

# 3D Object Recognition with Convolutional Neural Networks

Autor:

Albert García García — agg180@alu.ua.es

Supervisores:

Jorge Pomares Baeza — jpomares@ua.es

José García Rodríguez — jgarcia@dtic.ua.es

2016/06/06

Universidad de Alicante

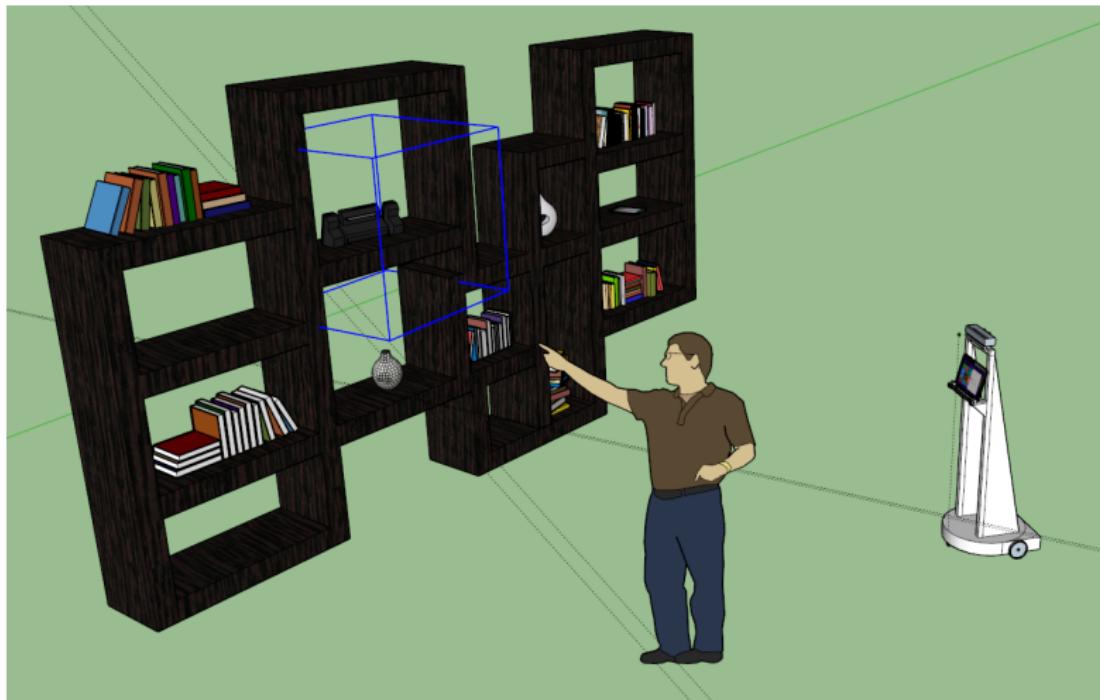
Máster en Automática y Robótica

1. Introducción
2. Marco Teórico
3. Representaciones Volumétricas
4. Red Neuronal Convolucional para Reconocimiento de Objetos
5. Conclusión

# Introducción

---

# Motivación: SIRMAVED

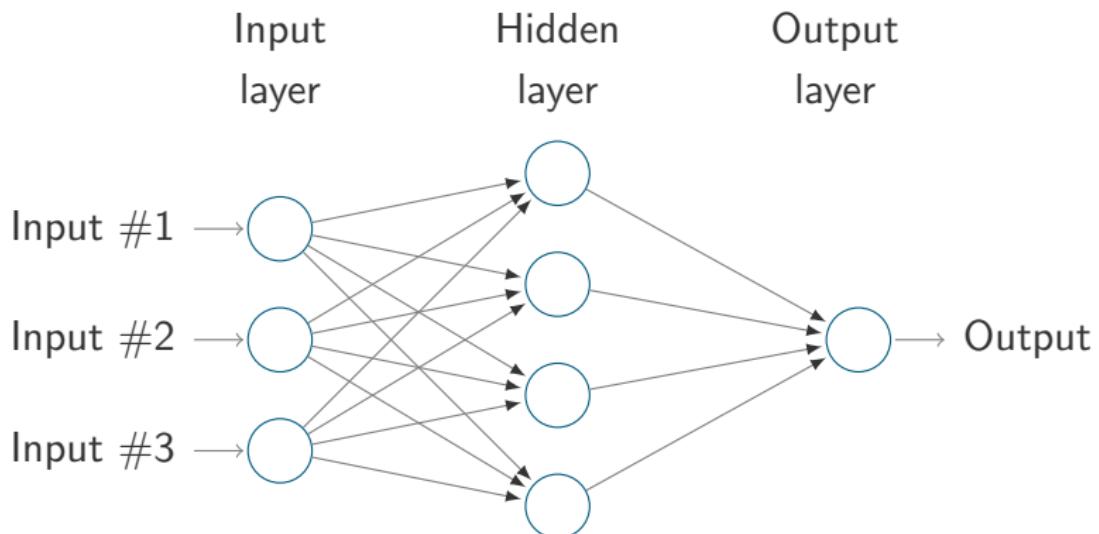


# Introducción y Objetivos

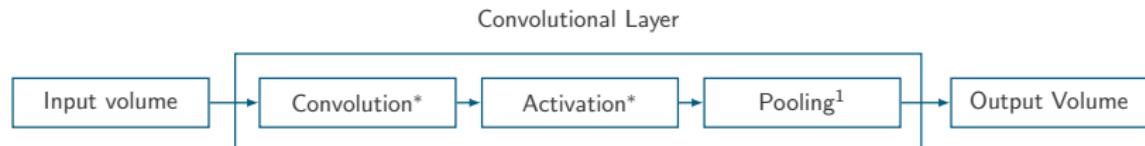
- Reconocimiento de clases de objetos 3D
- Aproximación tradicional con descriptores
- Deep Learning (aprendizaje de representaciones)
- Redes Neuronales Convolucionales
  - Convolutional Neural Networks (CNNs)
  - Mejoran por un amplio margen a métodos tradicionales
- Objetivos
  - Analizar el marco teórico de las CNN
  - Estudiar y proponer representaciones volumétricas 3D
  - Diseñar, implementar y testear una CNN 3D

## Marco Teórico

# Red Neuronal Tradicional



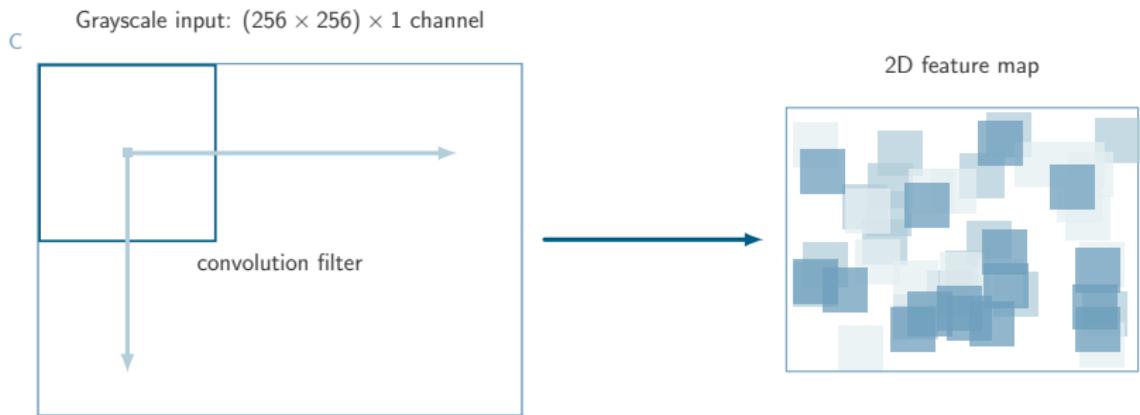
# Redes Neuronales Convolucionales



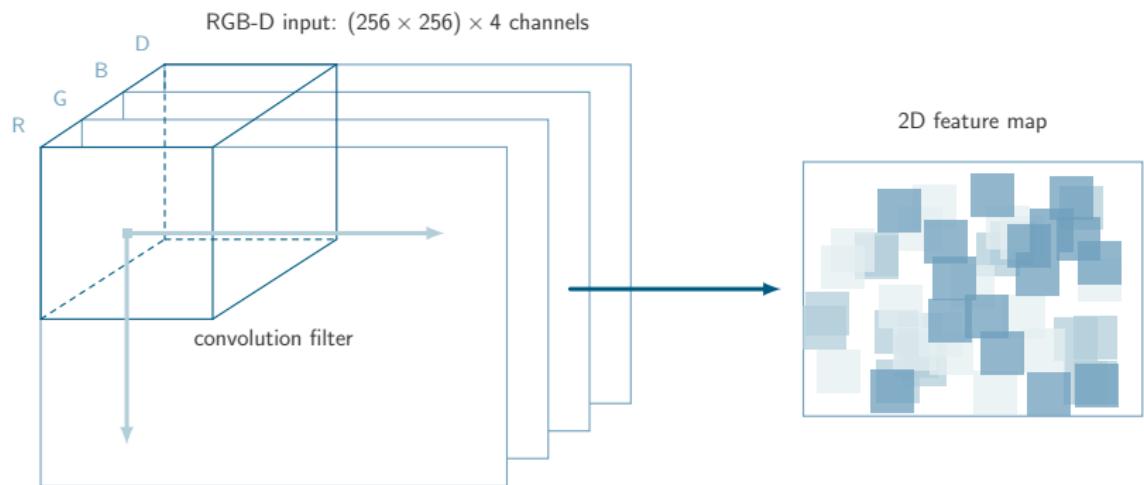
$$\begin{bmatrix} -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & 8 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 \end{bmatrix}$$



