



APRENDIZAJE PROFUNDO

El poder del aprendizaje automático unido al poder de cálculo de las computadoras actuales.

*Dra. María del Pilar Gómez Gil
Coordinación de computación
Instituto Nacional de Astrofísica, Óptica
y Electrónica.
pgomez@inaoep.mx
@pgomezgil*

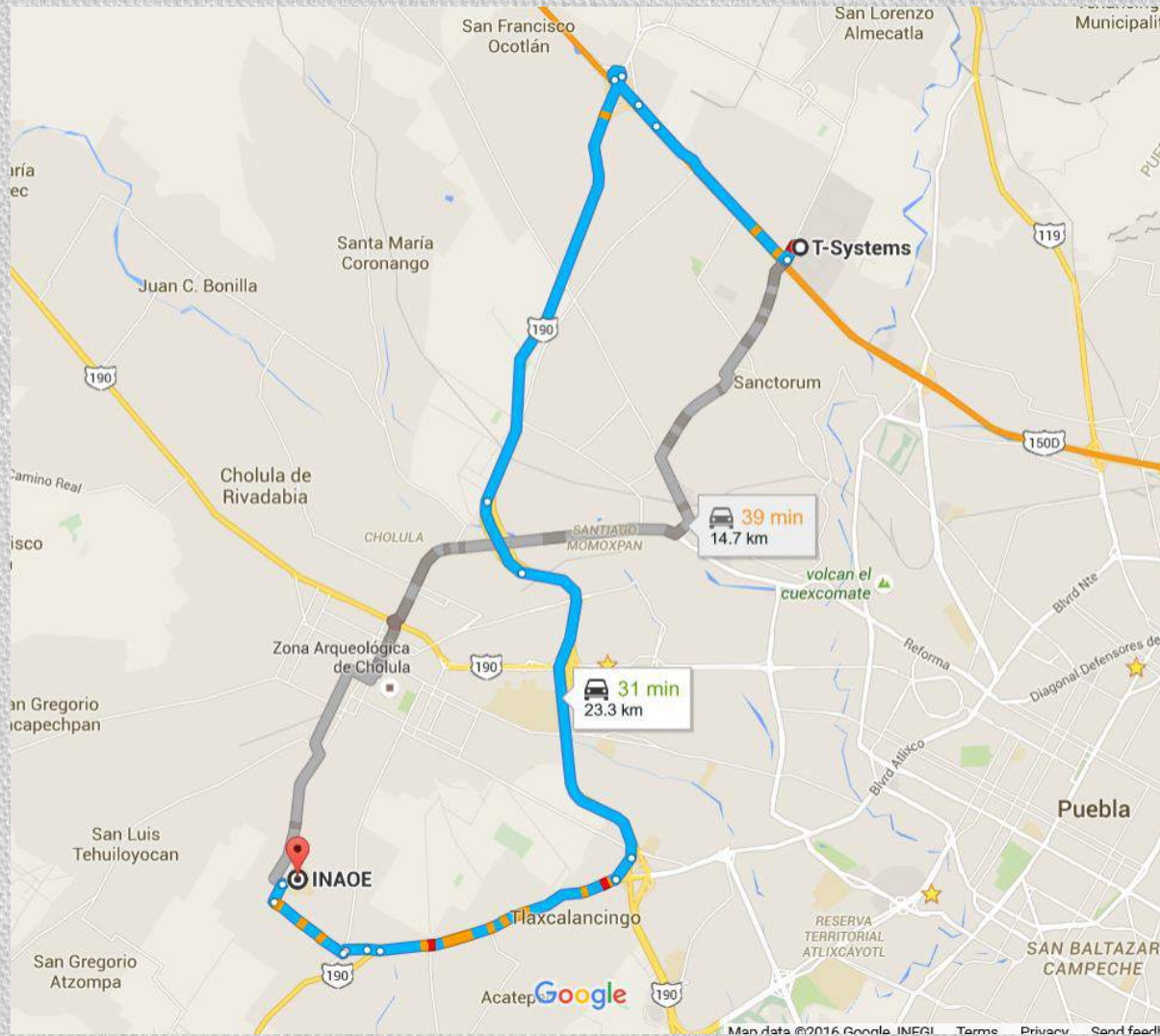
6 de Abril 2016

Un poco sobre INAOE (1/2)



- Es un centro público de investigación localizado en Tonantzintla, Puebla México .
- Su **misión** es contribuir a la generación, avance y difusión del conocimiento, para el desarrollo del país y de la humanidad.
- Identifica y busca la solución de problemas científicos y tecnológicos
- Participa en la formación de especialistas en las áreas de Astrofísica, Óptica, Electrónica, **Ciencias Computacionales** y áreas afines.

Un poco sobre INAOE (2/2)



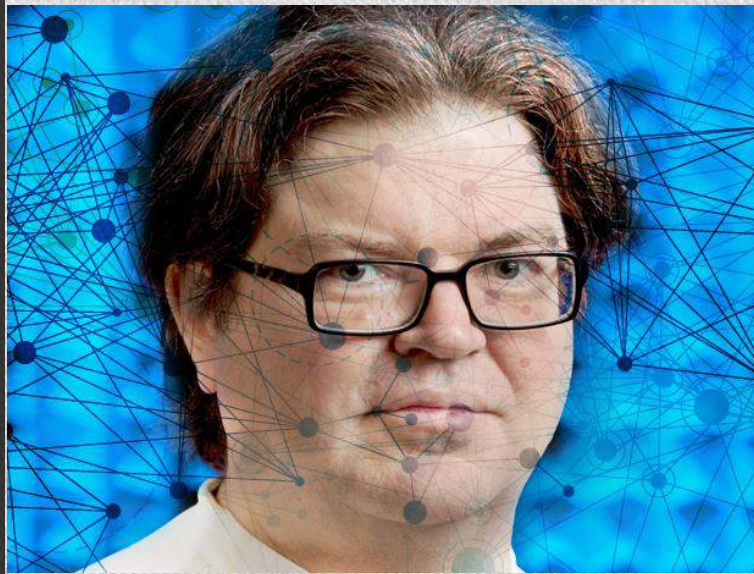
Esta presentación está disponible en:

<http://ccc.inaoep.mx/~pgomez/conferences/PggTSys16.pdf>



“Most of the knowledge in the world in the future is going to be extracted by machines and will reside in machines”

Yann LeCun, Director of AI Research, Facebook¹



Yan LeCun. Photo-illustration: Randi Klett

¹ “[The Five Tribes of Machine Learning \(And What You Can Learn from Each\)](#),” Pedro Domingos, Webminar produced by the Association of Computing Machinery (ACM). Dec. 29, 2015

¿Qué es aprendizaje profundo? (1/2)

- En el contexto de Inteligencia Artificial (IA), “aprendizaje profundo” (*Deep Learning* o **DL**) se refiere a la actividad **automática de adquisición de conocimiento**, a través del uso de máquinas que usan varios niveles para la extracción.
- El adjetivo “profundo” se aplica no en sí al conocimiento adquirido, sino a la forma en que el conocimiento se adquiere.

¿Qué es aprendizaje profundo? (2/2)

- La gran ventaja de DL es que no requiere de una definición “a mano” de las **características** que identifican a los patrones que se buscan, sino que automáticamente se generan dichas características, manipulando datos crudos
- Esto se lleva a cabo a través de construir automáticamente características de alto nivel, a través del uso de una gran cantidad de niveles jerárquicos de extractores, dentro de un sistema que **aprende automáticamente**.

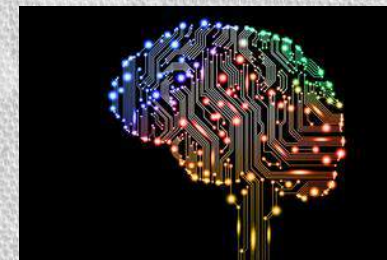


Foto tomada de:
<http://www.kodemaker.no/deeplearning/>

¿Qué es aprendizaje automático?

- Los sistemas basados en IA se caracterizan porque contienen una **representación del conocimiento**, que les permite tomar decisiones de forma autónoma.
- Algunas técnicas de IA utilizan *representaciones simbólicas* del conocimiento, como los sistemas expertos basados en reglas. Sin embargo, estas técnicas aunque son exactas, no son fácilmente escalables.
- La **inteligencia computacional** (IC) permite trabajar con sistemas imprecisos y encontrar soluciones en tiempos razonables, aunque no exactas, a través de crear *representaciones numéricas* del conocimiento.
- En el aprendizaje automático, se adquiere el conocimiento a través de analizar datos y manipularlos, usando estrategias basadas en teorías matemáticas

¿Cómo descubren conocimiento las computadoras?

- Hay muchas estrategias, las mas populares lo hacen:
 - Llenando “huecos” en el conocimiento existente,
 - Construyendo máquinas que intentan emular al cerebro,
 - Simulando el proceso de evolución,
 - Reduciendo incertidumbre de forma sistemática,
 - Encontrando similitudes entre diferencias temporales.

