribution. The authors’ version of the manuscript is available at <https://arxiv.org/abs/1805.11714>.

2018 ACM. This is not an authorized version of the manuscript. This version is not for distribution. The authors’ version of the manuscript is available at <https://arxiv.org/abs/1805.11714>.

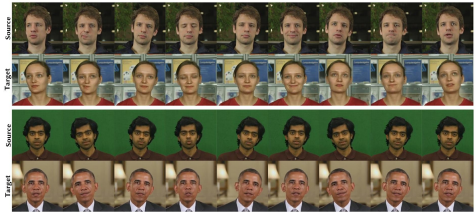


Fig. 5. Qualitative results of full-head reenactment: our approach enables full-frame target video portrait synthesis under full 3D head pose control. The output video portraits are photo-realistic and hard to distinguish from real videos. Note that even the shadow in the background of the second row 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



Figure 1: A blue 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 train dexterous in-hand manipulation policies which can perform vision-based object manipulation on a physical Shadow Hand. The training is performed in a simulated environment to learn a combination 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 more reliance toward human manipulation energy naturally including finger posing, wrist 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 [15]. We also include a video of our results. Images: /papers/167/abstract.html

1 Introduction

While dexterous manipulation of objects is an fundamental everyday task for humans, it is still challenging for autonomous robots. Modern-day robots are typically designed for specific tasks in controlled settings and are largely unable to adjust, adapt, and effectively. 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 generalist source of inspiration for robotic in-hand manipulation. The Shadow Hand (SH) is an example of a robotic hand designed for human-level dexterity. It has five fingers with a total of 38 degrees of freedom. The hand has been commercially available since 2015 (<http://www.shadow-robot.com>) and is open to OpenAI in alphabetical order.

Shao-An Huang, Benoit Bonet, Minh-Chau Do, Rishi Desai, Bob McKeown, Clark Barrett, Andrew Senior, Marko Popovic, Shuang Tang, Shi-Min Yue, James Tomlin, Yoram Singer, Bob Schölkopf, Peter Wabnitz, Li-Jian Wang, Wojciech Zarembki

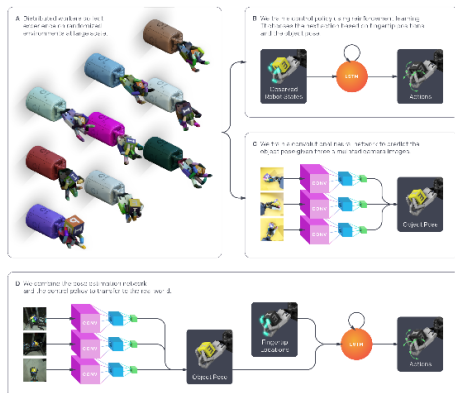


Figure 2: System Overview. (a) We use a large distribution of simulations with randomized parameters and appearances to collect data for both the control policy and vision-based pose estimator. (b) The control policy receives observed robot states and rewards from the distributed simulations and learns 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 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 Mikolov¹, 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 unsupervised weights. We find that these representations are surprisingly good at capturing syntactic and semantic regularities in language, and that such relationships are observed by a relation-specific vector offset. The above vector-oriented reasoning based on the offsets between words. For example, the multi-formal relationship is automatically learned, and with the learned vector representations, “King - Man + Queen” results in a vector very close to “Queen”. We demonstrate that the word co-occurrence system replicates the scores of syntactic analogy questions (provided with this paper), and is able to correctly answer almost 80% of the questions. We demonstrate that the word vector captures semantic regularities by using the vector offset method to answer Stanford 2012 task 2 questions. Remarkably, the method outperforms the best previous system.

