

Trabajo Fin de Grado

Grado en Ingeniería Informática
Mención en Ingeniería del Software

Análisis y Experimentación Práctica de Frameworks Deep Learning Aplicados a la Astronomía

Autor: **Jorge Barrio Conde**

Tutor: **Benjamín Sahelices Fernández**

Cotutor: **Fernando Buitrago Alonso**



Índice

1. Objetivos
2. Contexto
3. Tecnologías
4. Clasificación morfológica de galaxias
5. Detección de bordes de las galaxias
6. Demostración práctica
7. Conclusiones

Índice

1. Objetivos
2. Contexto
3. Tecnologías
4. Clasificación morfológica de galaxias
5. Detección de bordes de las galaxias
6. Demostración práctica
7. Conclusiones

Objetivos

- Aprender los **conocimientos** claves del **Deep Learning**
- Aplicar Deep Learning a **problemas astronómicos**:
 - Clasificación morfológica de galaxias
 - Detección de los bordes de las galaxias

Índice

1. Objetivos
2. Contexto
3. Tecnologías
4. Clasificación morfológica de galaxias
5. Detección de bordes de las galaxias
6. Demostración práctica
7. Conclusiones

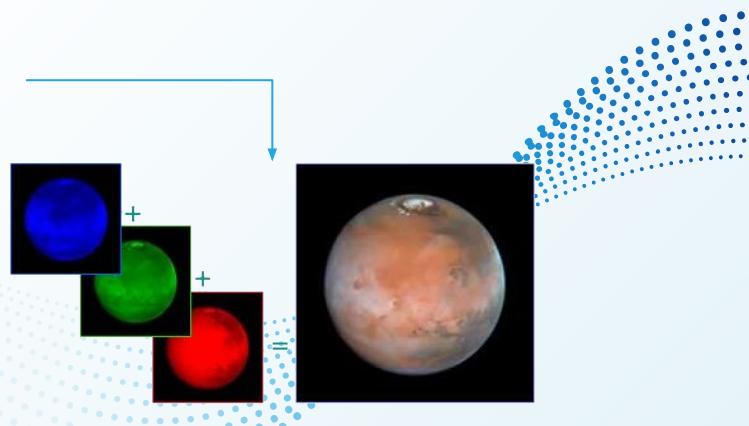
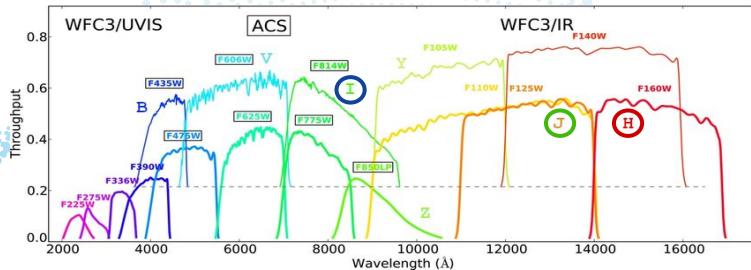
Telescopio Espacial Hubble

- Análisis de **galaxias distantes** captadas por el Telescopio Espacial Hubble



Cámaras y filtros del Hubble

- **ACS** se centra en el espectro **visible** (de 380 nm a 750 nm)
- **WFC3** capta el **infrarrojo** y el **ultravioleta**
- Las cámaras utilizan **filtros** para filtrar las longitudes de onda
- Imágenes obtenidas con técnica **RGB**