“ [The Five Tribes of Machine Learning \(And What You Can Learn from Each\)](#),”
Pedro Domingos, Webminar produced by the Association of Computing
Machinery (ACM). Dec. 29, 2015

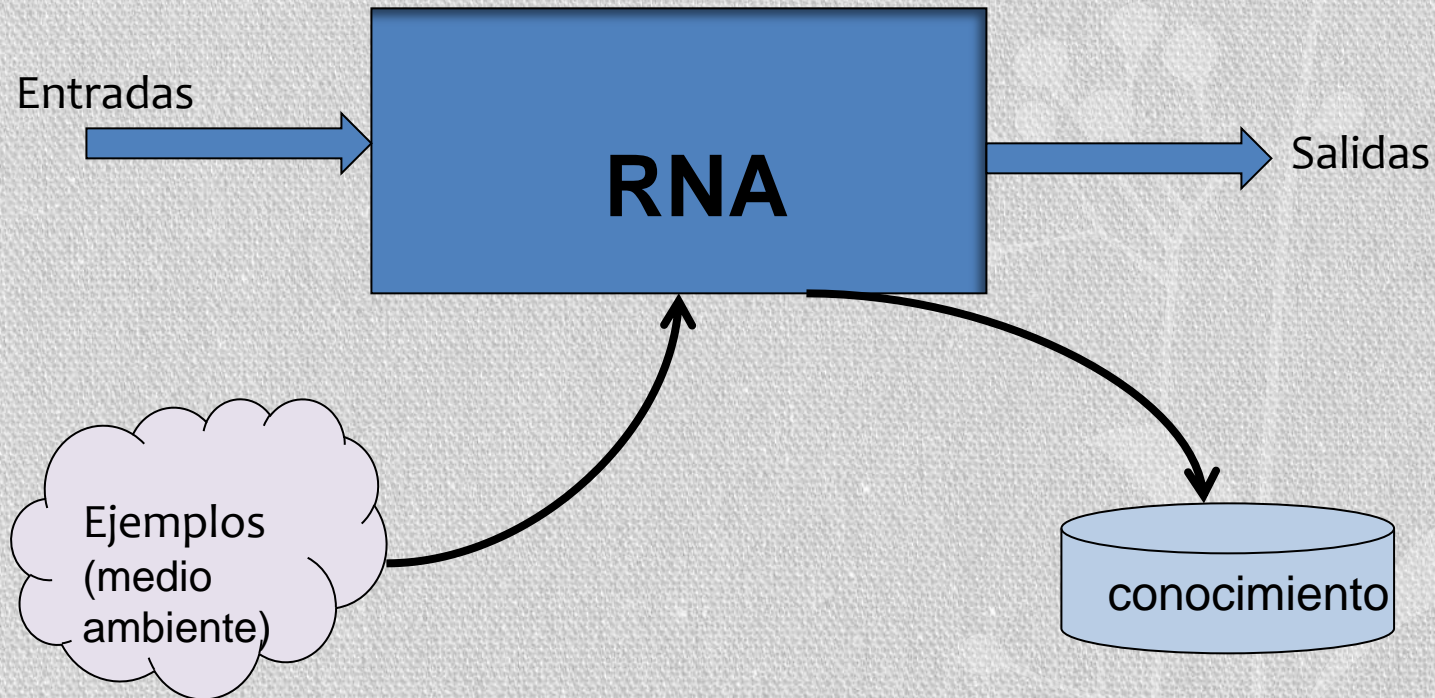
Emulando al cerebro: Redes Neuronales Artificiales (RNA)

- Son **modelos matemáticos** capaces de adaptar su comportamiento en respuesta a ejemplos tomados del medio ambiente
- Están **inspiradas** en la construcción del cerebro y las neuronas biológicas.
- Sus componentes son procesadores simples que funcionan de forma **paralela y distribuida**
- Pueden **almacenar y utilizar** ese conocimiento adquirido de la experiencia

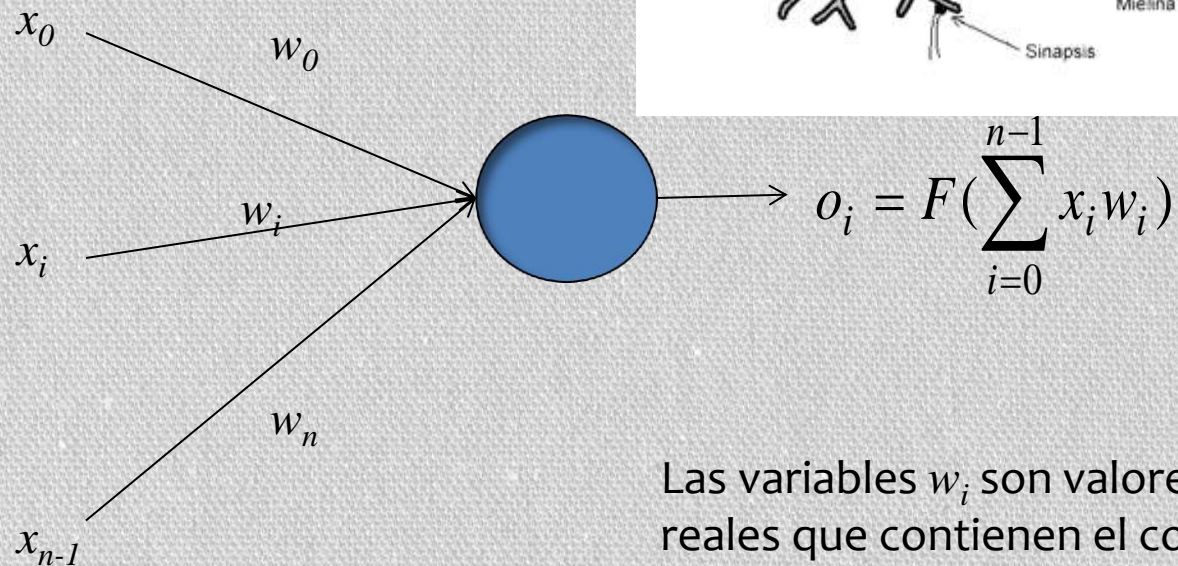
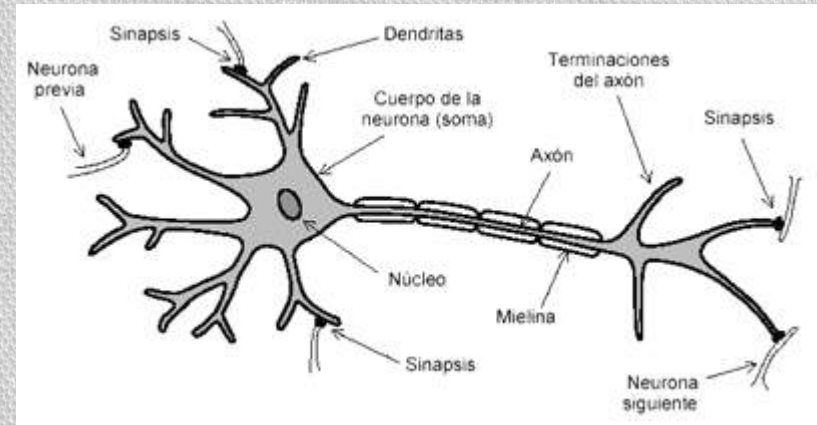


Haykin, S. (2009) Neural Networks and Learning Machines.
Third Edition. Pearson Education: Upper Saddle River, New Jersey