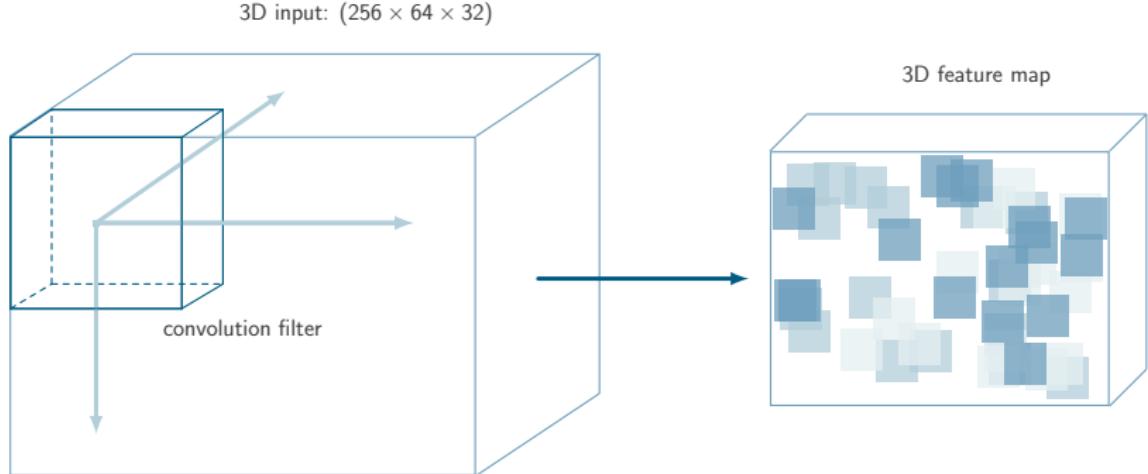
# Convolución 2D



# Convolución 2.5D



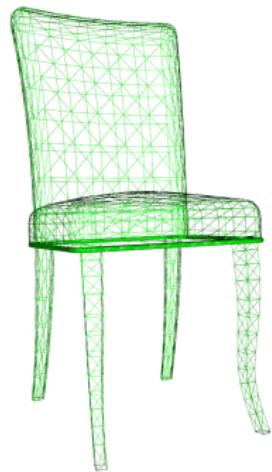
# Convolución 3D



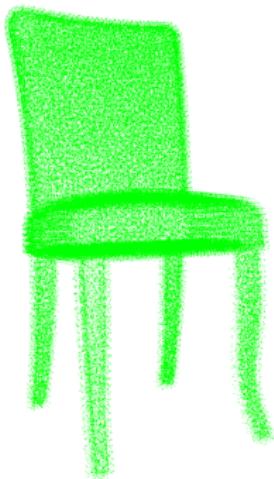
## Representaciones Volumétricas

---

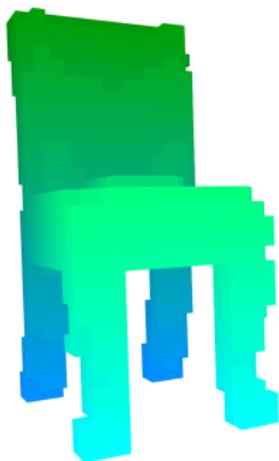
# Representaciones Volumétricas Discretas



(a) Malla

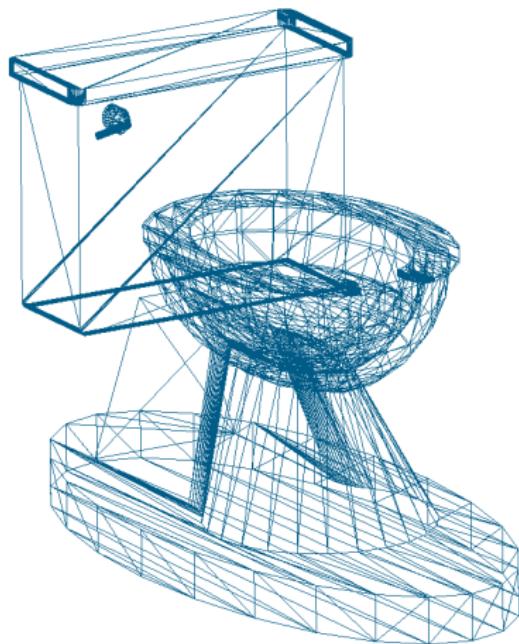


(b) Nube



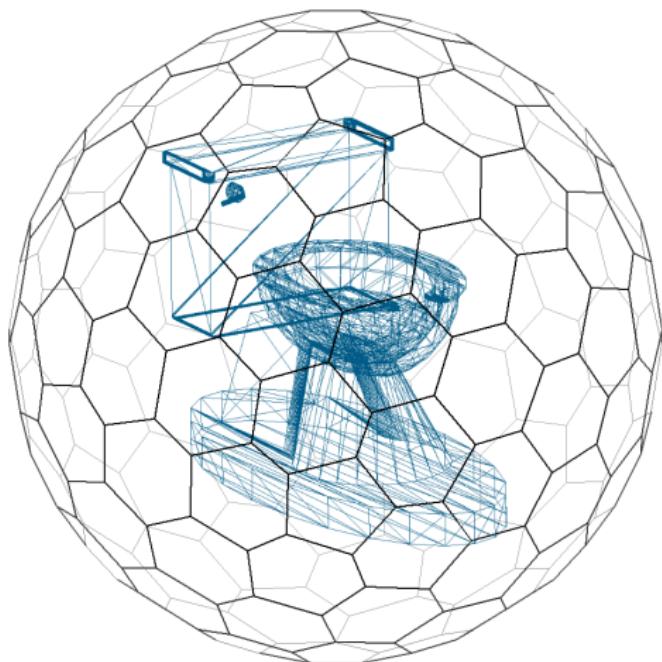
(c) Voxel Grid

# Adaptando el dataset



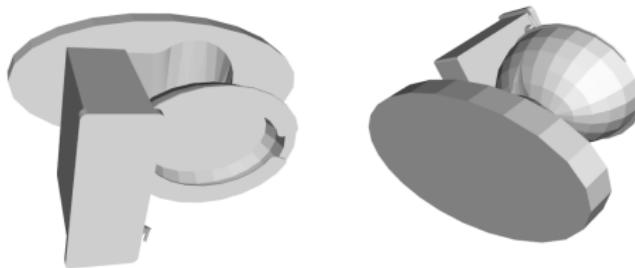
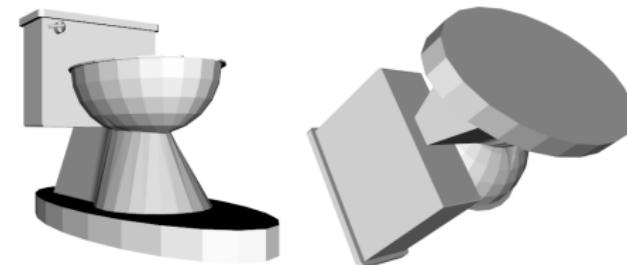
Malla 3D CAD como punto de partida

# Adaptando el dataset



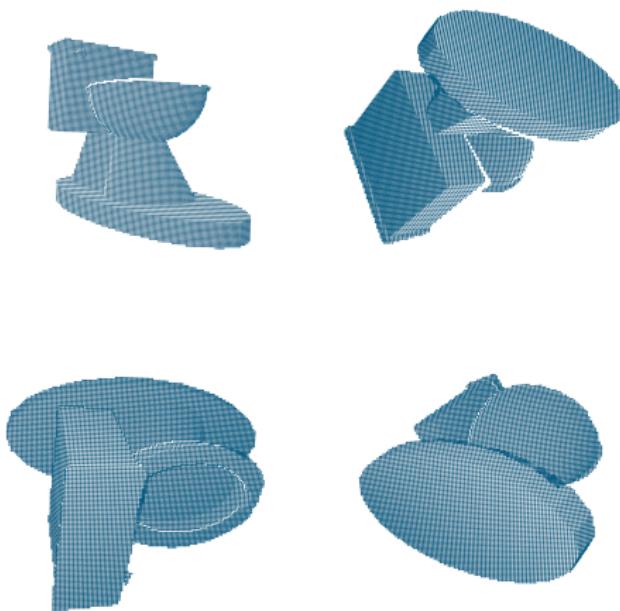
Icosaedro truncado con centro en el objeto