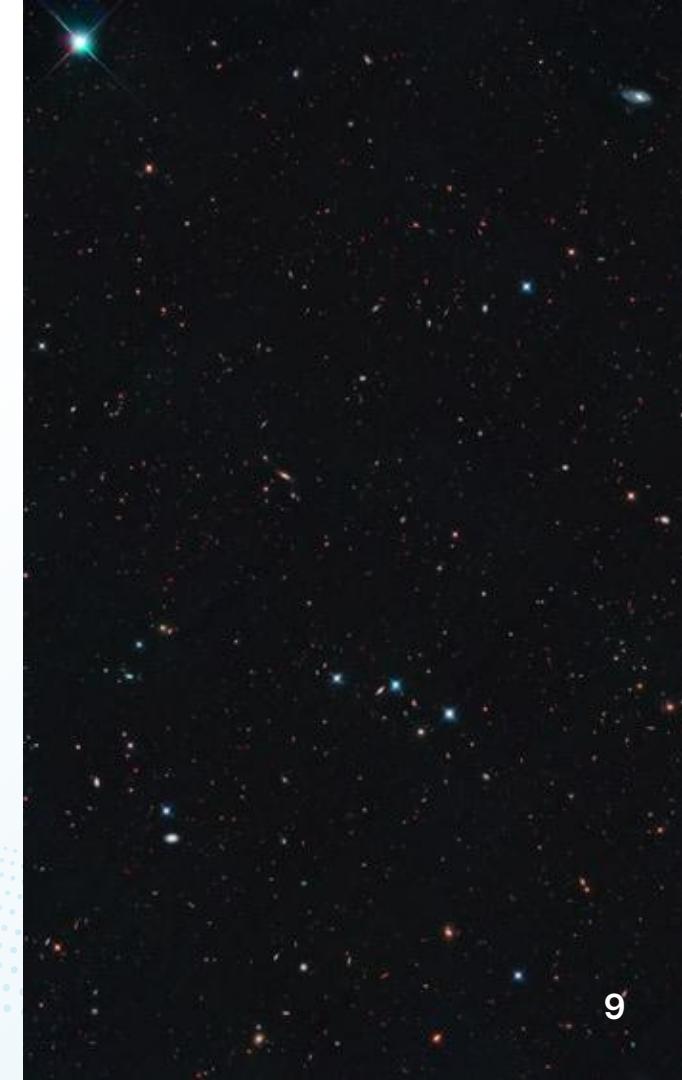


Cartografiado CANDELS

- CANDELS es uno de los proyectos más importantes del Hubble
- Galaxias distantes
- Representa gran parte de la historia del universo

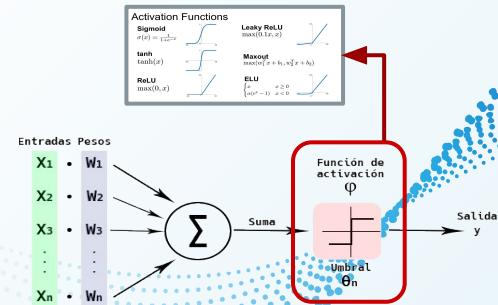
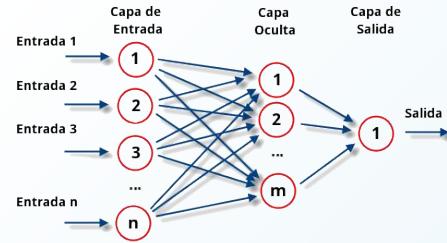
Cartografiado CANDELS

- Ejemplos de galaxias con las que se ha trabajado:



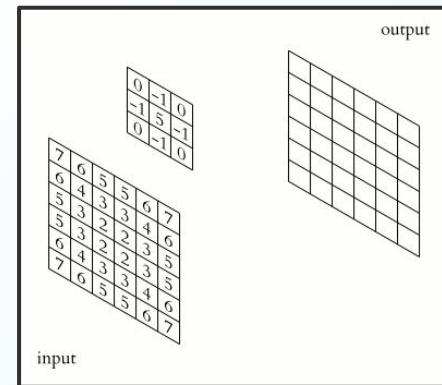
Redes Neuronales

- **Redes Neuronales Artificiales:** conjunto de neuronas conectadas entre sí y estructuradas en capas.
- **Neurona artificial:** genera una salida a partir de un conjunto de entradas. Frank Rosenblatt diseñó la primera neurona artificial en el año 1957.