El Contexto de Redes Neuronales

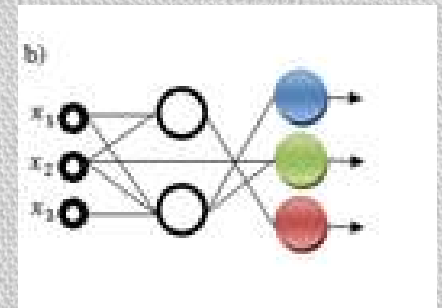
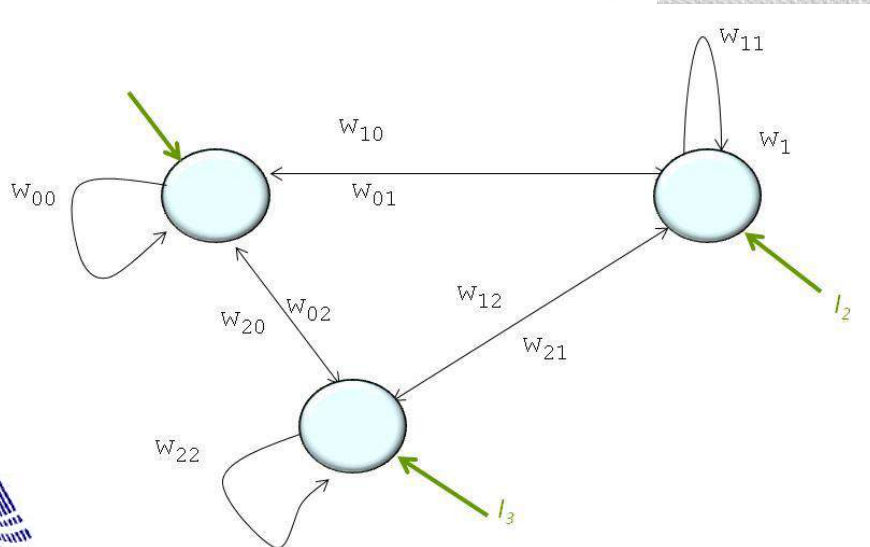
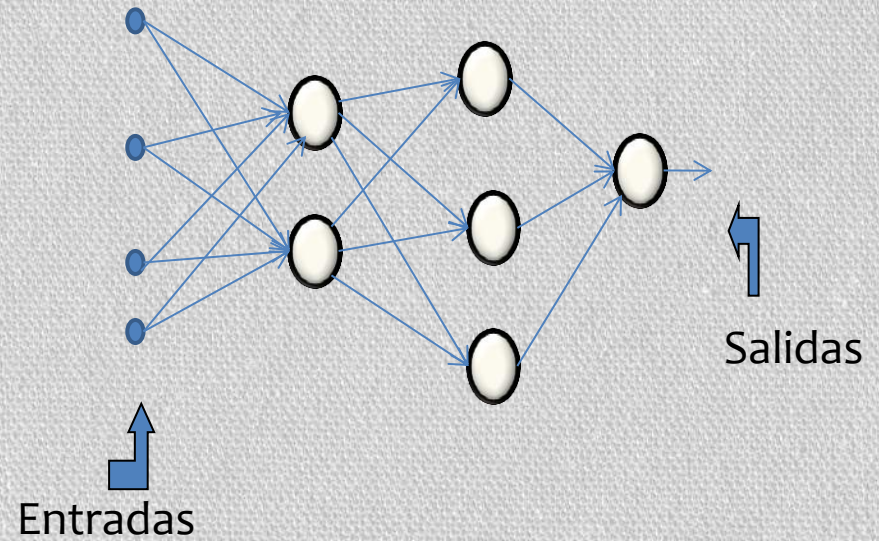
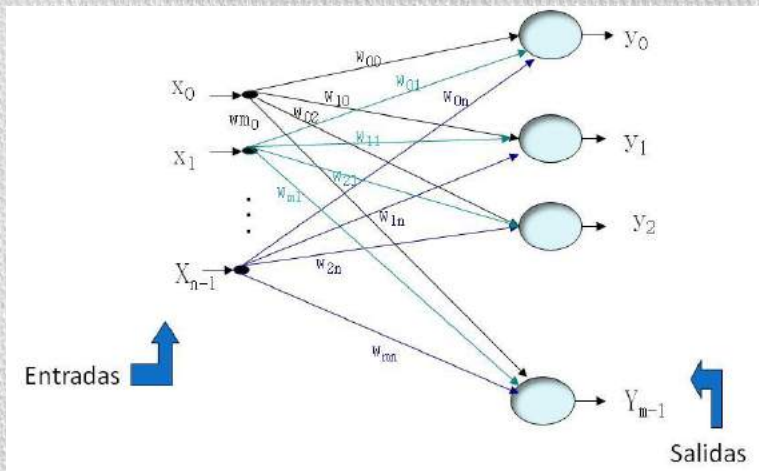


El componente fundamental: neurona

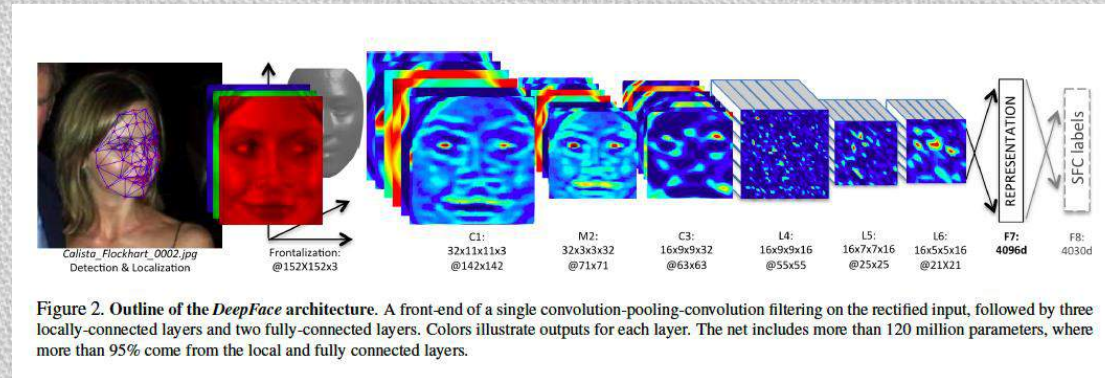
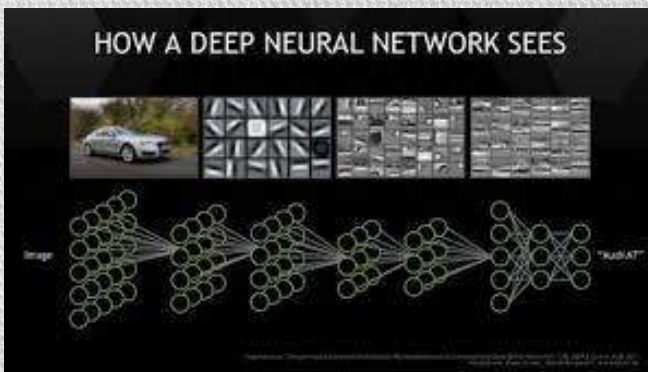
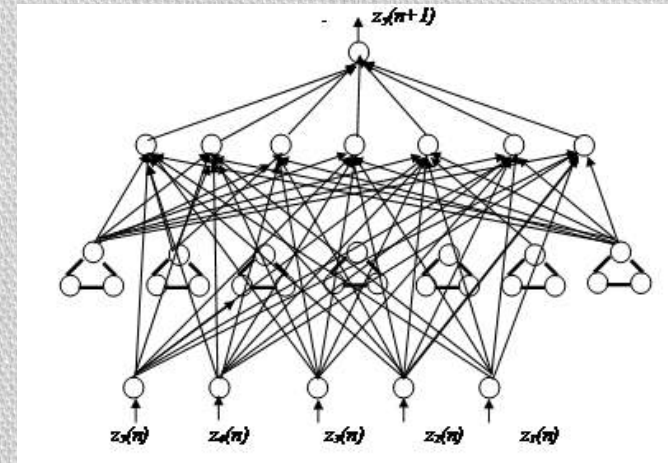
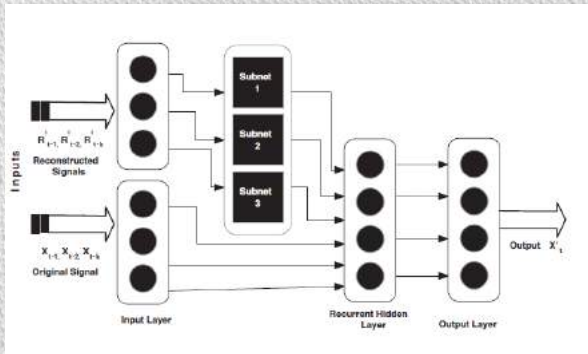


Las variables w_i son valores reales que contienen el conocimiento de la red neuronal

La conexión entre neuronas forma las RNA



Su topología puede ser bastante sofisticada



Matemáticas rigurosas solventan a las RNA

Neural Networks, Vol. 2, pp. 359–366, 1989
 Printed in the USA. All rights reserved.

ORIGINAL CONTRIBUTION

Multilayer Feedforward Networks are Universal Approximators

KUR HORNIK

Technische Universität Wien

MAXWELL STINCHCOMBE AND HALBER WHITE

University of California, San Diego

(Received 16 September 1988; revised and accepted 9 March 1989)

Abstract—This paper rigorously establishes that standard multilayer feedforward networks with one hidden layer using arbitrary squashing functions are capable of approximating any Borel measurable function from one finite dimensional space to another to any desired degree of accuracy, provided sufficiently many hidden units are available. In this sense, multilayer feedforward networks are a class of universal approximators.

Keywords—Feedforward networks, Universal approximation, Mapping networks, Network representation capability, Stone-Weierstrass Theorem, Squashing functions, Sigma-Pi networks, Back-propagation networks.

Lemma 2.1. All of the following are equivalent.