# Adaptando el dataset



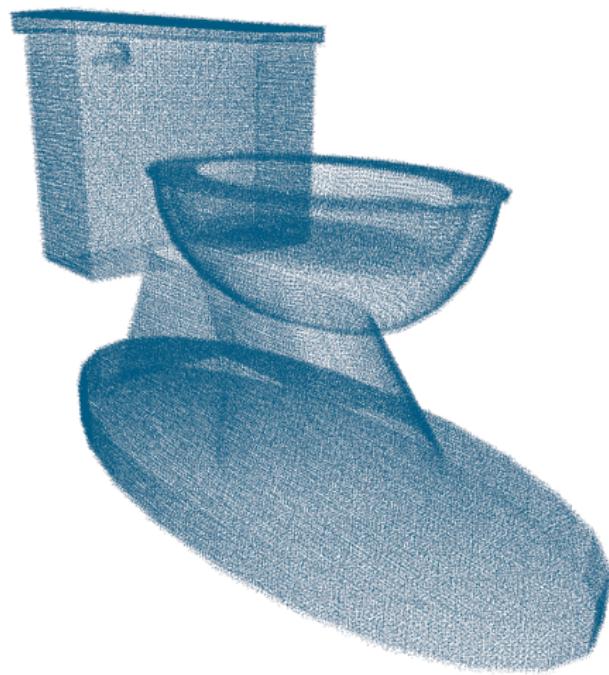
Mapas de profundidad desde los vértices

# Adaptando el dataset



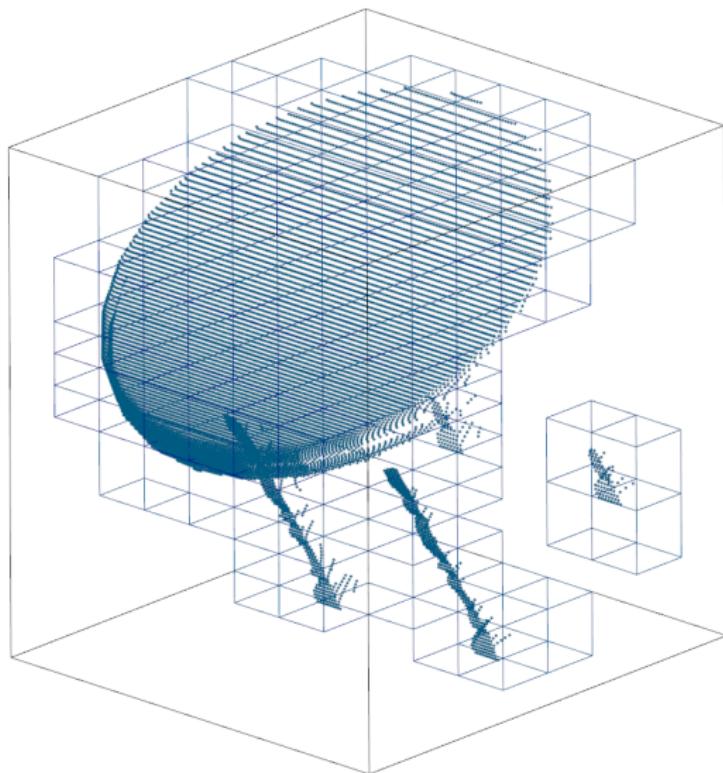
Nubes de puntos a partir del z-buffer

## Adaptando el dataset

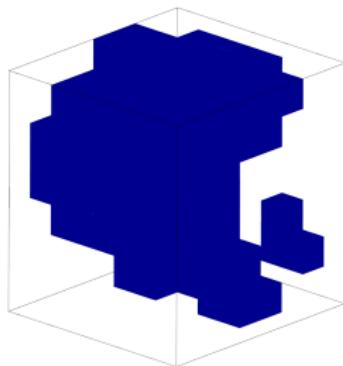


Nube de puntos completa registrada

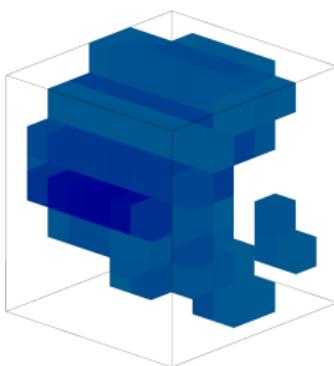
# Tensores Fijo y Adaptativo



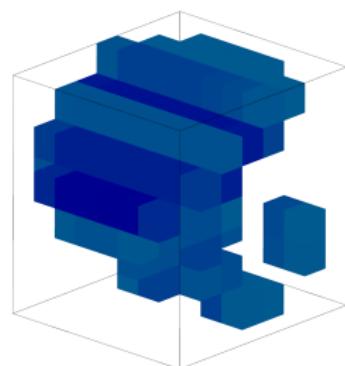
# Medidas de Ocupación



(a) Binaria

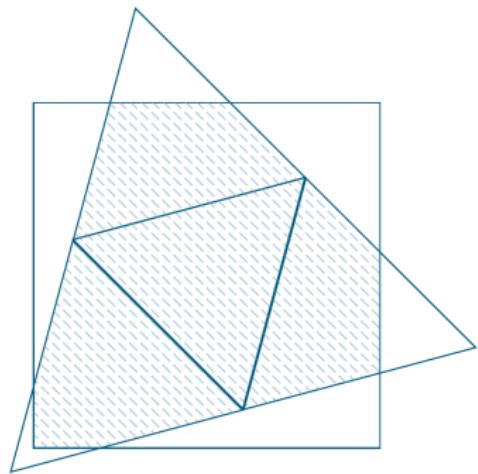
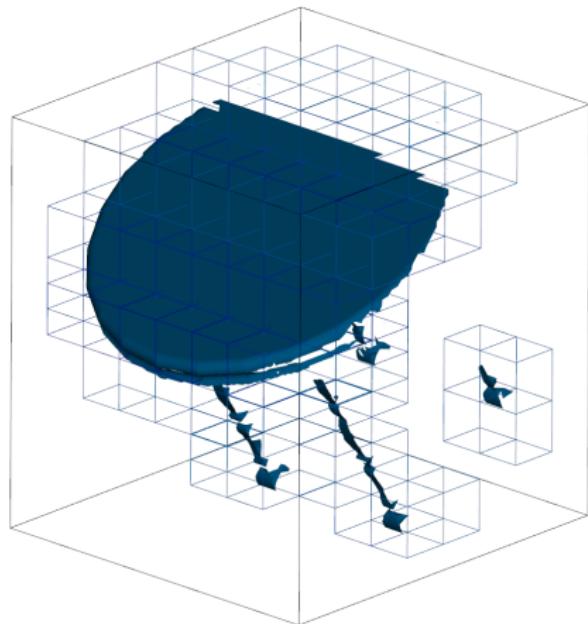


(b) Densidad

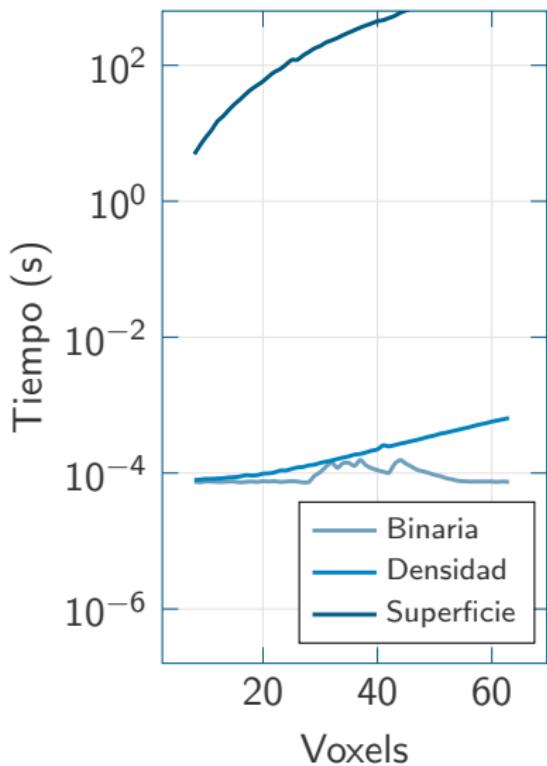
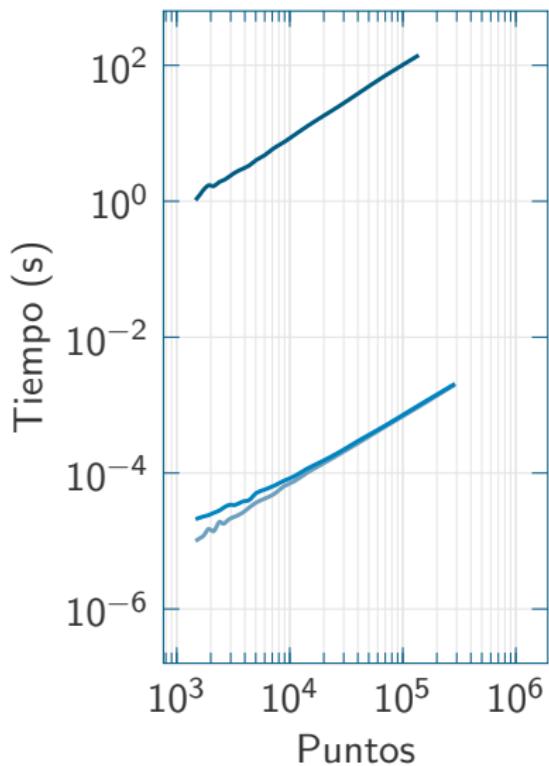


(c) Superficie

# Ocupación de Intersección de Superficie



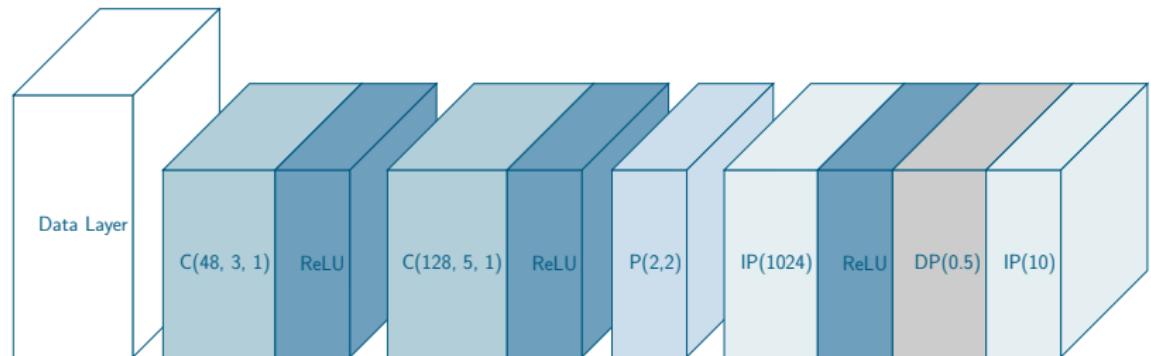
# Experimentación Temporal con Medidas de Ocupación



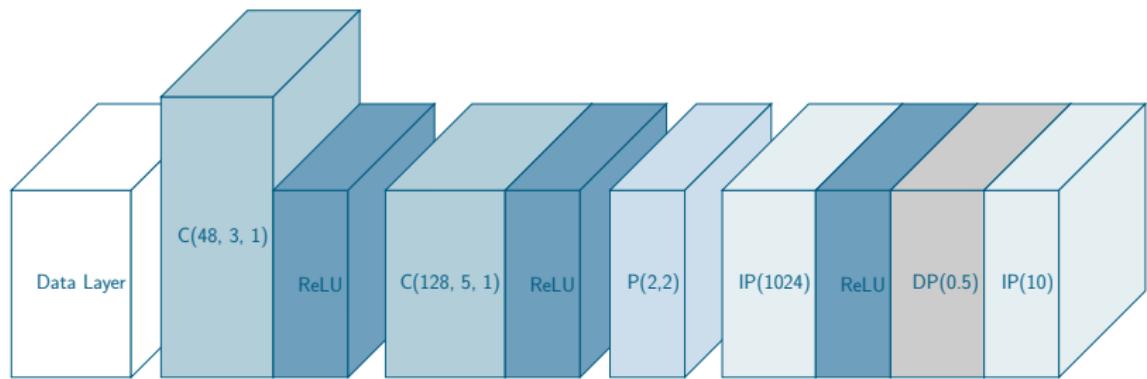
# Red Neuronal Convolucional para Reconocimiento de Objetos

