

# Inception-v1

## **Integrantes:**

- Perca Quispe Joel Cristian
- Espinoza Peñaloza Edgar Alfonso
- Vilchez Molina Misael Svante

# CONTENIDOS

01

INTRODUCCIÓN

02

ARQUITECTURA

03

CONJUNTO DE  
DATOS

04

APLICACIONES

05

DEMO



## Going deeper with convolutions

**Christian Szegedy**  
Google Inc.

**Wei Liu**  
University of North Carolina, Chapel Hill

**Yangqing Jia**  
Google Inc.

**Pierre Sermanet**  
Google Inc.

**Scott Reed**  
University of Michigan

**Dragomir Anguelov**  
Google Inc.

**Dumitru Erhan**  
Google Inc.

**Vincent Vanhoucke**  
Google Inc.

**Andrew Rabinovich**  
Google Inc.

### Abstract

We propose a deep convolutional neural network architecture codenamed Inception, which was responsible for setting the new state of the art for classification and detection in the ImageNet Large-Scale Visual Recognition Challenge 2014 (ILSVRC14). The main hallmark of this architecture is the improved utilization of the computing resources inside the network. This was achieved by a carefully crafted design that allows for increasing the depth and width of the network while keeping the computational budget constant. To optimize quality, the architectural decisions were based on the Hebbian principle and the intuition of multi-scale processing. One particular incarnation used in our submission for ILSVRC14 is called GoogLeNet, a 22 layers deep network, the quality of which is assessed in the context of classification and detection.

### 1 Introduction

In the last three years, mainly due to the advances of deep learning, more concretely convolutional networks [10], the quality of image recognition and object detection has been progressing at a dramatic pace. One encouraging news is that most of this progress is not just the result of more powerful hardware, larger datasets and bigger models, but mainly a consequence of new ideas, algorithms and improved network architectures. No new data sources were used, for example, by the top entries in the ILSVRC 2014 competition besides the classification dataset of the same competition for detection purposes. Our GoogLeNet submission to ILSVRC 2014 actually uses  $12\times$  fewer parameters than the winning architecture of Krizhevsky et al [9] from two years ago, while being significantly more accurate. The biggest gains in object-detection have not come from the utilization of deep networks alone or bigger models, but from the synergy of deep architectures and classical computer vision, like the R-CNN algorithm by Girshick et al [6].

Another notable factor is that with the ongoing traction of mobile and embedded computing, the efficiency of our algorithms – especially their power and memory use – gains importance. It is noteworthy that the considerations leading to the design of the deep architecture presented in this paper included this factor rather than having a sheer fixation on accuracy numbers. For most of the experiments, the models were designed to keep a computational budget of 1.5 billion multiply-adds at inference time, so that they do not end up to be a purely academic curiosity, but could be put to real world use, even on large datasets, at a reasonable cost.

# 1. INTRODUCCIÓN

- El enfoque es abordar el reconocimiento de imágenes utilizando redes convolucionales.
- El paper introduce la arquitectura Inception-v1, que se caracteriza por su enfoque modular y el uso de módulos Inception.

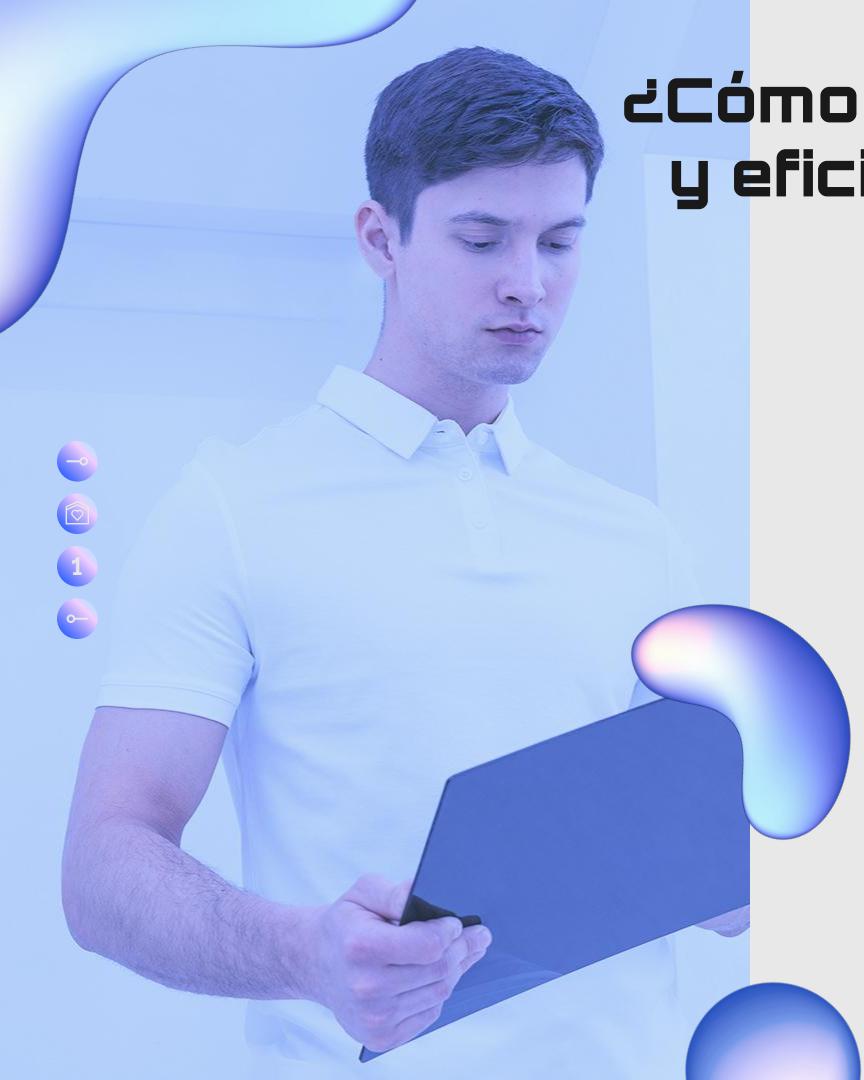
# Desafíos que aborda Inception-v1



Dos clases distintas de entre las 1000 clases del ILSVRC 2014.  
Se requiere conocimiento del dominio para distinguir entre estas  
clases.

- Aumento en la cantidad de parámetros
- Dificultad de entrenamiento a medida que se agregan capas adicionales.



A photograph of a young man with short brown hair, wearing a white button-down shirt, looking down at a tablet device he is holding in his hands. He is standing in front of a light blue wall. The image is partially overlaid by a large, semi-transparent white rectangle containing the main text.

# ¿Cómo se puede mejorar la precisión y eficiencia en el reconocimiento de imágenes?

La arquitectura de **Inception.v1** utiliza módulos que realizan convoluciones en paralelo con filtros de diferentes tamaños permitiendo: y combinando eficientemente la información Enriquecida en una sola capa.

- Capturar características en diferentes escalas al realizar convoluciones en paralelo con filtros de diferentes tamaños.
- Combinar eficientemente la información Enriquecida en una sola capa.

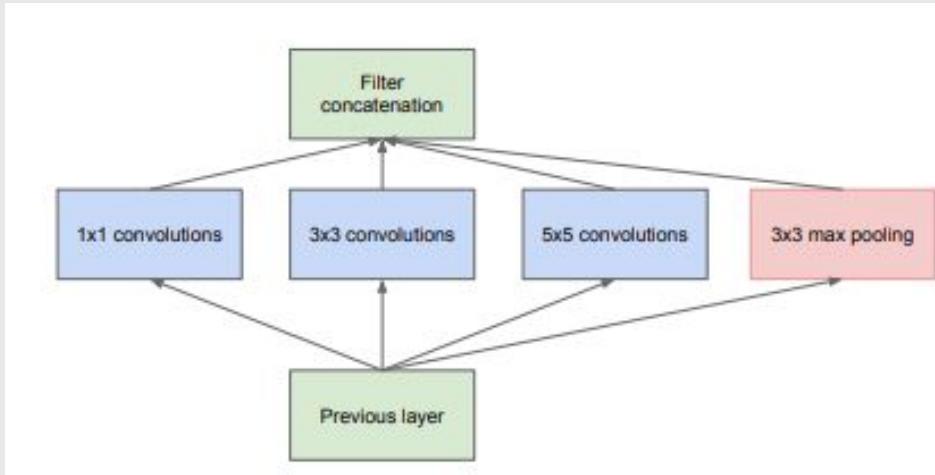
# Reducción de dimensionalidad: Capas 1x1

- Uso de capas de reducción de dimensionalidad en Inception-v1.
- Las convoluciones 1x1 se utilizan para disminuir la complejidad computacional y reducir la cantidad de parámetros en la red.
- La reducción de dimensionalidad permite que la red sea más profunda y eficiente sin sacrificar el rendimiento.



# ARQUITECTURA

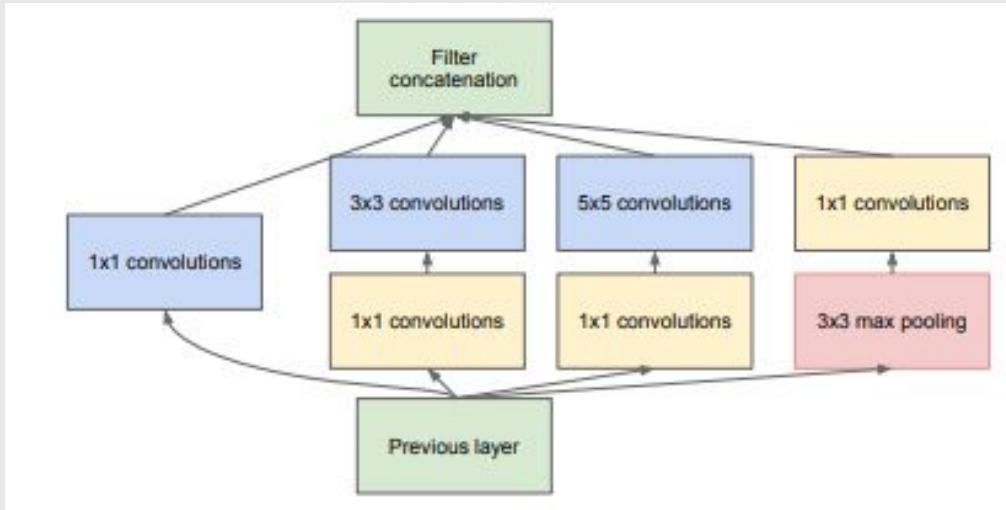
## Estructura principal



Módulo Inception naive-version

- Un módulo Inception está compuesto por múltiples rutas de convolución en paralelo con diferentes tamaños de filtro.
- Cada ruta utiliza convoluciones de 1x1, 3x3 y 5x5 para capturar características en diferentes escalas.

# ARQUITECTURA



Módulo Inception con reducción de dimensionalidad

## Factor de Escala: Convolución 1x1

- Las convoluciones 1x1 se utilizan para ajustar la dimensionalidad y el número de canales antes de las convoluciones de mayor tamaño.
- El factor de escala ayuda a controlar la cantidad de cálculos y parámetros de la red.

# ARQUITECTURA

12	20	30	0
8	12	2	0
34	70	37	4
112	100	25	12

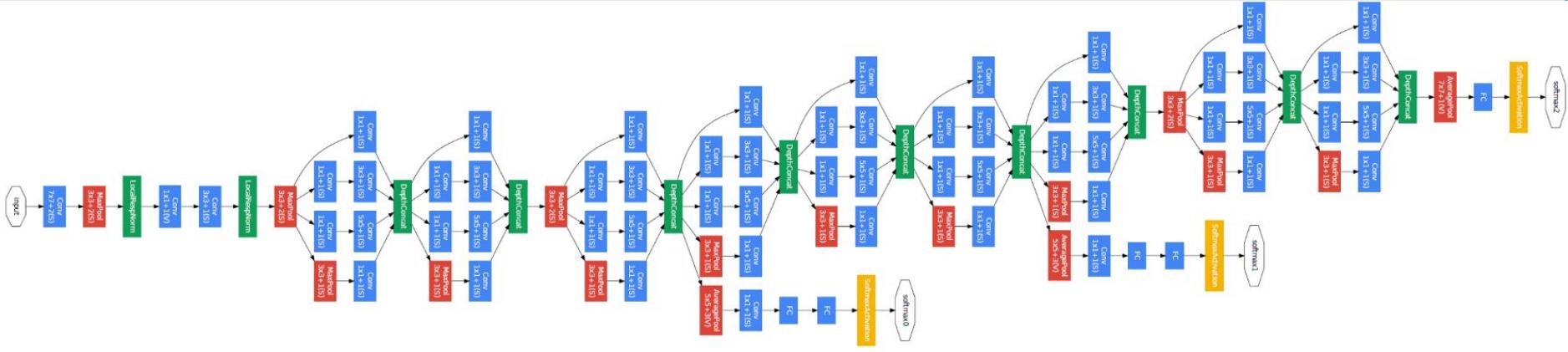
$2 \times 2$  Max-Pool

20	30
112	37

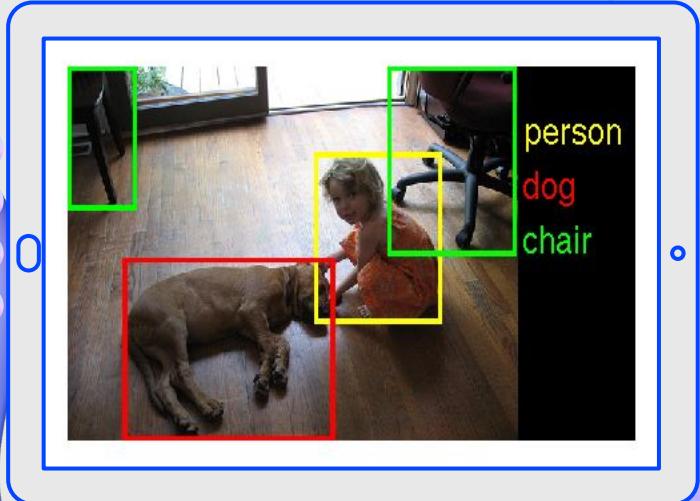
## Capa de Pooling: Reducción de Espacio

- La capa de pooling utilizada en Inception-v1 para reducir el tamaño espacial de los mapas de características.
- El pooling reduce la resolución espacial y el costo computacional en comparación con las convoluciones tradicionales.

# ARQUITECTURA



# CONJUNTO DE DATOS: DETECCIÓN



- 200 categorías de nivel básico completamente anotadas en los datos de prueba.
- Las categorías se eligieron cuidadosamente teniendo en cuenta distintos factores.
- Algunas de las imágenes de prueba no contienen ninguna de las 200 categorías.

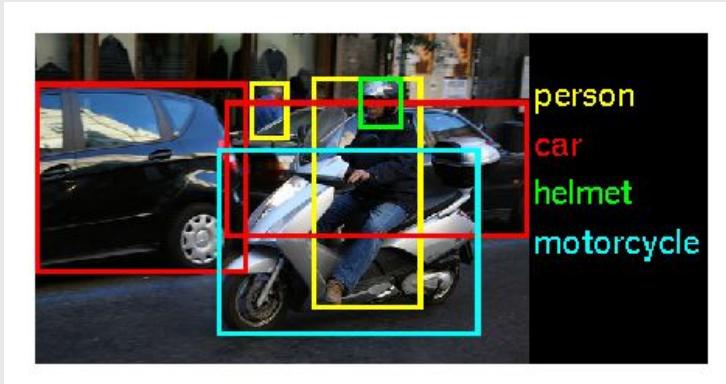
# CONJUNTO DE DATOS: DETECCIÓN

	ENTRENAMIENTO		VALIDACIÓN		PRUEBA	
NÚMERO DE CLASES DE OBJETOS	NÚMERO DE IMÁGENES	NÚMERO DE OBJETOS	NÚMERO DE IMÁGENES	NÚMERO DE OBJETOS	NÚMERO DE IMÁGENES	NÚMERO DE OBJETOS
200	395909	345854	20121	55502	40152	---





- **Average image resolution:** 482x415 pixels
- **Average object classes per image:** 1.534
- **Average object instances per image:** 2.758
- **Average object scale:** 0.170



# **CONJUNTO DE DATOS: CLASIFICACIÓN Y CLASIFICACIÓN CON LOCALIZACIÓN**

- Conjunto de validación y test consisten en 150000 fotografías.
- Etiquetas con la presencia o ausencia de 1000 categorías de objetos.
- Subconjunto aleatorio de 50000 de las imágenes con etiquetas se utilizaron como datos de validación junto con una lista de las 1.000 categorías.
- Las imágenes restantes se utilizaron para la evaluación sin proveer sus etiquetas.

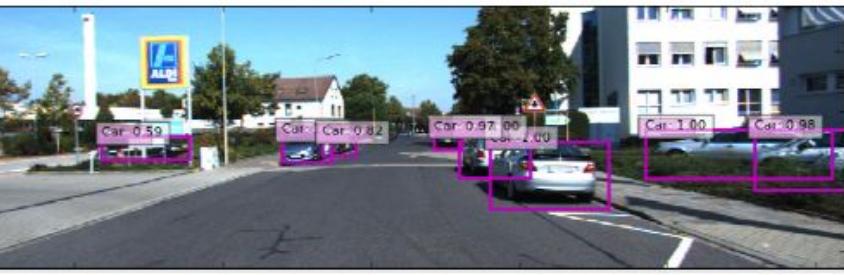
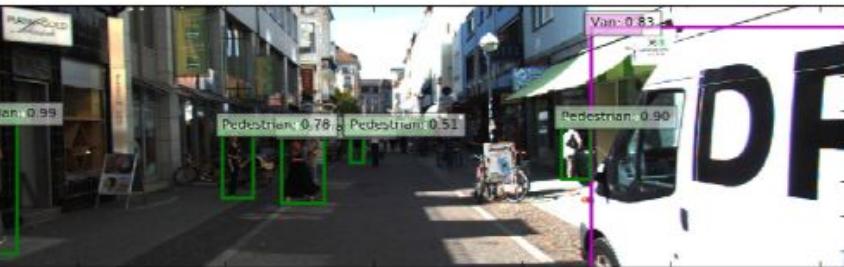
<b>Team</b>	<b>Year</b>	<b>Place</b>	<b>Error (top-5)</b>	<b>Uses external data</b>
SuperVision	2012	1st	16.4%	no
SuperVision	2012	1st	15.3%	Imagenet 22k
Clarifai	2013	1st	11.7%	no
Clarifai	2013	1st	11.2%	Imagenet 22k
MSRA	2014	3rd	7.35%	no
VGG	2014	2nd	7.32%	no
GoogLeNet	2014	1st	6.67%	no

<b>Number of models</b>	<b>Number of Crops</b>	<b>Cost</b>	<b>Top-5 error</b>	<b>compared to base</b>
1	1	1	10.07%	base
1	10	10	9.15%	-0.92%
1	144	144	7.89%	-2.18%
7	1	7	8.09%	-1.98%
7	10	70	7.62%	-2.45%
7	144	1008	6.67%	-3.45%

# Desafío ImageNet

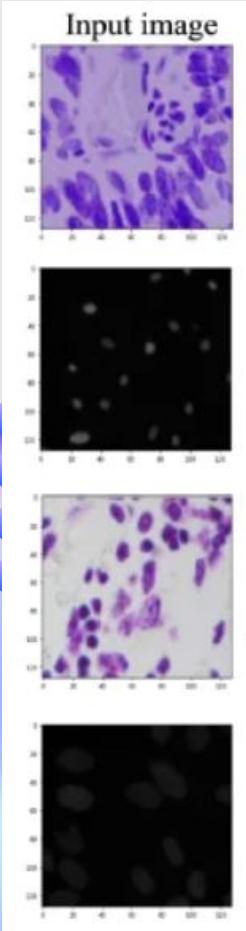
- Inception-v1 ha sido evaluado en el desafío ImageNet, una competencia de clasificación de imágenes.
- Inception-v1 logró una precisión de top-1 del 69.8% y una precisión de top-5 del 89.8% en este desafío.
- Inception-v1 es capaz de clasificar imágenes en diversas categorías, como personas, animales, vehículos, edificios, paisajes, entre otros.

# Detección de objetos



- Inception-v1 se ha utilizado para la detección precisa de objetos en imágenes.
- Puede combinarse con técnicas de detección para localizar y etiquetar objetos en imágenes con alta precisión.

# Segmentación semántica



- Inception-v1 es aplicable en tareas de segmentación semántica.
- Puede proporcionar mapas detallados y precisos de segmentación al fusionar características de diferentes escalas.

# Conclusiones

- La arquitectura Inception-v1 ha logrado un rendimiento sobresaliente en el desafío de clasificación de imágenes ILSVRC 2014, superando a los enfoques existentes.
- Su enfoque de diseño, como el uso de módulos de Inception y la reducción de dimensionalidad, ha demostrado ser efectivo en la captura de características y la mejora de la precisión.
- Inception-v1 ha demostrado ser eficiente en términos de uso de recursos computacionales en comparación con otras arquitecturas.
- La escalabilidad de la arquitectura permite adaptarse a diferentes tamaños de imágenes y requisitos de recursos, lo que facilita su implementación en diversas aplicaciones.