- (a) $\rho_\mu(f_n, f) \rightarrow 0$.
- (b) For every $\varepsilon > 0$ $\mu\{x: |f_n(x) - f(x)| > \varepsilon\} \rightarrow 0$.
- (c) $\int \min\{|f_n(x) - f(x)|, 1\} \mu(dx) \rightarrow 0$. □

Definition 2.4

For any measurable function $G(\cdot)$ mapping R to R and $r \in N$, let $\Sigma\Pi^r(G)$ be the class of functions

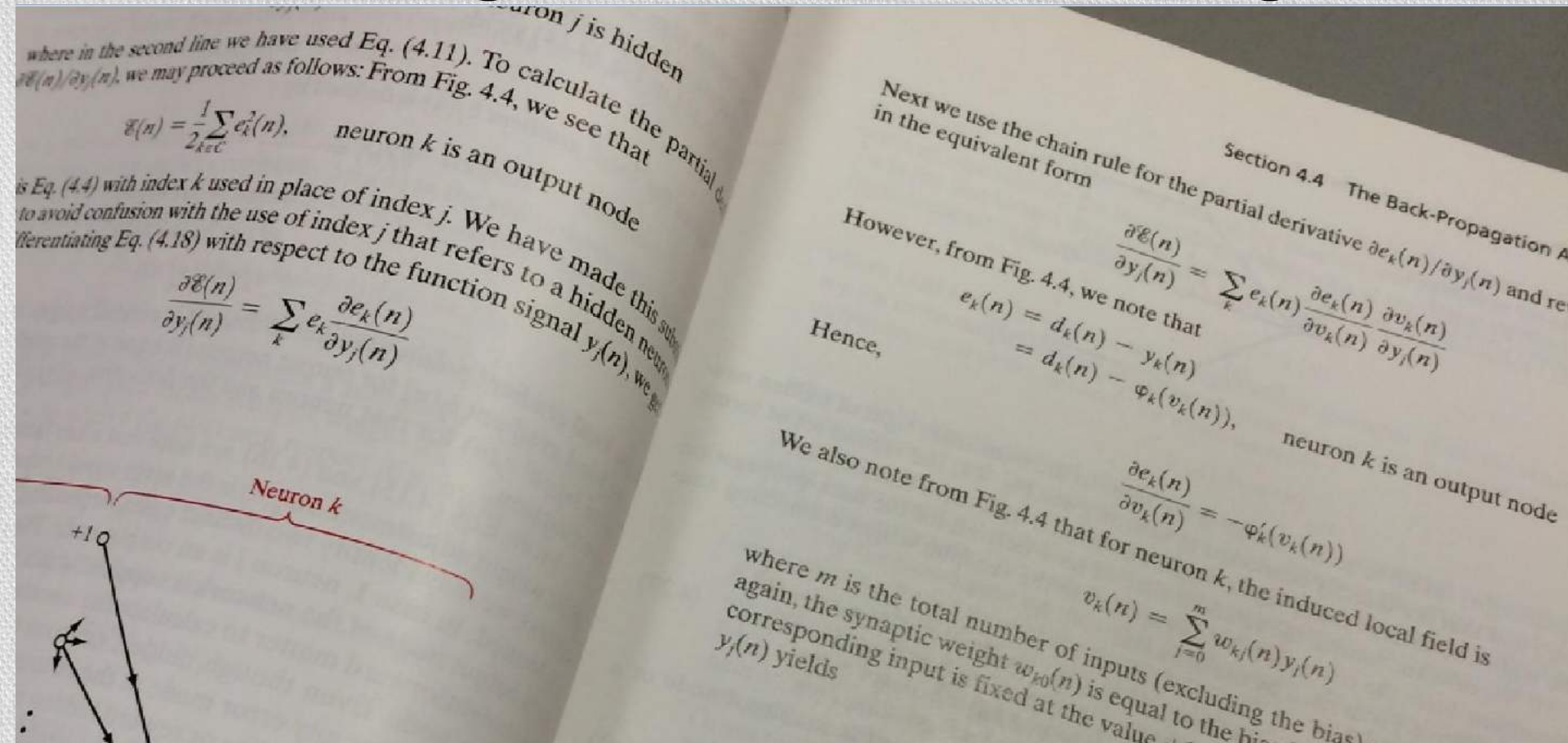
$\{f: R^r \rightarrow R: f(x) =$

$$\sum_{j=1}^q \beta_j \cdot \prod_{k=1}^{l_j} G(A_{jk}(x)), x \in R^r, \beta_j \in R, A_{jk} \in A^r, l_j \in N,$$

$q =$

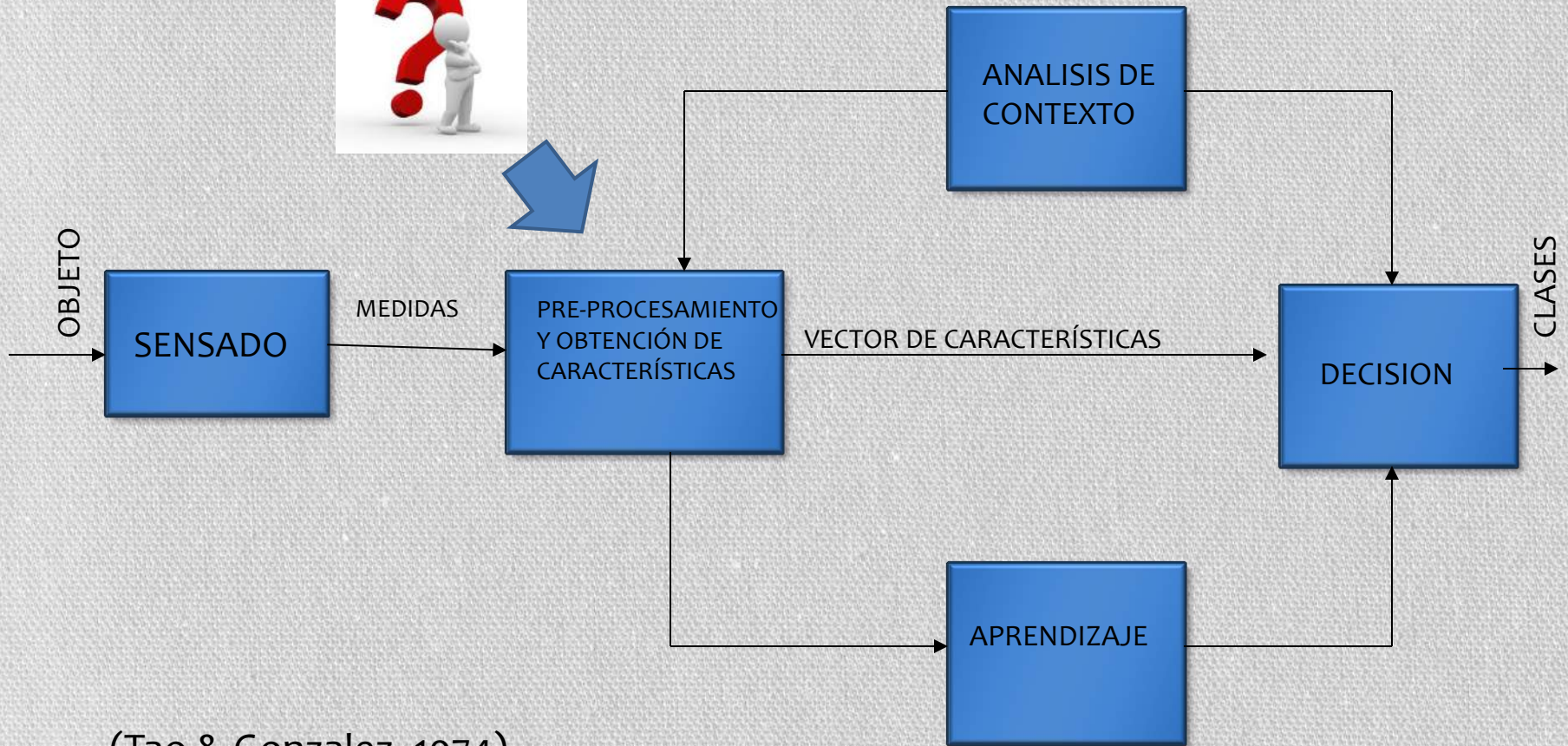


El famoso algoritmo de Retropropagación



Haykin, S. S. (2009). *Neural networks and learning machines (Vol. 3)*. Upper Saddle River: Pearson Education.