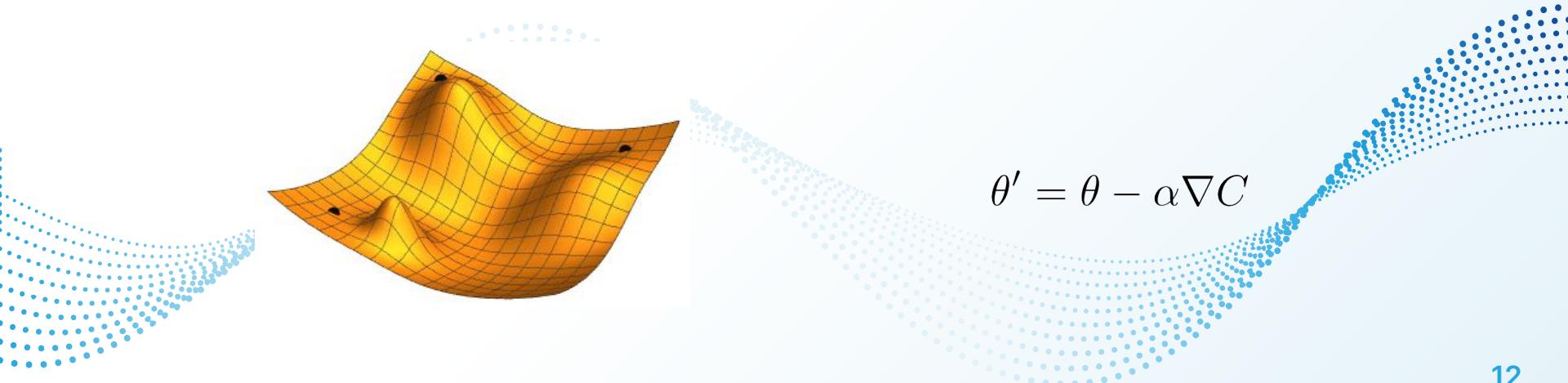
Redes Neuronales Convolucionales

- Enfocadas a **imágenes**
- **Ventajas:**
 - Detección de **características**
 - Dependencia de píxeles vecinos
- **Convolución:** operación para obtener un mapa de características de una imagen



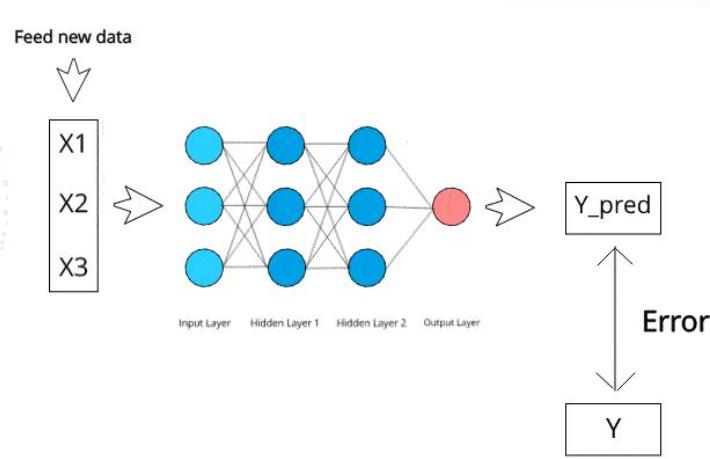
Descenso de Gradiente

- **Proceso iterativo** cuyo objetivo es **minimizar** el valor de la **función** de coste
- En cada iteración actualiza los parámetros de la función en sentido contrario al vector gradiente (∇C)