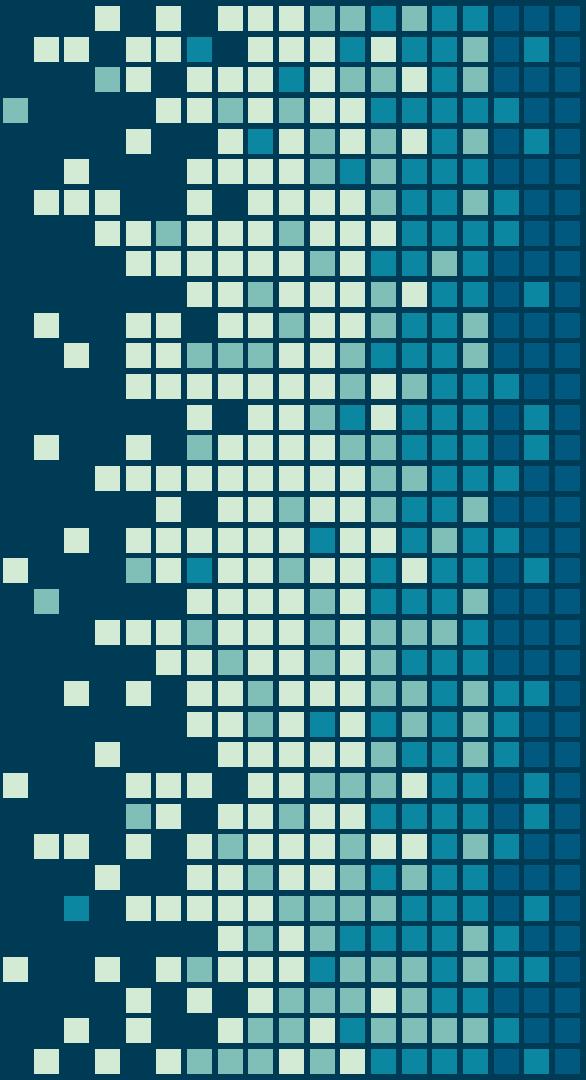
# Thanks

# Redes Neuronales Convolucionales:LeNet-5

Integrantes:

- **Quispe Huanca, Joselyn Lizeth**
- **Macedo Huaman Vanessa Mayra**

# 1. INTRODUCCIÓN



LeNet se refiere a LeNet-5 y es una red neuronal convolucional simple.

LeNet-5 fue una de las primeras redes neuronales convolucionales y promovió el desarrollo del deep learning.

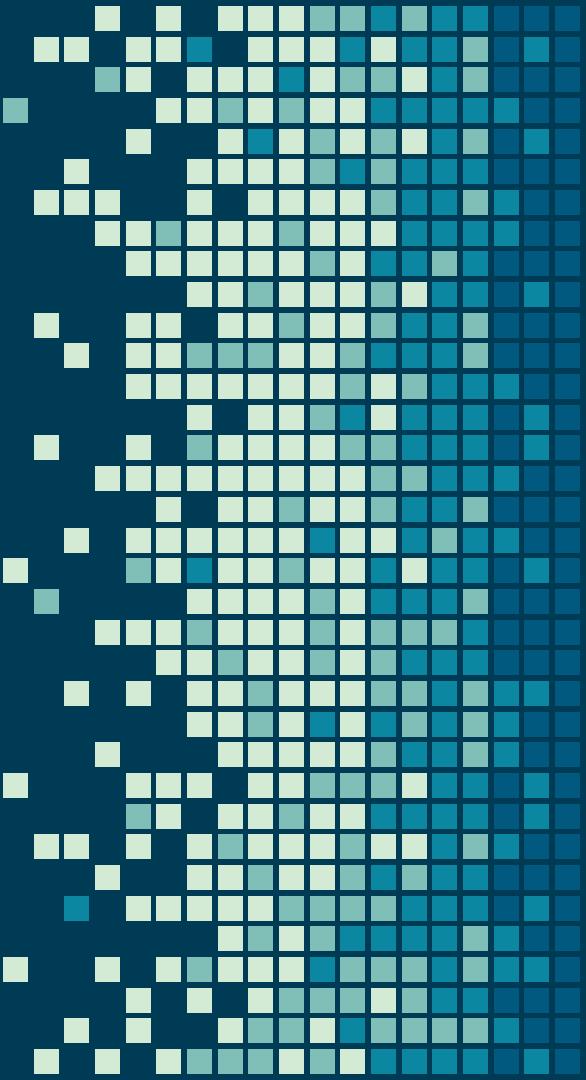
Fue propuesto en 1989 por Yann LeCun



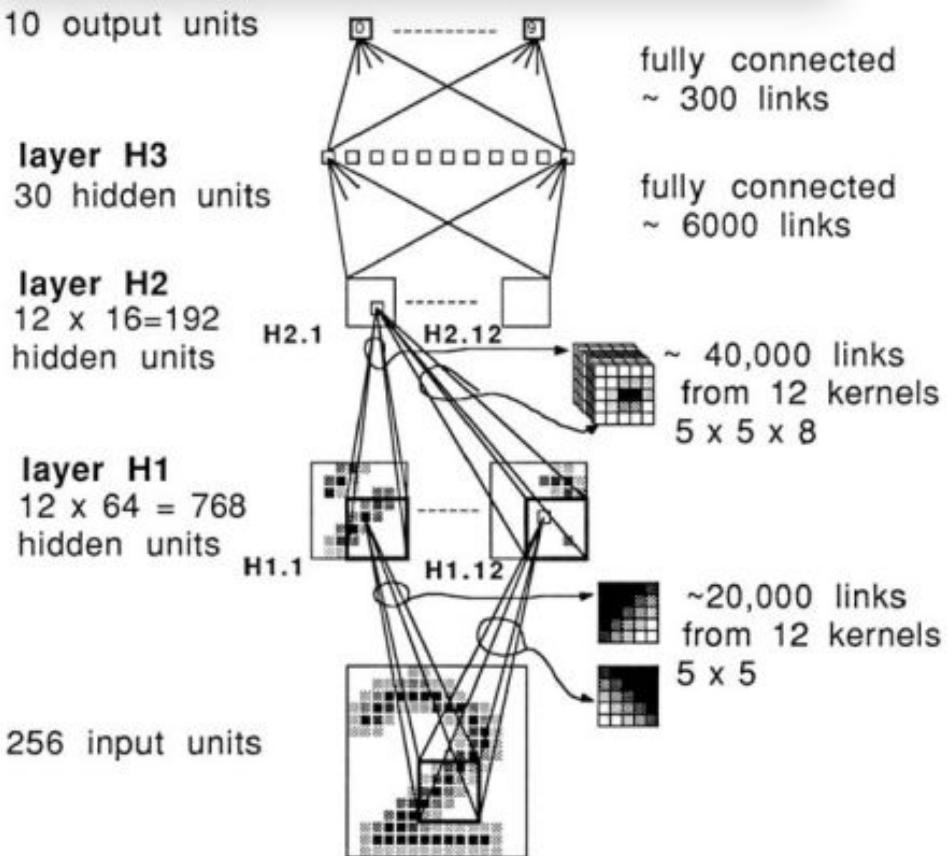
**Yann LeCun**

# 2. PAPER

Backpropagation Applied to  
Handwritten Zip Code Recognition

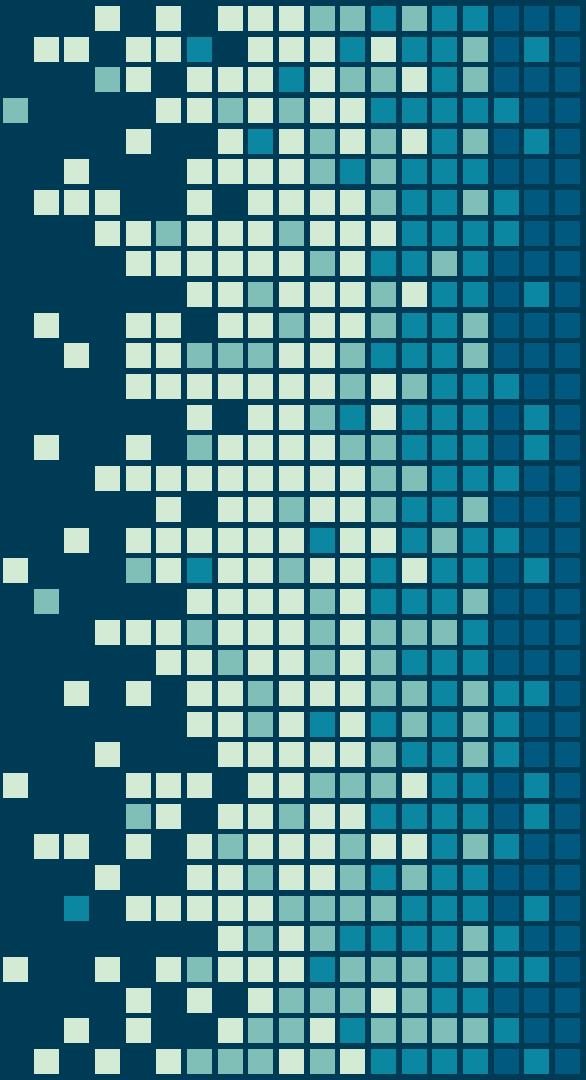


1611915485726803224414186  
 6359720299299722510044701  
 3084111591010615406103631  
 1064111030475262001179966  
 8912056708557131427955460  
 2018730187112993089970984  
 0109707597331972015519055  
 1075318255182814358090943  
 1787541655460354603546055  
 18255108503047520439401

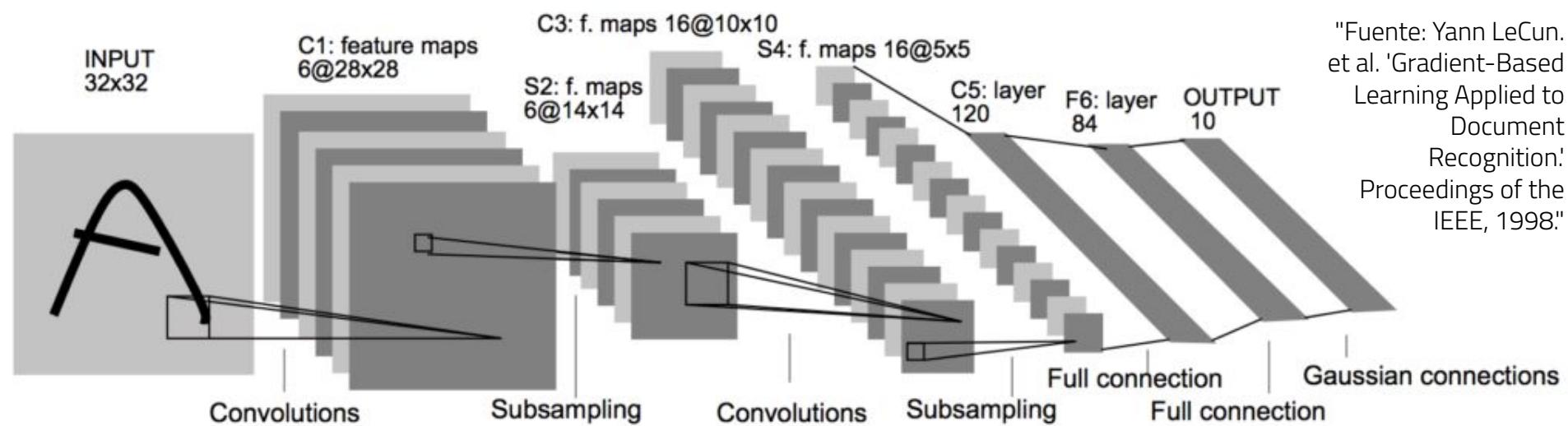


# 3. ARQUITECTURA

LeNet-5

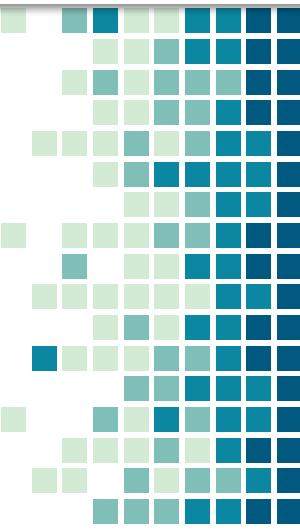


"Fuente: Yann LeCun,  
et al. 'Gradient-Based  
Learning Applied to  
Document  
Recognition:  
Proceedings of the  
IEEE, 1998."

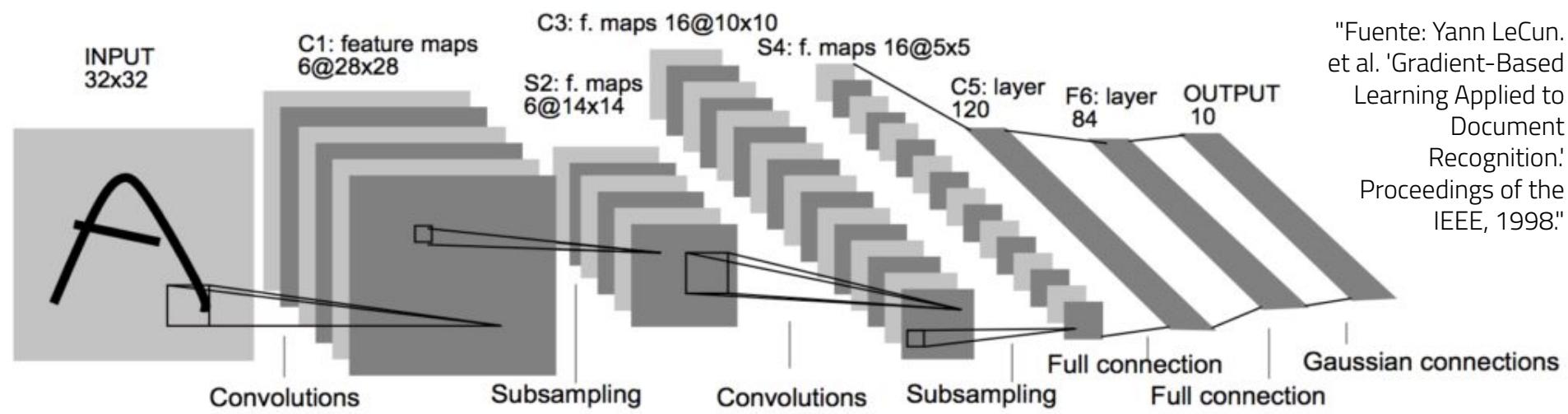


## Contiene:

- Capa de entrada
- Capa convolucional (C1)
- Capa de pooling (S2)
- Capa convolucional (C3)
- Capa de pooling (S4)
- Capa completamente conectada (C5)
- Capa completamente conectada (F6)
- Capa de salida



"Fuente: Yann LeCun, et al. 'Gradient-Based Learning Applied to Document Recognition: Proceedings of the IEEE, 1998."



## Capa de entrada

- Imágenes en escala de grises de 32x32 píxeles.

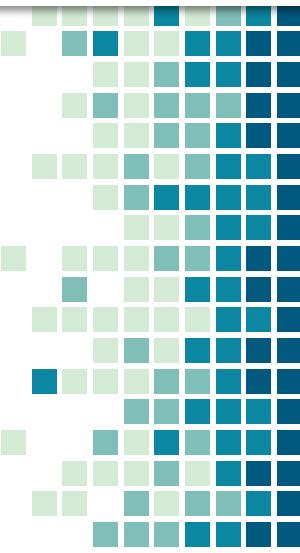
## Capa convolucional (C1)

- 6 filtros de tamaño 5x5
- Tangente hiperbólica ( $\tanh$ )

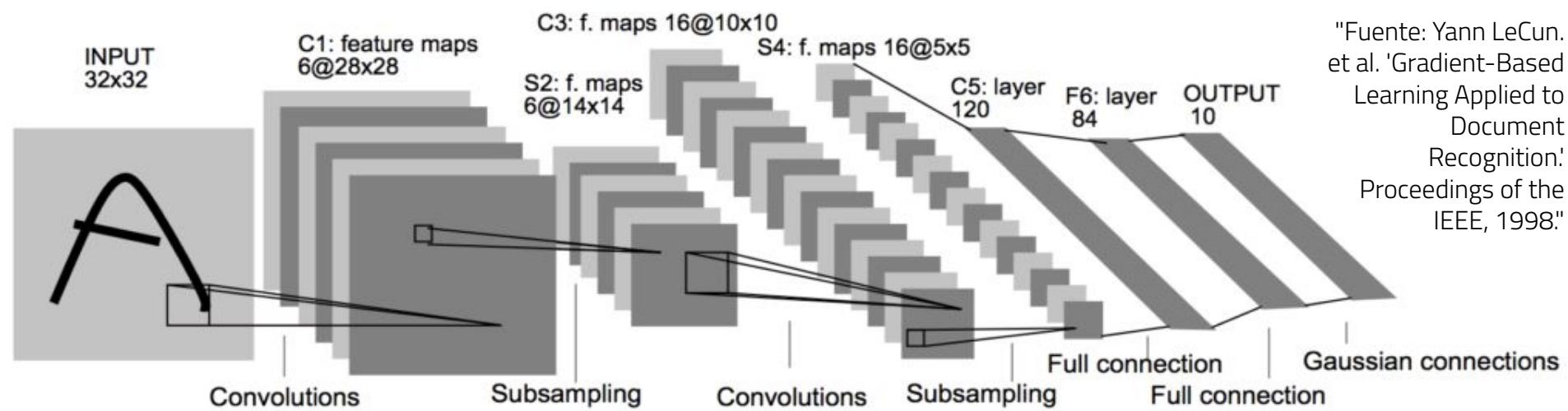
- Stride: 1
- 6 mapas de características de 28x28

## Capa de pooling (S2)

- Average pooling.
- Tamaño de ventana 2x2
- Stride: 2
- 6 mapas de características de 14x14



"Fuente: Yann LeCun,  
et al. 'Gradient-Based  
Learning Applied to  
Document  
Recognition:  
Proceedings of the  
IEEE, 1998."



## Capa convolucional (C3)

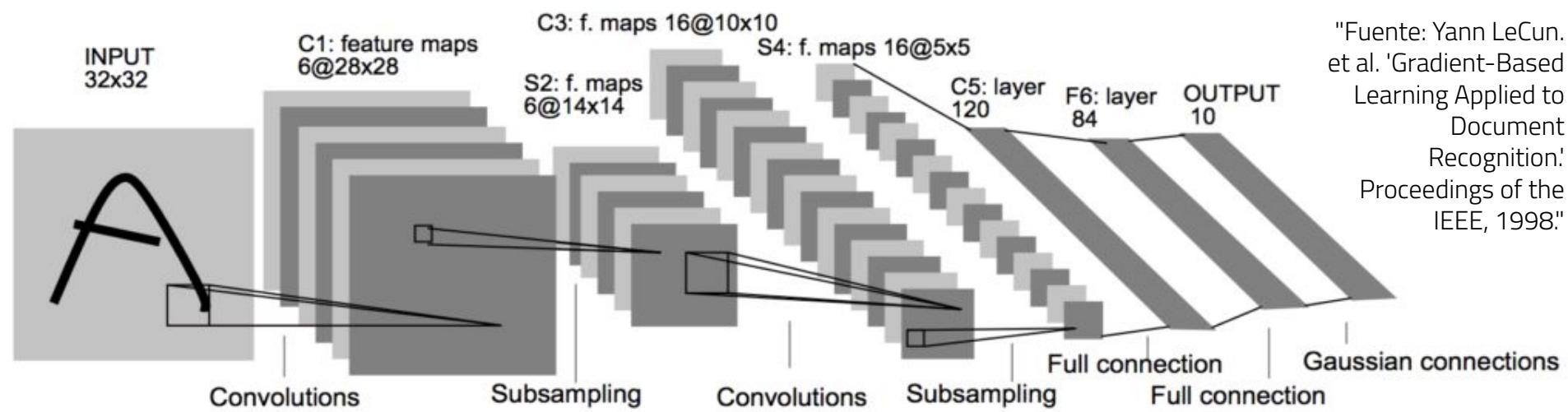
- 16 filtros de tamaño 5x5
- Tangente hiperbólica (tanh)
- Stride: 1
- 16 mapas de características de 10x10

## Capa de pooling (S4)

- Average pooling.
- Tamaño de ventana 2x2
- Stride: 2
- 16 mapas de características de 5x5.



"Fuente: Yann LeCun,  
et al. 'Gradient-Based  
Learning Applied to  
Document  
Recognition:  
Proceedings of the  
IEEE, 1998."