El problema de clasificar



(Tao & Gonzalez ,1974)

El neocognitron de Fukushima (1980)

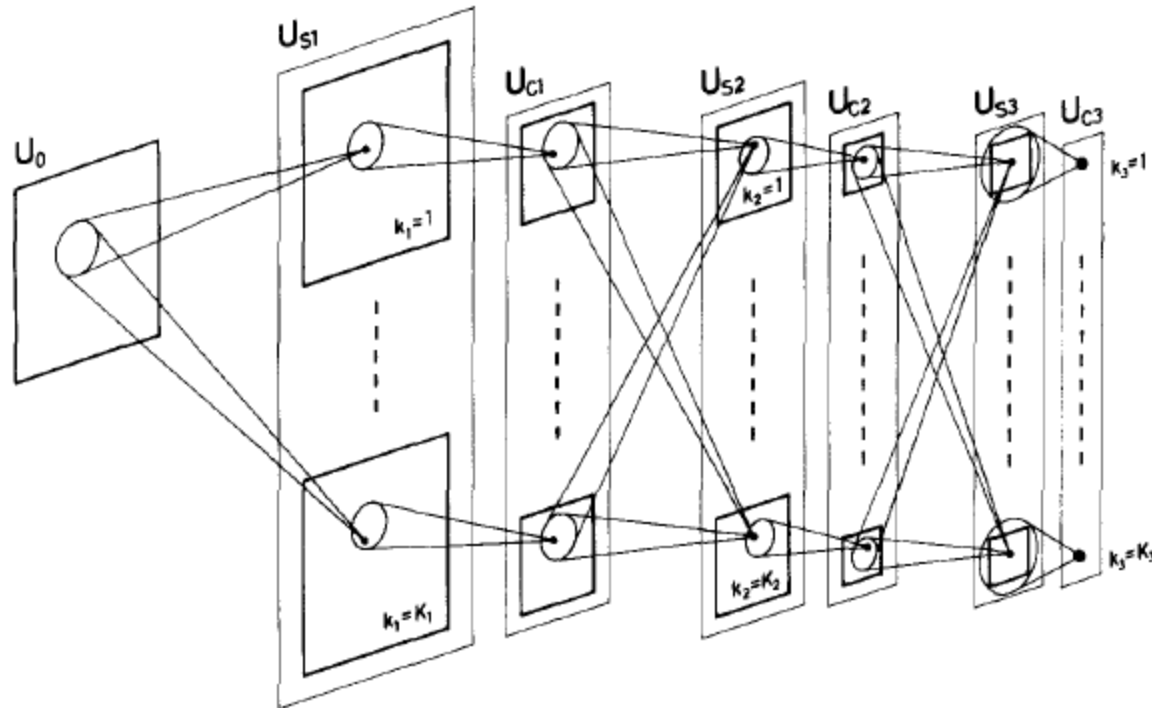


Fig. 2. Schematic diagram illustrating the interconnections between layers in the neocognitron

Fukushima, K. (1980). [Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position](#). *Biological cybernetics*, 36(4), 193-202.

Extrayendo características en 1989

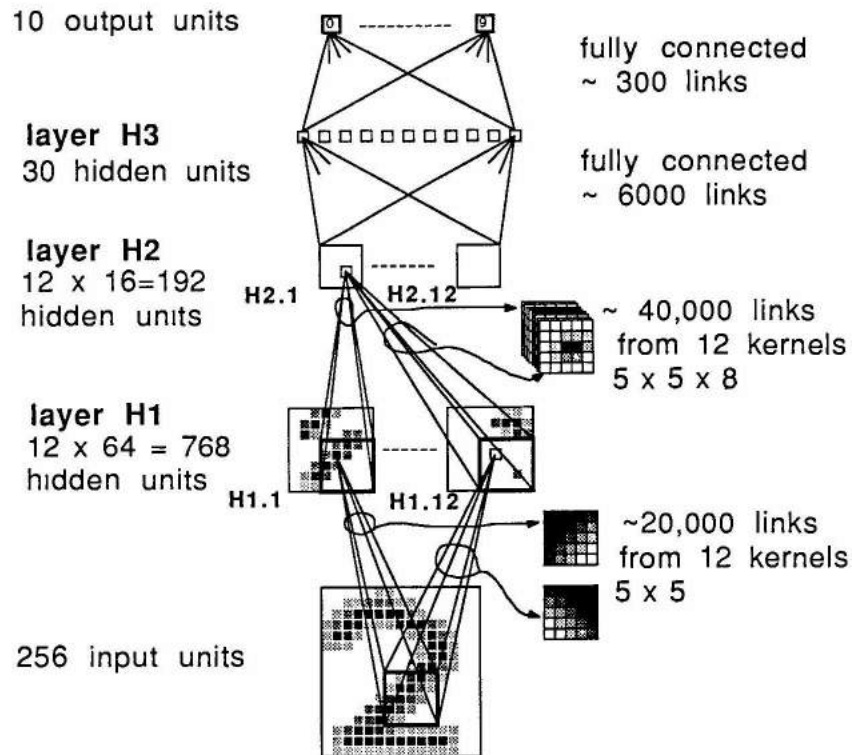


Figure 3 Log mean squared error (MSE) (top) and raw error rate (bottom) versus number of training passes

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). [propagation applied to handwritten zip code recognition](#). *Neural Computation*, 1(4), 541-551.

Extrayendo características actualmente

Samoyed (16); Papillon (5.7); Pomeranian (2.7); Arctic fox (1.0); Eskimo dog (0.6); white wolf (0.4); Siberian husky (0.4)

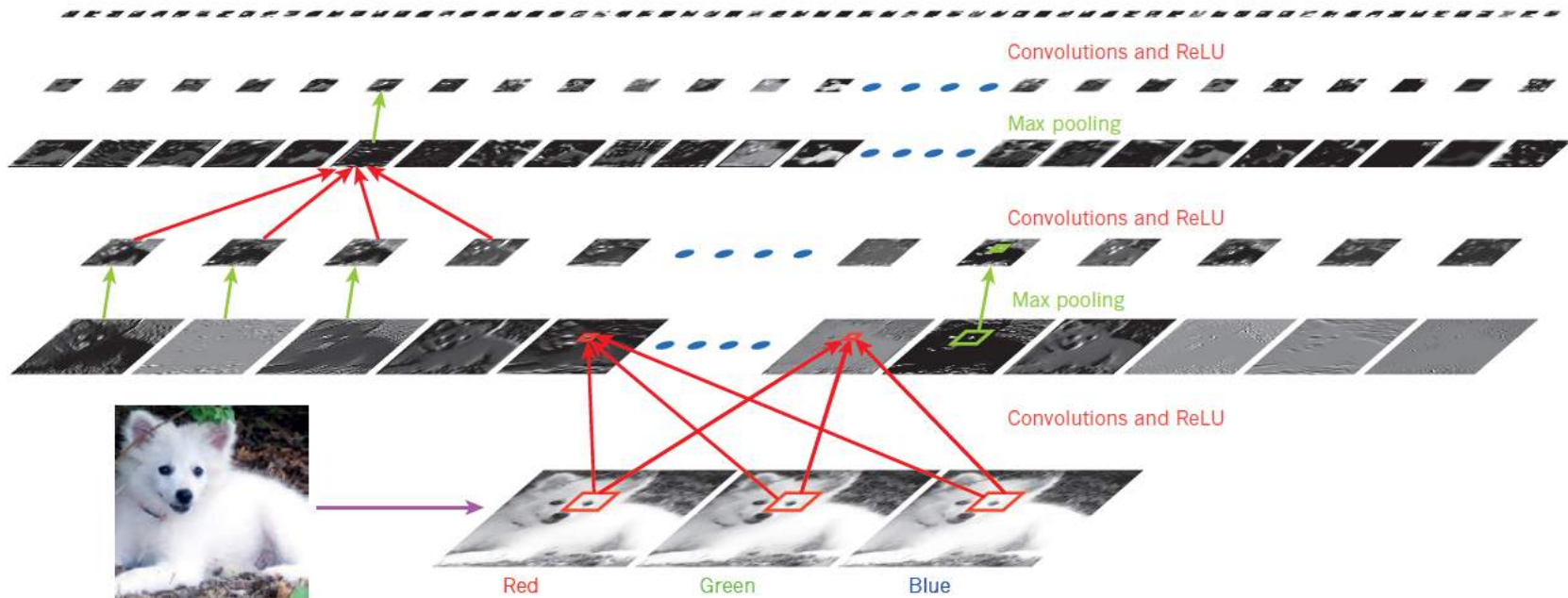


Figure 2 | Inside a convolutional network. The outputs (not the filters) of each layer (horizontally) of a typical convolutional network architecture applied to the image of a Samoyed dog (bottom left; and RGB (red, green, blue) inputs, bottom right). Each rectangular image is a feature map

corresponding to the output for one of the learned features, detected at each of the image positions. Information flows bottom up, with lower-level features acting as oriented edge detectors, and a score is computed for each image class in output. ReLU, rectified linear unit.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). [Deep learning](#). *Nature*, 521(7553), 436-444.

RNA y DL

- DL permite que redes neuronales con varios niveles de neuronas aprendan a representar características, sin decirle cuales son.
- A este campo se le conoce como “**aprendizaje de representaciones**” (*representational learning*)
- Utilizan combinaciones de aprendizaje supervisado y no supervisado, en los diferentes niveles.
- Las arquitecturas mas comunes son:
 - Red profunda de convolución (convolutional Net)
 - Red de creencias profundas (deep-belief network)
 - Redes recurrentes de memoria corta y larga (*long-short term memory, LSTM*)

Principales aplicaciones de DL

- Reconocimiento de voz
- Reconocimiento de imágenes
- Detección de fraudes
- Reconocimiento de caracteres manuscritos
- Descubrimiento de componentes farmacéuticos
- Procesamiento de lenguaje Natural

LeCun, Y., Bengio, Y., & Hinton, G. (2015). [Deep learning](#). *Nature*, 521(7553), 436-444.