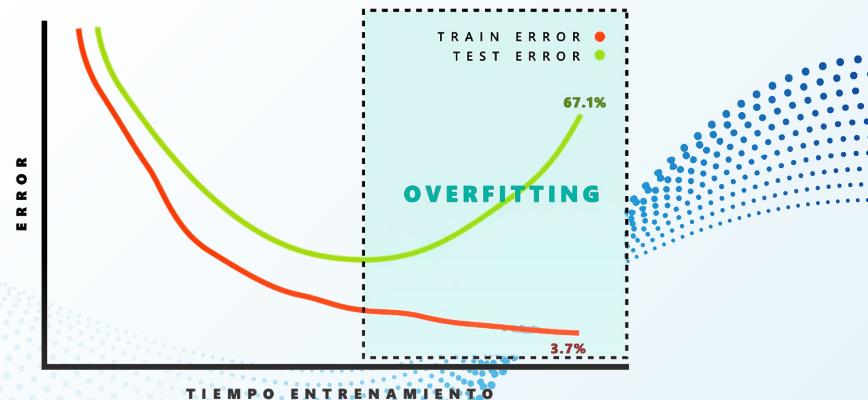
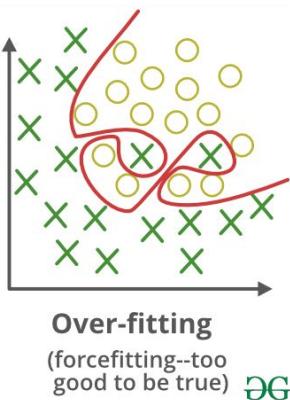
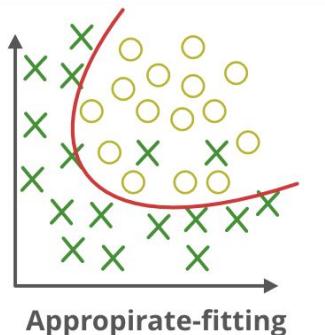
Algoritmo de Backpropagation

- Calcula las derivadas parciales de los parámetros de la red para que el Descenso del Gradiente optimice el coste.



Overfitting o sobreajuste

- La red genera conocimiento específico de los datos que conoce
- Incapacidad de generalizar conocimiento a nuevos datos



Transfer learning

- Red neuronal **pre-entrenada** para una **tarea genérica**
- Ventajas:
 - Precisiones más altas
 - Menor tiempo de entrenamiento
- Redes **ResNet**: redes convolucionales pre-entrenadas con ImageNet. Existen varias versiones nombradas según su número de capas (18, 34, 50, 101, etc.).

Índice

1. Objetivos
2. Contexto
3. Tecnologías
4. Clasificación morfológica de galaxias
5. Detección de bordes de las galaxias
6. Demostración práctica
7. Conclusiones

Tecnologías: Deep Learning



Python

Lenguaje de
programación



PyTorch

Biblioteca de
autodiferenciación



fastai

Biblioteca de
Deep Learning

Tecnología: Astroinformática



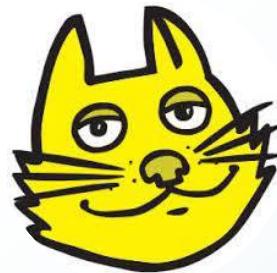
SAOImage DS9

Visualización de imágenes
astronómicas



SExtractor

Generador de catálogo de
los objetos de una imagen
astronómica



Topcat

Visualización y
procesamiento de datos
tabulares



Aladin

Visualización de imágenes
astronómicas

Tecnología: Otros



Ubuntu 20

Sistema operativo



Jupyter Notebook

Documentos



Google Colab

Plataforma de cómputo

Índice

1. Objetivos
2. Contexto
3. Tecnologías
4. Clasificación morfológica de galaxias
5. Detección de bordes de las galaxias
6. Demostración práctica
7. Conclusiones

Problema

- Problema inspirado en el paper de M. Huertas-Company et al. 2015
- Predecir 5 valores, en el rango [0, 1], correspondientes a la fracción de expertos que determinarían una característica específica de la galaxia:
 - Existencia de un disco (f_{disk})
 - Existencia de un esferoide ($f_{spheroid}$)
 - Detección de alguna irregularidad (f_{irr})
 - Es una fuente puntual (f_{ps})
 - Es inclasificable (f_{unc})

Interpretación de las fracciones

- La interpretación de estos valores depende de:
 - Propósitos científicos
 - Propiedades que se quieran resaltar de las galaxias
- El artículo propone una clasificación en 6 categorías morfológicas en base a estas fracciones:

```
if f_sph > 0.667 and f_disk < 0.667 and f_irr < 0.1 then
    morphology := "SPH"
else if f_sph < 0.667 and f_disk > 0.667 and f_irr < 0.1 then
    morphology := "DISK"
else if f_sph > 0.667 and f_disk > 0.667 and f_irr < 0.1 then
    morphology := "DISKSPH"
else if f_sph < 0.667 and f_disk > 0.667 and f_irr > 0.1 then
    morphology := "DISKIRR"
else if f_sph < 0.667 and f_disk < 0.667 and f_irr > 0.1 then
    morphology := "IRR"
else
    morphology := "NONE"
```