## Capa completamente conectada (C5)

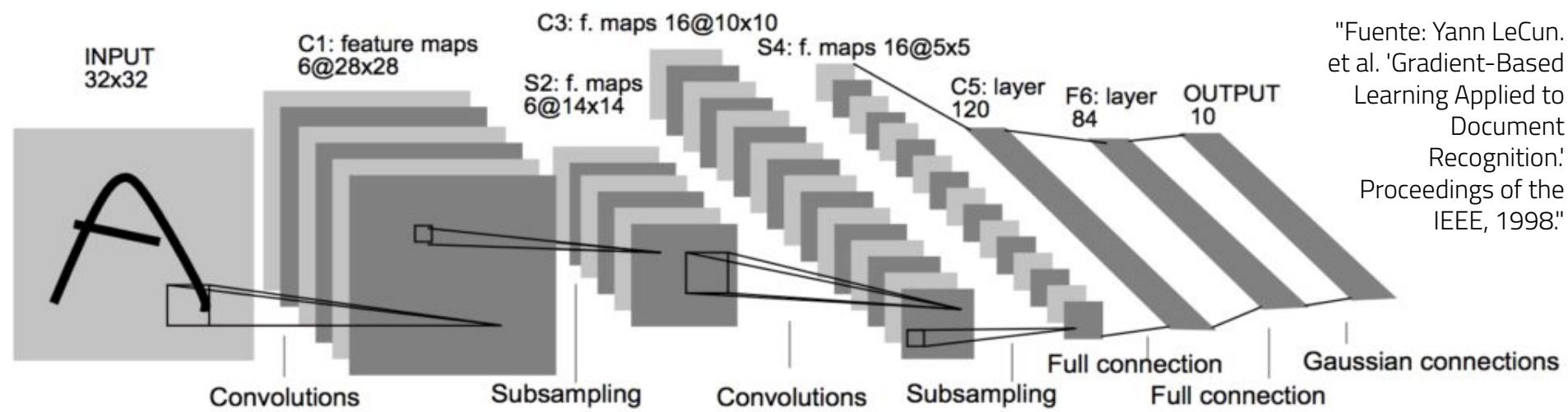
- 120 neuronas totalmente conectadas.
- Tangente hiperbólica (tanh).
- Un vector de 120 dimensiones.

## Capa completamente conectada (F6)

- 84 neuronas totalmente conectadas.
- Tangente hiperbólica (tanh).
- Un vector de 84 dimensiones.

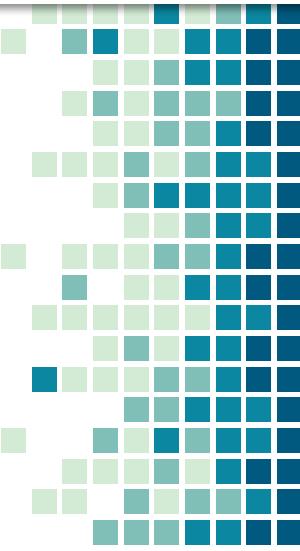


"Fuente: Yann LeCun,  
et al. 'Gradient-Based  
Learning Applied to  
Document  
Recognition:  
Proceedings of the  
IEEE, 1998."

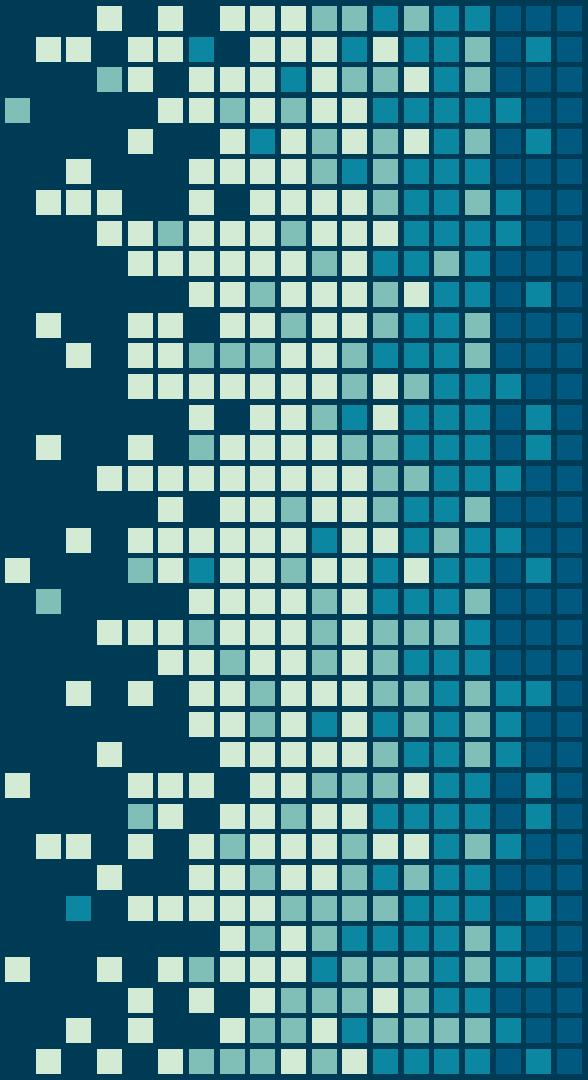


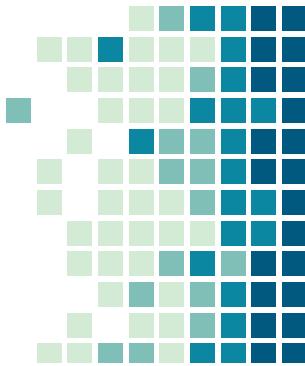
## Capa de salida

- 10 neuronas (una para cada clase en el conjunto de datos utilizado en el paper).
- Función de activación: Softmax.
- Un vector de probabilidades de 10 dimensiones, representando la probabilidad de cada clase.
- **Gaussian Connections:** Se refiere a la capa de normalización o suavizado, o a la función de activación gaussiana en la capa de salida softmax



# 4. DATASET MNIST



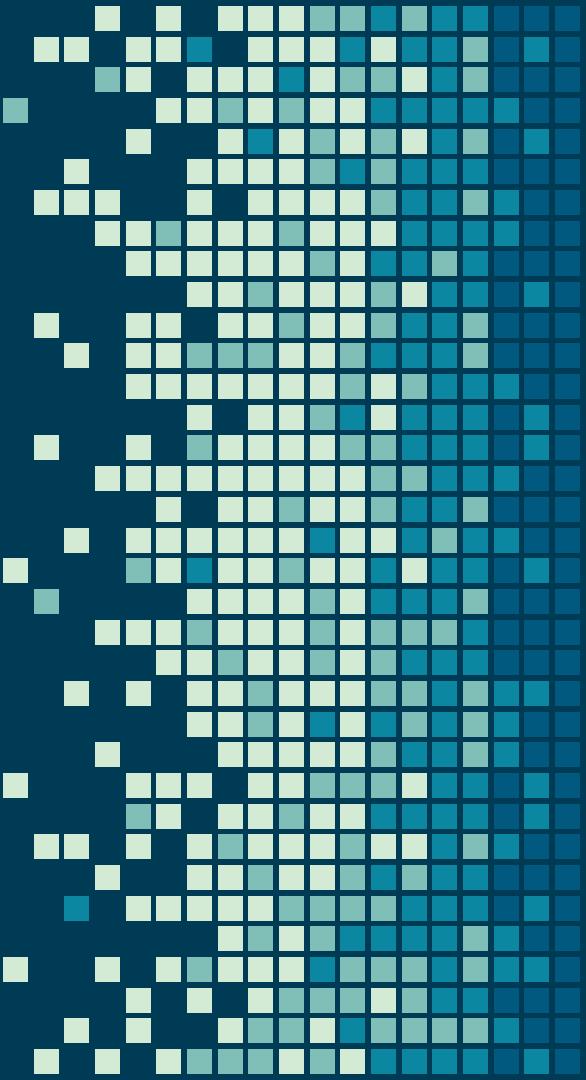


# MNIST (Modified National Institute of Standards and Technology database)

- Consiste en imágenes de dígitos escritos a mano, en escala de grises de 28x28 píxeles, donde cada imagen representa un dígito del 0 al 9.
- Conjunto de entrenamiento: 60,000.
- Conjunto de prueba: 10,000.

3	6	8	1	7	9	6	6	9	1
6	7	5	7	8	6	3	4	8	5
2	1	7	9	7	1	2	8	4	6
4	8	1	9	0	1	8	8	9	4
7	6	1	8	6	4	1	5	6	0
7	5	9	2	6	5	8	1	9	7
2	2	2	2	3	4	4	8	0	
0	2	3	8	0	7	3	8	5	7
0	1	4	6	4	6	0	2	4	3
7	1	2	8	7	6	9	8	6	1

# 5. APLICACIÓN



# APLICACIONES

## Reconocimiento de dígitos y caracteres

Reconocimiento de matrículas de vehículos, reconocimiento de códigos de barras, etc.

## Reconocimiento de objetos

Se utiliza para tareas de clasificación de objetos, detección de objetos y segmentación

## Reconocimiento facial

Detectar y reconocer caras en imágenes o videos

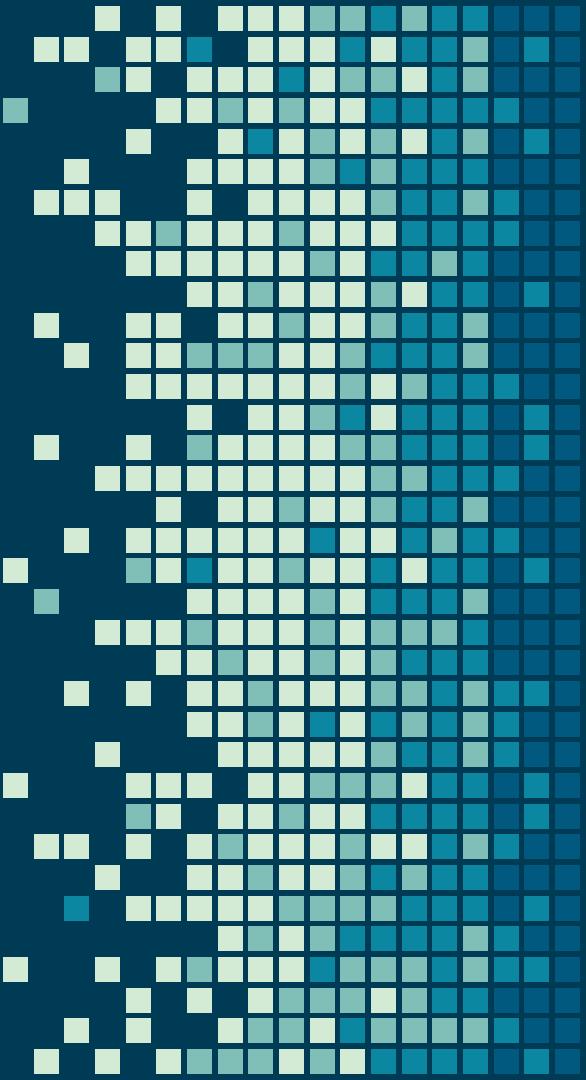
## Visión por computadora en general

Detección de anomalías, reconocimiento de gestos, seguimiento de objetos, reconstrucción 3D a partir de imágenes, entre otros.

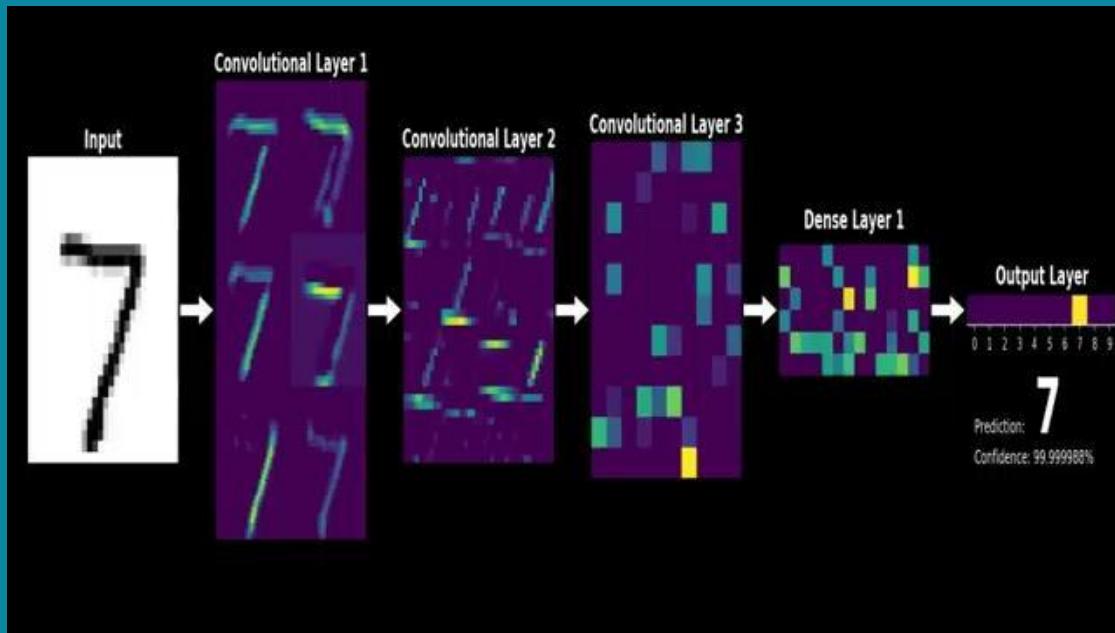
## Análisis de imágenes médicas

Detectar cáncer de mama, la retinopatía diabética, la detección de tumores cerebrales, entre otros

# 6. Demo



# “ DEMO





# CONCLUSIONES

- LeNet fue una de las primeras arquitecturas exitosas de redes neuronales convolucionales, diseñada específicamente para el reconocimiento de caracteres escritos a mano.
- La estructura jerárquica de LeNet, permitió extraer características relevantes de las imágenes de entrada y lograr un alto rendimiento en la clasificación de dígitos..

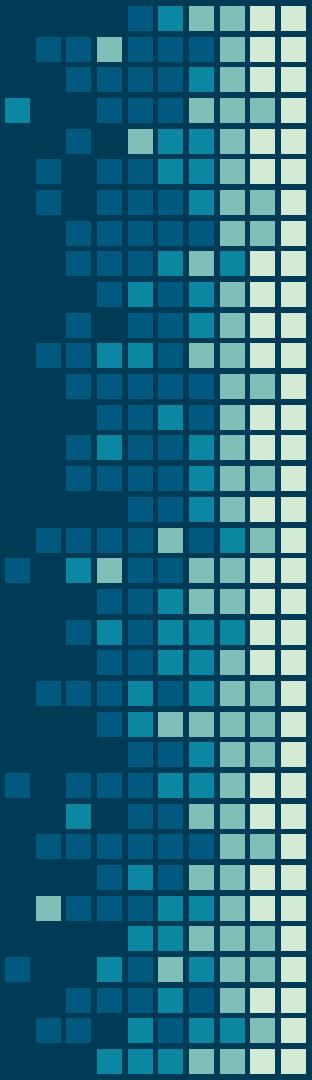


# CONCLUSIONES

- El entrenamiento y evaluación de LeNet se realizó utilizando el dataset MNIST, demostrando su eficacia en la clasificación de imágenes de dígitos escritos a mano.
- LeNet sentó las bases para el desarrollo y avance de otras arquitecturas más sofisticadas de redes neuronales convolucionales, y ha sido ampliamente utilizada en aplicaciones de reconocimiento de patrones y visión por computadora.

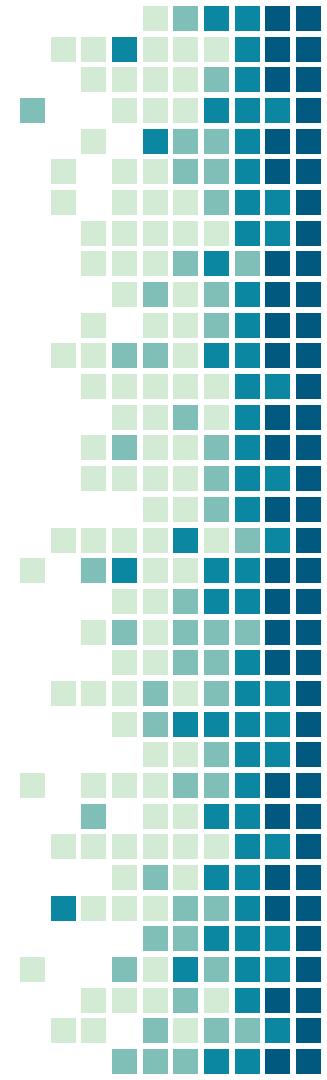
# THANKS!

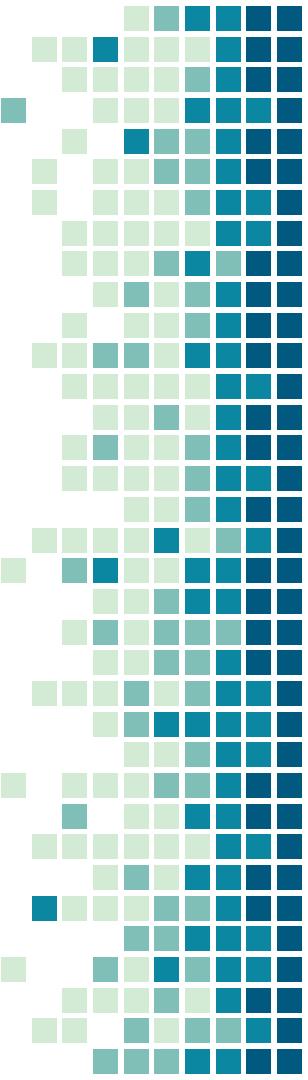
Any questions?



# BIBLIOGRAFÍA

"Fuente: Yann LeCun. et al. 'Gradient-Based Learning Applied to Document Recognition.' Proceedings of the IEEE, 1998."





# THIS IS A SLIDE TITLE

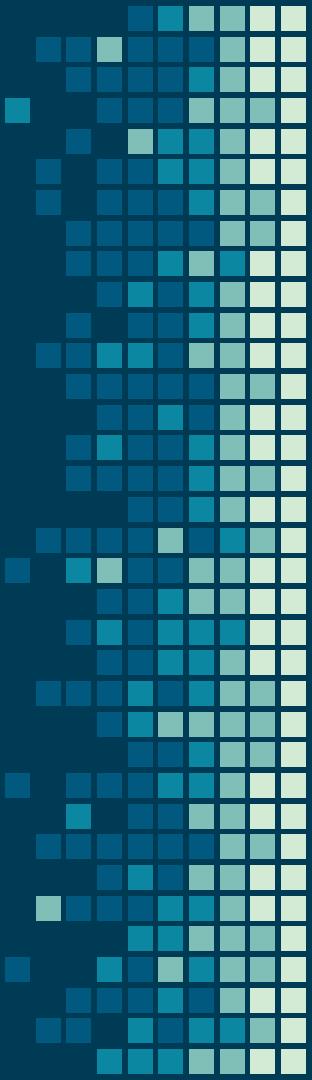
- Here you have a list of items
- And some text
- But remember not to overload your slides with content

Your audience will listen to you or read the content,  
but won't do both.



# BIG PT

Bring the attention of your audience over  
a key concept using icons or illustrations



# IN TWO OR THREE COLUMNS

## Yellow

Is the color of gold, butter and ripe lemons. In the spectrum of visible light, yellow is found between green and orange.

## Blue

Is the colour of the clear sky and the deep sea. It is located between violet and green on the optical spectrum.

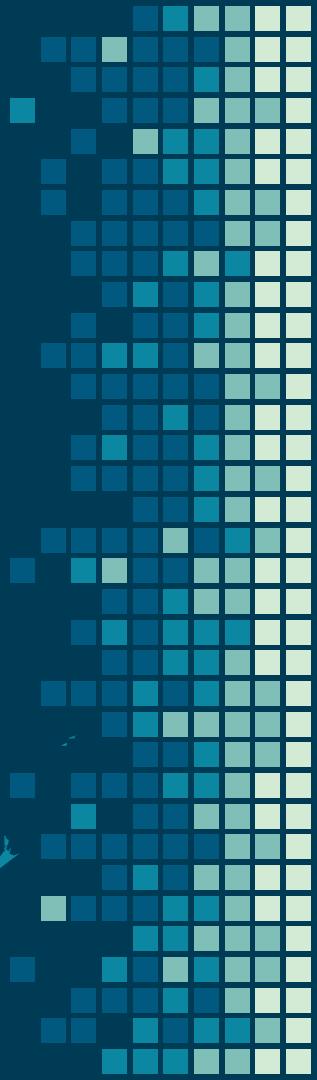
## Red

Is the color of blood, and because of this it has historically been associated with sacrifice, danger and courage.