1 Introduction

A defining feature of neural network language models is their representation of words in high-dimensional vector spaces. In these models (Bengio et al., 2003; Mikolov, 2005; 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 needs to learn a discrete vector that has an inherent relationship to one another, a continuous space model needs to learn a word vector 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 (Schubert and Weston, 2008; Tenen et al., 2010) where learned word representations are used with specialized 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 v_i , and focus on the straightforward relation, we observe that $v_{\text{Queen}} - v_{\text{Man}} = v_{\text{King}} - v_{\text{Man}} = v_{\text{Queen}} - v_{\text{King}}$ and so on. Perhaps more surprisingly, we find that this also the case for a variety of semantic relations, as measured by the Stanford 2012 task of measuring relation similarity.

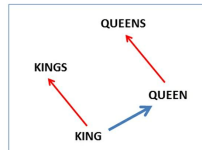
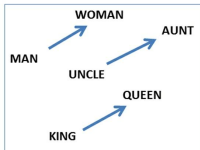


Figure 2: Left panel shows vector offsets for three word pairs illustrating the gender relation. Right panel shows a different projection, and the singular/plural relation for two words. In high-dimensional space, multiple relations can be embedded for a single word.

2013-6-10

Inteligencia Artificial

Estado del Arte

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, pingduo_jiang, satori, zffeng, qyl, marcin@google.com

Wolfgang Macherey, Miroslav Miklavic, Yanzhao Chen, Qin Gao, Klaus Moravets, Jeff Klingner, Apurva Shah, Mikko Juhanen, Xiaohang Li, Lukasz Kaiser, Stephan Gouws, Yoshitaka Kato, Tamas Kocz, Hideto Kozuma, Keith Stevens, George Szepesvari, Nishant Prasad, Wei Yang, Chifeng Yang, Jason Riesa, Jason Riesa, Alex Buzhko, Oriol Vinyals, Georgy Corrado, Marc'Antonio Frete, Jeffrey Dean

Abstract

Neural Machine Translation (NMT) is an end-to-end learning approach for automatic 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 argued 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 Google's Neural Machine Translation system, which attempts to address many of these issues. Our model consists of a deep LSTM network with a residual unit's flexible layers using residual connections as well as attention connections from the decoder network to the encoder. To improve translation and inference efficiency, we use an attention mechanism to compute the hidden layer of the decoder to use for each step of the encoder. In addition to these modifications, we employ low-precision arithmetic during inference computation. To improve handling of rare words, we divide words into a hierarchy of difficulty and model each word's "difficulty" for both input and output. This method provides a good balance between the flexibility of "classical" statistical models and the efficiency of "neural" statistical models, notably enabling translation of rare words, and ultimately improves the overall accuracy of the system. Our beam search decoder employs a length-normalization procedure and uses a coverage penalty, which encourages generation of shorter sentences that is more likely to cover all the words in the source sentence. We directly optimize the translation BLEU scores, we consider replacing the model by using reinforcement learning, but we found that the improvement in the BLEU scores did not reflect the human evaluation. On the WMT14 English-to-French and English-to-German benchmarks, Google's NMT system consistently results in superior BLEU. Using a human-like evaluation on a set of unrelated phrase sentences, a human translator rates an average of 80% compared to Google's phrase-based production system.

1 Introduction

Neural Machine Translation (NMT) [1, 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 separate neural networks (NNs), one to generate the input text sequence and one to generate the associated output. NMT is often accompanied by an encoder-decoder architecture [2] which helps it cope effectively with long input sequences.

An advantage of Neural Machine Translation is that it achieves more holistic designs than traditional phrase-based machine translation [3]. In practice, however, NMT systems need to be aware of accuracy, data efficiency and translation speed when training. In this work, we report on our efforts to improve the very best publicly available translation systems. Three related weaknesses of Neural Machine Translation are