Soluciones propuestas

- **Regresión** de las fracciones de expertos

Imagen → f_{sph} , f_{disk} , f_{irr} , f_{PS} , f_{Unc}

- **Clasificación** de las galaxias en las categorías morfológicas

Imagen → morfología (SPH, DISK, DISKSPH, DISKIRR, IRR O NONE)

Conjunto de datos

- 3460 imágenes RGB etiquetadas
- Galaxias distantes más masivas de tipo espiral de CANDELS

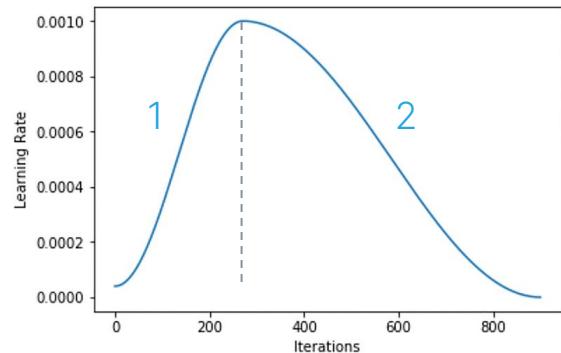


Enmascaramiento



Entrenamiento: fine_tune

- Política de un ciclo
- Dos fases:
 1. **Incrementar** learning rate hasta un valor máximo
 2. **Reducir** del valor máximo hasta el valor mínimo
- Ventajas:
 - Mayor precisión
 - Menor tiempo



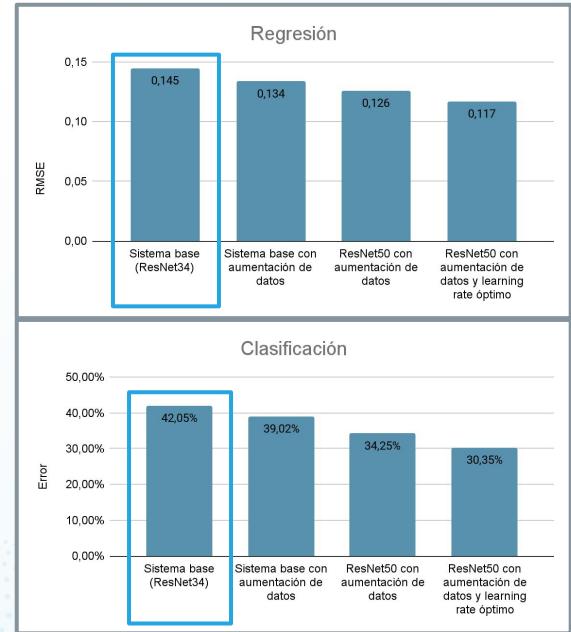
Label Smoothing

- “Suavización de las etiquetas”
- Valores de la regresión proceden de opiniones subjetivas
- Ruido de ± 0.05 a los objetivos de la regresión
- Reduce el overfitting

Diseño de la experimentación

1. Sistema base compuesto por una resnet34

- Errores relativamente bajos pero bastante mejorables
- Raíz del error cuadrático medio de 0.15
- Precisión de 58% de acierto en la clasificación morfológica