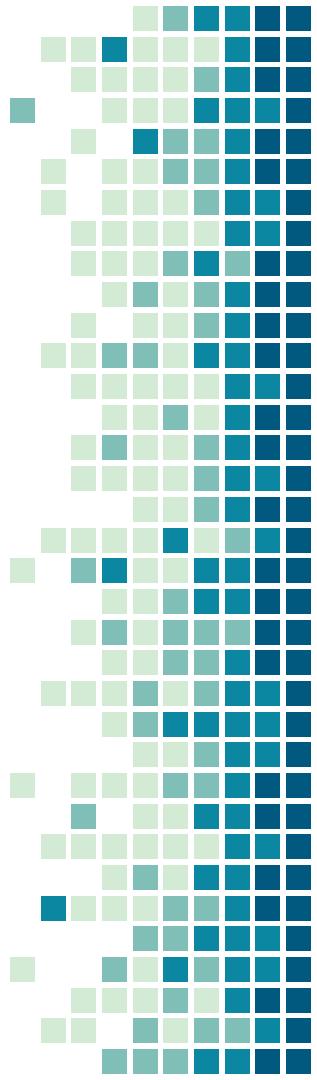
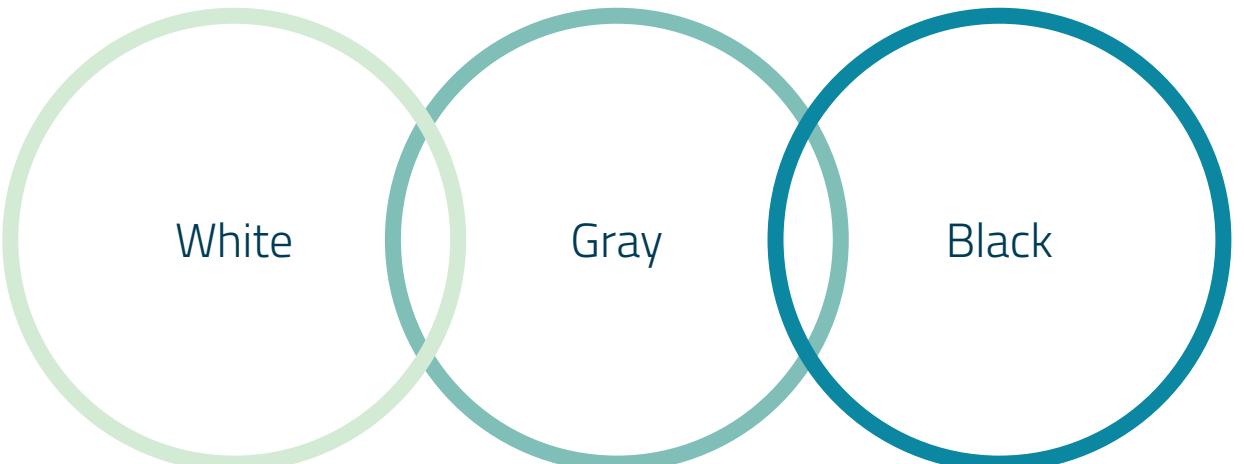
A photograph of a woman with long brown hair, seen from behind, sitting at a wooden desk. She is holding a silver spoon and stirring a dark liquid in a grey mug. A small Tazo tea bag hangs from the bottom of the mug. Her right hand rests on the trackpad of an open white laptop. The laptop screen displays a website with a grid of images and the word "Design".

Want big impact?  
Use big image.

# MAPS

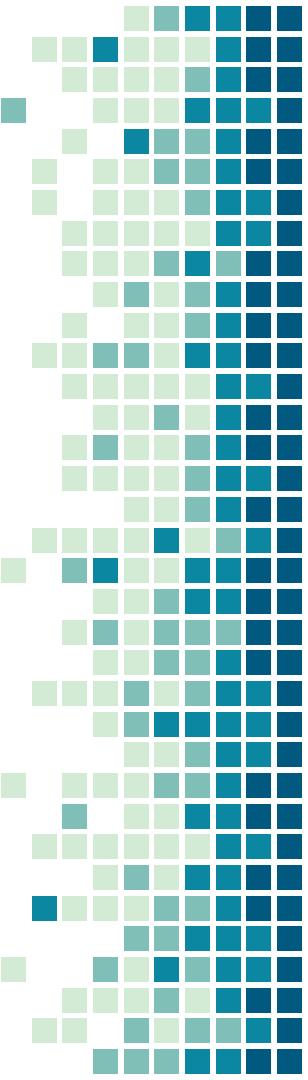


# USE CHARTS TO EXPLAIN YOUR IDEAS



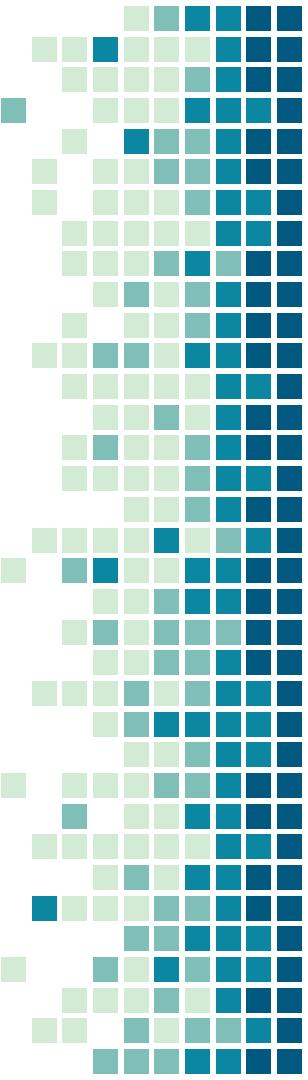
# AND TABLES TO COMPARE DATA

	A	B	C
Yellow	10	20	7
Blue	30	15	10
Orange	5	24	16



89,526,124

Whoa! That's a big number, aren't you proud?



89,526,124\$

That's a lot of money

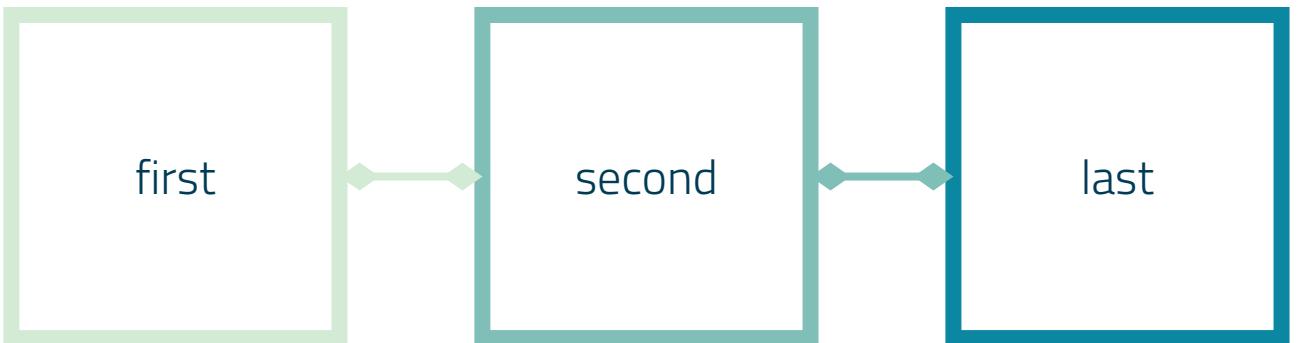
185,244 users

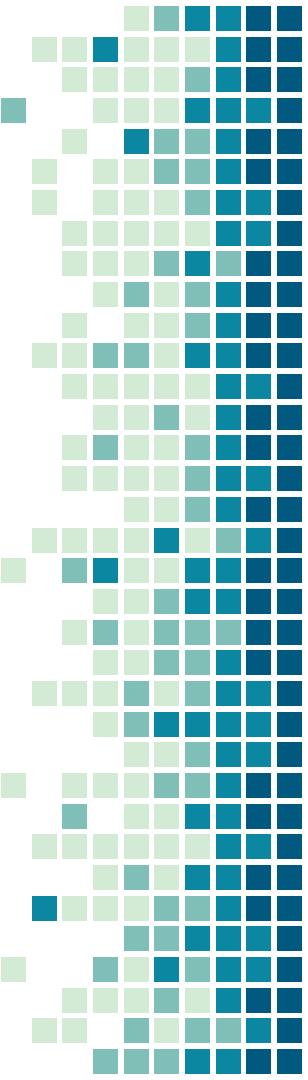
And a lot of users

100%

Total success!

# OUR PROCESS IS EASY





# LET'S REVIEW SOME CONCEPTS

## Yellow

Is the color of gold, butter and ripe lemons. In the spectrum of visible light, yellow is found between green and orange.

## Yellow

Is the color of gold, butter and ripe lemons. In the spectrum of visible light, yellow is found between green and orange.

## Blue

Is the colour of the clear sky and the deep sea. It is located between violet and green on the optical spectrum.

## Blue

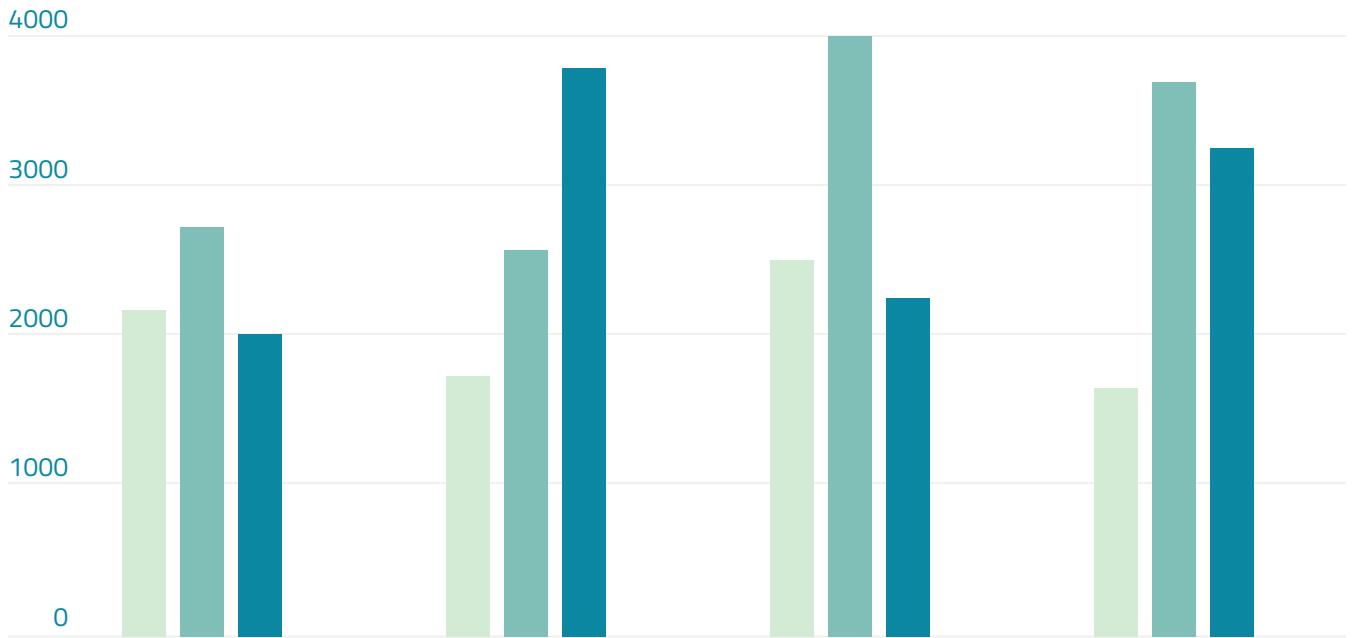
Is the colour of the clear sky and the deep sea. It is located between violet and green on the optical spectrum.

## Red

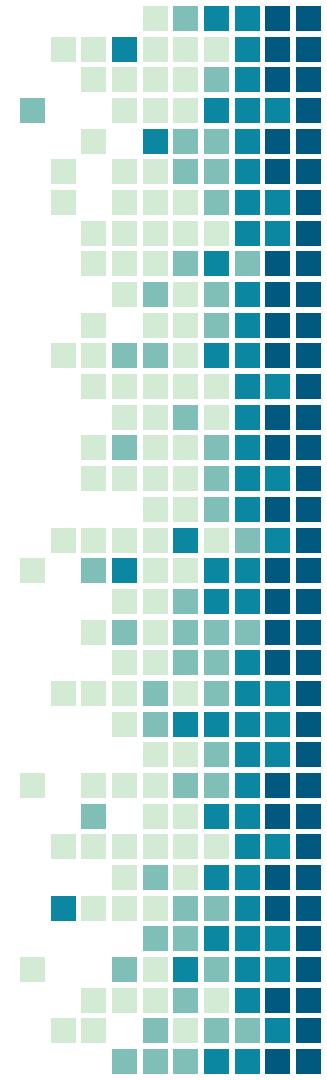
Is the color of blood, and because of this it has historically been associated with sacrifice, danger and courage.

## Red

Is the color of blood, and because of this it has historically been associated with sacrifice, danger and courage.

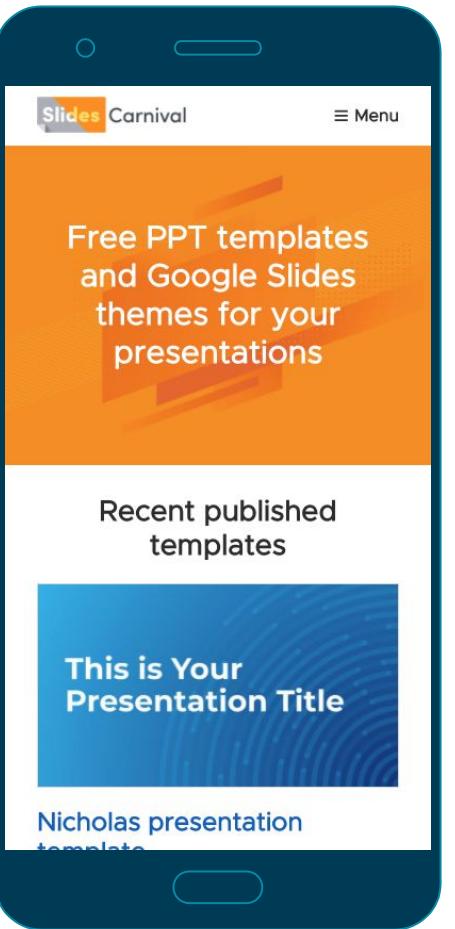


You can insert graphs from Excel or Google Sheets



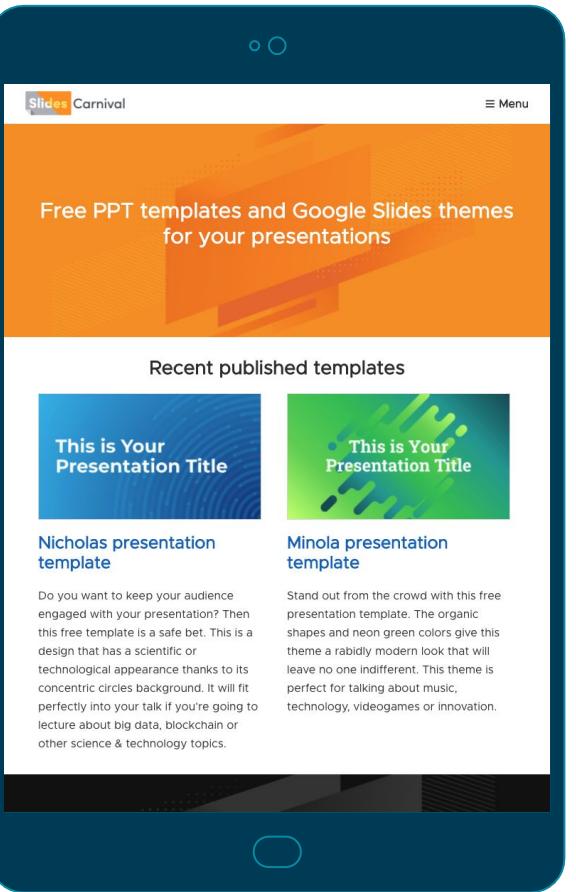
# MOBILE PROJECT

Show and explain your web, app or software projects using these gadget templates.



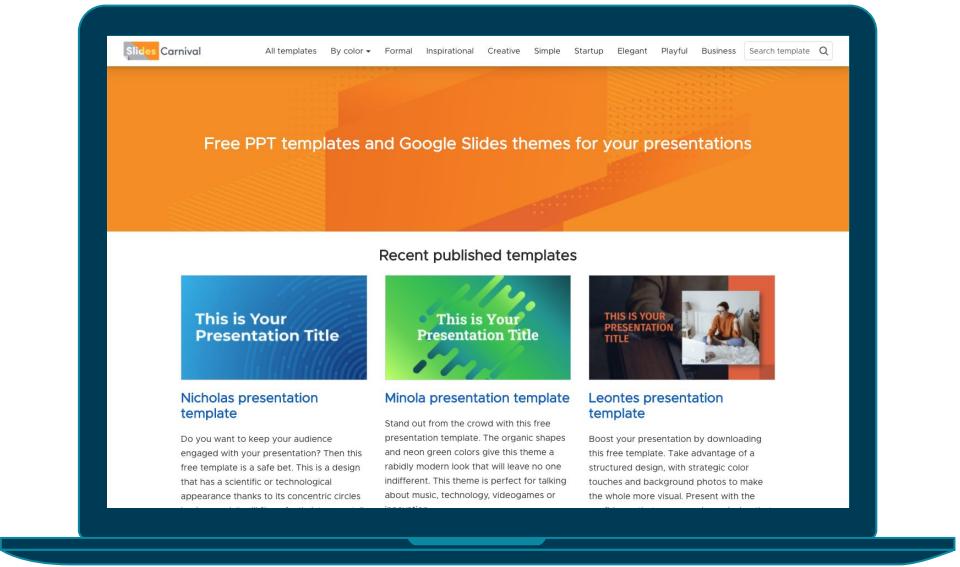
# TABLET PROJECT

Show and explain your web, app or software projects using these gadget templates.



# DESKTOP PROJECT

Show and explain your web, app or software projects using these gadget templates.



# CREDITS

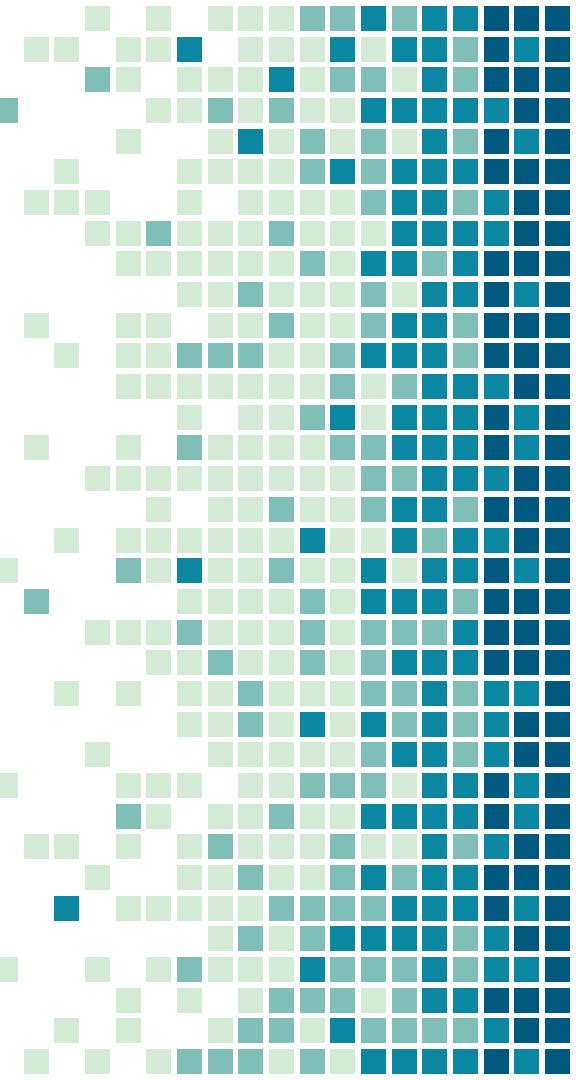
Special thanks to all the people who made and released these awesome resources for free:

- Presentation template by [SlidesCarnival](#)
- Photographs by [Unsplash](#)

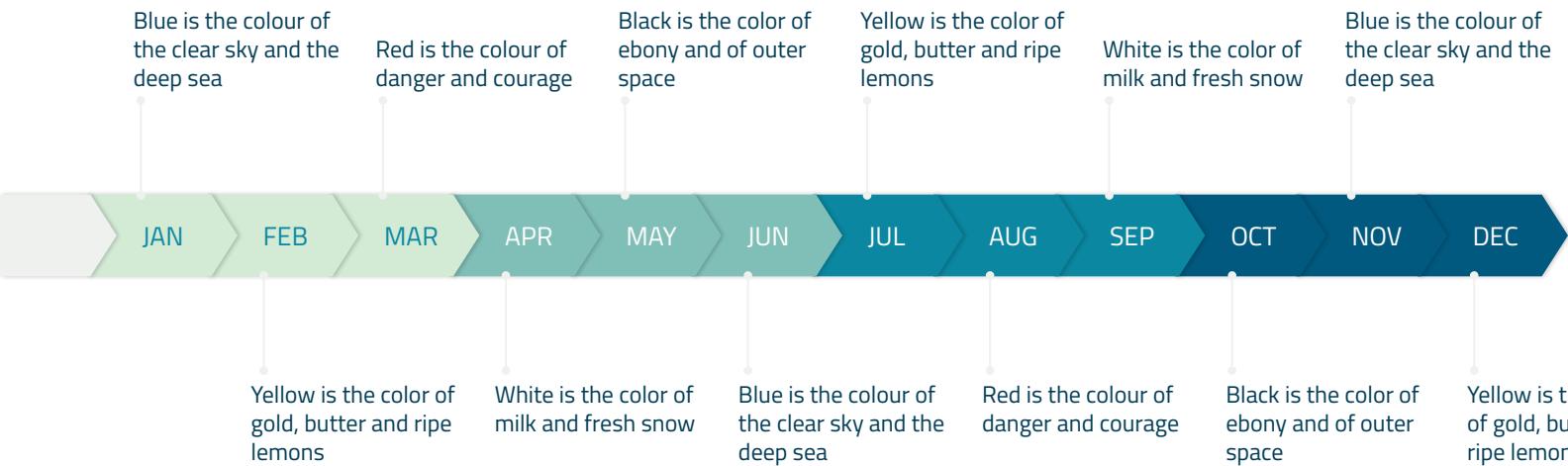
2.