---

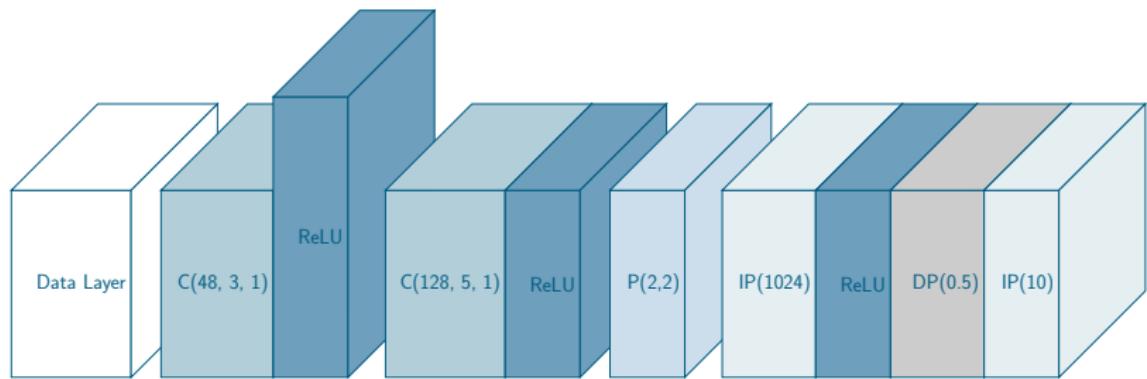
# Arquitectura



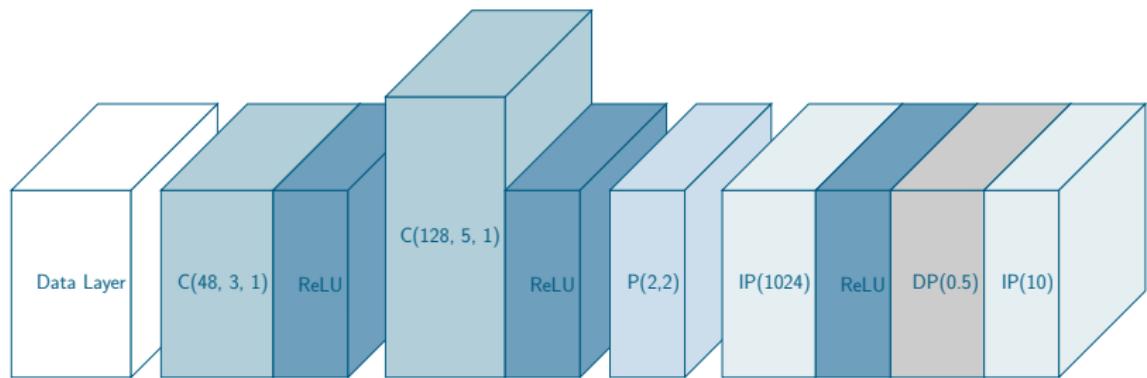
Capa de Datos



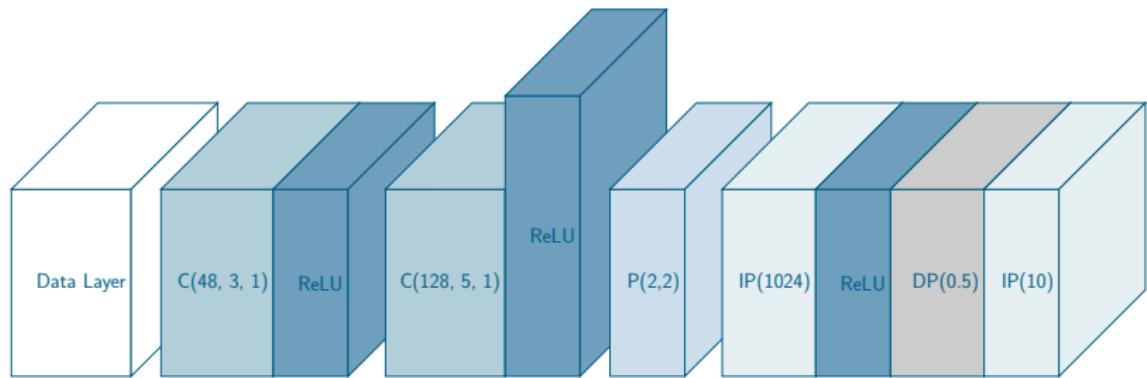
Capa de Convolución ( $3 \times 3, 1$ )



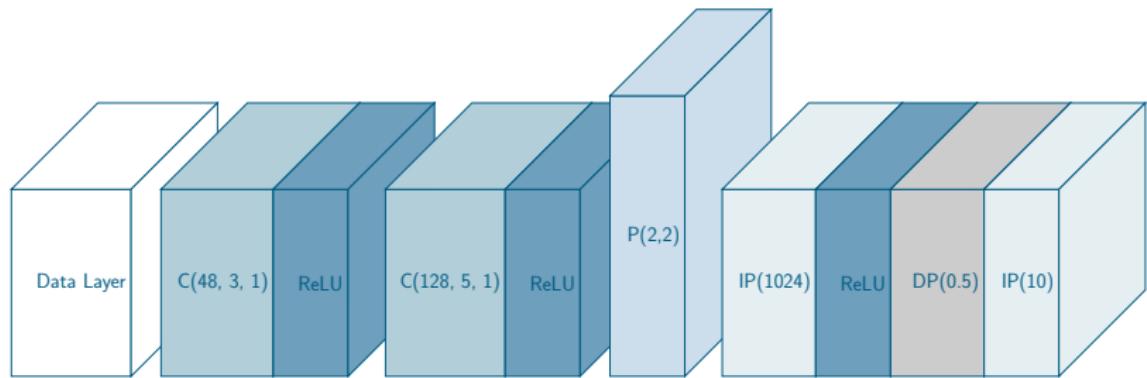
Capa de Activación (ReLU)



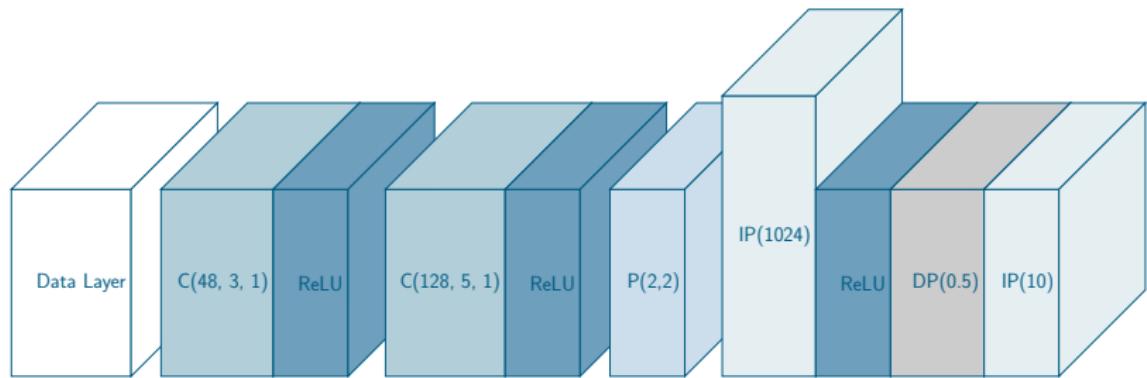
Capa de Convolución ( $5 \times 5, 1$ )



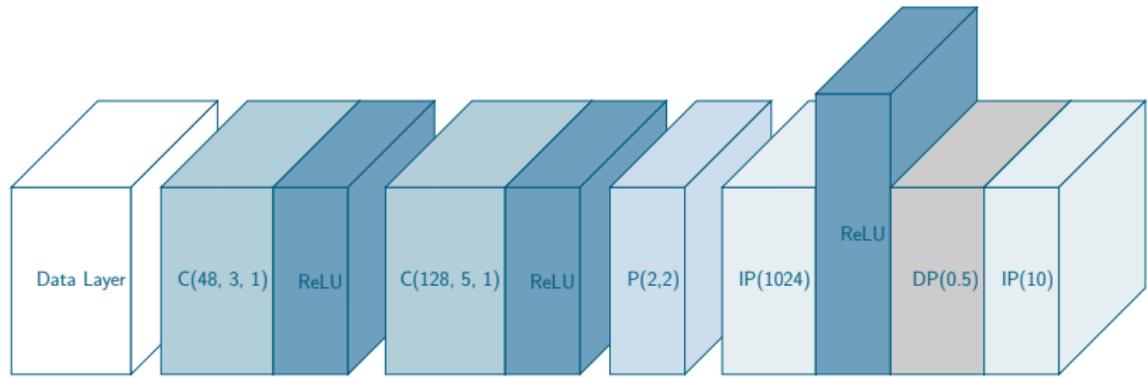
Capa de Activación (ReLU)



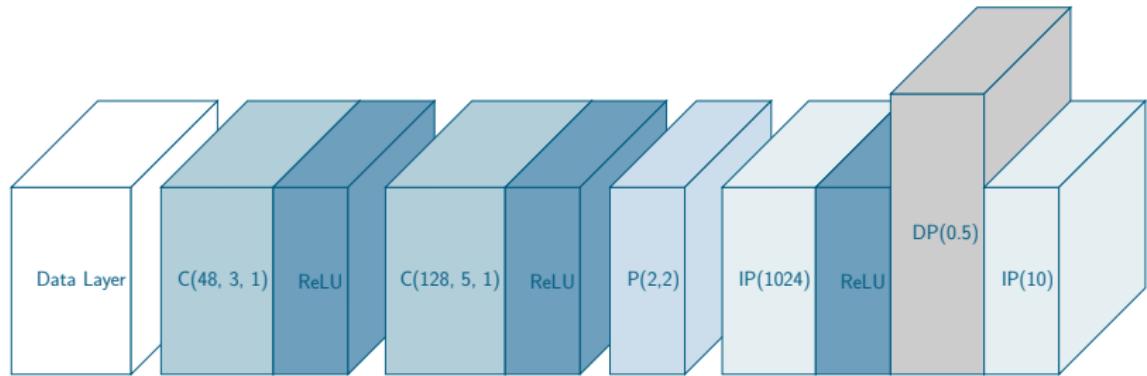
Capa de Pooling (2,2)



