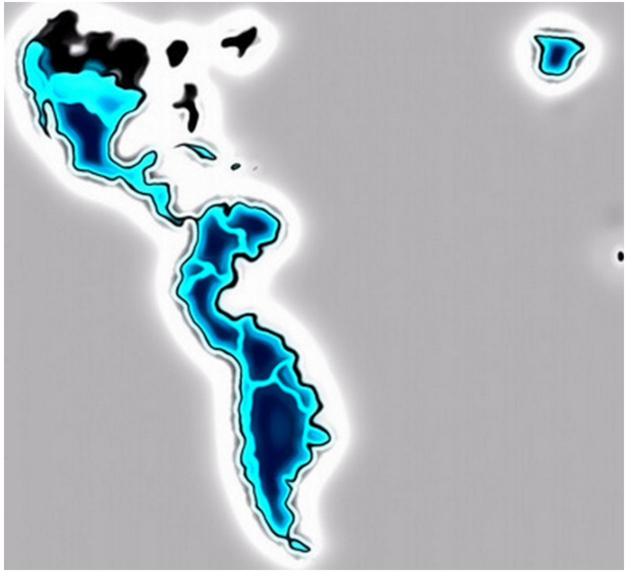
# Redes Neuronales Profundas en Python

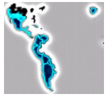
Dr. Erik Zamora
IPN, UPIITA, CIC



# Objetivos y contenido

Explicar

- ¿Por qué las redes neuronales son valiosas?
- ¿Qué aplicaciones hay en la investigación de la ciencia física?
- ¿Qué son las redes neuronales?
- ¿Cómo entrenar las redes neuronales? (para detectar y segmentar imágenes)



# Logros tecnológicos más impresionantes

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 Grew
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.

# Logros tecnológicos más impresionantes

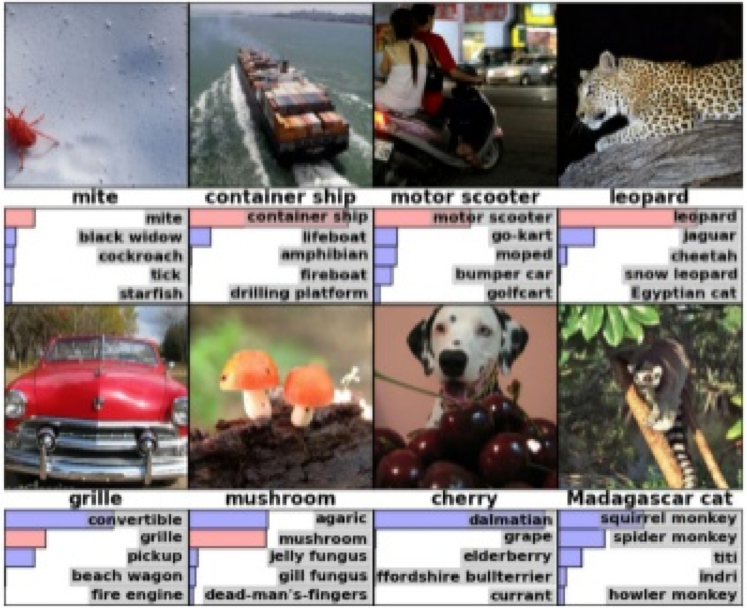

Nvidia usando redes neuronales

# Logros tecnológicos más impresionantes

## ImageNet Challenge

IMAGENET

- 1,000 object classes (categories).
- Images:
  - 1.2 M train
  - 100k test.



# Logros tecnológicos más impresionantes

## ImageNet Classification with Deep Convolutional Neural Networks

**Alex Krizhevsky**
University of Toronto
kriz@cs.utoronto.ca

**Ilya Sutskever**
University of Toronto
ilya@cs.utoronto.ca

**Geoffrey E. Hinton**
University of Toronto
hinton@cs.utoronto.ca

### Abstract

We trained a large, deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes. On the test data, we achieved top-1 and top-5 error rates of 37.5% and 17.0% which is considerably better than the previous state-of-the-art. The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers with a final 1000-way softmax. To make training faster, we used non-saturating neurons and a very efficient GPU implementation of the convolution operation. To reduce overfitting in the fully-connected layers we employed a recently-developed regularization method called "dropout" that proved to be very effective. We also entered a variant of this model in the ILSVRC-2012 competition and achieved a winning top-5 test error rate of 15.3%, compared to 26.2% achieved by the second-best entry.