## EXTRA RESOURCES

For Business Plans, Marketing Plans,  
Project Proposals, Lessons, etc



# TIMELINE



# ROADMAP

Blue is the colour of the clear sky and the deep sea

1

Red is the colour of danger and courage

3

Black is the color of ebony and of outer space

5

Yellow is the color of gold, butter and ripe lemons

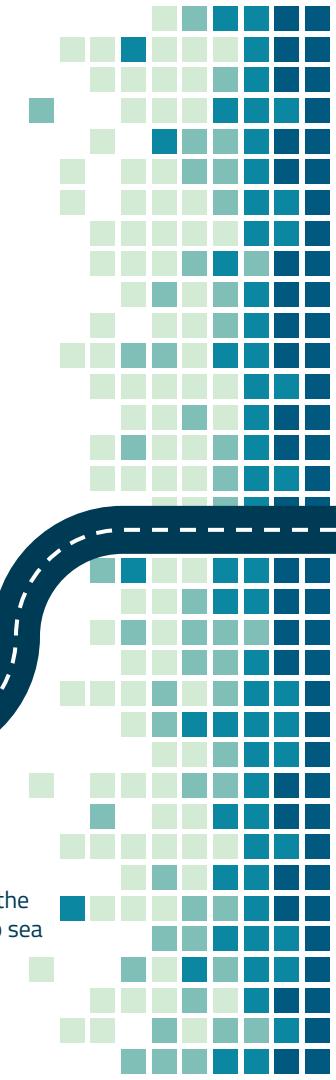
2

White is the color of milk and fresh snow

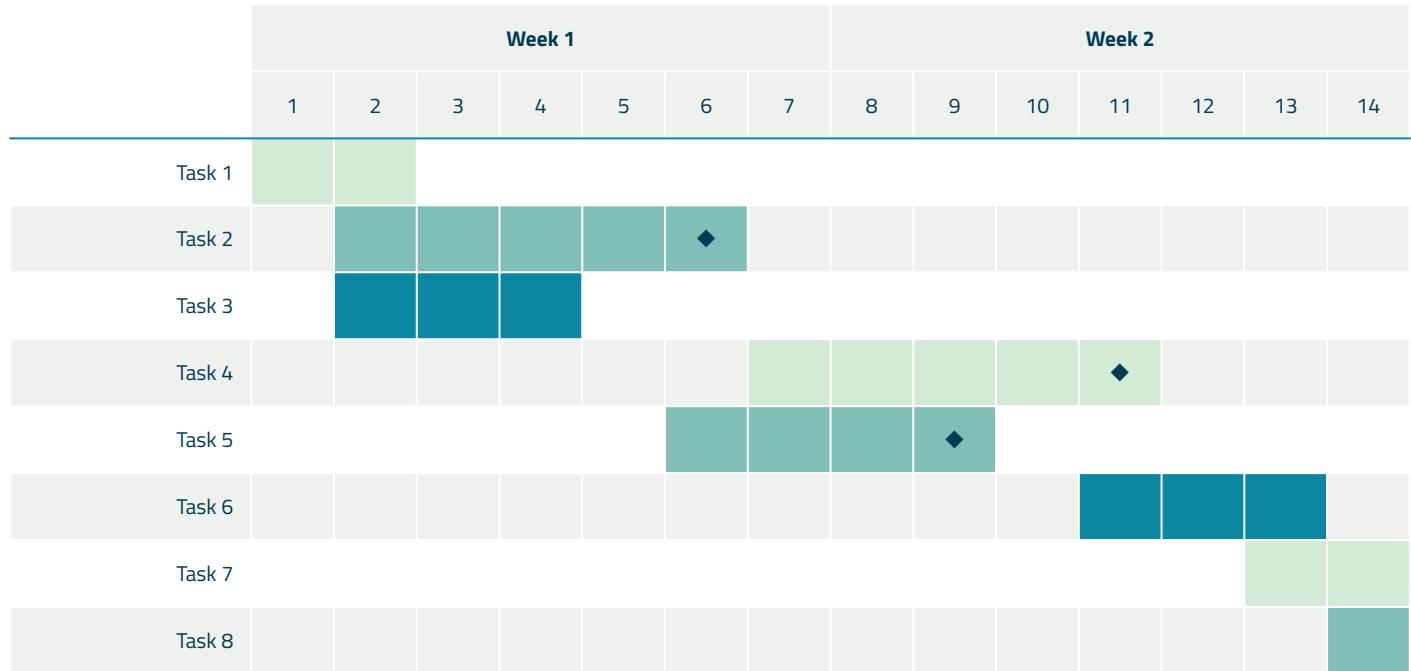
4

Blue is the colour of the clear sky and the deep sea

6



# GANTT CHART



# SWOT ANALYSIS

## STRENGTHS

Blue is the colour of the clear sky and the deep sea

S

Black is the color of ebony and of outer space

## OPPORTUNITIES

## WEAKNESSES

Yellow is the color of gold, butter and ripe lemons

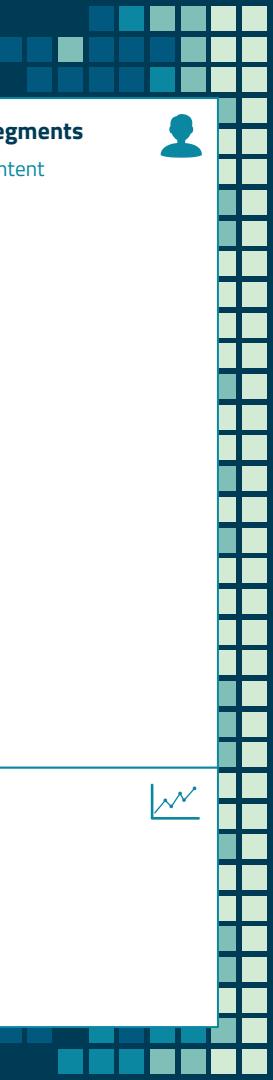
W

White is the color of milk and fresh snow

T

## THREATS

# BUSINESS MODEL CANVAS



<b>Key Partners</b> Insert your content		<b>Key Activities</b> Insert your content		<b>Value Propositions</b> Insert your content		<b>Customer Relationships</b> Insert your content		<b>Customer Segments</b> Insert your content	
<b>Key Resources</b> Insert your content						<b>Channels</b> Insert your content			
<b>Cost Structure</b> Insert your content				<b>Revenue Streams</b> Insert your content					

# FUNNEL



# TEAM PRESENTATION



**Imani Jackson**

JOB TITLE

Blue is the colour of the clear  
sky and the deep sea



**Marcos Galán**

JOB TITLE

Blue is the colour of the clear  
sky and the deep sea



**Ixchel Valdía**

JOB TITLE

Blue is the colour of the clear  
sky and the deep sea



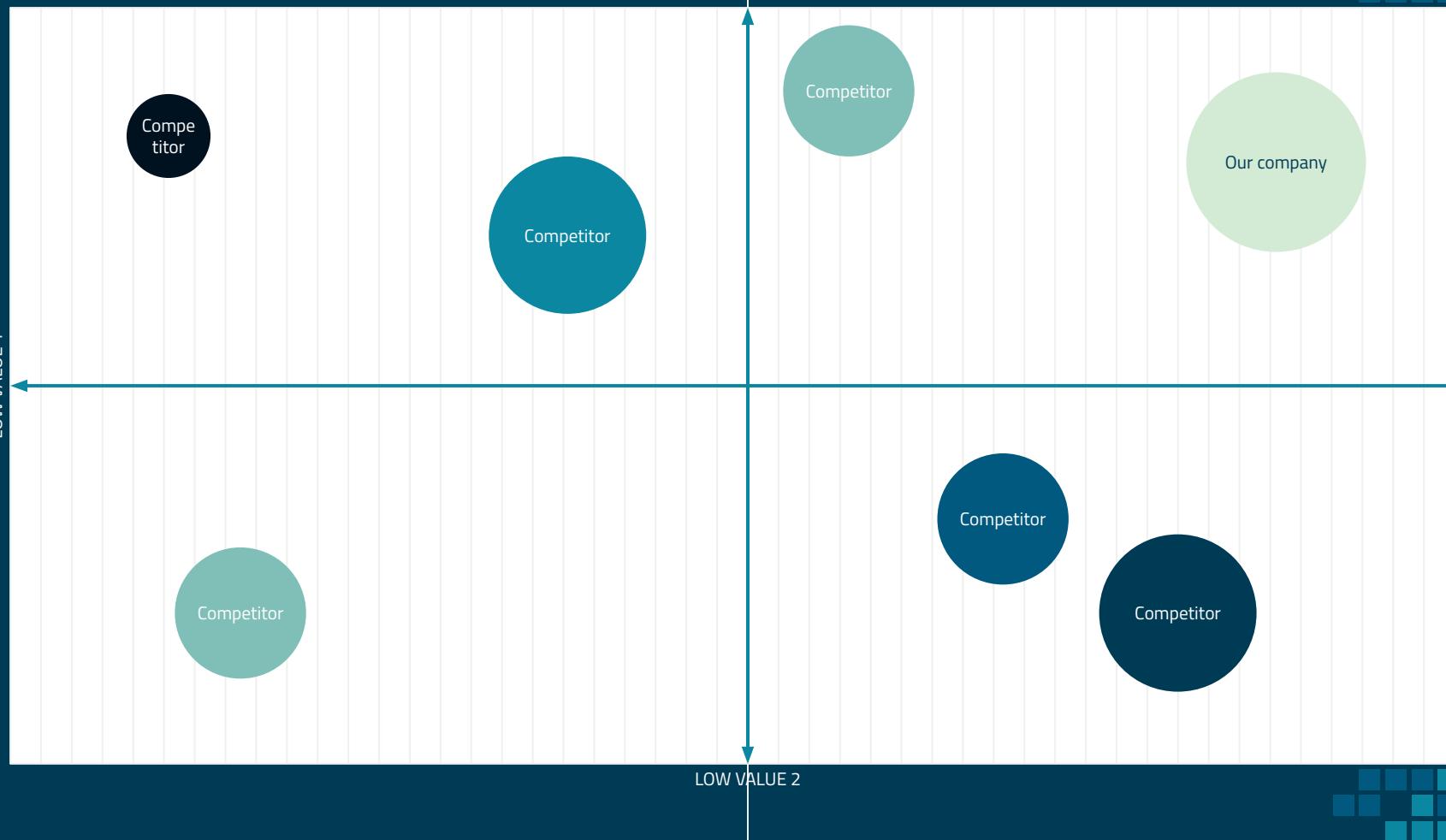
**Nils Årud**

JOB TITLE

Blue is the colour of the clear  
sky and the deep sea



## COMPETITOR MATRIX



# WEEKLY PLANNER

	SUNDAY	MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY
9:00 - 9:45	Task						
10:00 - 10:45	Task						
11:00 - 11:45	Task						
12:00 - 13:15	✓ Free time						
13:30 - 14:15	Task						
14:30 - 15:15	Task						
15:30 - 16:15	Task						



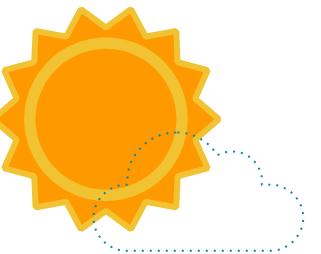
SlidesCarnival icons are editable shapes.

This means that you can:

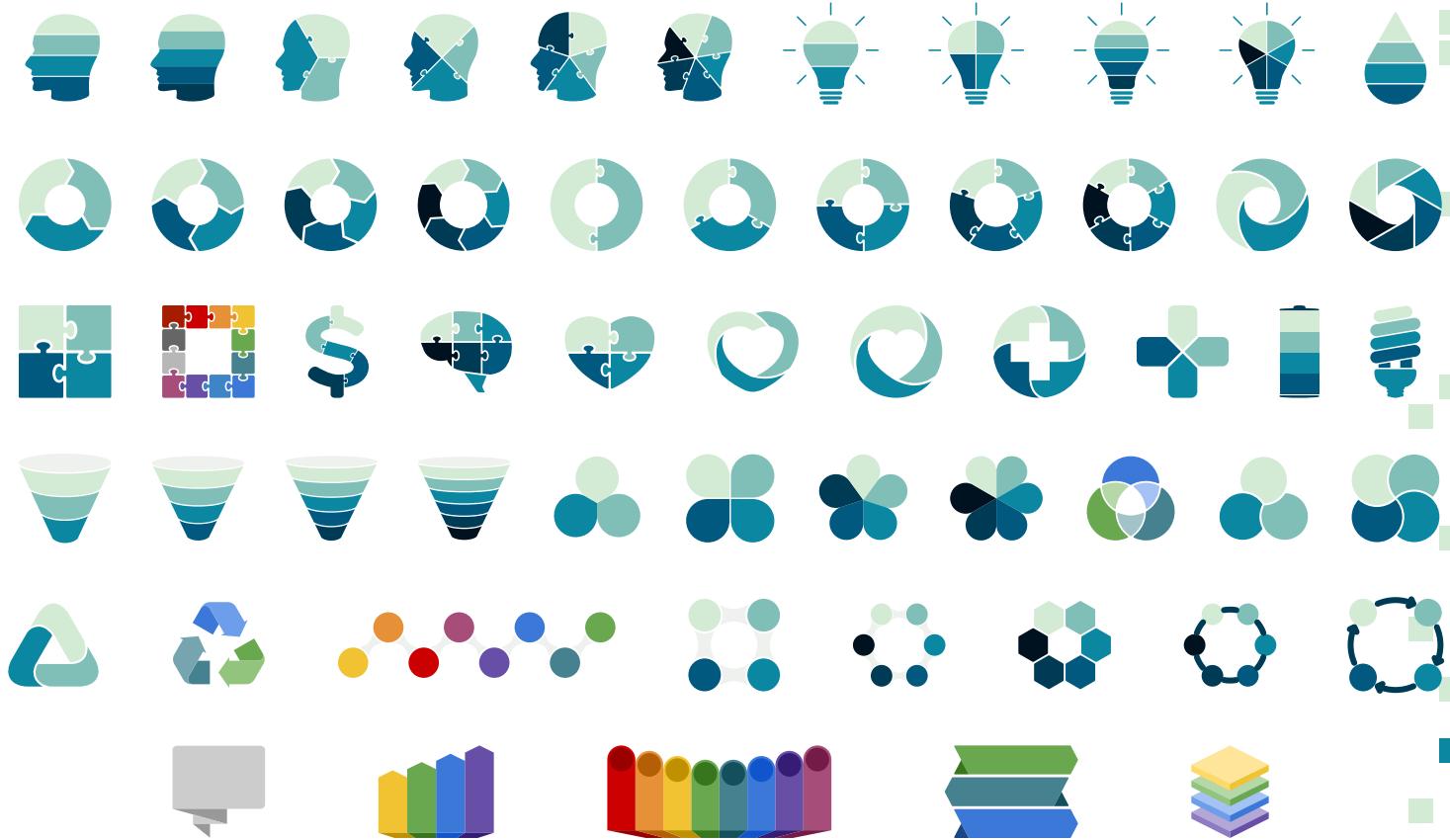
- Resize them without losing quality.
- Change fill color and opacity.
- Change line color, width and style.

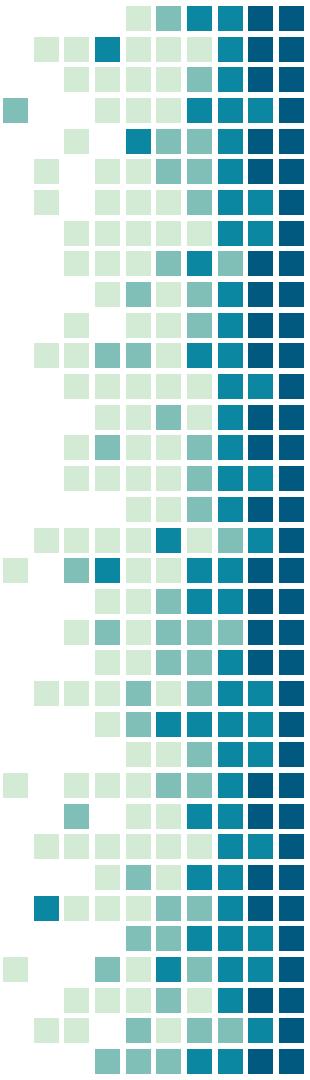
Isn't that nice? :)

Examples:



# DIAGRAMS AND INFOGRAPHICS





You can also use any emoji as an icon!

And of course it resizes without losing quality.

How? Follow Google instructions

<https://twitter.com/googledocs/status/730087240156643328>



and many more...



## Free templates for all your presentation needs



For PowerPoint and  
Google Slides



100% free for personal  
or commercial use



Ready to use,  
professional and  
customizable



Blow your audience  
away with attractive  
visuals

# LONG SHORT-TERM MEMORY NETWORK

- Jack Christopher Huaihua Huayhua
- Angel Tomas Concha Layme
- Jean Pierre Chavez Guevara

# CONTENIDO

01

Artículo

02

Estructura

03

Datos de  
entrenamiento

04

Áreas de  
aplicación

05

Ventajas y  
desventajas

05

Conclusiones





# O1

## Artículo

# LONG SHORT TERM MEMORY

Año: 1997

**Medio de publicación:** Journal Neural Computation.

**Autores:** Sepp Hochreiter y Jürgen Schmidhuber.

## LONG SHORT-TERM MEMORY

NEURAL COMPUTATION 9(8):1735–1780, 1997

Sepp Hochreiter

Fakultät für Informatik  
Technische Universität München  
80290 München, Germany  
[hochreit@informatik.tu-muenchen.de](mailto:hochreit@informatik.tu-muenchen.de)

<http://www7.informatik.tu-muenchen.de/~hochreit>

Jürgen Schmidhuber

IDSIA  
Corso Elvezia 36  
6900 Lugano, Switzerland  
[juergen@idsia.ch](mailto:juergen@idsia.ch)

<http://www.idsia.ch/~juergen>

## MOTIVO

Superar problemas de desvanecimiento y explosión de errores en RNN tradicionales.

## IMPACTO

Aplicaciones en reconocimiento de voz, procesamiento del lenguaje natural y predicciones secuenciales.

## PUBLICACIÓN

Revista científica especializada en inteligencia artificial, aprendizaje automático o procesamiento de lenguaje natural.

## RESULTADOS

Capacidad para resolver problemas complejos con largos retrasos temporales.