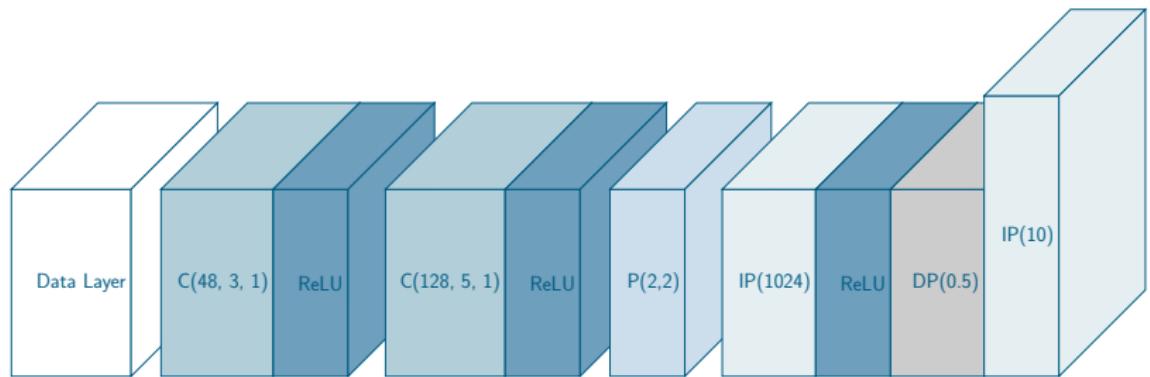
Capa Completamente Conectada (1024 neuronas)



Capa de Activación (ReLU)



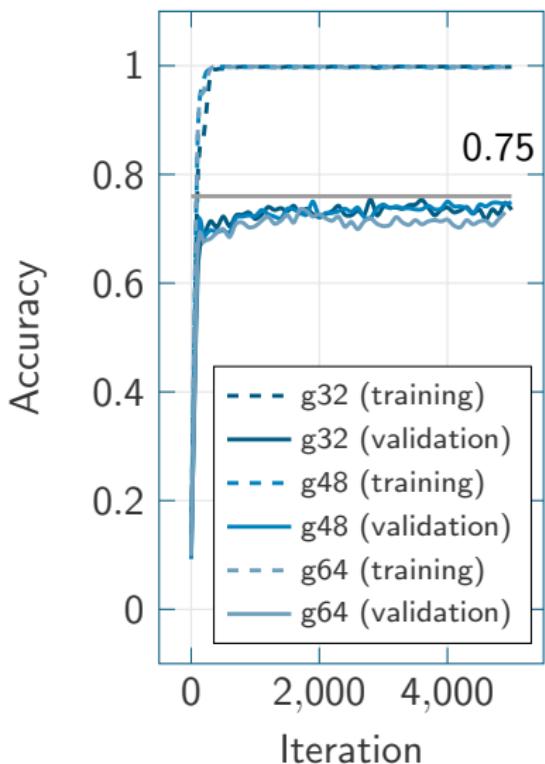
Capa de Dropout (0.5)



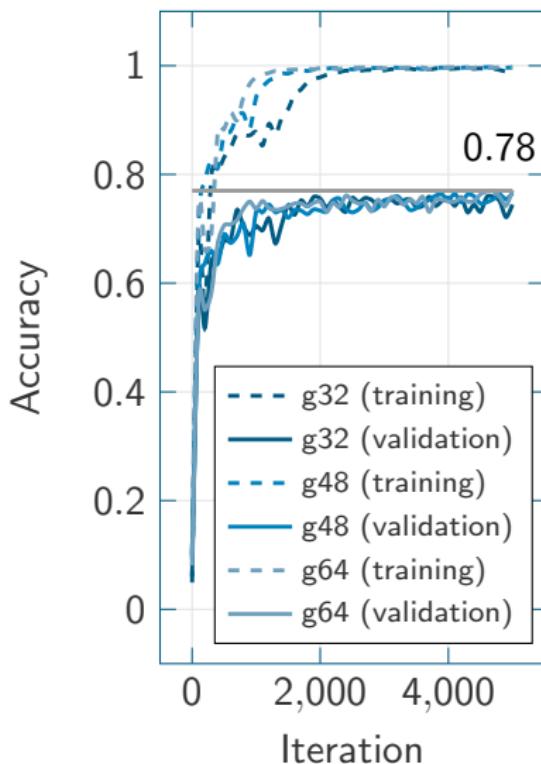
Capa Completamente Conectada (10 neuronas de salida)

# Mallas con ocupación de densidad normalizada en red 2.5D

Tensor Fijo



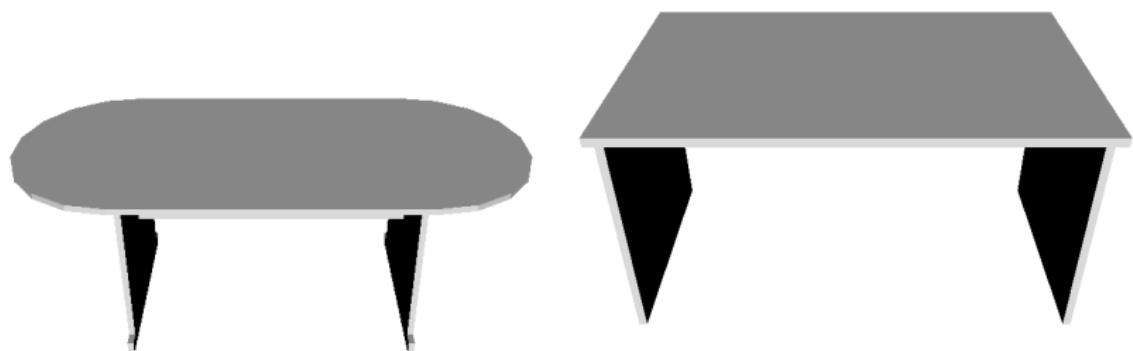
Tensor Adaptativo



# Matriz de Confusión

Desk	Table	Nstand	Bed	Toil.	Dresser	Bath.	Sofa	Moni.	Chair
52	9	1	4	0	5	1	5	0	9
25	69	0	1	0	0	0	0	1	4
1	2	60	1	4	8	0	0	2	8
4	0	0	80	0	0	3	11	1	1
1	0	3	1	84	0	1	3	2	5
3	0	14	0	0	61	0	1	6	1
0	1	0	3	0	0	34	8	3	1
1	0	1	4	1	2	0	88	1	2
1	1	1	1	0	5	1	1	87	2
1	2	1	2	1	1	0	1	1	90