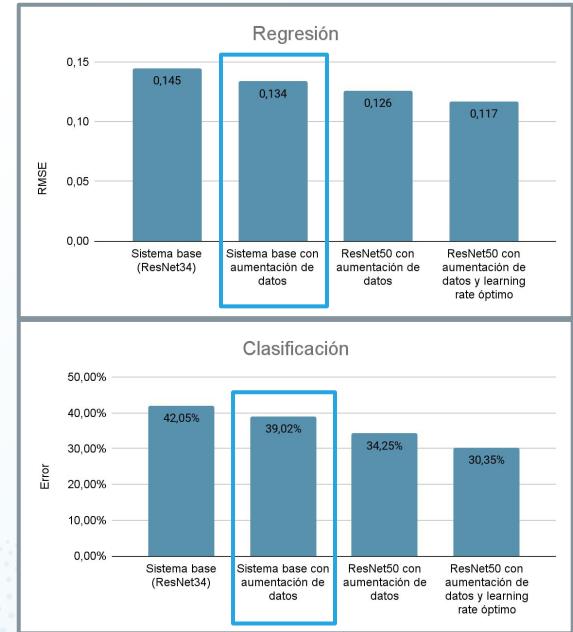


Diseño de la experimentación

2. Aumentación de datos sobre en el sistema base

- Transformaciones aplicadas:
 - Normalización
 - Ruido Gaussiano
 - Rotación en múltiplos de 90º

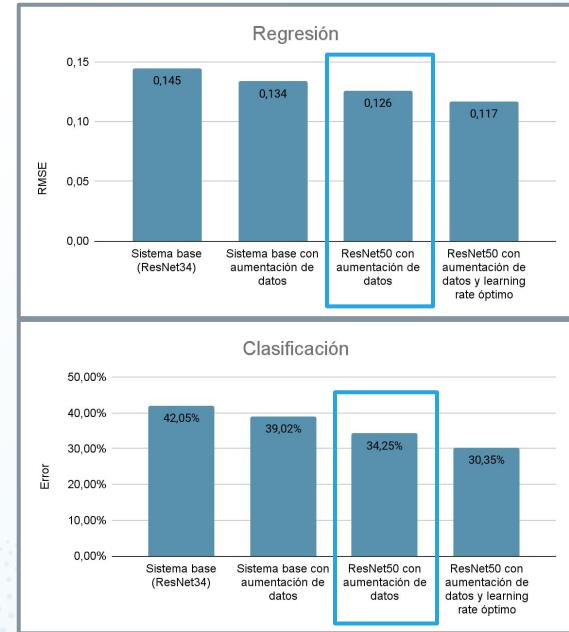
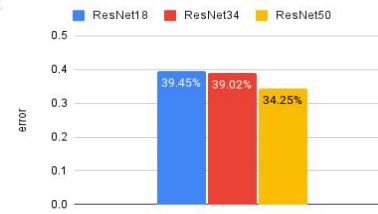
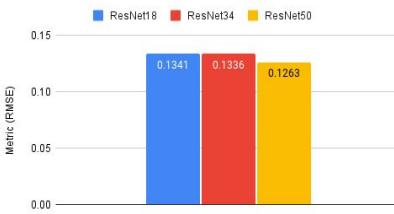
- Mejora la precisión y reduce el overfitting



Diseño de la experimentación

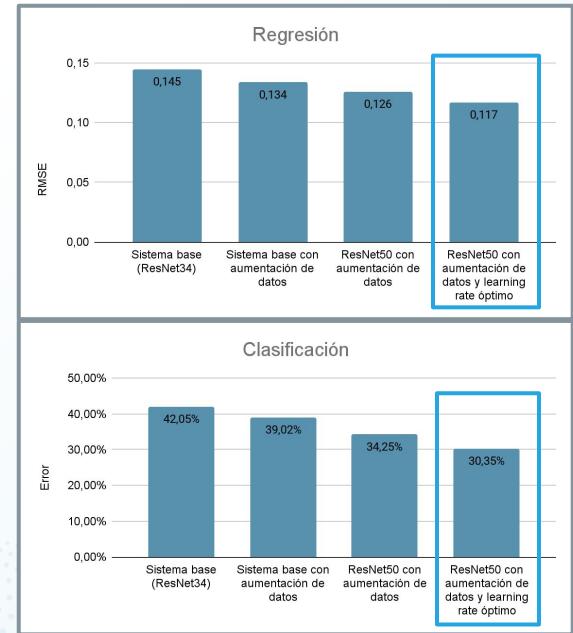
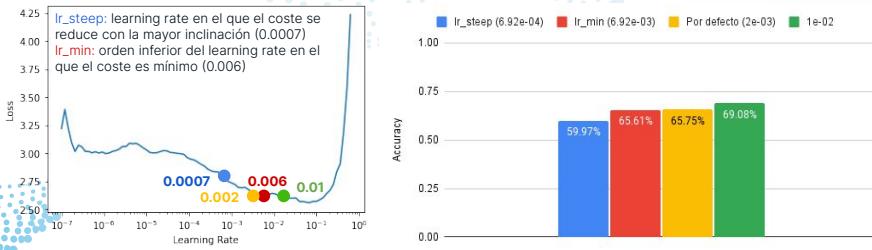
3. Análisis de distintas arquitecturas **ResNets**