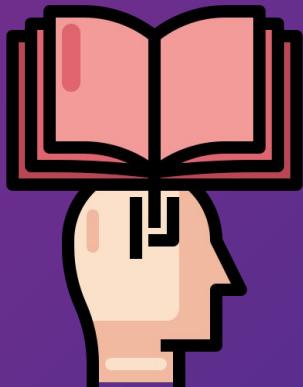
O2

# Estructura

# Idea Principal

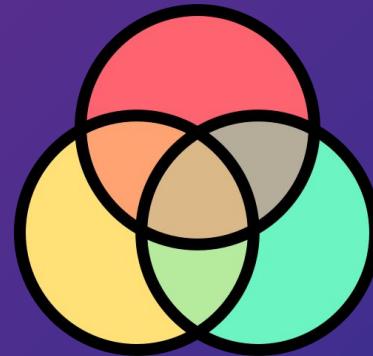
## Memorización

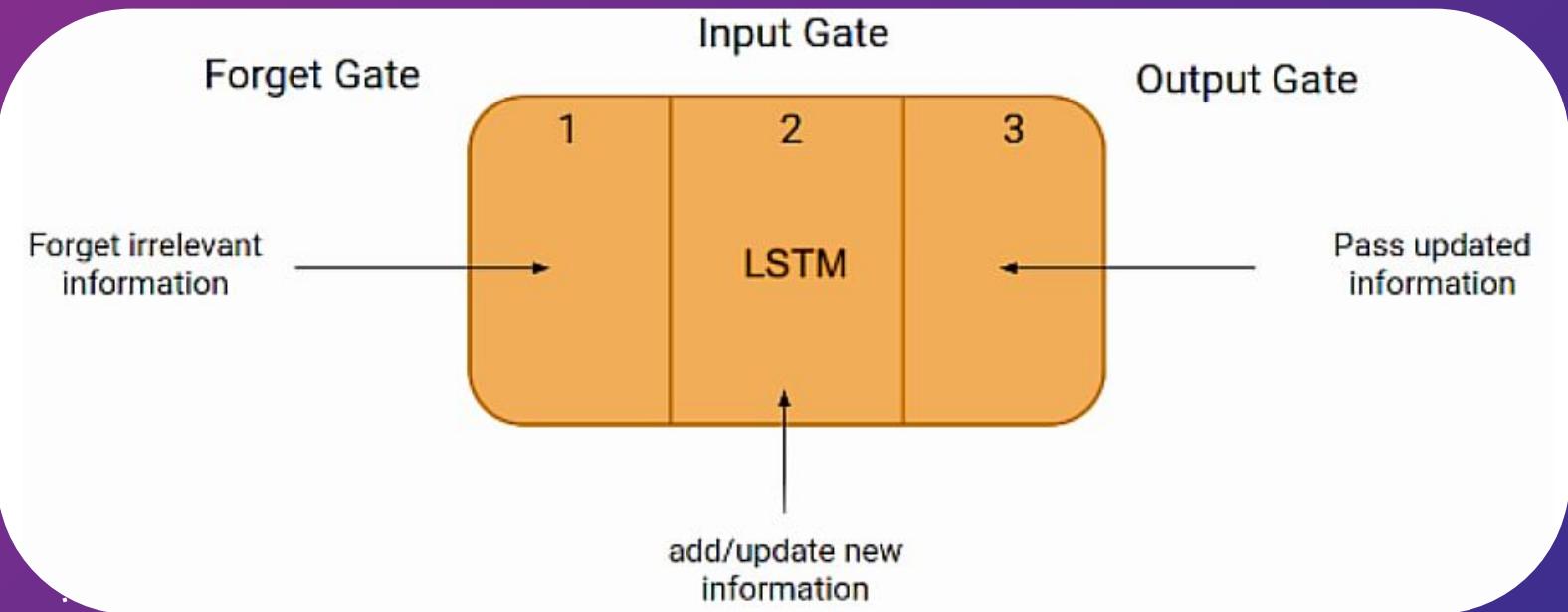
Se gana información → se almacena para uso futuro

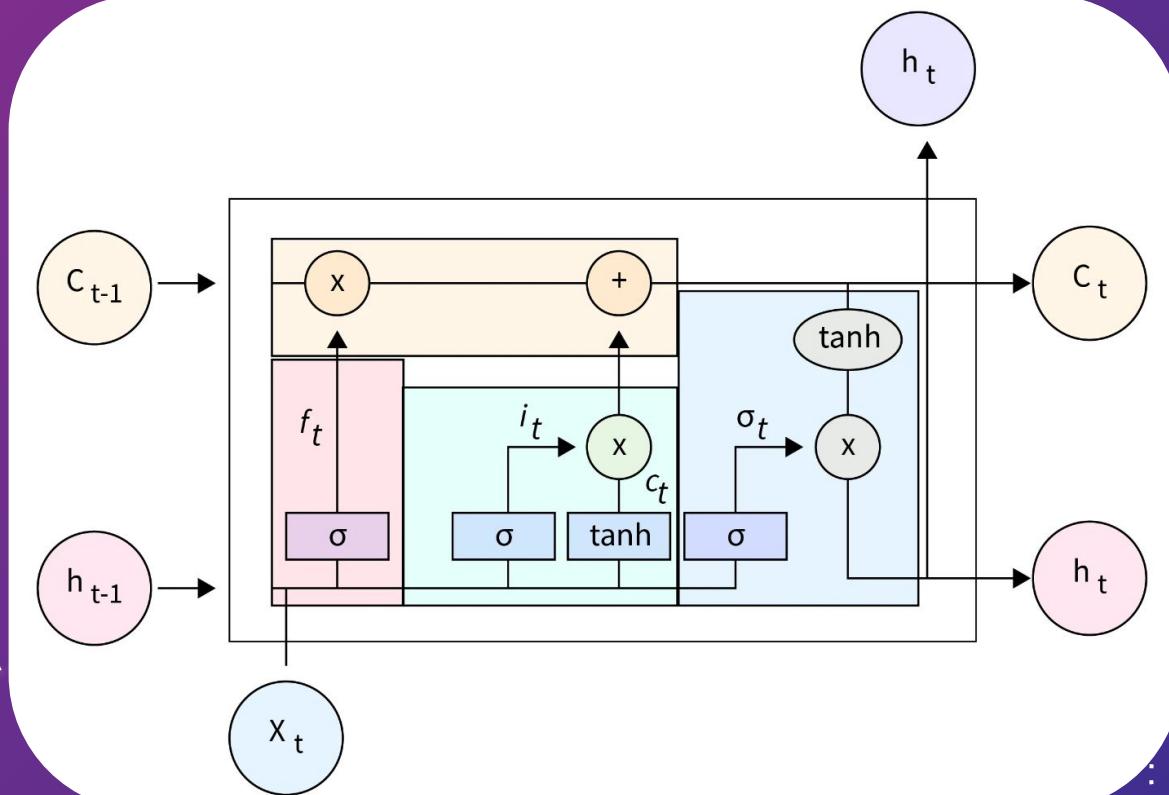


## Combinación

Información almacenada + habilidades analíticas





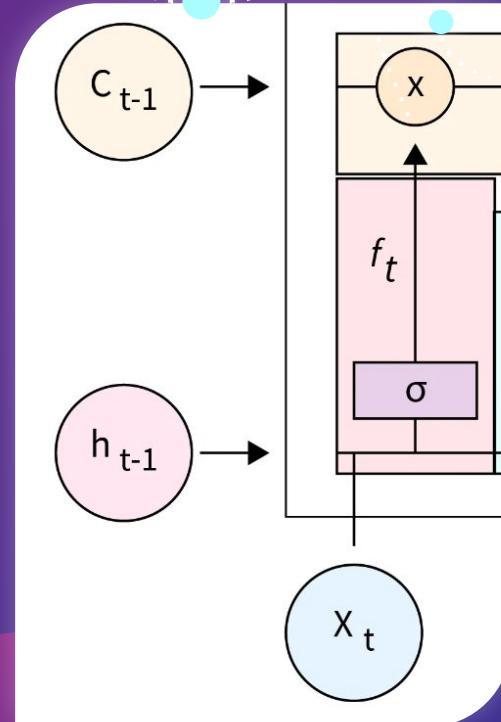


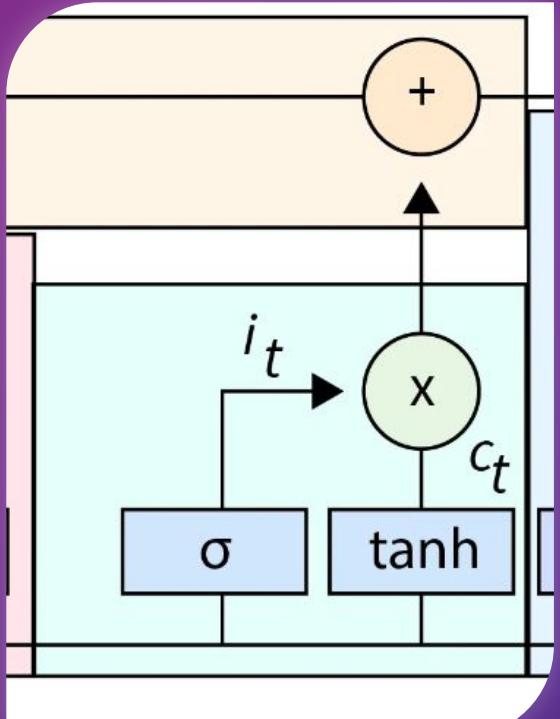
# FORGET GATE

Decide qué información debe eliminarse del **estado de la celda**.

Toma el **estado oculto** anterior y la entrada actual y lo pasa a una función de activación Sigmóide.

Genera un valor entre 0 y 1, donde 0 significa olvidar y 1 significa mantener.





## INPUT GATE

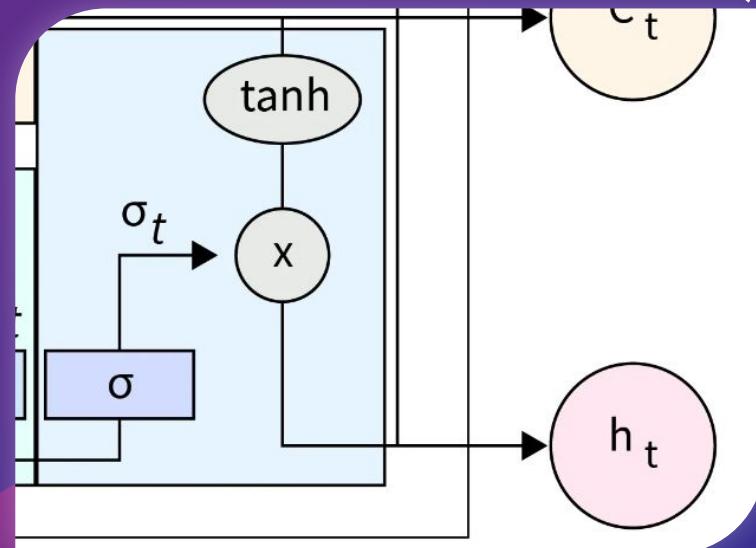
Considera la **entrada actual** y el **estado oculto** anterior para actualizar el valor del **estado de la celda**.

- Función de activación de Sigmoide → Decide qué porcentaje de la información se requiere.
- Función de activación de Tanh → Mapea los datos entre -1 y 1, luego la multiplica por la función Sigmoide.

# OUTPUT GATE

Devuelve el **estado oculto** para la próxima ocasión.

- Función de activación Sigmoide → Decide el porcentaje de información relevante requerida.
- Función de activación Tanh → Toma el **estado de la celda** actualizada y la multiplica por la función de activación Sigmoide.

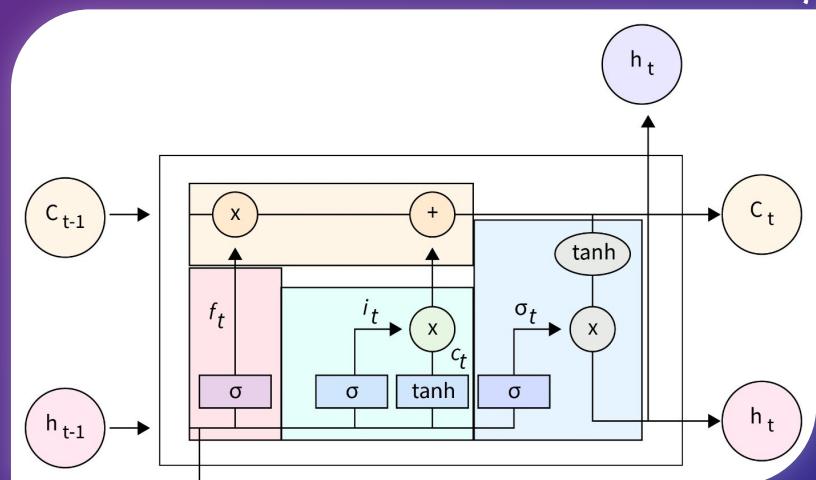


# CELL STATE

La **puerta de olvido** y la **puerta de entrada** actualizan la **celda de estado**.

La **celda de estado** anterior se multiplica por la salida de la **puerta de olvido**, se suma con la salida de la **puerta de entrada**.

Este valor se usa luego para calcular el estado oculto en la puerta de salida.





# 03

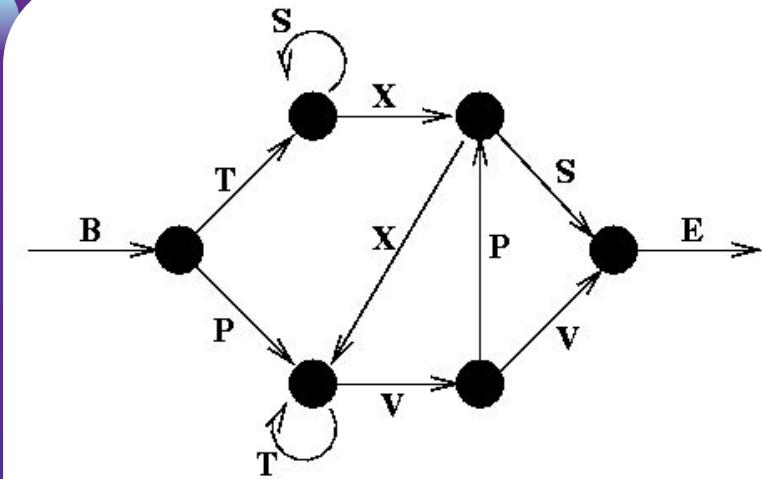
## Datos de entrenamiento

# Embedded Reber

Es una gramatica

Genera una secuencia de caracteres

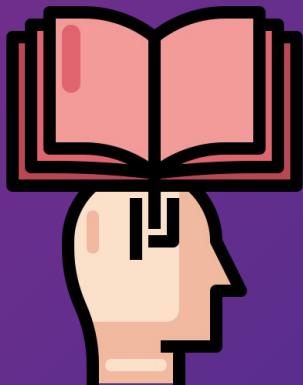
Sigue una logica determinista



# LSTM - REBER

## APRENDER

Las reglas que usaba la gramática



## PREDICIR

Los caracteres que siguen

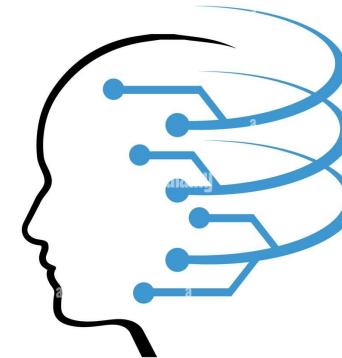


## Otros datos

Secuencias de datos como texto, audio, video.

Corpus de noticias, libros, artículos científicos, sitios web

Datos de entrenamiento de Stanford Question Answering Dataset (SQuAD)



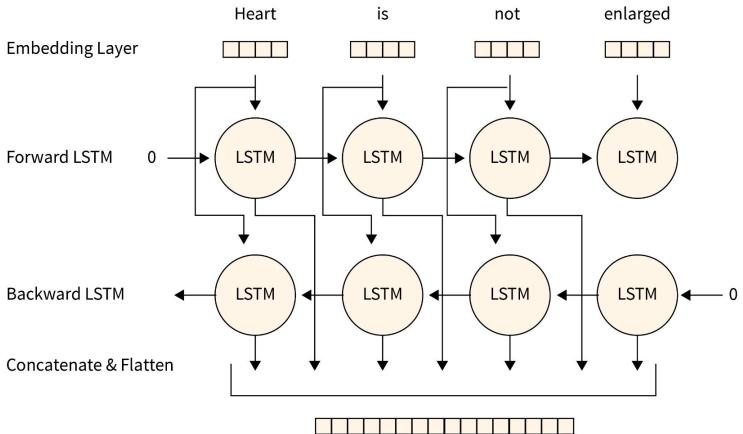
# | 04

## Áreas de aplicación

# Modelado de lenguaje

LSTM es capaz de: crear modelos de lenguaje

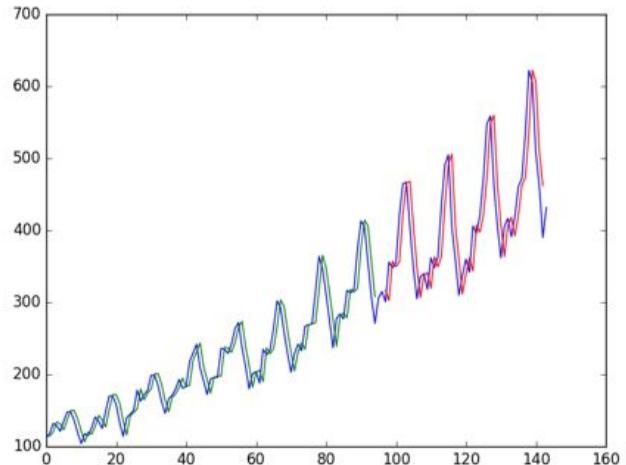
Son usados en: sistemas de traducción automática o chatbots



# Predicción de series temporales

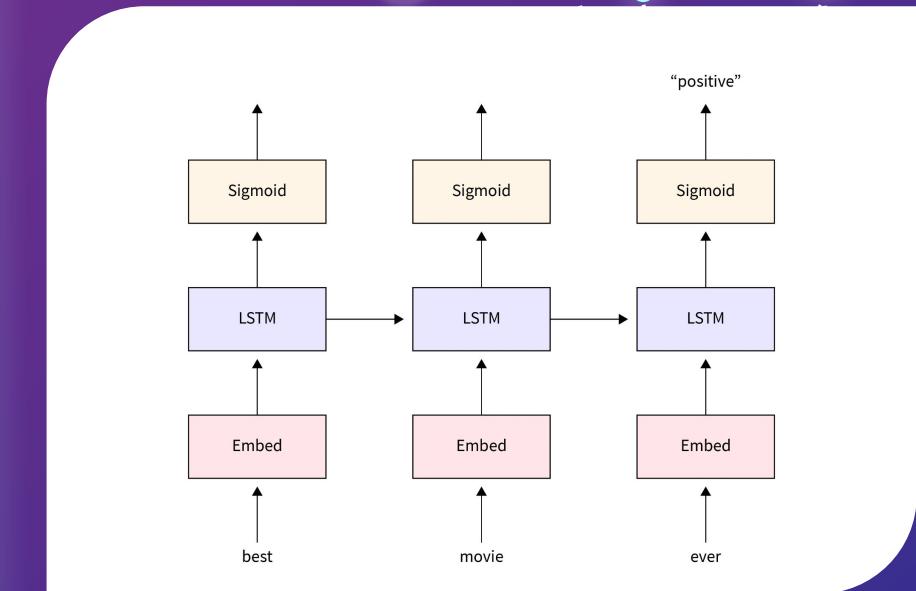
LSTM es capaz de: modelar datos de series temporales y predecir valores futuros en la serie

Son usados para: predecir precios de acciones o patrones de tráfico.



# Análisis de sentimientos

LSTM es utilizado para: analizar y clasificar las emociones expresadas en un texto.



# Reconocimiento de voz

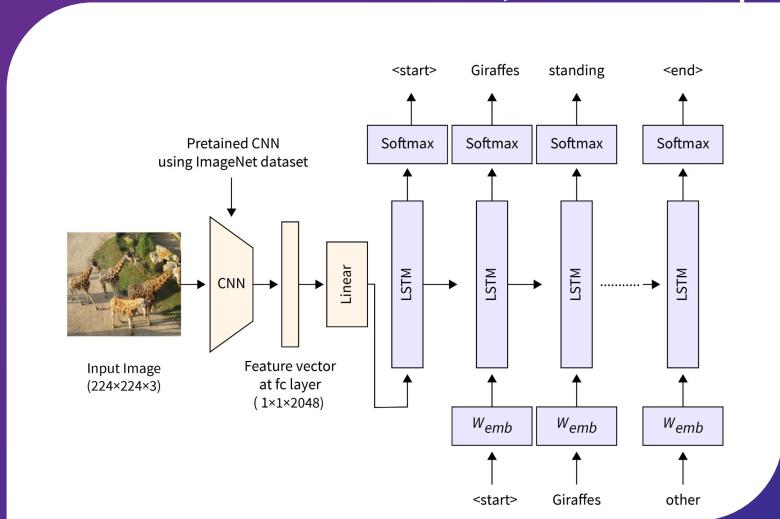
LSTM es capaz de: construir sistemas de reconocimiento de voz

Son usados los siguientes casos: asistentes virtuales , automatización de tareas, accesibilidad y seguridad



# Subtítulos de imágenes

LSTM es utilizado para: generar subtítulos descriptivos para imágenes, como en motores de búsqueda de imágenes



# 05

# VENTAJAS Y DESVENTAJAS

# VENTAJAS

Captura de dependencias a largo plazo.

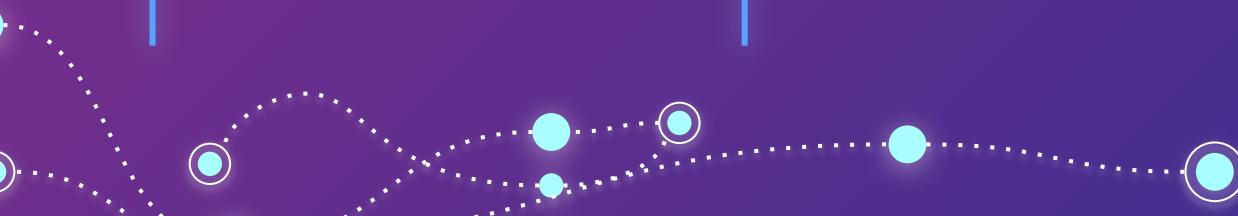
Manejo de secuencias de longitud variable.

Aprendizaje de representaciones jerárquicas.

Memoria de largo plazo.

Resistencia al ruido en los datos.

Adaptabilidad a diferentes dominios.



# DESVENTAJAS

Mayor complejidad computacional

Mayor dificultad de interpretación.

Requiere tiempo y experiencia para su configuración.

Mayor consumo de recursos.

Riesgo de sobreajuste.