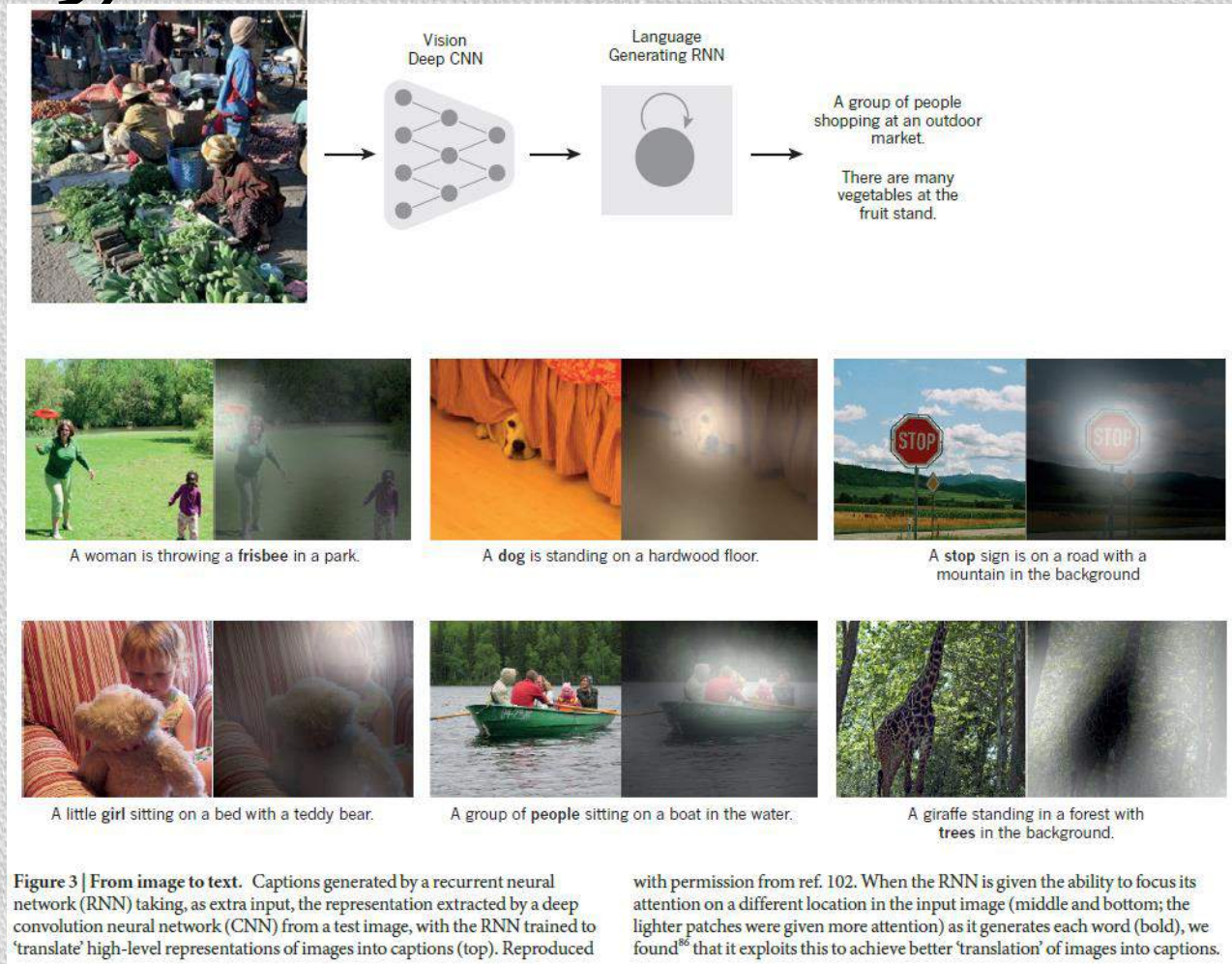
Procesando datos con el neocognitron (1980)



FIGURE 12. Experiment of handwritten numeral recognition on a minicomputer.

Fukushima, K. (1988). [Neocognitron: A hierarchical neural network capable of visual pattern recognition](#). *Neural networks*, 1(2), 119-130.

Procesando datos con la red de convolución (2015)



LeCun, Y., Bengio, Y., & Hinton, G. (2015). [Deep learning](#). *Nature*, 521(7553), 436-444.

El poder del cómputo actual es la diferencia

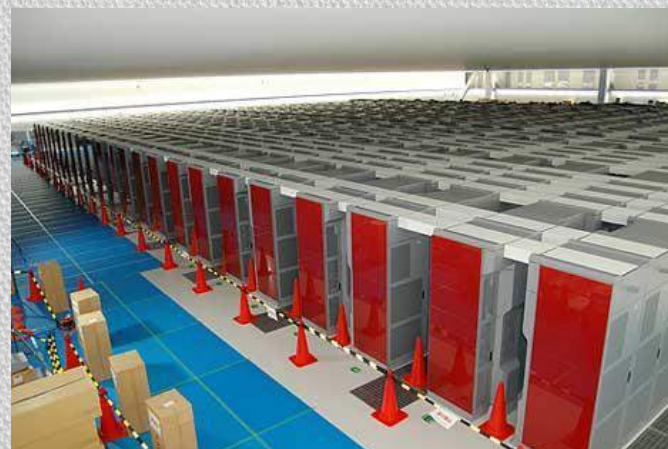


IBM'S Watson

<http://www.wired.com/2012/03/ibm-watson/>



IBM Blue Gene / Q systems powering the Fermi supercomputer at Cineca in Italy.



K SUPERCOMPUTER, RIKEN Advanced Institute for Computational Science (AICS), Japan

<http://www.datacenterknowledge.com/top-10-supercomputers-illustrated-june-2012-continued/>

Laboratorio Nacional de Supercómputo del sureste de México



<http://www.lns.buap.mx/>

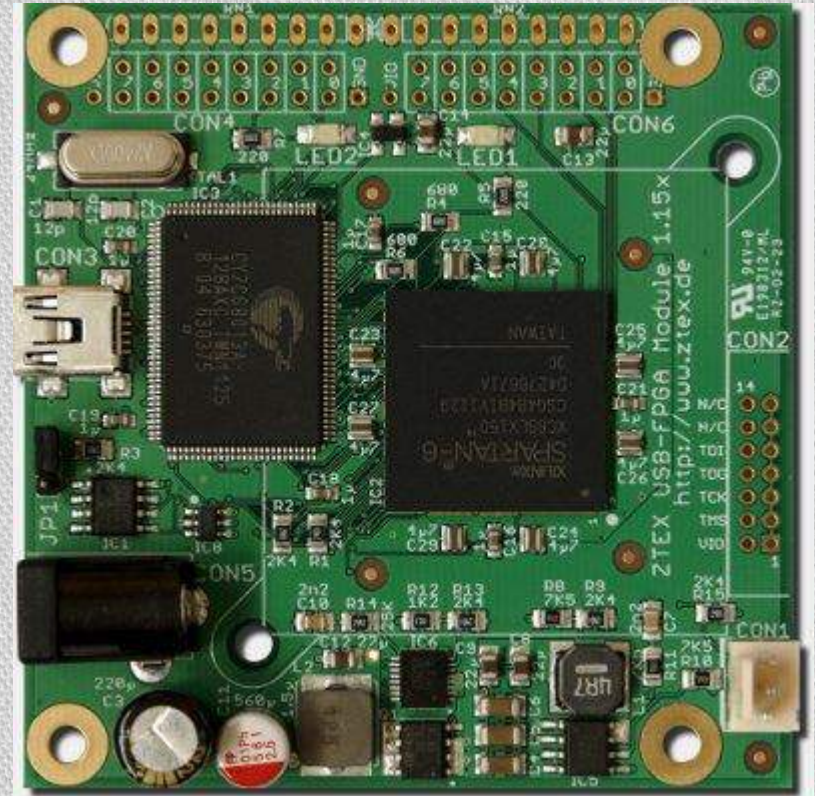


<http://www.conacytprensa.mx/index.php/tecnologia/tic/2723-Ins-del-sureste-de-mexico-un-centro-de-datos-de-clase-mundial>

Se pueden usar aplicaciones de DL con equipo de alto desempeño mas pequeño



<http://www.geforce.com/whats-new/articles/introducing-the-geforce-gtx-680m-mobile-gpu>



Spartan 6 LX150 FPGA Board

<http://www.ztex.de/usb-fpga-1/usb-fpga-1.15x.e.html>

Aplicación exitosa a la fecha: Verificación de Caras (1/2)