# Confusiones

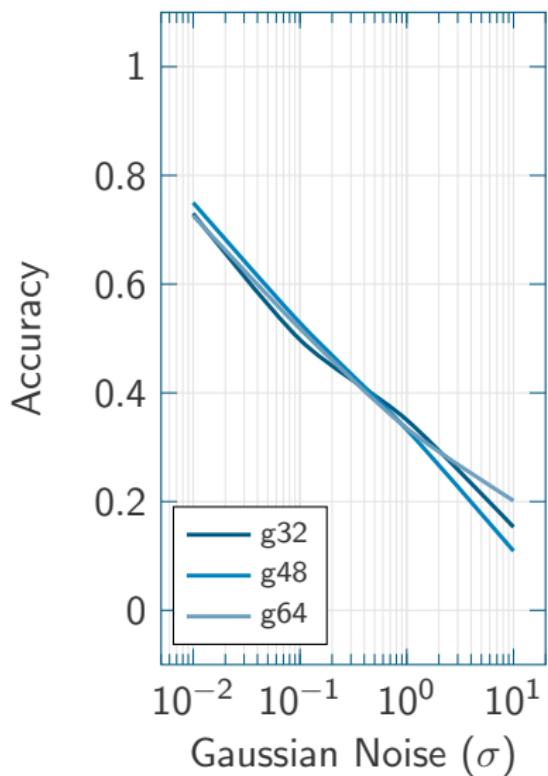


# Simulación de Ruido

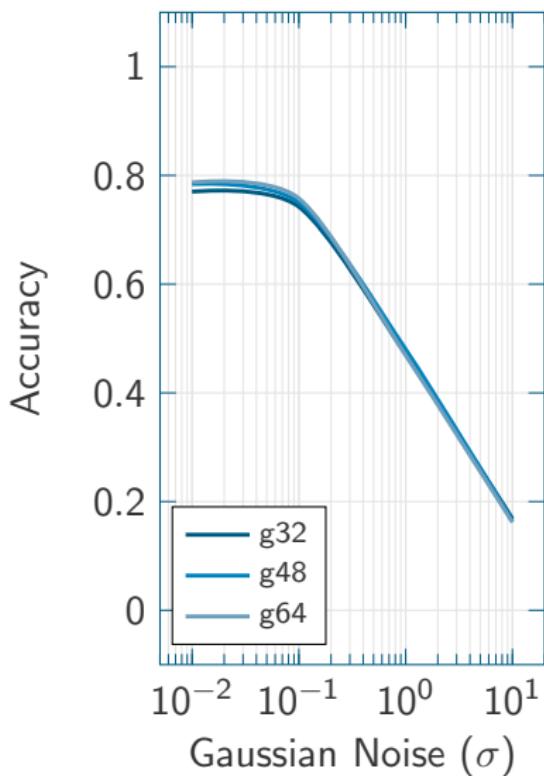


# Mallas con ocupación de densidad normalizada (ruido)

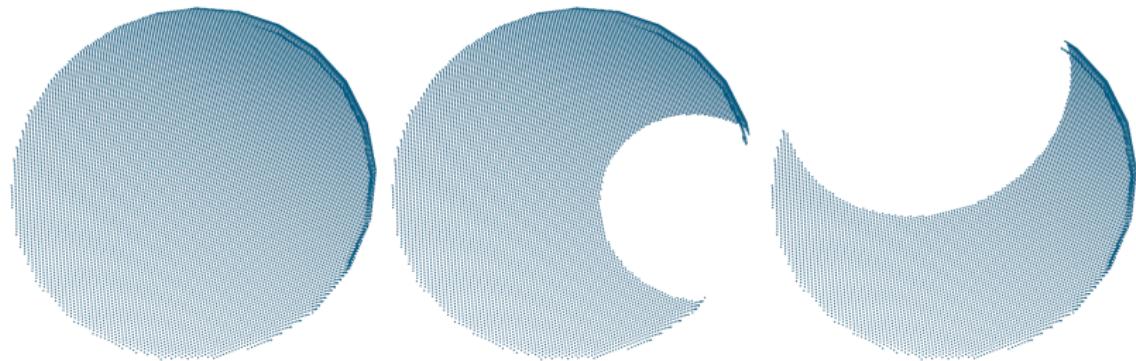
Tensor Fijo



Tensor Adaptativo

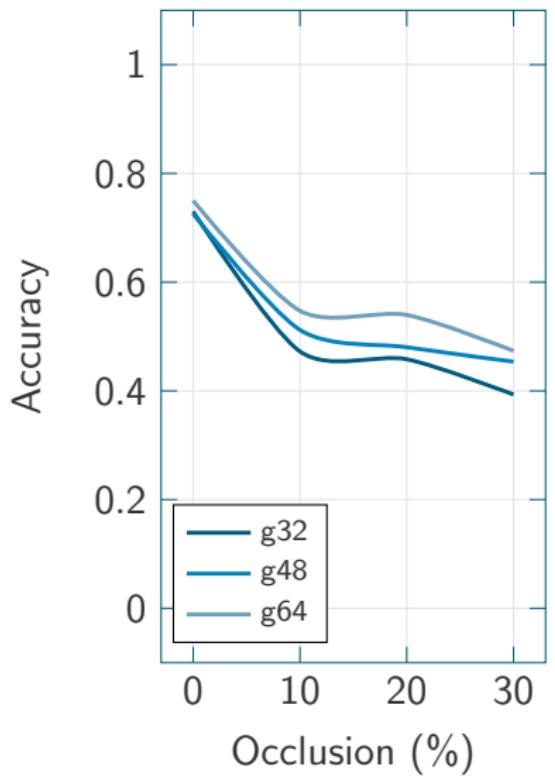


# Simulación de Oclusión

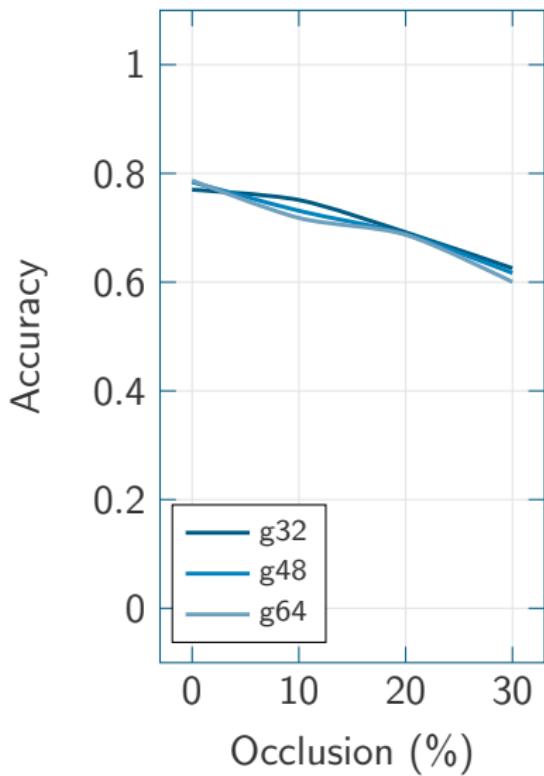


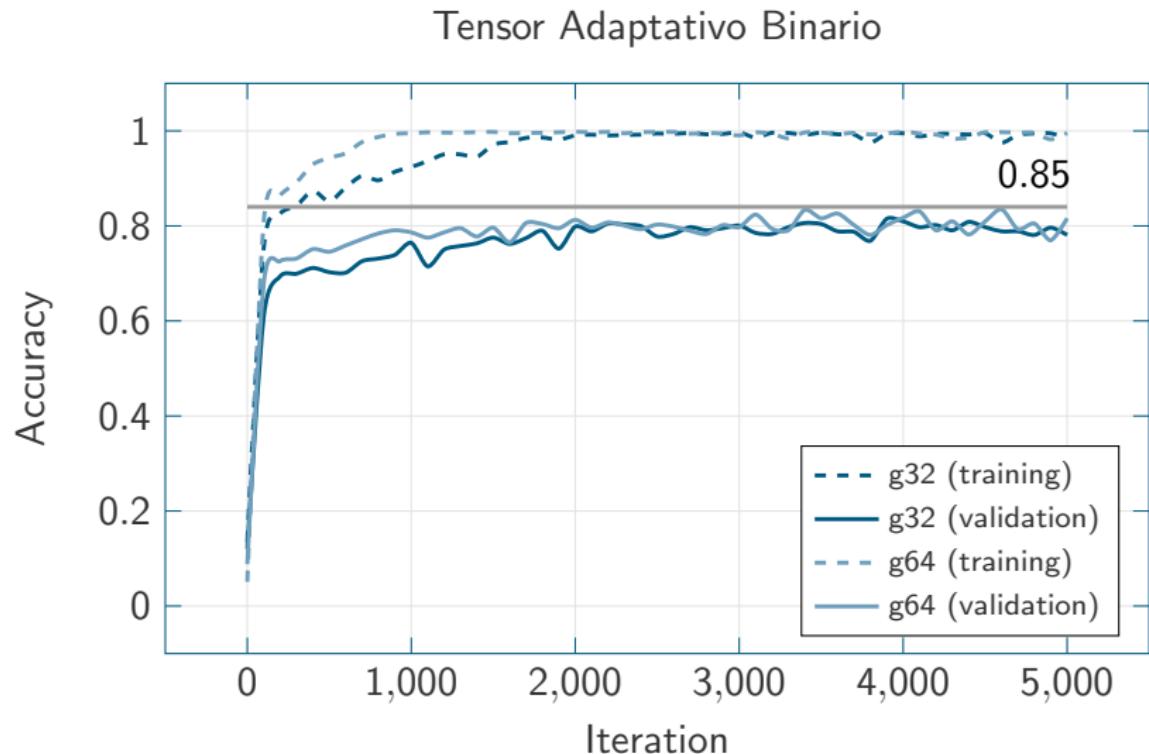
# Mallas con ocupación de densidad normalizada (occlusión)

Tensor Fijo

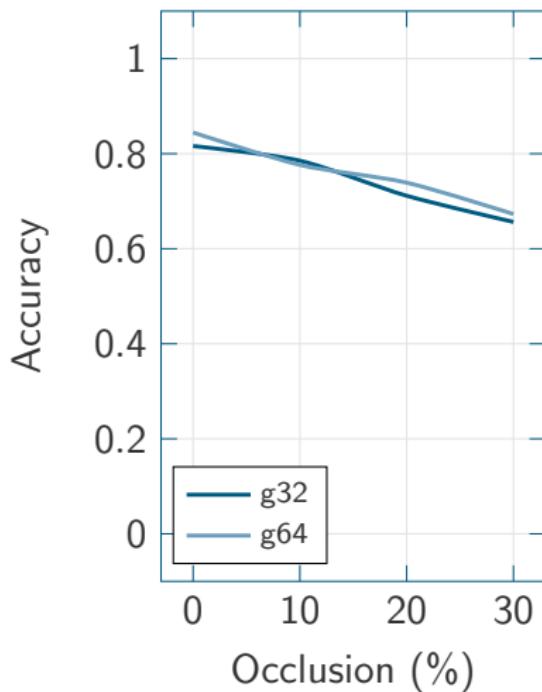
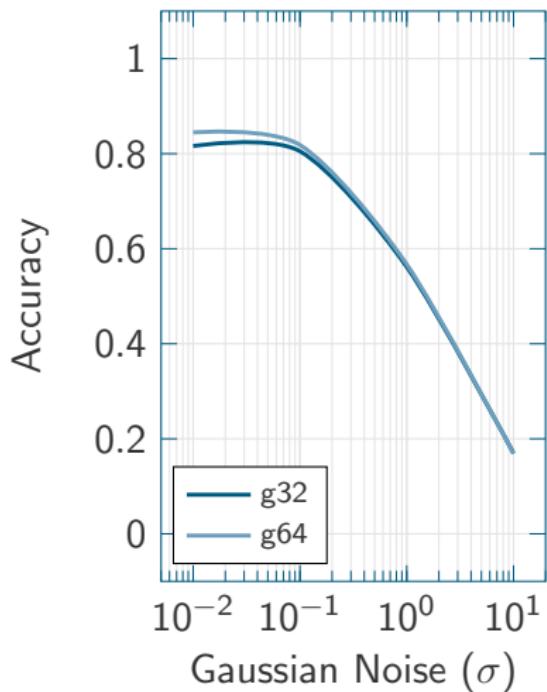


