

# 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.



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

La arquitectura de **Inception.v1** utiliza módulos 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

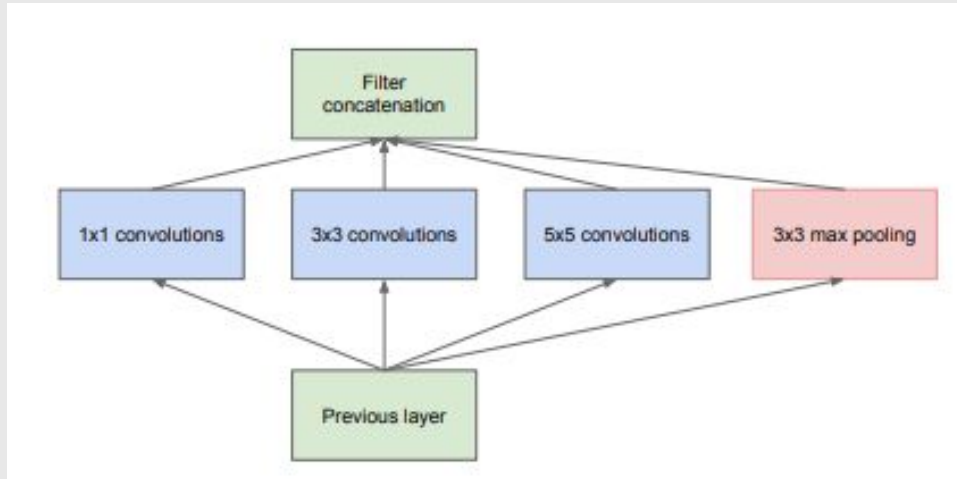
- Uso de capas de reducción de dimensionalidad en Inception-v1.
- Las convoluciones 1x1 se utilizan para disminuir la complejidad computacional y reducir la cantidad de parámetros en la red.
- La reducción de dimensionalidad permite que la red sea más profunda y eficiente sin sacrificar el rendimiento.