## Squeeze-and-Excitation Networks

Jie Hu*
Momenta
hujie@momenta.ai

Li Shen*
University of Oxford
lishen@robots.ox.ac.uk

Gang Sun*
Momenta
sungang@momenta.ai

### Abstract

*Convolutional neural networks are built upon the convolution operation, which extracts informative features by fusing spatial and channel-wise information together within local receptive fields. In order to boost the representational power of a network, much existing work has shown the benefits of enhancing spatial encoding. In this work, we focus on channels and propose a novel architectural unit, which we term the "Squeeze-and-Excitation"(SE) block, that adaptively recalibrates channel-wise feature responses by explicitly modelling interdependencies between channels. We demonstrate that by stacking these blocks together, we can construct SENet architectures that generalise extremely well across challenging datasets. Crucially, we find that SE blocks produce significant performance improvements for existing state-of-the-art deep architectures at slight computational cost. SENets formed the foundation of our ILSVRC 2017 classification submission which won first place and significantly reduced the top-5 error to 2.251%, achieving a ∼25% relative improvement over the winning entry of 2016.*

requiring additional supervision. One such approach was popularised by the Inception architectures [14, 39], which showed that the network can achieve competitive accuracy by embedding multi-scale processes in its modules. More recent work has sought to better model spatial dependence [1, 27] and incorporate spatial attention [17].

In contrast to these methods, we investigate a different aspect of architectural design - the channel relationship, by introducing a new architectural unit, which we term the "*Squeeze-and-Excitation*" (SE) block. Our goal is to improve the representational power of a network by explicitly modelling the interdependencies between the channels of its convolutional features. To achieve this, we propose a mechanism that allows the network to perform *feature recalibration*, through which it can learn to use global information to selectively emphasise informative features and suppress less useful ones.

The basic structure of the SE building block is illustrated in Fig. 1. For any given transformation $\mathbf{F}_{tr} : \mathbf{X} \rightarrow \mathbf{U}$, $\mathbf{X} \in \mathbb{R}^{W' \times H' \times C'}$, $\mathbf{U} \in \mathbb{R}^{W \times H \times C}$, (e.g. a convolution or a set of convolutions), we can construct a corresponding SE block to perform feature recalibration as follows.

# Aplicaciones en la investigación física

nature COMMUNICATIONS

ARTICLE

Received 19 Feb 2014 | Accepted 4 Jun 2014 | Published 2 Jul 2014

DOI: 10.1038/ncomms5308

## Searching for exotic particles in high-energy physics with deep learning

P. Baldi[1], P. Sadowski[1] & D. Whiteson[2]

Collisions at high-energy particle colliders are a traditionally fruitful source of exotic particle discoveries. Finding these rare particles requires solving difficult signal-versus-background classification problems, hence machine-learning approaches are often used. Standard approaches have relied on 'shallow' machine-learning models that have a limited capacity to learn complex nonlinear functions of the inputs, and rely on a painstaking search through manually constructed nonlinear features. Progress on this problem has slowed, as a variety of techniques have shown equivalent performance. Recent advances in the field of deep learning make it possible to learn more complex functions and better discriminate between signal and background classes. Here, using benchmark data sets, we show that deep-learning methods need no manually constructed inputs and yet improve the classification metric by as much as 8% over the best current approaches. This demonstrates that deep-learning approaches can improve the power of collider searches for exotic particles.

**Table 1 | Performance for Higgs benchmark.**

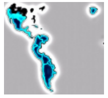
| Technique | Low-level | High-level | Complete |
|---|---|---|---|
| *AUC* | | | |
| BDT | 0.73 (0.01) | 0.78 (0.01) | 0.81 (0.01) |
| NN | 0.733 (0.007) | 0.777 (0.001) | 0.816 (0.004) |
| DN | 0.880 (0.001) | 0.800 (<0.001) | 0.885 (0.002) |
| *Discovery significance* | | | |
| NN | $2.5\sigma$ | $3.1\sigma$ | $3.7\sigma$ |
| DN | $4.9\sigma$ | $3.6\sigma$ | $5.0\sigma$ |

Comparison of the performance of several learning techniques: boosted decision trees (BDT), shallow neural networks (NN), and deep neural networks (DN) for three sets of input features: low-level features, high-level features and the complete set of features. Each neural network was trained five times with different random initializations. The table displays the mean area under the curve (AUC) of the signal-rejection curve in Fig. 7, with s.d. in parentheses as well as the expected significance of a discovery (in units of Gaussian $\sigma$) for 100 signal events and $1{,}000 \pm 50$ background events.