2014 IEEE Conference on Computer Vision and Pattern Recognition

DeepFace: Closing the Gap to Human-Level Performance in Face Verification

Yaniv Taigman Ming Yang Marc'Aurelio Ranzato Lior Wolf

Facebook AI Research
Menlo Park, CA, USA
{yaniv, mingyang, ranzato}@fb.com

Tel Aviv University
Tel Aviv, Israel
wolf@cs.tau.ac.il

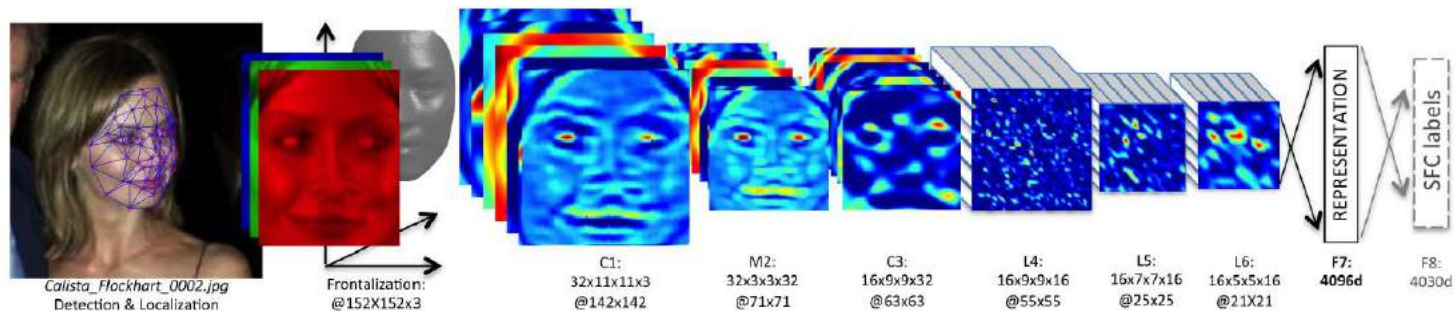


Figure 2. **Outline of the DeepFace architecture.** A front-end of a single convolution-pooling-convolution filtering on the rectified input, followed by three locally-connected layers and two fully-connected layers. Colors illustrate feature maps produced at each layer. The net includes more than 120 million parameters, where more than 95% come from the local and fully connected layers.

Y. Taigman, M. Yang, M. Ranzato and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," *Computer Vision and Pattern Recognition (CVPR)*, 2014 IEEE Conference on, Columbus, OH, 2014, pp. 1701-1708.

Aplicación exitosa a la fecha: Verificación de Caras (2/2)

Method	Accuracy (%)	AUC	EER
MBGS+SVM- [31]	78.9 ±1.9	86.9	21.2
APEM+FUSION [22]	79.1 ±1.5	86.6	21.4
STFRD+PMML [9]	79.5 ±2.5	88.6	19.9
VSOFF+OSS [23]	79.7 ±1.8	89.4	20.0
DeepFace-single	91.4 ±1.1	96.3	8.6

Table 4. Comparison with the state-of-the-art on the *YTF* dataset.

<http://dx.doi.org/10.1109/CVPR.2014.220>

Y. Taigman, M. Yang, M. Ranzato and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, Columbus, OH, 2014, pp. 1701-1708.



Otras aplicaciones del aprendizaje de representaciones: clasificación de series de tiempo (1/2)

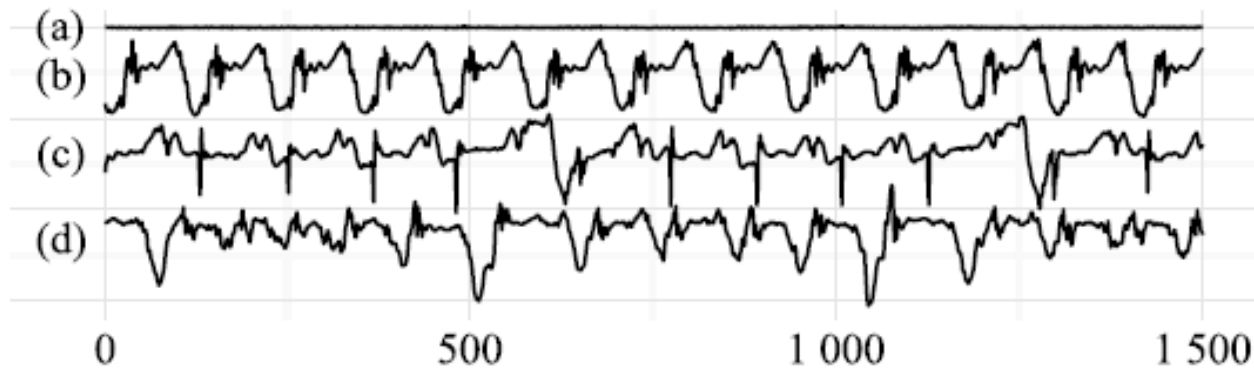


Fig. 2 Four 1D samples of 3D weakly labeled physical activities: (a) “standing”; (b) “walking”; (c) “ascending stairs”; (d) “descending stairs”

Zheng, Y., Liu, Q., Chen, E., Ge, Y., & Zhao, J. L. (2016). Exploiting multi-channels deep convolutional neural networks for multivariate time series classification. *Frontiers of Computer Science*, 10(1), 96-112.

Otras aplicaciones del aprendizaje de representaciones: clasificación de series de tiempo (1/2)

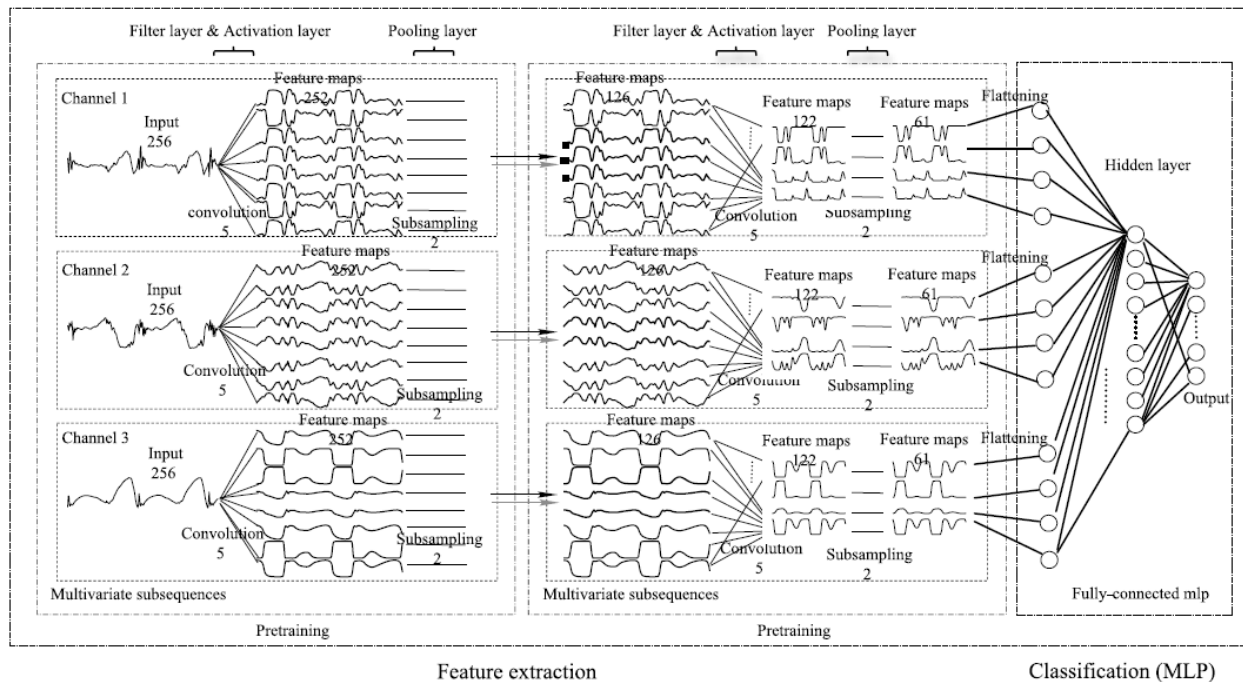
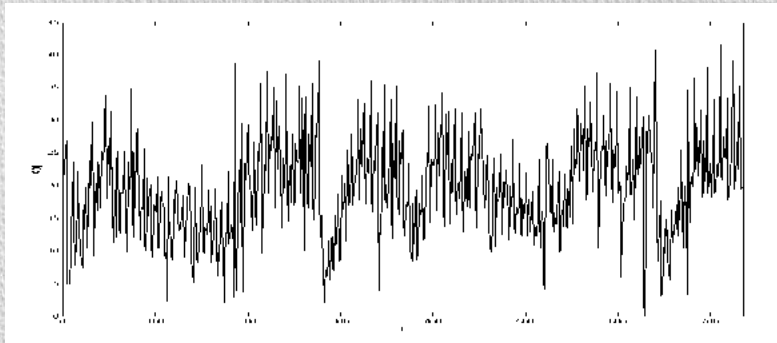


Fig. 3 A two-stage MC-DCNN architecture for activity classification. This architecture consists of three channels input, two filter layers, two pooling layers and two fully-connected layers. Pretraining is performed for two stages gradually and then supervised learning is applied. This architecture is denoted as $8(5)-2-4(5)-2-732-4$ based on the template $C1(Size)-S1-C2(Size)-S2-H-O$, where $C1$ and $C2$ are numbers of filters in the first and second stage, $Size$ denotes the kernel size, $S1$ and $S2$ are subsampling factors, H and O denote the numbers of units in hidden and output layers of MLP