# ARQUITECTURA

## Estructura principal



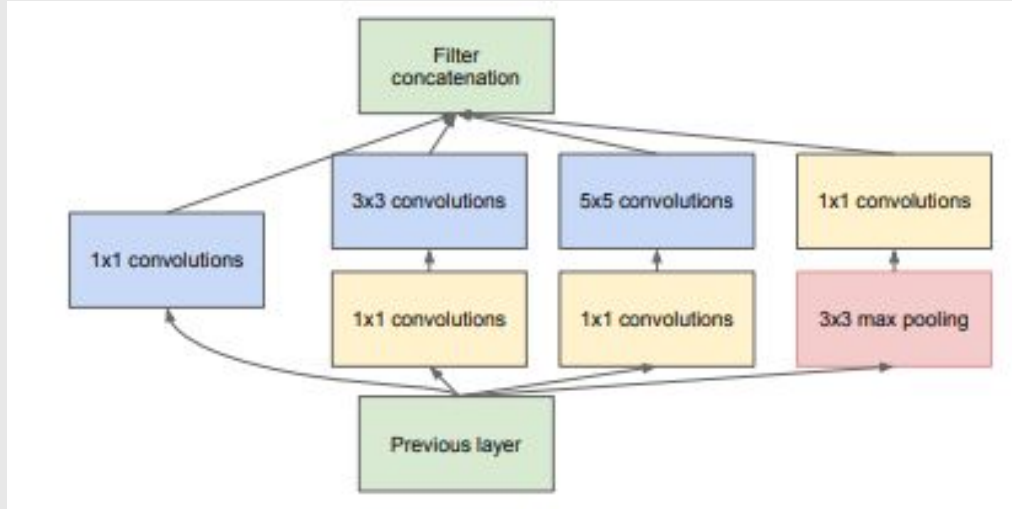
Módulo Inception naive-version

- 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

## 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.

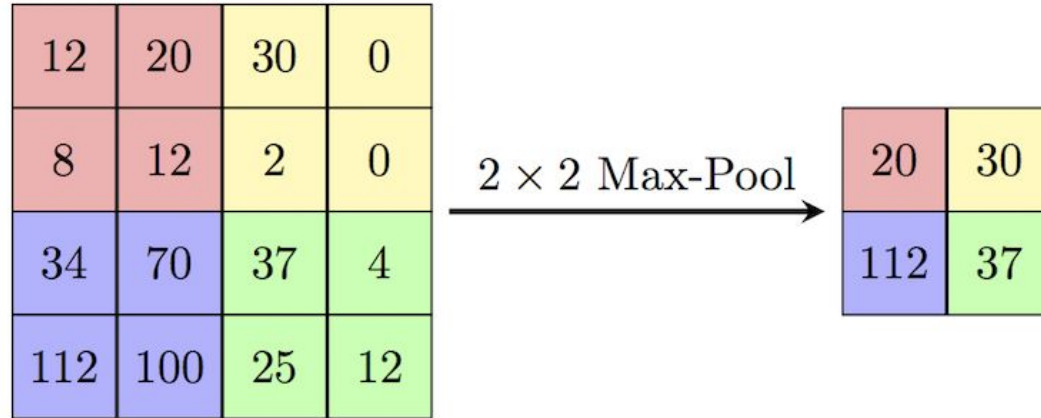


Módulo Inception con reducción de dimensionalidad



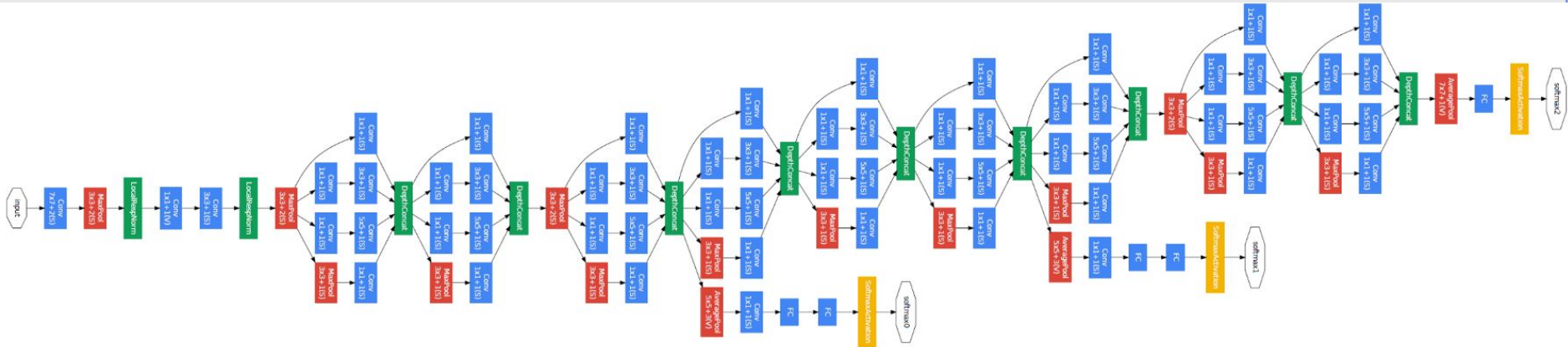
# ARQUITECTURA

## Capa de Pooling: Reducción de Espacio

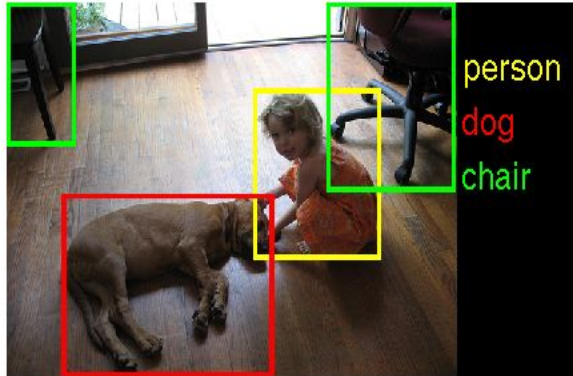


- La capa de pooling utilizada en Inception-v1 para reducir el tamaño espacial de los mapas de características.
- El pooling reduce la resolución espacial y el costo computacional en comparación con las convoluciones tradicionales.