**Table 2 | Performance comparison for the SUSY benchmark.**

| Technique | Low-level | High-level | Complete |
|---|---|---|---|
| *AUC* | | | |
| BDT | 0.850 (0.003) | 0.835 (0.003) | 0.863 (0.003) |
| NN | 0.867 (0.002) | 0.863 (0.001) | 0.875 (<0.001) |
| $NN_{dropout}$ | 0.856 (<0.001) | 0.859 (<0.001) | 0.873 (<0.001) |
| DN | 0.872 (0.001) | 0.865 (0.001) | 0.876 (<0.001) |
| $DN_{dropout}$ | 0.876 (<0.001) | 0.869 (<0.001) | 0.879 (<0.001) |
| *Discovery significance* | | | |
| NN | $6.5\sigma$ | $6.2\sigma$ | $6.9\sigma$ |
| DN | $7.5\sigma$ | $7.3\sigma$ | $7.6\sigma$ |

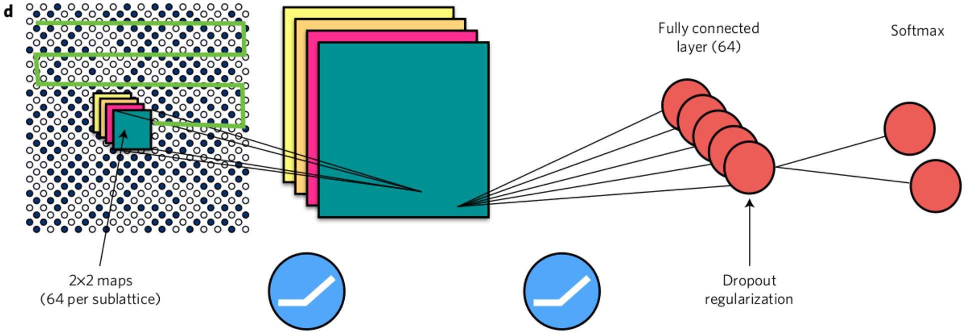
BDT, boosted decision tree; DN, deep neural network; NN, shallow neural network; SUSY, supersymmetry particle.
Each model was trained five times with different weight initializations. The mean area under the curve (AUC) is shown with s.d. in parentheses as well as the expected significance of a discovery (in units of Gaussian $\sigma$) for 100 signal events and $1{,}000 \pm 50$ background events.



# Aplicaciones en la investigación física

nature physics

PUBLISHED O



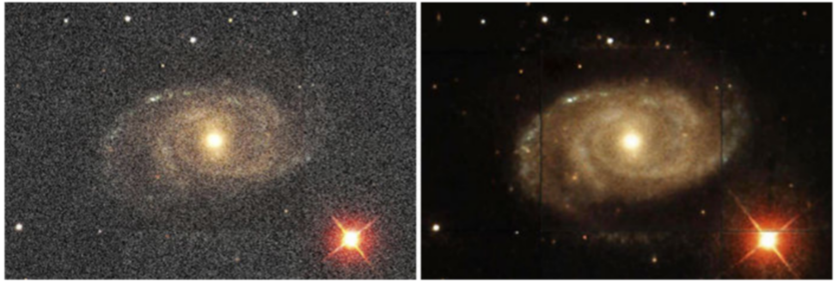

# Machine learning phases of matter

Juan Carrasquilla[1*] and Roger G. Melko[1,2]

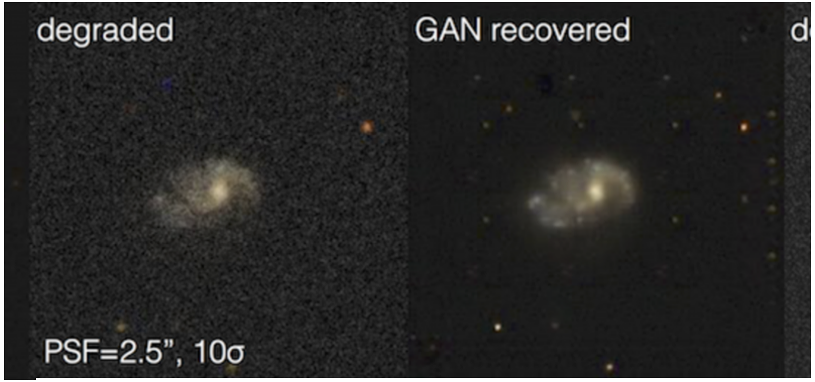
**Condensed-matter physics is the study of the collective behaviour of infinitely complex assemblies of electrons, nuclei, magnetic moments, atoms or qubits[1]. This complexity is reflected in the size of the state space, which grows exponentially with the number of particles, reminiscent of the 'curse of dimensionality' commonly encountered in machine learning[2]. Despite this curse, the machine learning community has developed techniques with remarkable abilities to recognize, classify, and characterize complex sets of data. Here, we show that modern machine learning architectures, such as fully connected and convolutional neural networks[3], can identify phases and phase transitions in a variety of condensed-matter Hamiltonians. Readily programmable through modern software libraries[4,5], neural networks can be trained to detect multiple types of order parameter, as well as highly non-trivial states with no conventional order, directly from raw state configurations sampled with Monte Carlo[6,7].**