Sensibilidad a la selección de hiperparámetros

# 106

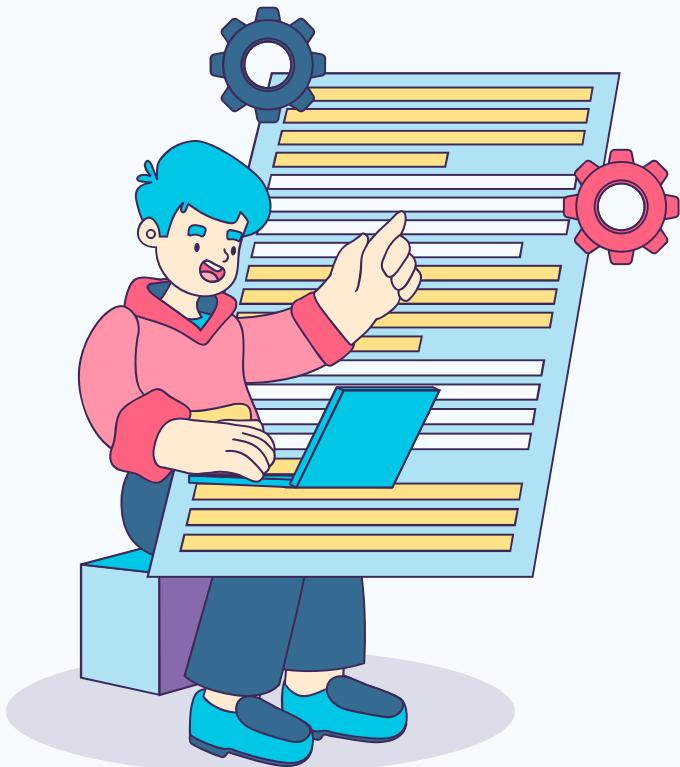
# CONCLUSIONES

(LSTM) ha sido una arquitectura revolucionaria

Sus logros respaldan su aplicabilidad en diversas áreas

Implementación y tiempo computacional son retos a considerar.

...  
...



# ResNet-50

Deep Residual Learning  
for Image Recognition



# Table of contents

01

Introducción

02

Arquitectura

03

Dataset

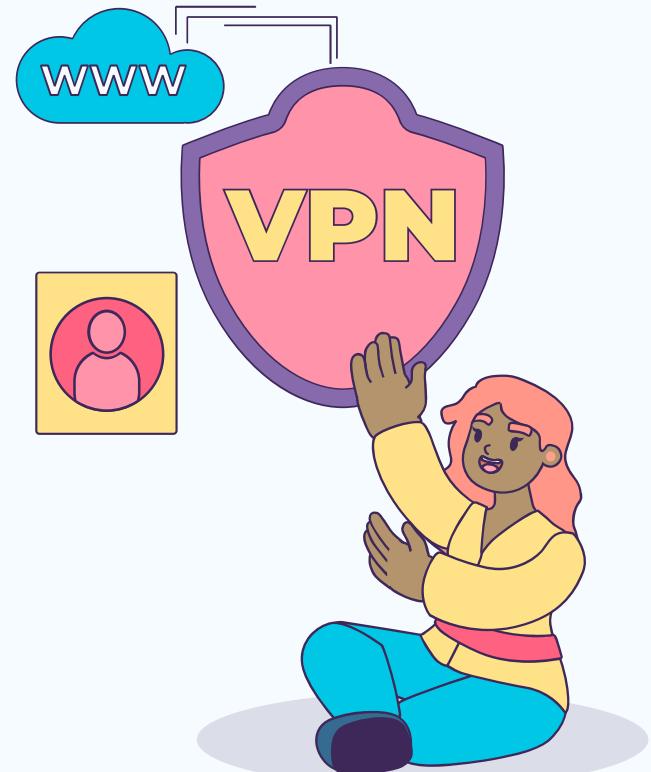
04

Áreas de aplicación

000

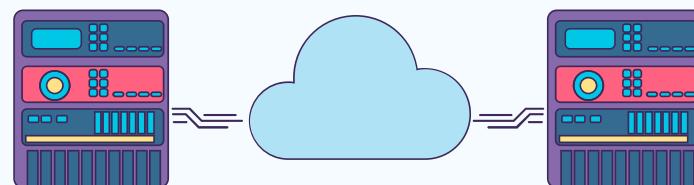
# 01

# Introducción



# Introducción

- Resnet50 se refiere a Residual Network que utiliza 50 capas convolucionales.
- Introducido en 2015 por He Kaiming, Zhang Xiangyu, Ren Shaoqing, and Sun Jian en el paper “Deep Residual Learning for Image Recognition”.
- Demostrar que estas redes residuales son más fáciles de optimizar y pueden ganar precisión con una profundidad considerablemente mayor.



## Deep Residual Learning for Image Recognition

Kaiming He    Xiangyu Zhang    Shaoqing Ren    Jian Sun

Microsoft Research

{kahe, v-xiangz, v-shren, jiansun}@microsoft.com

### Abstract

*Deeper neural networks are more difficult to train. We present a residual learning framework to ease the training of networks that are substantially deeper than those used previously. We explicitly reformulate the layers as learning residual functions with reference to the layer inputs, instead of learning unreferenced functions. We provide comprehensive empirical evidence showing that these residual networks are easier to optimize, and can gain accuracy from*

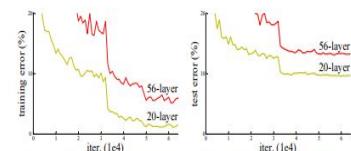


Figure 1. Training error (left) and test error (right) on CIFAR-10 with 20-layer and 56-layer “plain” networks. The deeper network has higher training error, and thus test error. Similar phenomena on ImageNet is presented in Fig. 4.

# ¿Aprender mejores redes es tan fácil como apilar más capas?

- Agregar más capas introduce algunos problemas durante el entrenamiento como:



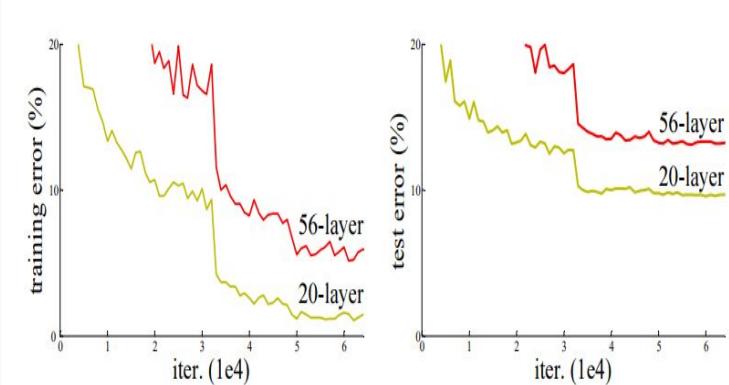
## Problema de los gradientes desvanecidos/explorosivos

- Dificulta la convergencia desde el principio.
- Solución:
  - Inicialización normalizada
  - Capas de normalización intermedias



## Problema de degradación

- Las redes más profundas son capaces de empezar a converger.
- Añadir más capas al modelo conduce un mayor error de entrenamiento.
- No es causada por el overfitting.

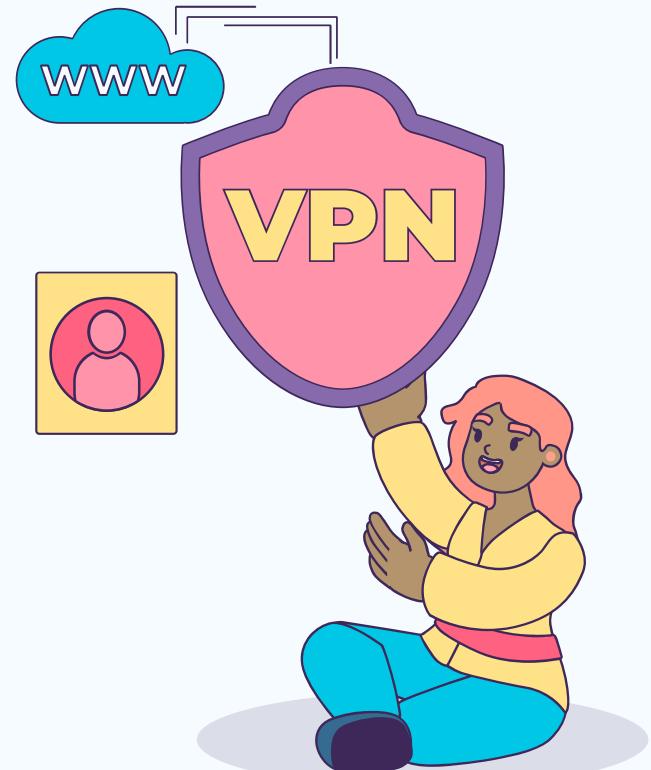


Dataset: CIFAR - 10

000

# 02

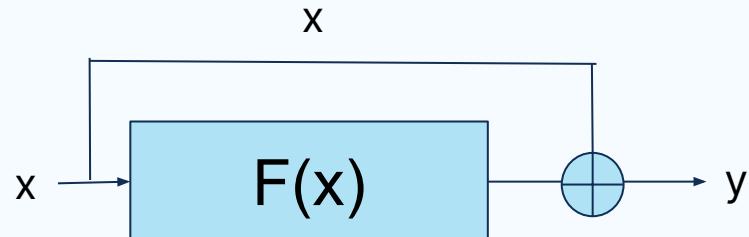
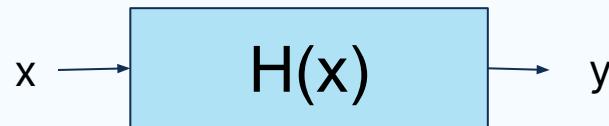
# Arquitectura



# Residual Learning

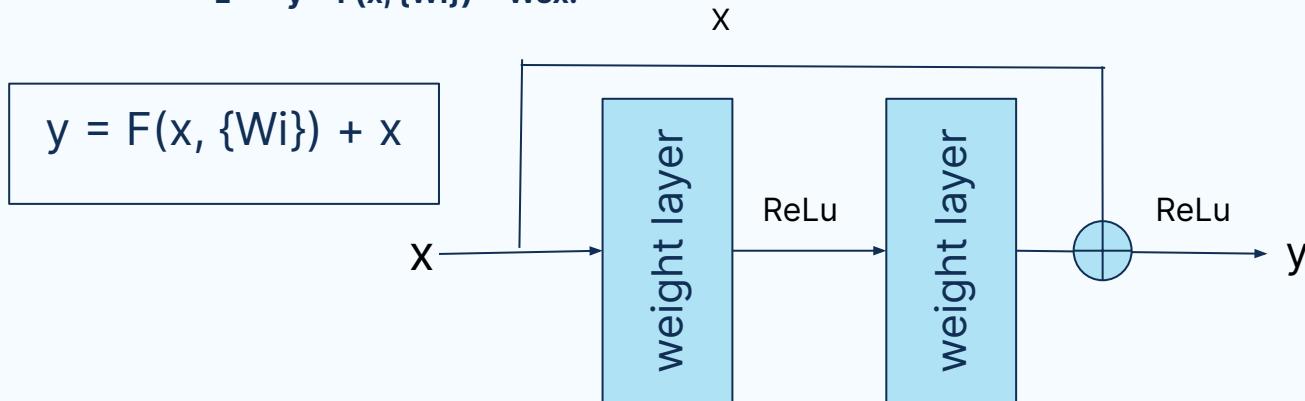
- $H(x)$ : mapeo subyacente que debe ser ajustada por unas cuantas capas apiladas
- $x$ : Entradas a la primera de estas capas
- En lugar de esperar que las capas apiladas aproximen  $H(x)$ , dejamos explícitamente que estas capas aproximen una función residual:

$$F(x) := H(x) - x$$



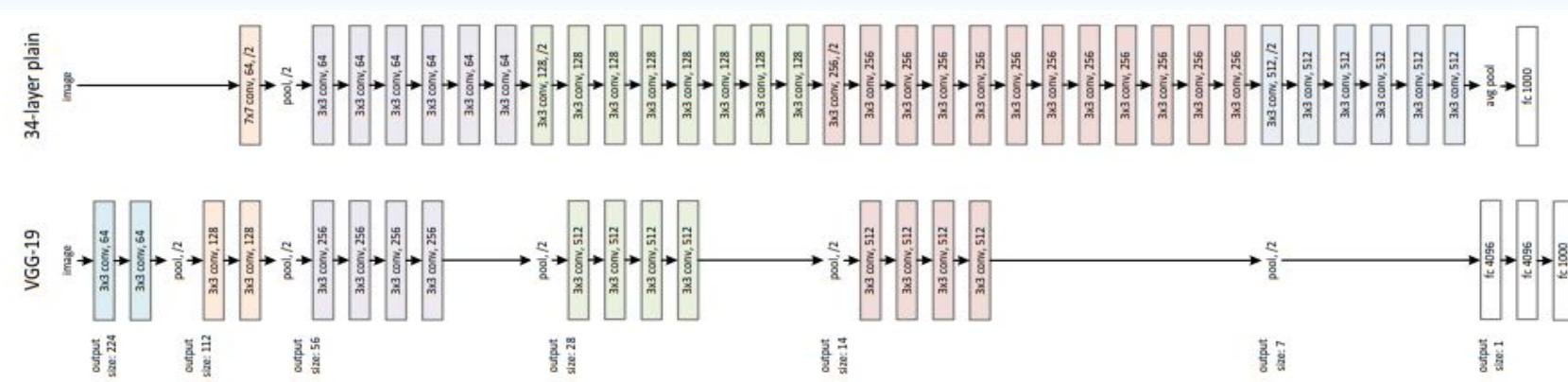
# Residual Block

- **$F(x, \{W_i\})$ :** Representa el mapeo residual que debe aprenderse.
- La operación  $F + x$  se realiza mediante una conexión abreviada y una suma de elementos.
- Dimensiones diferentes (entrada/salida):
  - Realizar una proyección lineal  $W_s$  por las conexiones de acceso directo para igualar las dimensiones:
    - $y = F(x, \{W_i\}) + W_s x.$



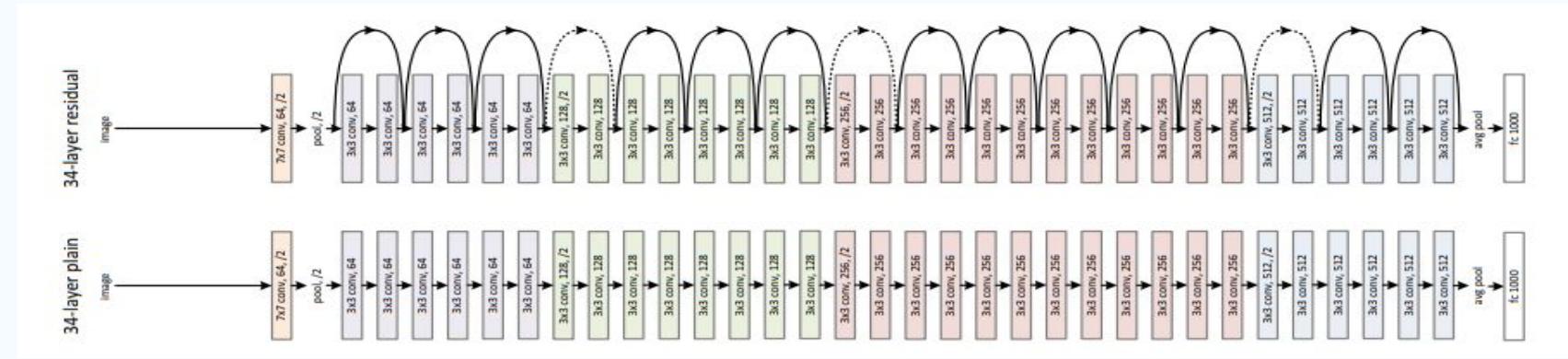
# Arquitectura

- **PLAIN NETWORK:**
- Las capas convolucionales tienen en su mayoría filtros de  $3 \times 3$ 
  - Para el mismo tamaño del mapa de características de salida, las capas tienen el mismo número de filtros.
  - Si el tamaño del mapa de características se reduce a la mitad, el número de filtros se duplica para preservar la complejidad temporal por capa.
- El modelo básico de 34 capas tiene 3.600 millones de FLOPs, en cambio, VGG-19 tiene 19.600 millones de FLOPs.



# Arquitectura

- **RESIDUAL NETWORK:**
- Inserta conexiones de atajo que convierten la red en su versión residual equivalente.
- Cuando hay diferencia en las dimensiones, existen dos opciones:
  - El atajo sigue realizando el mapeo de identidad, rellena con entradas cero adicionales para aumentar la dimensionalidad.
  - Se utiliza la ecuación del atajo de proyección, igualando las dimensiones (mediante convoluciones  $1 \times 1$ ).

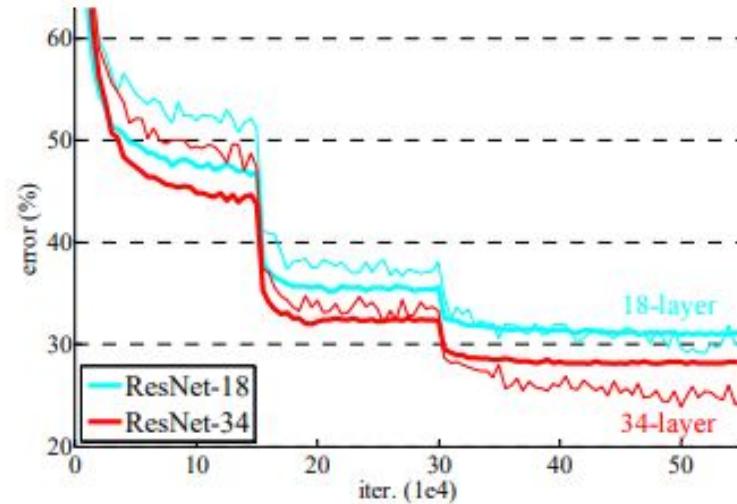
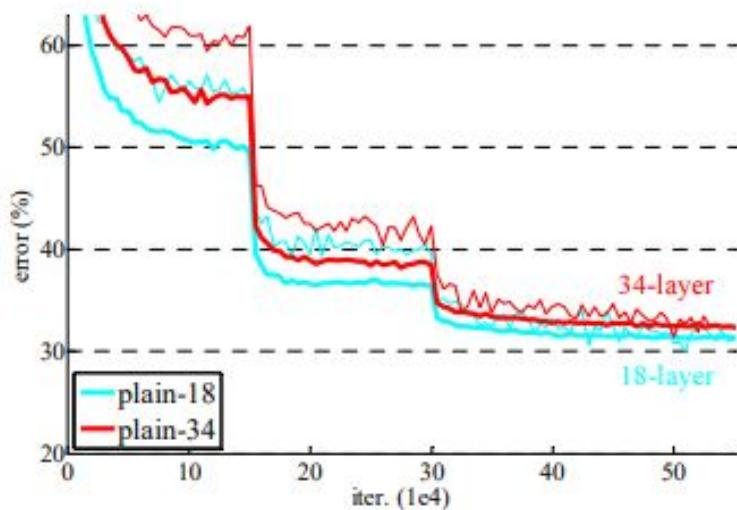


# Implementación

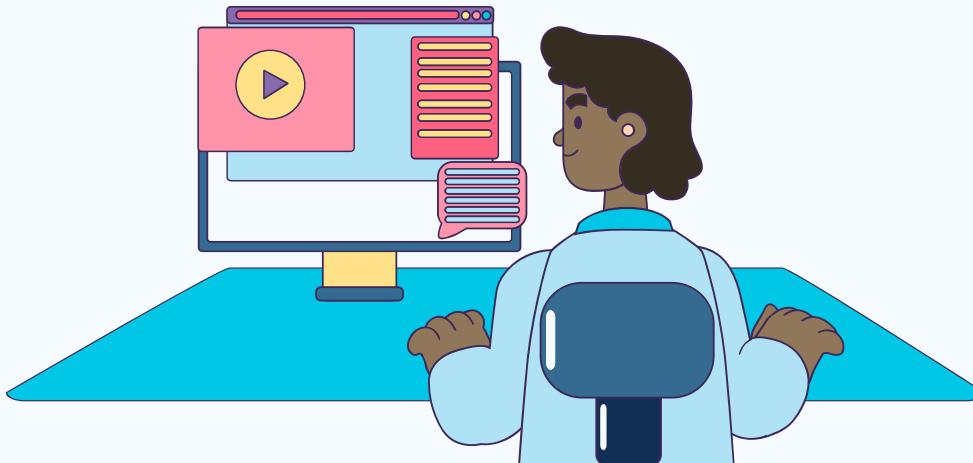
La arquitectura ResNet incluye los siguientes elementos:

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112			7×7, 64, stride 2		
conv2_x	56×56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		$1.8 \times 10^9$	$3.6 \times 10^9$	$3.8 \times 10^9$	$7.6 \times 10^9$	$11.3 \times 10^9$

# Resultados



# 03 DATASET





14,197,122 images, 21841 synsets indexed

[Home](#) [Download](#) [Challenges](#) [About](#)Not logged in. [Login](#) | [Signup](#)

## ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

### Competition

The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) evaluates algorithms for object detection and image classification at large scale. One high level motivation is to allow researchers to compare progress in detection across a wider variety of objects -- taking advantage of the quite expensive labeling effort. Another motivation is to measure the progress of computer vision for large scale image indexing for retrieval and annotation.

For details about each challenge please refer to the corresponding page.