# ARQUITECTURA



# CONJUNTO DE DATOS: DETECCIÓN



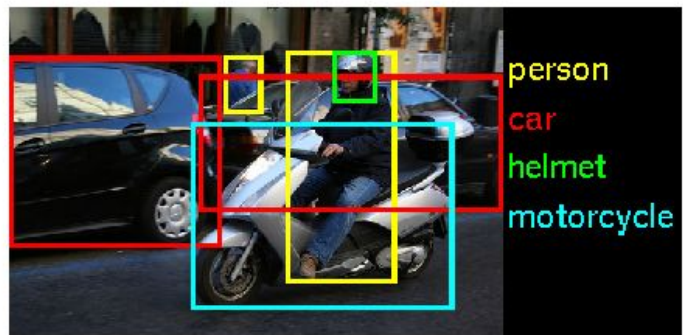
- 200 categorías de nivel básico completamente anotadas en los datos de prueba.
- Las categorías se eligieron cuidadosamente teniendo en cuenta distintos factores.
- Algunas de las imágenes de prueba no contienen ninguna de las 200 categorías.

# 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.

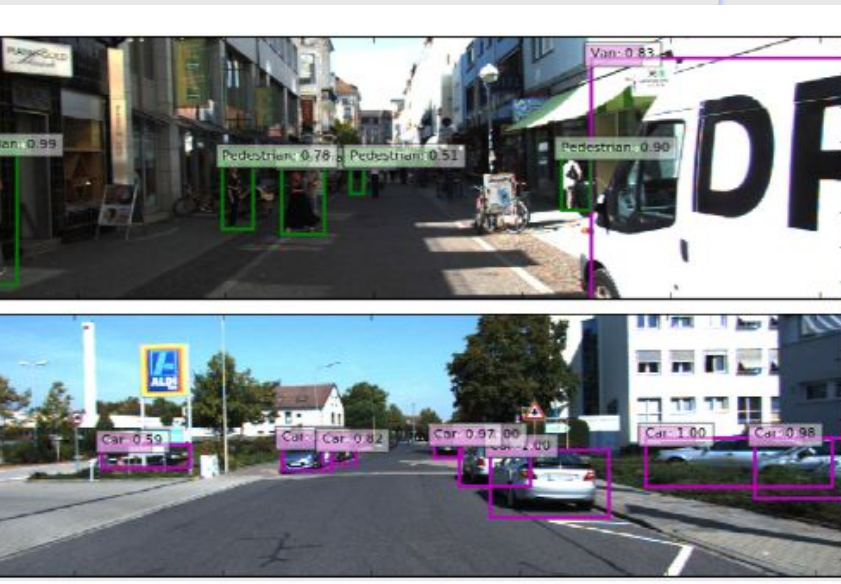
Team	Year	Place	Error (top-5)	Uses external data
SuperVision	2012	1st	16.4%	no
SuperVision	2012	1st	15.3%	Imagenet 22k
Clarifai	2013	1st	11.7%	no
Clarifai	2013	1st	11.2%	Imagenet 22k
MSRA	2014	3rd	7.35%	no
VGG	2014	2nd	7.32%	no
GoogLeNet	2014	1st	6.67%	no

Number of models	Number of Crops	Cost	Top-5 error	compared to base
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.

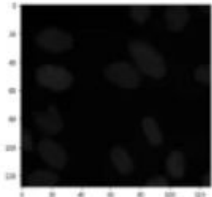
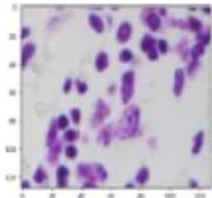
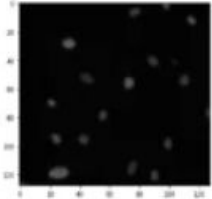
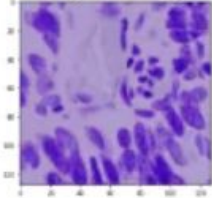




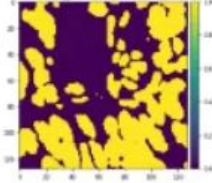
# Detección de objetos

- 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.

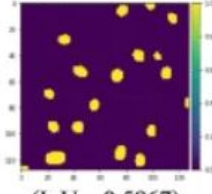
Input image



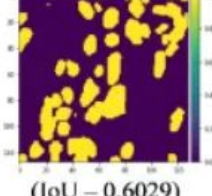
Inception U-Net



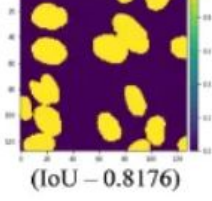
(IoU – 0.5522)



(IoU – 0.5967)



(IoU – 0.6029)



(IoU – 0.8176)

# 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**