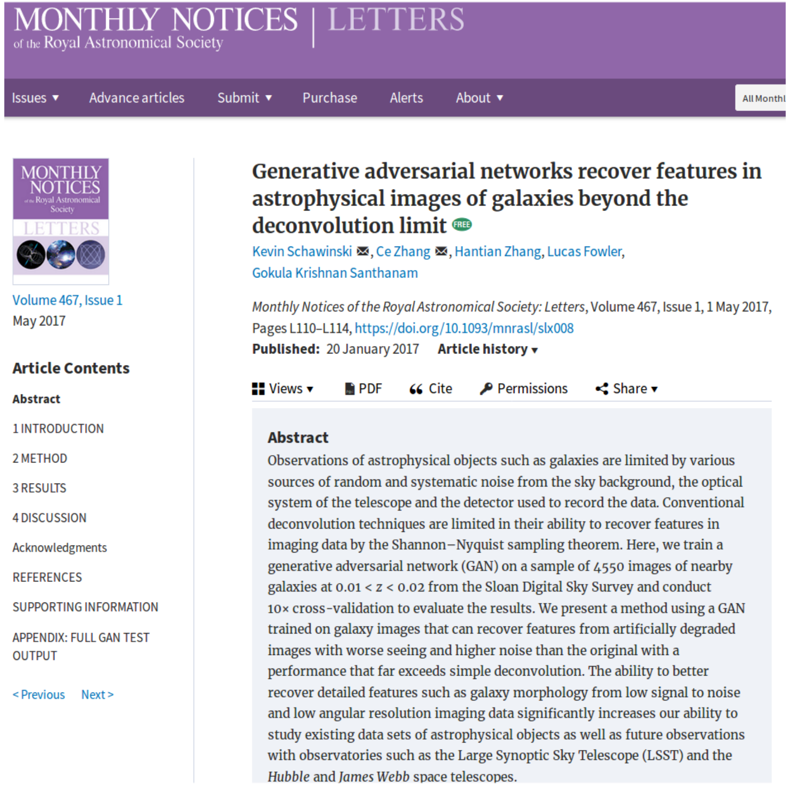
phase transitions in correlated many-body systems. In particular, we have argued that neural networks encode information about conventional ordered phases by learning the order parameter of the phase, without knowledge of the energy or locality conditions of the Hamiltonian. Furthermore, we have shown that neural networks can encode basic information about unconventional phases such as the ones present in the square-ice model and the Ising lattice gauge theory, as well as Anderson localized phases. These results indicate that neural networks have the potential to represent ground state wavefunctions. For instance, ground states of the toric code[1,8]

# Aplicaciones en la investigación física



AI that "knows" what a galaxy should look like transforms a fuzzy image (left) into a crisp one (right).
KIYOSHI TAKAHASE SEGUNDO/ALAMY STOCK PHOTO



MONTHLY NOTICES of the Royal Astronomical Society | LETTERS

Issues ▾ Advance articles Submit ▾ Purchase Alerts About ▾

Volume 467, Issue 1
May 2017

**Article Contents**



< Previous Next >

## Generative adversarial networks recover features in astrophysical images of galaxies beyond the deconvolution limit

Kevin Schawinski, Ce Zhang, Hantian Zhang, Lucas Fowler, Gokula Krishnan Santhanam

*Monthly Notices of the Royal Astronomical Society: Letters*, Volume 467, Issue 1, 1 May 2017, Pages L110–L114, https://doi.org/10.1093/mnrasl/slx008
**Published:** 20 January 2017 **Article history** ▾