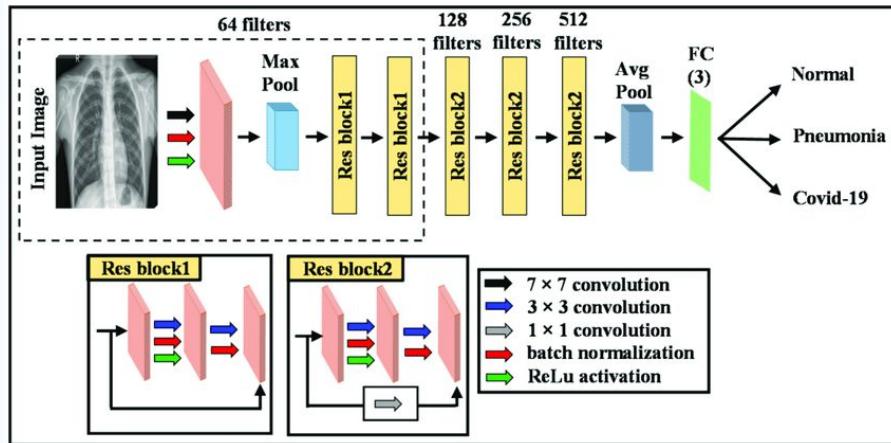
- [ILSVRC 2017](#)
- [ILSVRC 2016](#)
- [ILSVRC 2015](#)
- [ILSVRC 2014](#)
- [ILSVRC 2013](#)
- [ILSVRC 2012](#)
- [ILSVRC 2011](#)
- [ILSVRC 2010](#)

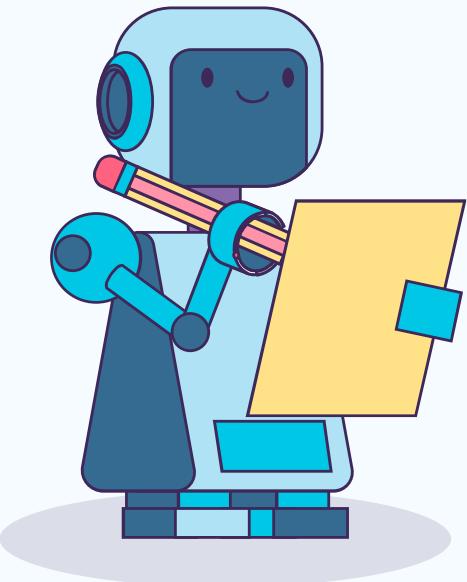
# 04

## Áreas de aplicación



- Clasificación de imágenes
- Detección de objetos
- Segmentación semántica
- Reconocimiento facial
- Procesamiento de imágenes médicas
- Análisis de imágenes en tiempo real



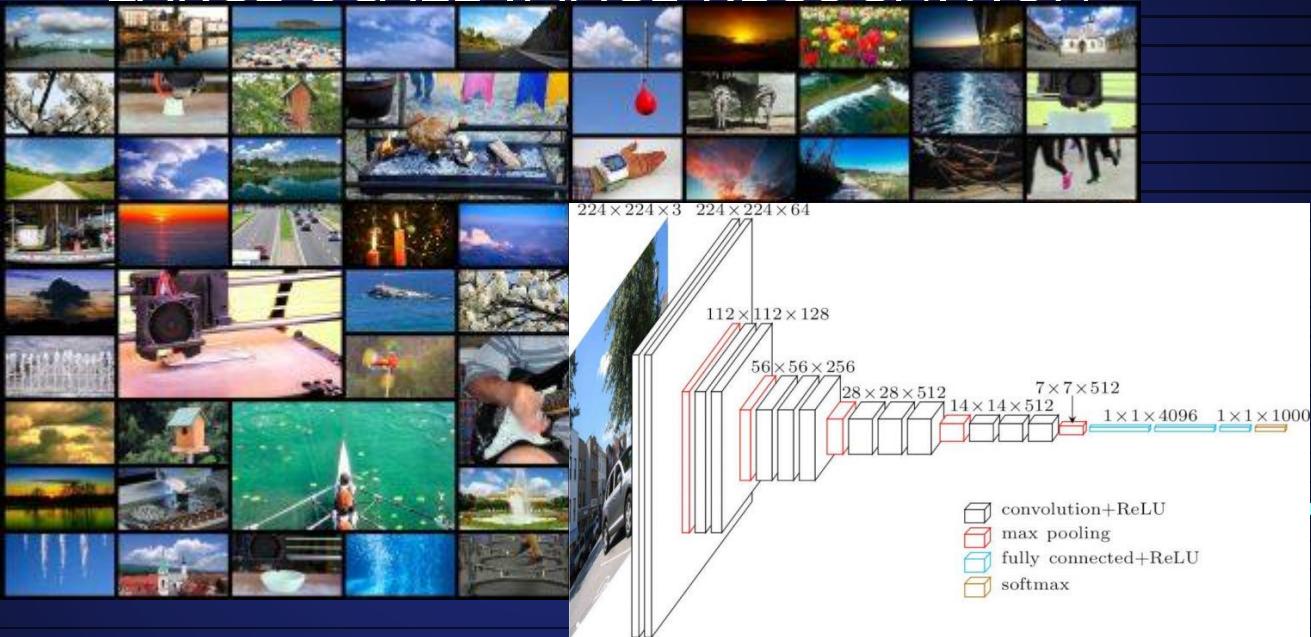


# Thanks!

# VGG - 16

Chipana Perez, Gabriela Angel  
Vizcarra Vargas, Piero Emiliano  
Hilares Angelo, Maryori Lizeth

# VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION



Karen Simonyan y Andrew Zisserman

# Artículo de presentación

Very deep convolutional networks for large-scale image recognition

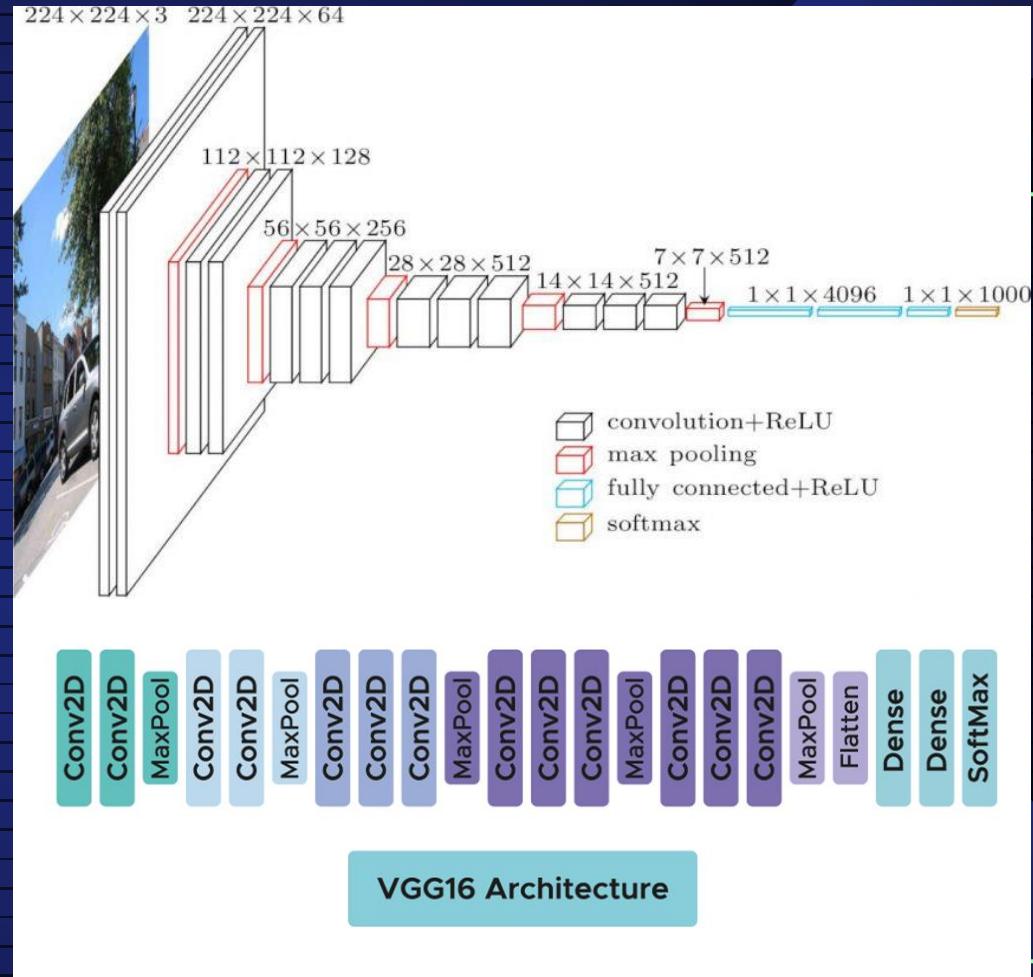
Autores:  
Karen Simonyan  
Andrew Zisserman

- CNN alto rendimiento en el reconocimiento de imágenes a gran escala.
- Menciona la dificultad de entrenar redes más profundas y la necesidad de explorar arquitecturas que puedan manejar esta complejidad
- Objetivo: Presentar una arquitectura de CNN VGG-16 (y también VGG-19), que tiene una profundidad significativamente mayor que otras arquitecturas.  
El objetivo es demostrar que una mayor profundidad de la red puede mejorar el rendimiento en la clasificación de imágenes en el conjunto de datos ImageNet.



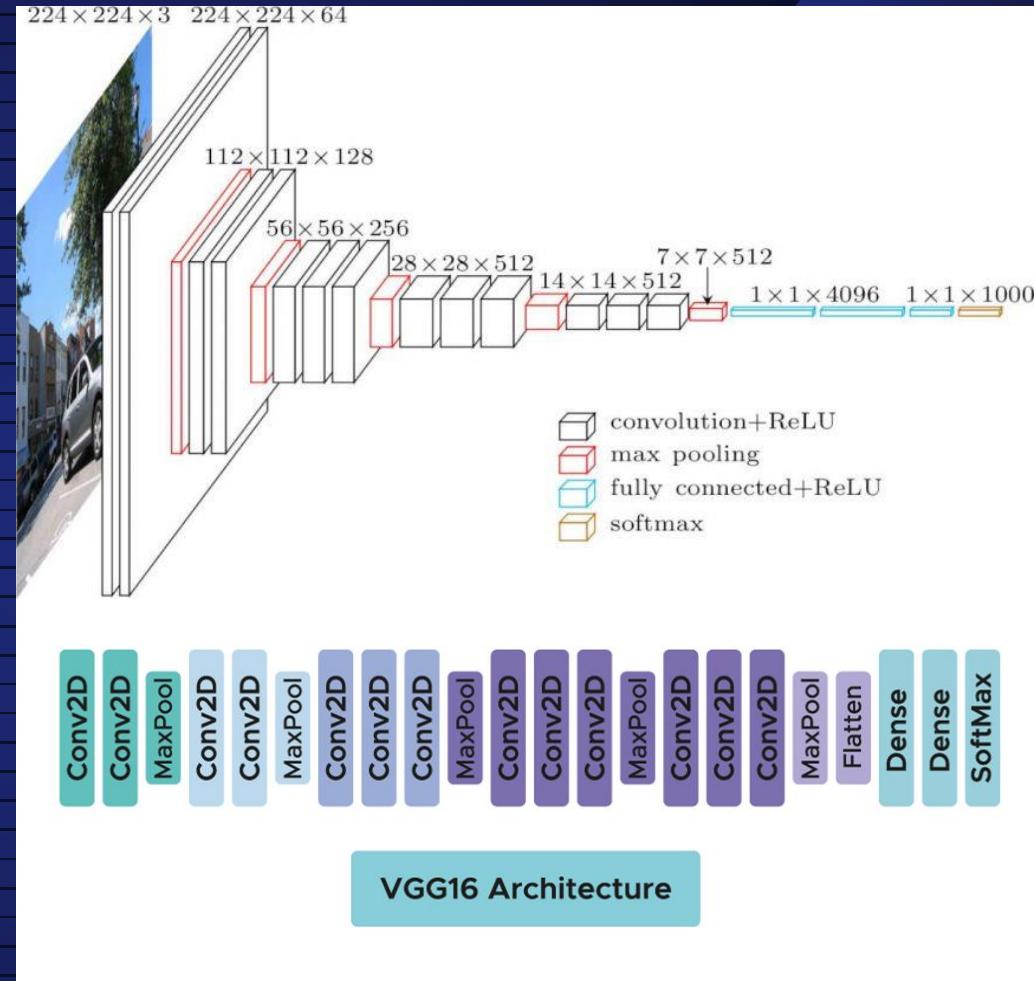
# Arquitectura/estructura

- **Capas de entrada:** imágenes de tamaño  $224 \times 224$
- **Capas convolucionales:** filtros de tamaño  $3 \times 3$  con un paso de 1 píxel .  
Aplican la función de activación ReLU después de cada convolución.  
Objetivo: extraer características relevantes de las imágenes en diferentes niveles de abstracción.
- **Capas de agrupación (pooling):** Se aplica cada 2 capas convolucionales, (Max Pooling) con un tamaño de ventana de  $2 \times 2$  y un paso de 2 píxeles.  
Objetivo: Reducir la dimensionalidad y ayuda a obtener una representación más compacta de las características extraídas.



# Arquitectura/estructura

- **Capas totalmente conectadas:** Cada capa con 4096 unidades.  
Objetivo: transformar las características extraídas en una salida clasificada. Después de cada capa totalmente conectada, se aplica la función de activación ReLU, excepto en la última capa.
- **Capa de salida:** La capa de salida consiste en una capa totalmente conectada con 1000 unidades, correspondientes a las 1000 clases diferentes en el conjunto de datos ImageNet. La función de activación utilizada en esta capa es la función Softmax, que asigna probabilidades a cada clase.

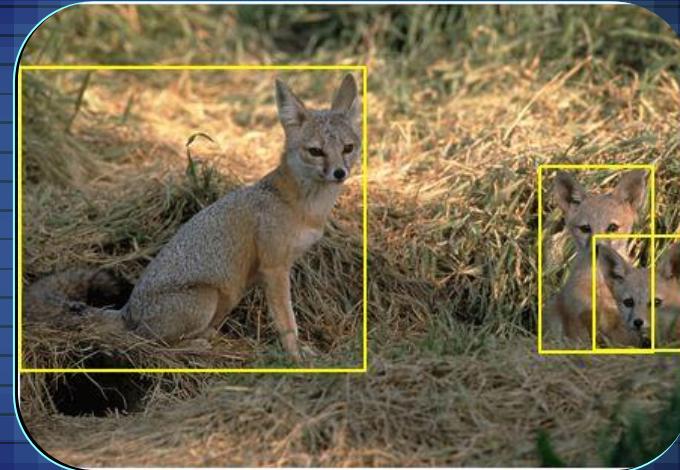


# Datos

## ILSVRC

Contiene los datos de imagen y la realidad del terreno para los conjuntos de entrenamiento y validación, y los datos de imagen para el conjunto de prueba.

Las anotaciones se ordenan por sus synsets (por ejemplo, "gato persa", "bicicleta de montaña" o "perro caliente") como su wnid. Estas identificaciones se ven como n00141669. El nombre de cada imagen tiene una correspondencia directa con el nombre del archivo de anotación.



ImageNet Large-Scale Visual  
Recognition Challenge

Table 7: Comparison with the state of the art in ILSVRC classification. Our method is denoted as “VGG”. Only the results obtained without outside training data are reported.

Method	top-1 val. error (%)	top-5 val. error (%)	top-5 test error (%)
VGG (2 nets, multi-crop & dense eval.)	<b>23.7</b>	<b>6.8</b>	<b>6.8</b>
VGG (1 net, multi-crop & dense eval.)	24.4	7.1	7.0
VGG (ILSVRC submission, 7 nets, dense eval.)	24.7	7.5	7.3
GoogLeNet (Szegedy et al., 2014) (1 net)	-		<b>7.9</b>
GoogLeNet (Szegedy et al., 2014) (7 nets)	-		<b>6.7</b>
MSRA (He et al., 2014) (11 nets)	-	-	8.1
MSRA (He et al., 2014) (1 net)	<b>27.9</b>	9.1	9.1
Clarifai (Russakovsky et al., 2014) (multiple nets)	-	-	11.7
Clarifai (Russakovsky et al., 2014) (1 net)	-	-	12.5
Zeiler & Fergus (Zeiler & Fergus, 2013) (6 nets)	36.0	14.7	14.8
Zeiler & Fergus (Zeiler & Fergus, 2013) (1 net)	37.5	16.0	16.1
OverFeat (Sermanet et al., 2014) (7 nets)	34.0	13.2	13.6
OverFeat (Sermanet et al., 2014) (1 net)	35.7	14.2	-
Krizhevsky et al. (Krizhevsky et al., 2012) (5 nets)	38.1	16.4	16.4
Krizhevsky et al. (Krizhevsky et al., 2012) (1 net)	40.7	18.2	-

# Áreas de aplicación

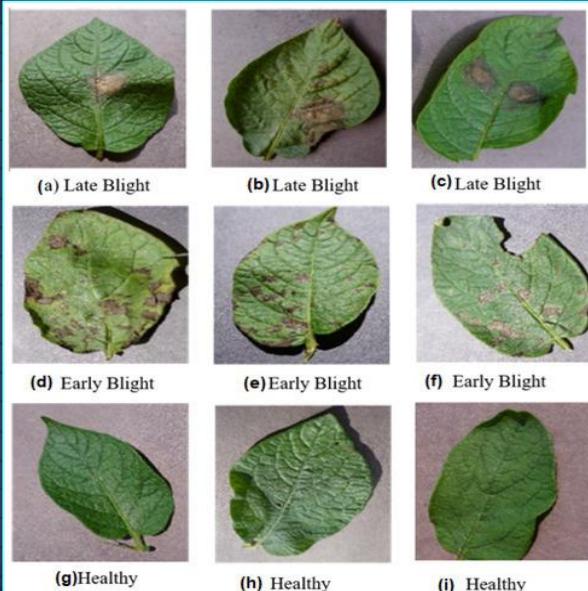
## Clasificación de imágenes

VGG-16 ha sido ampliamente utilizada para tareas de clasificación de imágenes en diferentes dominios, como reconocimiento de objetos, detección de enfermedades en imágenes médicas, clasificación de escenas, etc.



# Predicción de edad ósea con red basada en VGG-16 y transfer learning

# Áreas de aplicación



## Detección de objetos

Junto con técnicas de detección como R-CNN, Faster R-CNN y YOLO, la VGG-16 se ha utilizado para la detección de objetos en imágenes, lo que implica identificar la presencia y ubicación de objetos específicos en una imagen.

Application of convolutional neural  
networks for detection of the late blight  
Phytophthora infestans in potato *Solanum*...  
*tuberosum*

# Áreas de aplicación

## Recuperación de información visual

La VGG-16 se ha utilizado en aplicaciones de recuperación de información visual, donde el objetivo es buscar imágenes similares en grandes bases de datos de imágenes. La extracción de características de la VGG-16 permite representar las imágenes de manera compacta y compararlas eficientemente.

## Segmentación semántica

La VGG-16 se ha utilizado en aplicaciones de segmentación semántica, donde el objetivo es asignar una etiqueta semántica a cada píxel de una imagen. Esta tarea es útil en aplicaciones como la conducción autónoma, el etiquetado automático de imágenes y la realidad aumentada.



# Demo

# Conclusiones

1. VGG-16 es una arquitectura de red convolucional popular y ampliamente utilizada en el campo de la visión por computadora. Fue desarrollada por el grupo de investigación Visual Geometry Group en la Universidad de Oxford.
2. La arquitectura VGG-16 se caracteriza por su simplicidad y eficacia. Está compuesta por 16 capas, incluyendo 13 capas convolucionales y 3 capas completamente conectadas.
3. VGG-16 ha demostrado un rendimiento impresionante en tareas de clasificación de imágenes, logrando resultados cercanos al estado del arte en conjuntos de datos desafiantes como ImageNet.
4. La estructura de VGG-16, con capas convolucionales en cascada, permite aprender características visuales jerárquicas de una imagen, desde características de bajo nivel (como bordes y texturas) hasta características de alto nivel (como formas y objetos).



# Conclusiones

5. La capacidad de transferencia de aprendizaje de VGG-16 es una de sus características más destacadas. Al pre-entrenar la red en grandes conjuntos de datos, se pueden transferir los conocimientos aprendidos a tareas específicas con conjuntos de datos más pequeños, ahorrando tiempo y recursos de entrenamiento.
6. VGG-16 se ha utilizado en una variedad de aplicaciones, como clasificación de objetos, detección de objetos, segmentación semántica, recuperación de información visual, entre otras. Su capacidad para aprender características visuales complejas la hace valiosa en diversos problemas de visión por computadora.
7. Aunque VGG-16 ha sido una arquitectura influyente, también tiene algunas limitaciones. Es una red profunda y relativamente pesada, lo que puede dificultar su implementación en dispositivos con recursos limitados. Además, puede requerir conjuntos de datos grandes para un entrenamiento efectivo.



# Conclusiones

En general, la arquitectura VGG-16 ha dejado una huella significativa en el campo de la visión por computadora. Su simplicidad, eficacia y capacidad de transferencia de aprendizaje la han convertido en una opción popular para muchas aplicaciones de procesamiento de imágenes y ha sentado las bases para el desarrollo de arquitecturas más avanzadas.