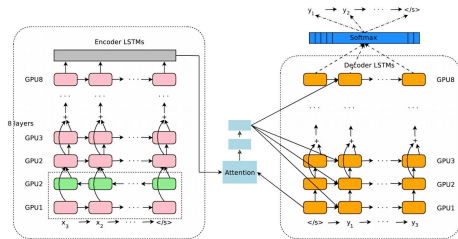


Figure 1: The model architecture of GNMT, Google's Neural Machine Translation system. On the left is the encoder network, on the right is the decoder network, in the middle is the attention module. The bottom encoder layer is bi-directional: the pink nodes gather information from left to right while 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

Romana Pfaffen, Caisheng Xiong and Richard Socher
{rpfaffen, cxiong, rsocher}@stanford.edu

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 reinforcement and a new training method. This method combines standard supervised word prediction and reinforcement learning (RL). Models trained only with the former often exhibit “stop-and-stare” – they assume good truth is provided at each step during training. However, when standard word prediction is combined with the global reward provided by testing 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 41.6 BLEU@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 parts.

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 extracts facts summaries by copying parts of the input (Nesher 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 BLEU@1 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-2000 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 to the CNN/Daily Mail dataset (Hermann et al., 2015), which contains input sequences of up to 500 tokens and multi-sentence summaries of up to 100 tokens. The analysis by Nallapati et al. (2016) illustrates a key problem with attentional encoder-decoder models: they often generate redundant 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) (Stanford, 2006). To our knowledge, this is the first model for abstractive summarization on the NYT dataset. We introduce a key attention mechanism and a new training 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 country's 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.

ML₁ 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.

RL₁ 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+RL₁ 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 has 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

Inteligencia Artificial

Estado del Arte

arXiv:1412.2306v2 [cs.CV] 14 Apr 2015

Deep Visual-Semantic Alignments for Generating Image Descriptions

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

Abstract

We present a model that generates natural language descriptions of images and their regions. Our approach leverages distant supervision and their anchor 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 Multiresolution Neural Network architecture that uses the regional alignments to learn to generate natural descriptions of image regions. We demonstrate that our alignment model produces state-of-the-art results in several experiments on Flickr3K, Flickr30K and MSCOCO datasets. We also show that the generated descriptions significantly outperform textual baselines on both full images and on a new dataset of region-level annotations.



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

generating these descriptions of image regions (Fig. 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. Ideally, the model should be free of assumptions about specific hand-crafted templates, rules or capacities 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 [1, 76, 77], but these descriptions analyze variations of several hundred visual instances in the images on captions. One can imagine that we can leverage these large image-sentence datasets by treating the sentences as weak labels, to which candidate regions of words correspond to some particular, but unknown location in the image. Our approach is to learn 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 regions of the image that they describe.

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 [10]. However, the remarkable ability has proven to be a herculean 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 concepts and prior progress has been achieved on these endeavors [4, 11]. However, while closed vocabularies of visual concepts contain a convenient modeling assumption, they are rarely intuitive when compared to the common sense of multi-observations that a human can compare.

Some pioneering approaches that address the challenge of generating image descriptions have been developed [70, 13]. However, these models rely only on learning natural 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

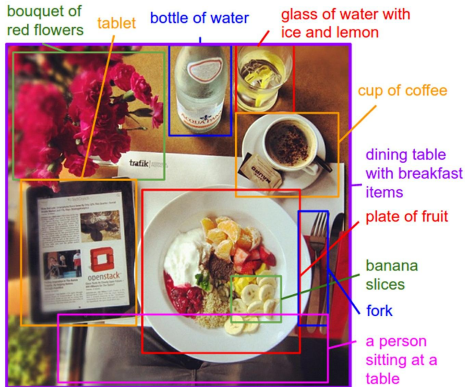


Figure 1. Motivation/Concept Figure: Our model treats language as a rich label space and generates descriptions of image regions.