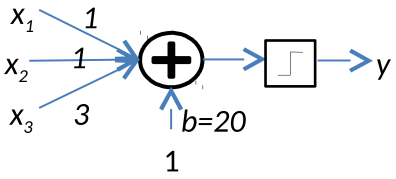
Views ▾ PDF Cite Permissions Share ▾

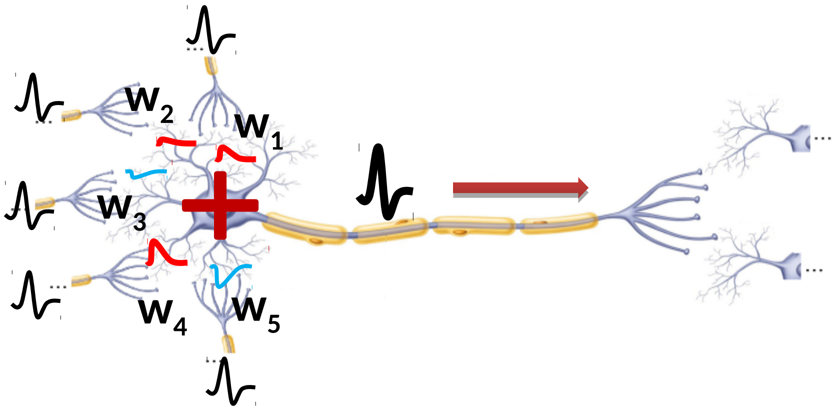
**Abstract**

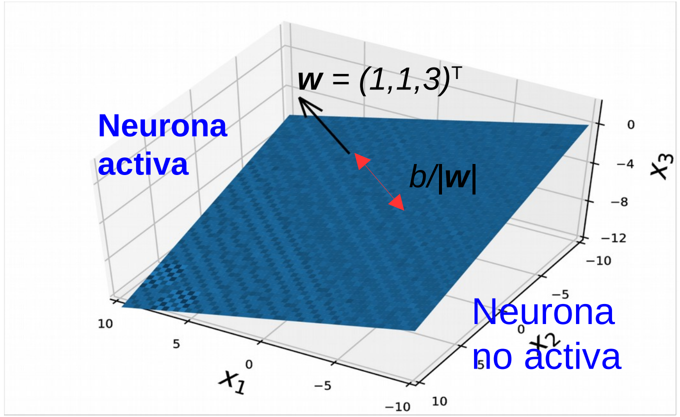
Observations of astrophysical objects such as galaxies are limited by various sources of random and systematic noise from the sky background, the optical system of the telescope and the detector used to record the data. Conventional deconvolution techniques are limited in their ability to recover features in imaging data by the Shannon–Nyquist sampling theorem. Here, we train a generative adversarial network (GAN) on a sample of 4550 images of nearby galaxies at $0.01 < z < 0.02$ from the Sloan Digital Sky Survey and conduct 10× cross-validation to evaluate the results. We present a method using a GAN trained on galaxy images that can recover features from artificially degraded images with worse seeing and higher noise than the original with a performance that far exceeds simple deconvolution. The ability to better recover detailed features such as galaxy morphology from low signal to noise and low angular resolution imaging data significantly increases our ability to study existing data sets of astrophysical objects as well as future observations with observatories such as the Large Synoptic Sky Telescope (LSST) and the *Hubble* and *James Webb* space telescopes.

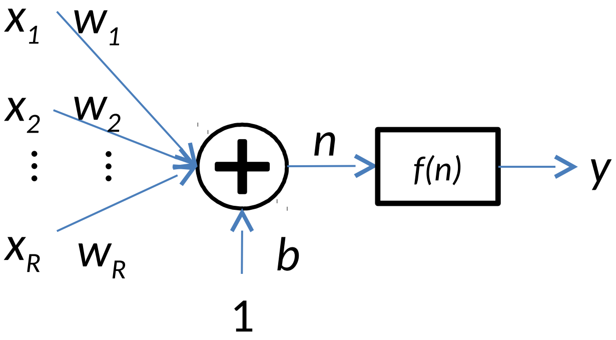
The success of the project points to a more "data-driven" future for astrophysics in which information is learned automatically from data, instead of manually crafted physics models. ETH Zurich is hosting this work on the space.ml cross-disciplinary astrophysics/computer-science initiative, where the code is available to the general public.