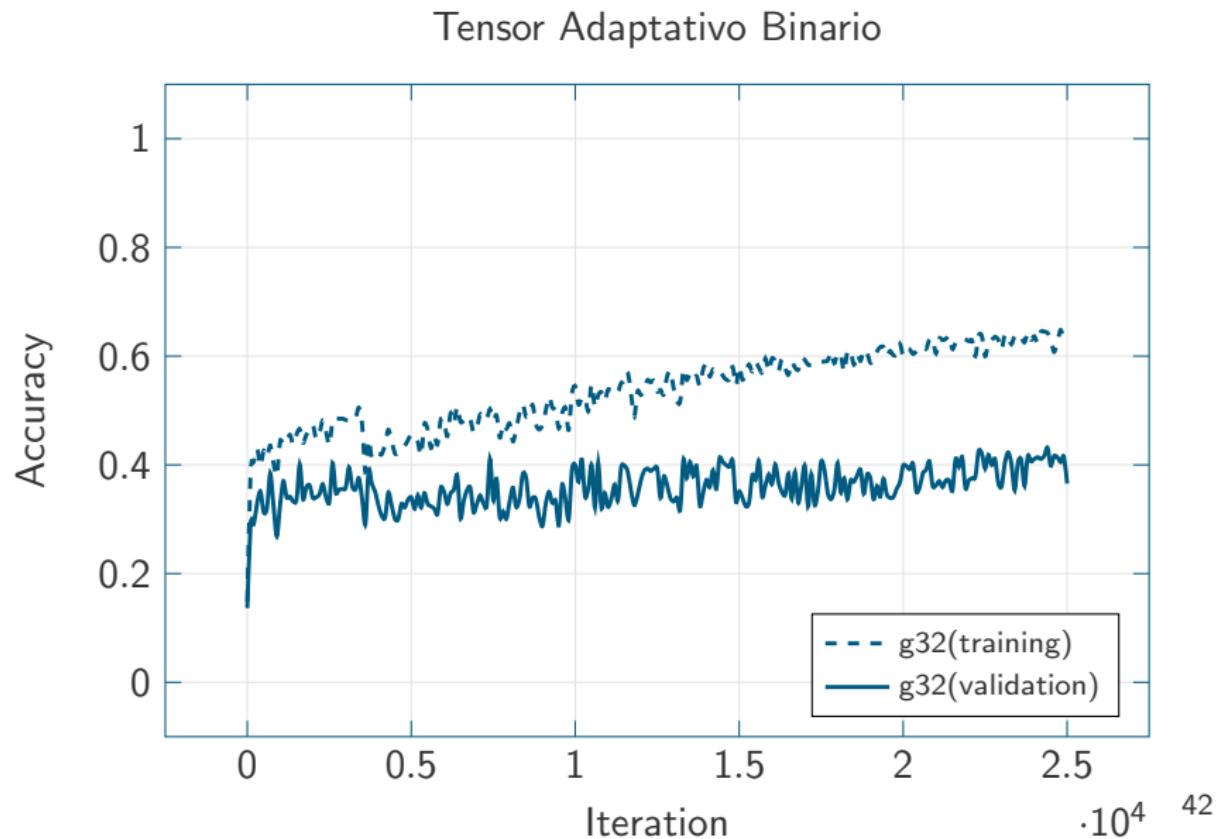
Tensor Adaptativo





Tensor Adaptativo Binario





## Conclusión

---

- Estudio teórico sobre las Redes Neuronales Convolucionales
- Estudio sobre frameworks y datasets existentes
- Estado del arte sobre Redes Neuronales Convolucionales 3D
- Estado del arte de representaciones volumétricas
- Propuesta de representaciones volumétricas
- Propuesta de arquitecturas 2.5D y 3D
- Estudio del efecto del ruido y oclusión en las redes
- **Reconocimiento de objetos 3D con un 85% de acierto**

- **Colaboración en el proyecto nacional SIRMAVED** (DPI2013-40534-R) del Programa Estatal de I+D+i Orientada a los Retos de la Sociedad.
- **Publicación aceptada en congreso Core A:** *International Joint Conference on Neural Networks (IJCNN 2016, Vancouver). PointNet: A 3D Convolutional Neural Network for Real-Time Object Class Recognition.*
- **iV&L Net Training School 2016:** Integrating Vision and Language: Deep Learning (Malta).

- Aceleración del cálculo de representaciones con CUDA
- Propuesta de representaciones volumétricas más adecuadas
- Control más estricto de overfitting (aumento de dataset)
- Experimentación exhaustiva con la red 3D
- Despliegue de la red en un sistema con escenas reales
- Identificación de pose de los objetos

# 3D Object Recognition with Convolutional Neural Networks

Autor:

Albert García García — agg180@alu.ua.es

Supervisores:

Jorge Pomares Baeza — jpomares@ua.es

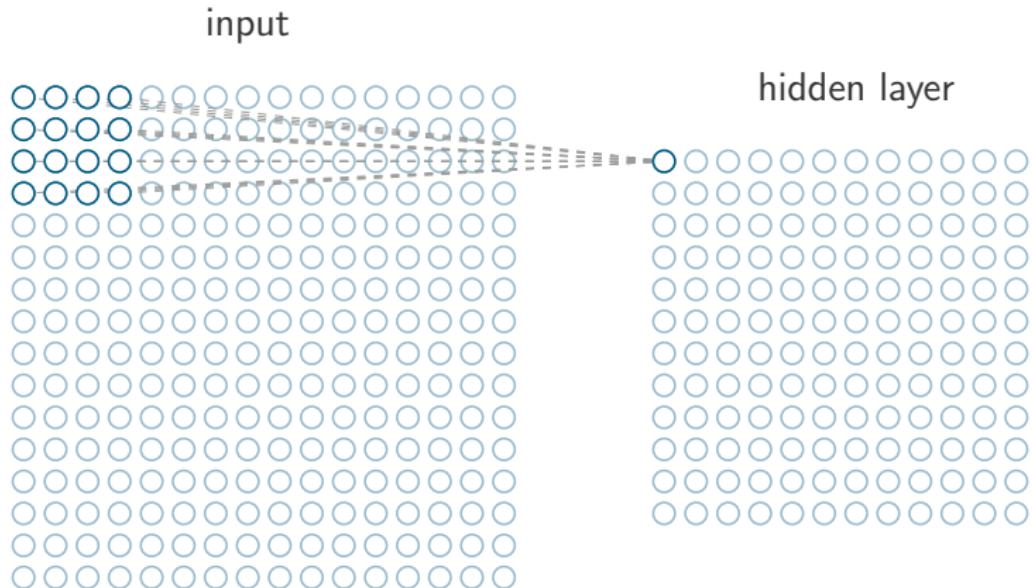
José García Rodríguez — jgarcia@dtic.ua.es

2016/06/06

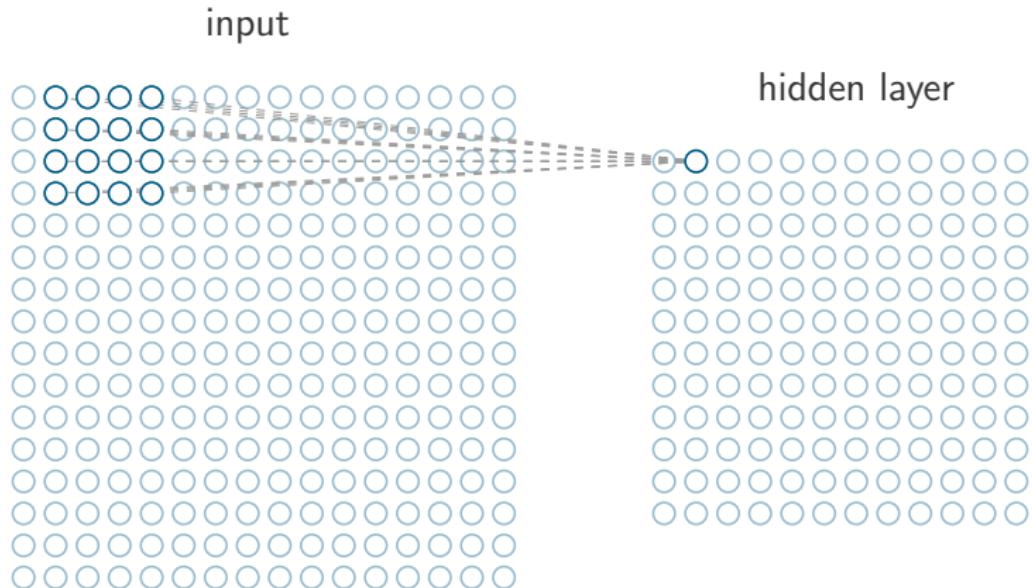
Universidad de Alicante

Máster en Automática y Robótica

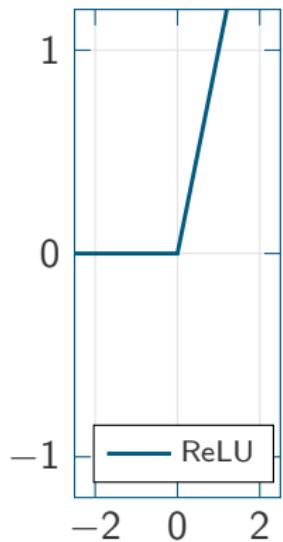
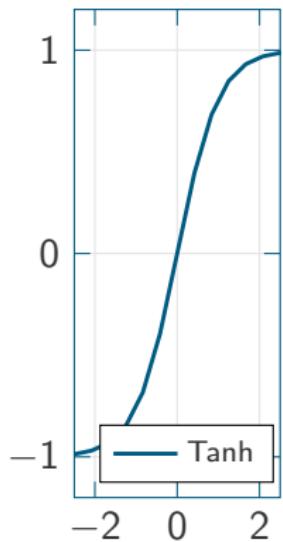
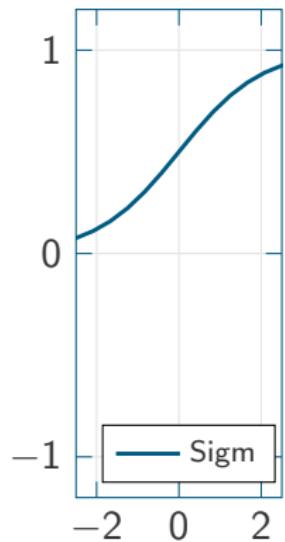
# Capa de Convolución