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

Han Zhang¹, Tao Xu¹, Hongsheng Li², Shouping Zhang¹, Xiaohu Huang¹, Xiangqun Wang¹, Dimitris Metaxas³

¹Department of Computer Science, Rutgers University

²Department of Computer Science and Engineering, Lehigh University

³Department of Electronic Engineering, The Chinese University of Hong Kong

⁴Department 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 accurate 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 shapes and basic colors of the objects 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 modify objects and add compelling details with the refinement process. Samples generated by StackGAN are more plausible than those generated by existing approaches. Arguably, our StackGAN is the first text-to-image generator that can generate high-resolution images with photo-realistic details.



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.

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 description is multi-modal. There are a larger number of images that correspond to the given text descriptions.

Recently, Generative Adversarial Networks (GAN) [1, 2, 10] have shown promising results in modeling complex multi-modal data and generating high-quality images. Reed et al. [11, 12] demonstrated that GAN can effectively gen-

1. Introduction

Generating photo-realistic images from text descriptions is a challenging problem. It has tremendous applications including photo editing and computer-aided image search.

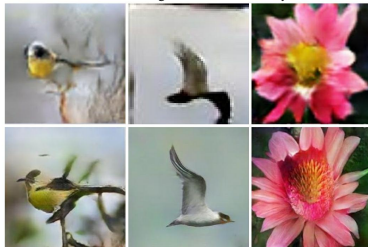
¹han@cs.rutgers.edu

This bird has a yellow belly and tarsus, grey back, wings, and brown throat, nape with a black face

This bird is white with some black on its head and wings, and has a long orange beak

This flower has overlapping pink pointed petals surrounding a ring of short yellow filaments

(a) Stage-I images



(b) Stage-II images

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:10.1016/j.artint.2016.01.002

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

David Silver^a, Alex Shaw^a, Clark S. Ballester^a, Arthur Guez^a, Laurent Eliezer^a, George Siu^a, Daniel Hui^a, Matt Botvinick^a, Sander Sutskever^a, Ilya Sutskever^a, Mark Denker^b, Simon Osindero^b, Mike Hvan^a, Nir Kalchbrenner^a, Tom Silverthorn^a, David Greaves^a, Markus Luck^c, Kevin Leventkov^d, Thore Graepel^e & Demis Hassabis^a

The game of Go has long been viewed as the most challenging of board 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 a neural network to evaluate board positions and a policy network to select moves. This neural network is trained by a self-supervised learning process that involves learning from human expert games, and generalized to novel situations by a novel, multi-step reinforcement learning process. The neural network plays Go at the level of state-of-the-art professional human players. Without any handicaps, the neural network plays Go at the level of state-of-the-art professional human players. We combine this neural network with a policy network to play Go at the level of state-of-the-art professional human players. We combine this neural network with a value network to play Go at the level of state-of-the-art professional human players. We combine this neural network with a self-play program to play Go at the level of state-of-the-art professional human players. We combine this neural network with a self-play program to play Go at the level of state-of-the-art professional human players.

All games of perfect information have an optimal value function V^* , which determines the outcome of the game. In a zero-sum two-player game, this value function is the value of the game for the player to move. In a non-zero-sum game, the value function is the value of the game for the player to move, given the actions of the other players. In a game of perfect information, the value function is the value of the game for the player to move, given the actions of the other players. In a game of imperfect information, the value function is the value of the game for the player to move, given the actions of the other players, and the beliefs of the player about the hidden information. In a game of imperfect information, the value function is the value of the game for the player to move, given the actions of the other players, and the beliefs of the player about the hidden information. In a game of imperfect information, the value function is the value of the game for the player to move, given the actions of the other players, and the beliefs of the player about the hidden information.

Supervised learning of policy networks
For the first stage of the training program, we build on prior work on supervised learning of policy networks. The goal is to learn a policy network that outputs a probability distribution over legal moves given a board position. This is done by training a neural network to predict the next move in a sequence of moves from a set of human expert games. The neural network is trained to predict the next move in a sequence of moves from a set of human expert games. The neural network is trained to predict the next move in a sequence of moves from a set of human expert games.