# Pero...¿Qué es una neurona artificial?



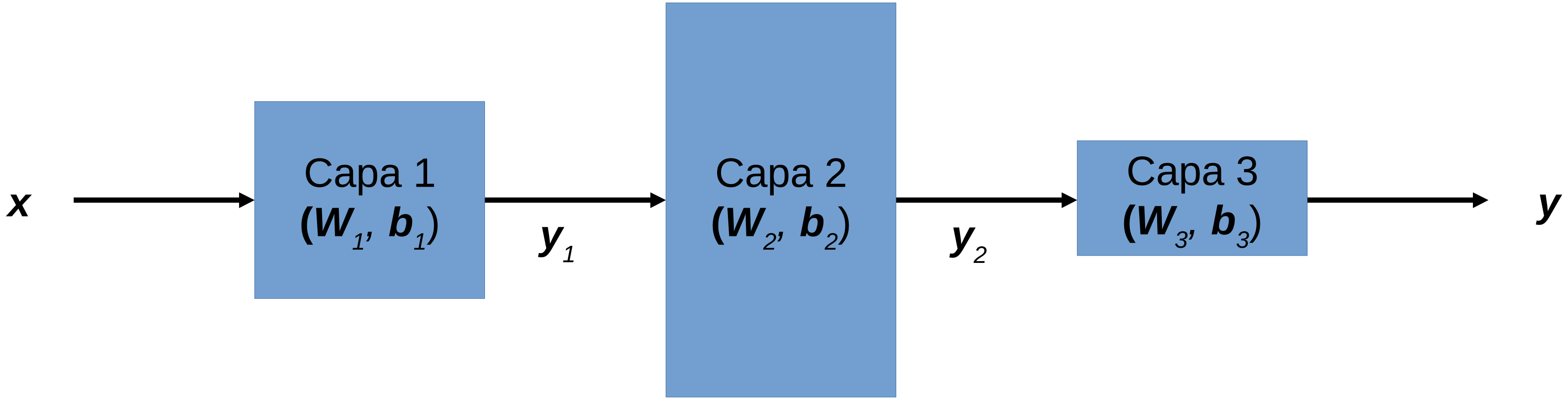




$$y = f\left(\mathbf{w}^T \mathbf{x} + b\right)$$

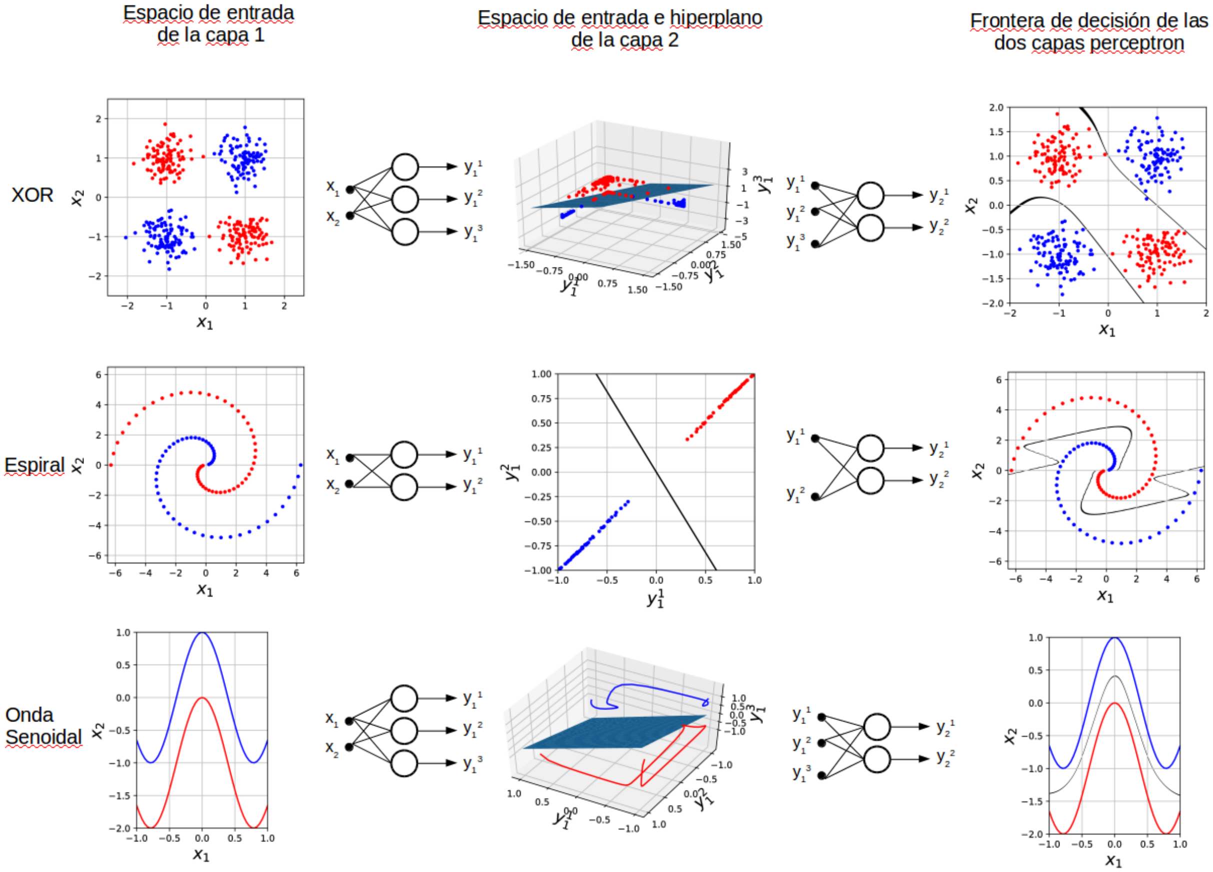
## Pero…¿Qué es una red neuronal artificial?