- Relación entre la profundidad de la red y la precisión final de la misma
- Mejor caso: ResNet50



Diseño de la experimentación

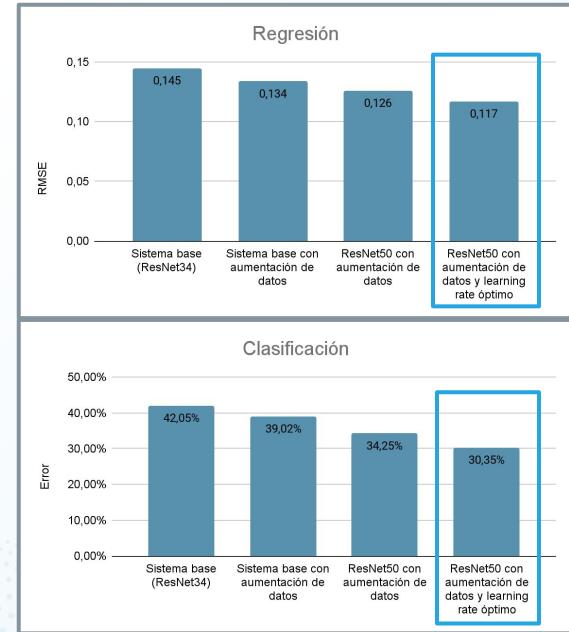
4. Búsqueda de un **learning rate** óptimo
 - Uso del buscador de learning rates de fastai
 - Comparación de diversos learning rates
 - Mejor caso: 0.01



Análisis de los mejores resultados

- **Regresión** → RMSE = 0.117
 - Huertas-Company et al. 2015 → RMSE = 0.13
 - Error medio (MAE) = 0.07
- **Clasificación** → Precisión ≈ 70%
 - Mayores confusiones en morfologías que comparten características

Confusion matrix							
		Predicted					
		DISK	DISKIRR	DISKSPPH	IRR	NONE	SPH
Actual		114	17	33	2	2	2
DISK		28	29	1	2	0	2
DISKSPPH		17	1	135	0	1	27
IRR		5	12	2	8	2	1
NONE		9	4	4	1	6	8
SPH		0	0	23	0	4	190

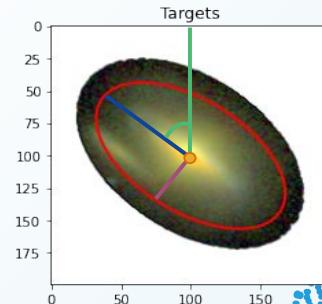


Índice

1. Objetivos
2. Contexto
3. Tecnologías
4. Clasificación morfológica de galaxias
5. Detección de bordes de las galaxias
6. Demostración práctica
7. Conclusiones

Problema

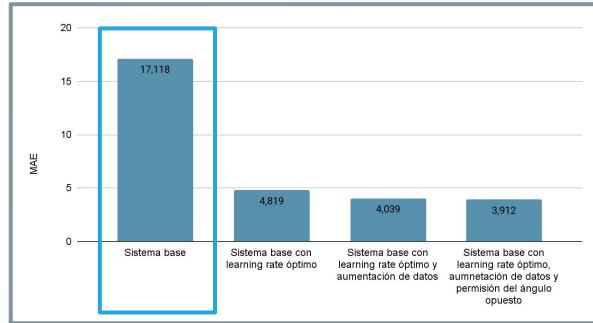
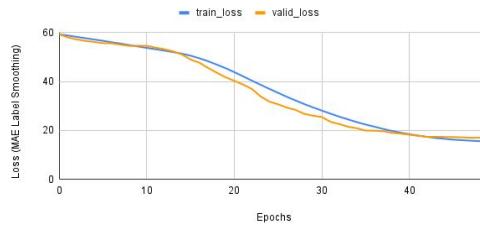
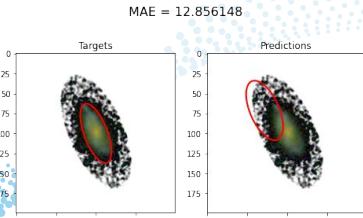
- Regresión de **5 valores** para formar una elipse del borde:
 - **Ángulo** de la galaxia con respecto al eje “y”
 - **Centro** de la galaxia en el eje “x”
 - **Centro** de la galaxia en el eje “y”
 - Longitud del **semieje menor**
 - Longitud del **semieje mayor**
- Valores subjetivos → Label Smoothing