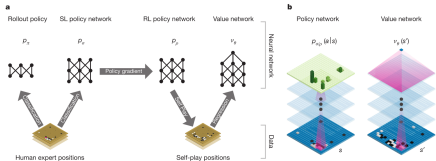


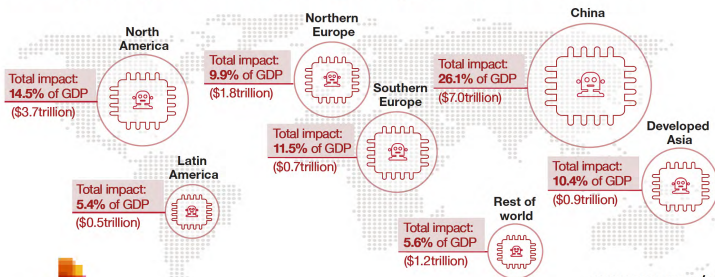
Figure 1 | Neural network training pipeline and architecture. a. A fast rollout policy p_1 and supervised learning (SL) policy network p_2 are trained to predict human expert moves in a data set of positions. A reinforcement learning (RL) policy network p_3 is initialized to the SL policy network, and 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_1 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 θ (SL policy network) or μ (RL policy network), and outputs a probability distribution $p_{\theta}(a|s)$ or $p_{\mu}(a|s)$ 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

Sizing the prize – Which regions gain the most from AI?



www.pwc.com/ai
#AIrevolution

© 2017 PricewaterhouseCoopers LLP. All rights reserved.

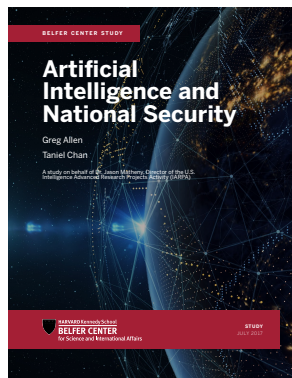
Inteligencia Artificial

La Fiebre de la Inteligencia Artificial



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.



● 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

Psychological targeting as an effective approach to digital mass persuasion

L. S. Mars*¹, M. Reissman², G. Nave³, and J. S. Bailly⁴*

¹Yale University School of Medicine, University of Pennsylvania, New York City; ²NYU Stern School of Business, University of Pennsylvania, Philadelphia, PA; ³Wharton University of Pennsylvania, Philadelphia, PA; ⁴Harvard University, Cambridge, MA; ⁵University of Michigan, Ann Arbor, Michigan

Received July 14, 2016; Accepted September 1, 2016; Published October 11, 2016 (first published as a preprint in *medRxiv* on July 15, 2016)

People are exposed to persuasive communication across many contexts, and many contexts are designed to encourage people to use health, education, or political products or to take the specific action. Evidence shows that such persuasive appeals are more effective in influencing behavior when they are tailored to a target user's psychological characteristics. However, the emergence of large-scale psychological persuasion in the real world has been limited by the opportunity costs of targeted digital advertisements, such as their Facebook Likes or Tweets. Capitalizing on the ease of psychological persuasion from digital devices, we tested the effects of psychological persuasion on people's actual behavior in an experimentally valid setting, in three field experiments. The results show that targeted psychological persuasion can be used in a wide variety of settings to influence behavior in the real world. Psychological persuasion can be used to influence behavior in an experimentally valid setting, in three field experiments. The results show that targeted psychological persuasion can be used in a wide variety of settings to influence behavior in the real world. Psychological persuasion can be used to influence behavior in an experimentally valid setting, in three field experiments. The results show that targeted psychological persuasion can be used in a wide variety of settings to influence behavior in the real world.