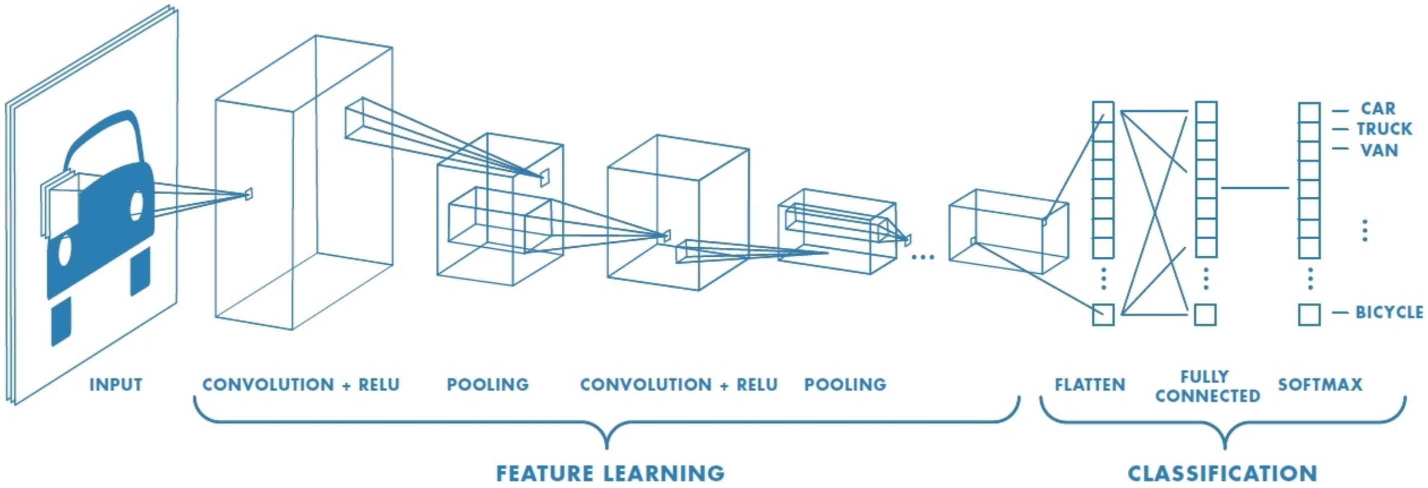

$$\mathbf{y} = f_3\left(\mathbf{W}_3 f_2\left(\mathbf{W}_2 f_1\left(\mathbf{W}_1 \mathbf{x} + \mathbf{b}_1\right) + \mathbf{b}_2\right) + \mathbf{b}_3\right)$$

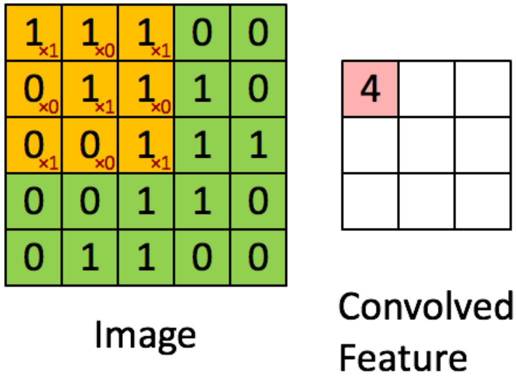
¿Qué es el entrenamiento?