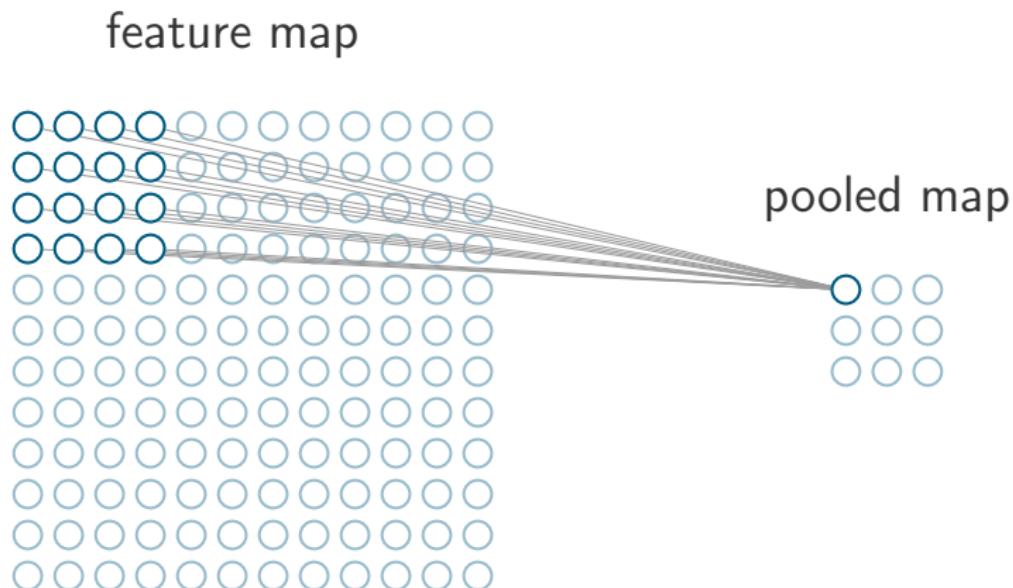
# Capa de Convolución



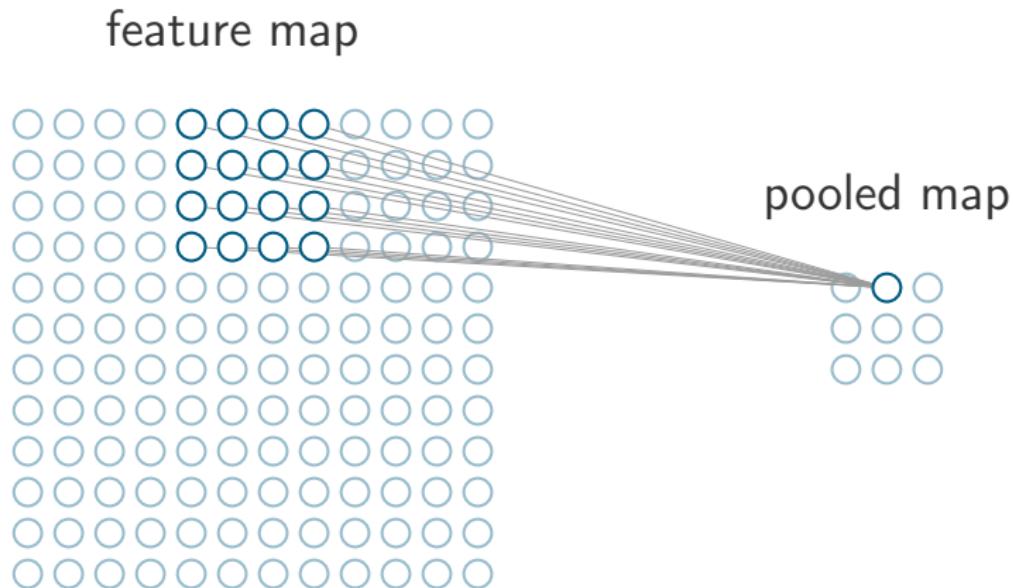
# Capa de Activación



# Capa de Pooling



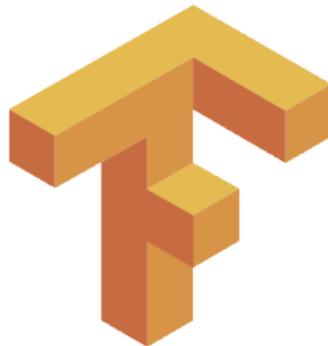
# Capa de Pooling



Caffe



theano



Microsoft  
CNTK

# Dataset



10 clases,  $\approx$  4000 modelos de entrenamiento,  $\approx$  700 de test



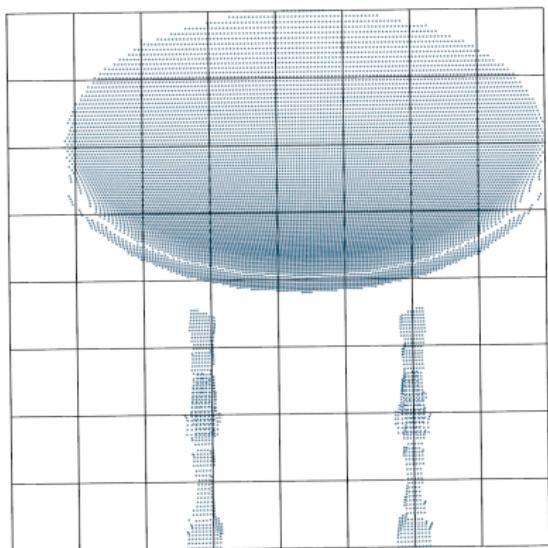
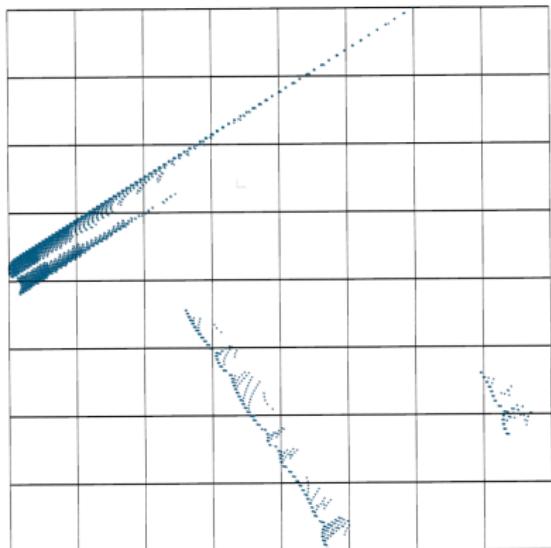
# Hardware

---

OS	Ubuntu 16.04 Xenial Xerus
Motherboard	Asus X99-A
CPU	<b>Intel(R) Core(TM) i7-5820K CPU @ 3.30GHz</b> 3.3 GHz (3.6 GHz Turbo Boost) — 6 cores (12 threads)
GPU (visualization)	<b>NVIDIA GeForce GT730</b> 96 CUDA cores — 1024 MiB of DDR3 Video Memory
GPU (deep learning)	<b>NVIDIA GeForce Titan X</b> 3072 CUDA cores — 12 GiB of GDDR5 Video Memory
GPU (compute)	<b>NVIDIA Tesla K40c</b> 2880 CUDA cores — 12 GiB of GDDR5 Video Memory
RAM	4 × 8 GiB Kingston Hyper X DDR4 2666 MHz CL13
Storage (Data)	(RAID1) Seagate Barracuda 7200rpm 3TiB SATA III HDD
Storage (OS)	Samsung 850 EVO 500GiB SATA III SSD

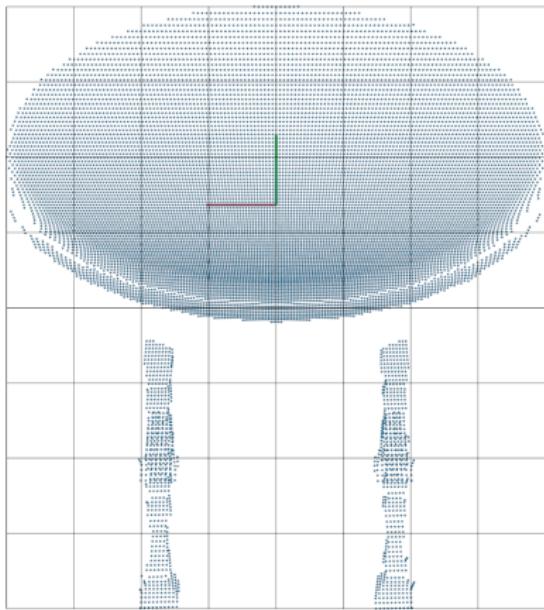
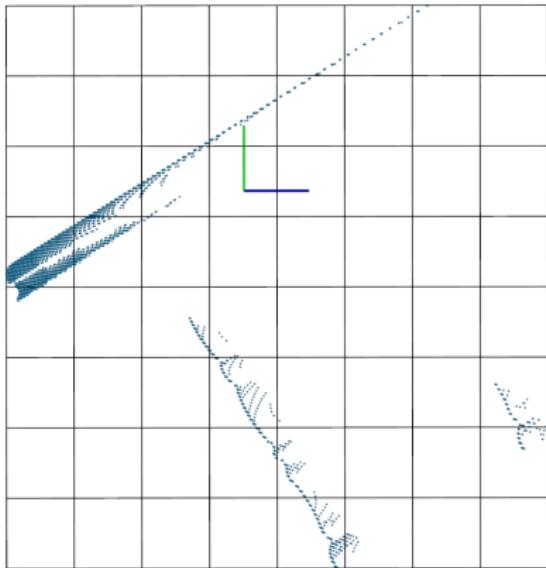
---

# Tensor Fijo



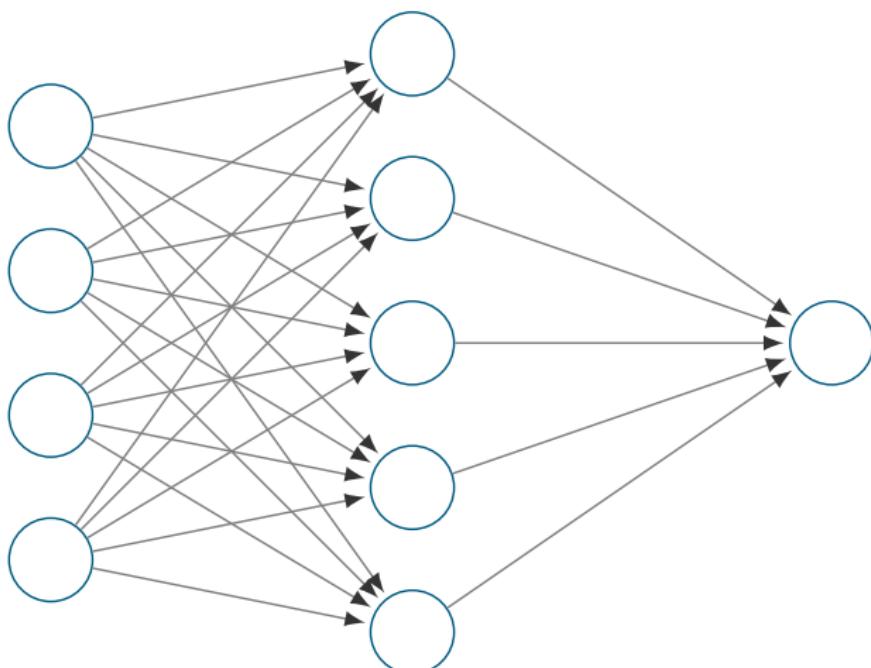
Tamaño de voxel fijo, la nube es escalada

# Tensor Adaptativo



Tamaño de voxel adaptativo, el tensor se adapta

# Dropout Desactivado



# Dropout Activado