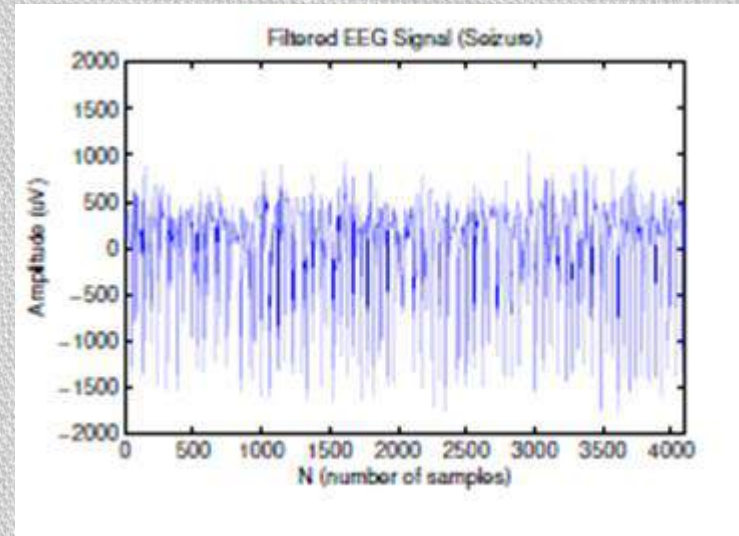
Zheng, Y., Liu, Q., Chen, E., Ge, Y., & Zhao, J. L. (2016). Exploiting multi-channels deep convolutional neural networks for multivariate time series classification. *Frontiers of Computer Science*, 10(1), 96-112.

Ejemplos de señales complejas



Retiros de Cajeros automáticos. Fragmento De la Serie 1 de Competencia Internacional NN5-001

S. F. Crone, “NN5 forecasting competition for artificial neural networks & computational intelligence,” 2008, available in <http://www.neural-forecasting-competition.com/NN5/index.htm>



Estado Ictal en un paciente con epilepsia

Juarez-Guerra E, Alarcon-Aquino V and Gomez-Gil P. “[Epilepsy Seizure Detection in EEG Signals Using Wavelet Transforms and Neural Networks.](#)” New Trends in Networking, Computing, E-learning, Systems Sciences, and Engineering. Lecture Notes in Electrical Engineering. Eds: K. Elleithy, T. Sobh. Vol 312, 2015, pp 261-269.

Algunas Desventajas de DL

- En la mayoría de los casos se requieren grandes cantidades de datos
- Los algoritmos de aprendizaje aún necesitan mejorarse
- La migración/adecuación para programar aplicaciones en paralelo puede ser tediosa
- El porcentaje de éxito obtenido aún es bajo para poder usarse en muchas aplicaciones prácticas, pero va mejorando
- Puede ser que se identifiquen patrones que no son útiles

Los sistemas encuentran lo que aprendieron antes - Google Deep Dream

Puerto Escondido, Oaxaca

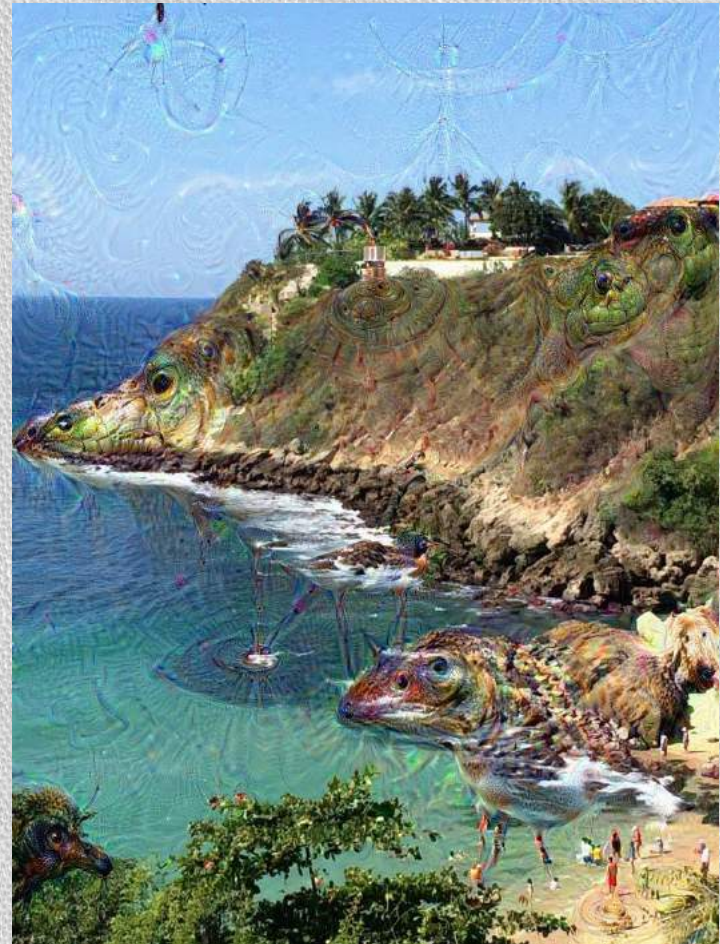
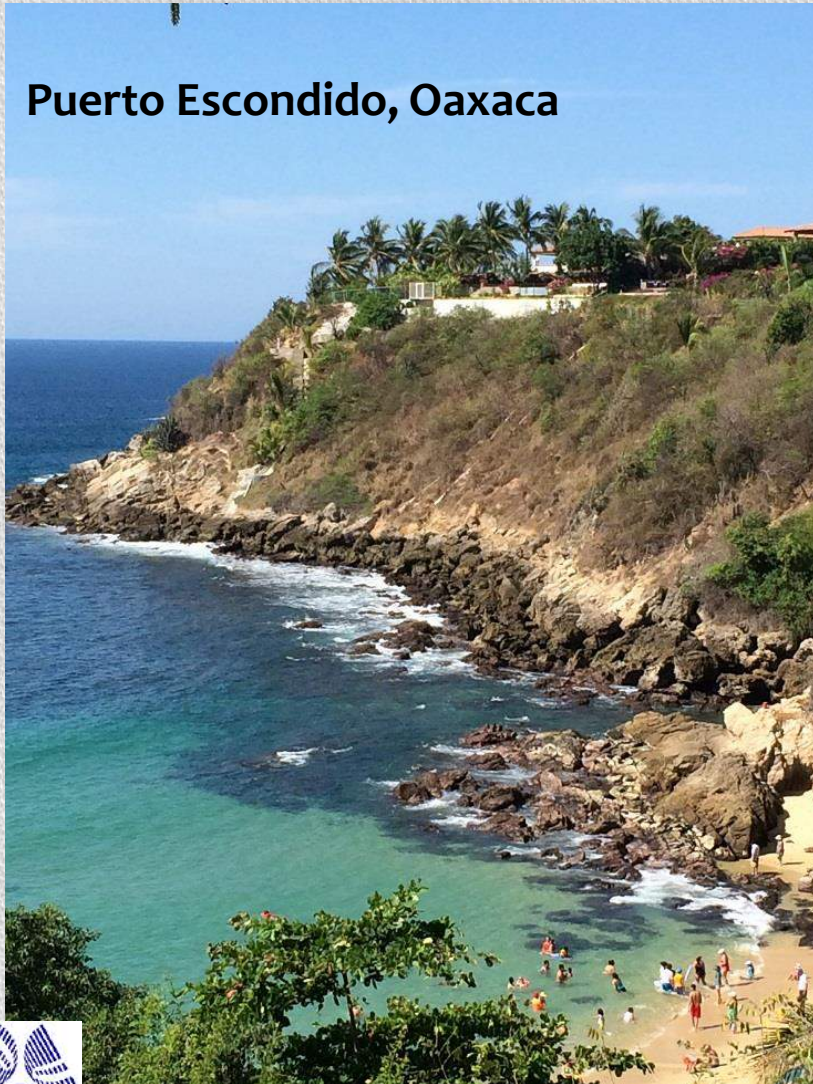


Imagen generada usando
<http://deepdreamgenerator.com/>

Una playa de puerto escondido procesada con *Deep Dream*



Imagen generada usando
<http://deepdreamgenerator.com/>

Consideraciones finales

- DL y *big-data* van de la mano
- Esta tecnología aún no está madura, pero está teniendo avances muy rápidos, y pronto se verá reflejada en otras aplicaciones
- El futuro de DL, según Yan LeCun¹ gira alrededor de:
 - Aprendizaje no supervisado
 - Redes Neuronales Recurrentes
 - Entendimiento de Lenguaje Natural
 - Combinación de aprendizaje de representaciones con razonamiento complejo

1. LeCun, Y., Bengio, Y., & Hinton, G. (2015). [Deep learning](#). *Nature*, 521(7553), 436-444.

Muchas gracias por su atención!



Imagen tomada de: [https://www.tumblr.com/search/deep dream art](https://www.tumblr.com/search/deep%20dream%20art)

Esta presentación está disponible en:

<http://ccc.inaoep.mx/~pgomez/conferences/PggTSys16.pdf>