# Entendiendo las capas ocultas

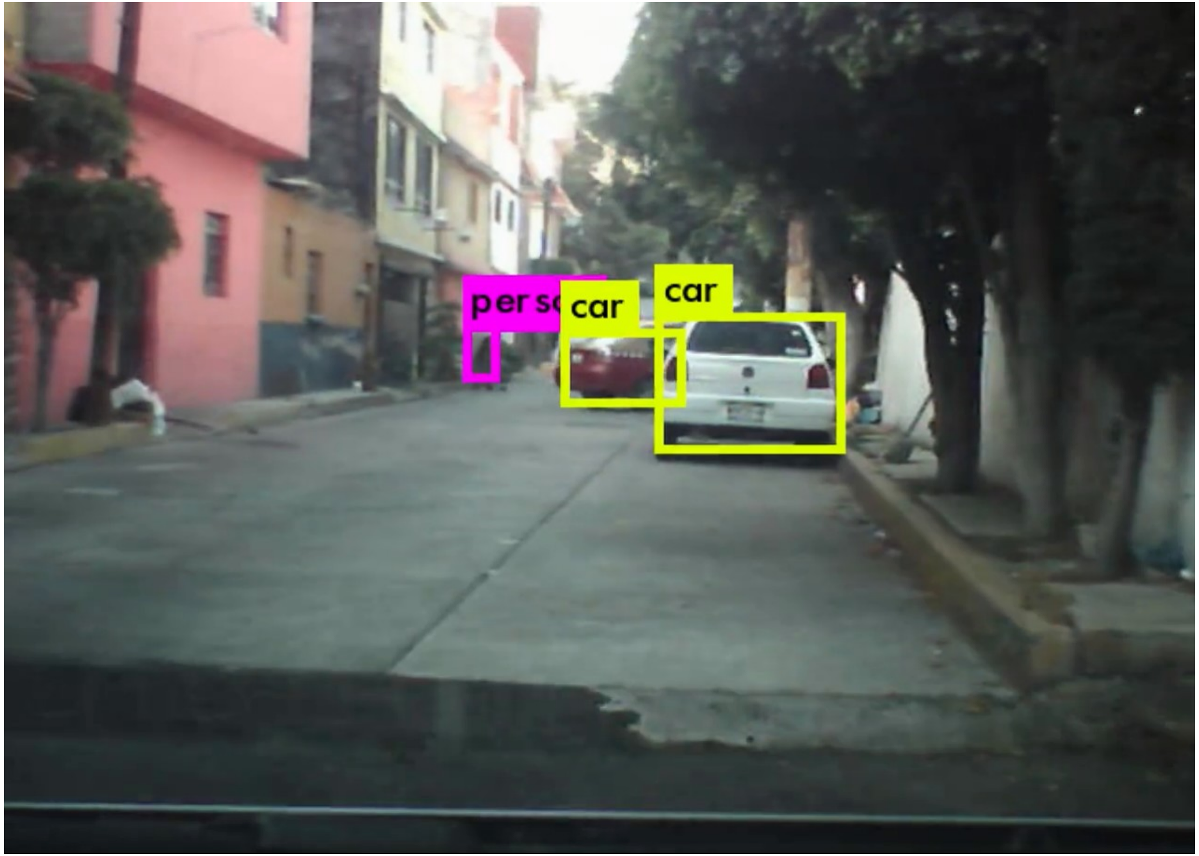


# ¿Qué son las redes neuronales convolucionales?



# Detección de imágenes



# Segmentación de imágenes

# Ejemplos en Python

## Conclusiones

- Las redes neuronales sirven para **automatizar procesos**, aumentando la productividad en las empresas y en la ciencia.

- Son un medio para **transferir la experiencia** de los seres humanos expertos a las maquinas mediante ejemplos (datos).

- **Cuellos de botella**: computadoras con GPUs, recolección y etiquetado de datos de ejemplo, (y habilidad humana para entrenar).

Gracias

# ¿Comentarios o Preguntas?

https://github.com/ezamorag/Minitaller-Aprendizaje-Profundo


## Erik Zamora

Contacto:

ezamora1981@gmail.com

Oficina 31, tercer nivel del edificio del CIC-IPN.

Laboratorio de Robotica y Mecatrónica.

Horario: entre 7hrs y 15hrs.

Erik Zamora

## ¿Por qué investigo?

Poner en tela de juicio el Status Quo del conocimiento dominante  y saciar mi curiosidad por entender

## ¿Para qué?

Para descubrir nuevas ideas e inventar mejores cosas.

## ¿Cómo lo hágo?

Por ejemplo, buscando defectos en los pilares del aprendizaje profundo: el descenso por gradiente y el hiperplano ¿Por qué no usar otros?

## ¿Qué hago?

- **Ciencia Básica.** Buscando mejores estructuras matemáticas y métodos de entrenamiento para redes neuronales artificiales.
- **Ciencia aplicada.** Desarrollando software para la navegación autónoma de los robots.

Erik Zamora

## Productividad como investigador:

- 2.5 años
- Revista JCR: 4 publicados, 4 en revisión, 3 en preparación.
- Conferencia Internacional: 7 publicados, 0 en revisión, 1 en preparación
- Conferencia Nacional: 3 publicados
- Software: 1 registrado IMPI, 1 en desarrollo
- 2 sistemas de navegación

## Formación de Capital Humano como profesor:

- 9.5 años de docente
- 1067 alumnos de ingeniería
- >5300 suscriptores HackeandoTec con > 450,000 vistas y > 2 años tiempo de visualización
- Coordinador del equipo coches autonomos (en 2o lugar )
- TT: 11 graduados, 10 en proceso
- MC: 4 en proceso
- PhD: 6 en proceso

# ¿Por qué los proyectos académicos son lentos?

Por la fricción en tiempo:

- Los estudiantes van y vienen
- Los estudiantes tienen otras ocupaciones (cursos, servicio social, practicas, tareas)
- Los profesores tienen otros motivadores (mantener sus estímulos económicos)
- Los profesores tienen otras ocupaciones (clases, revisar tareas y tesis, formar recursos humanos, publicar artículos, solicitar patentes, etc)
- Pocos son disciplinados y constantes.

Por la fricción en dinero:

- Para acceder a los recursos económicos se requiere ser reconocido por su productividad, esto hace la mayor parte del dinero vaya hacia las mismas pocas personas.
- La administración de los recursos es lenta y torpe por los controles de administrativos.
- Hay que pagar más caro para poder comprar más rápido.
- No hay dinero suficiente para contratar profesionales, así que lo se hace graduar estudiantes. Y los estudiantes requieren tiempo para aprender, cuando ya aprendieron, se van.  No hay continuidad.
- No hay personal técnico dedicado solo a prestar servicios en los laboratorios.