Significance
Building on recent advances in the assessment of people's actual behavior from digital devices, this paper demonstrates the effectiveness of psychological persuasion that is, the use of persuasive appeals to a user's specific psychological characteristics. On the one hand, this form of persuasion can be used to help people take more healthful and positive actions. On the other hand, it could be used to help people take more harmful actions. Psychological persuasion can be used to help people take more healthful and positive actions. On the other hand, it could be used to help people take more harmful actions. Psychological persuasion can be used to help people take more healthful and positive actions. On the other hand, it could be used to help people take more harmful actions.

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 unlimited number of crossword puzzles!

Low Openness



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

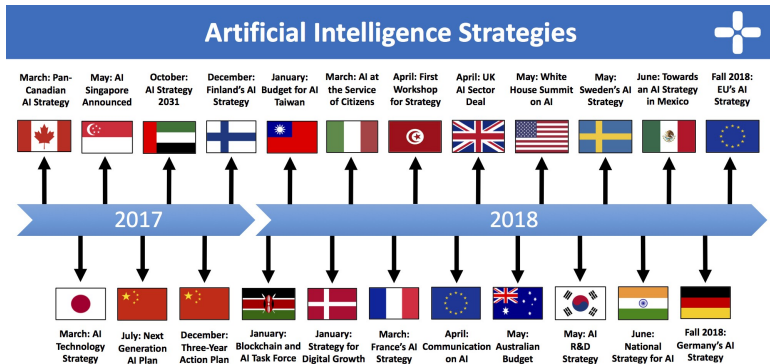
PNAS

2017-11-28

- **"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



2018-07-13 | Politics + AI | Tim Dutton

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.

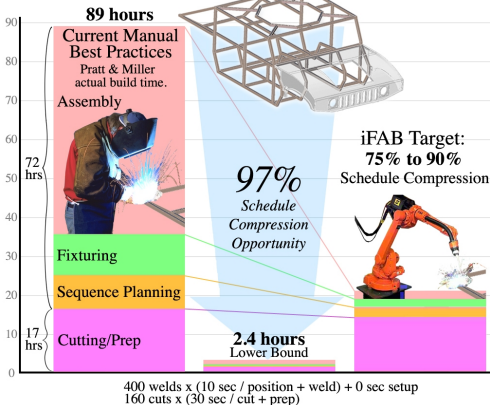
La Sociedad Humana

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.

iFab Opportunity



Instruction Generation

Robotic Assembly

MACHINE INSTRUCTION



Augmented Fixturing

HUMAN INSTRUCTION



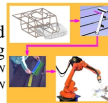
Virtual Layout

HUMAN INSTRUCTION



Automated Process Planning

MACHINE & HUMAN INSTRUCTION



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

Machine earning

Mentions of AI and machine learning on earnings calls of public companies



Source: Bloomberg

Economist.com

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

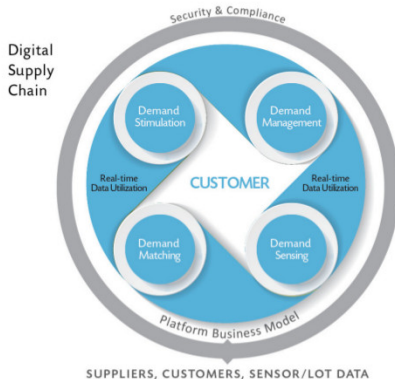


YuMi de ABB.

- ▶ **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

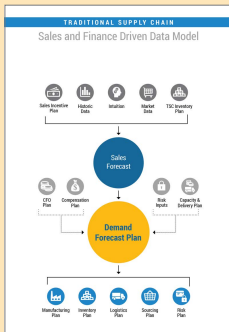


EXHIBIT 1

Driving Demand in the Digital Supply Chain

38

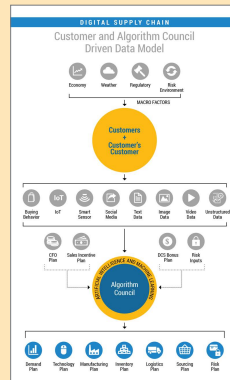


EXHIBIT 2

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?