Diseño de la experimentación

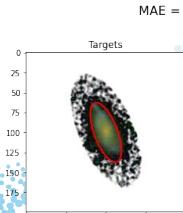
1. Sistema base compuesto por una resnet34

- Error medio relativamente alto: 17.1
- Lo que peor predice la red es el centro de la galaxia

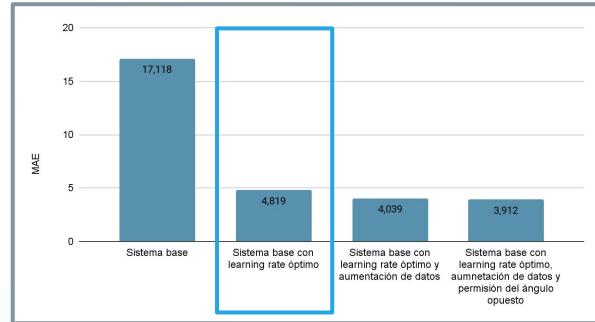
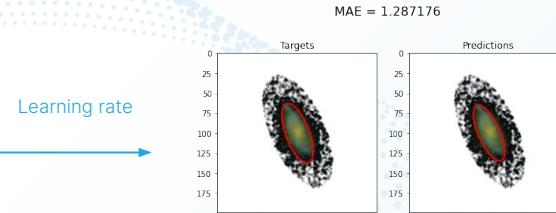


Diseño de la experimentación

2. Entrenar el sistema base con un **learning rate** óptimo
 - Se escogió 0.1 como tasa de aprendizaje
 - Mejora sustancial del error de 17.1 a 4.8



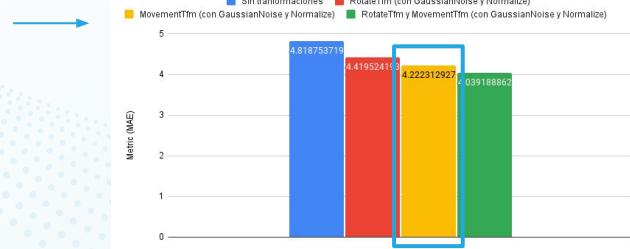
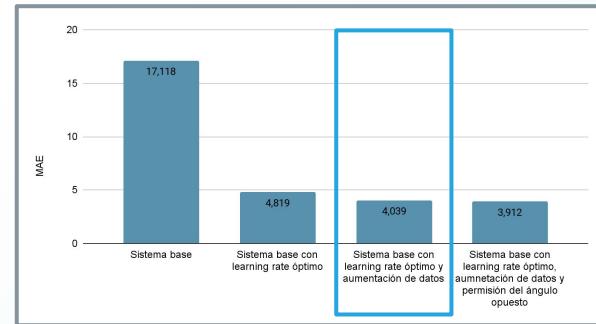
Learning rate



Diseño de la experimentación

3. Aplicar transformaciones de datos

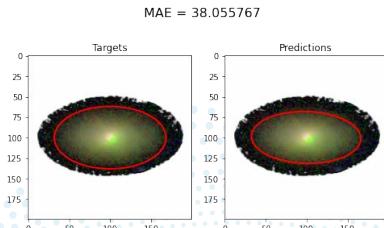
- Estas transformaciones son:
 - Normalización
 - Ruido Gaussiano
 - Rotación en múltiplos de 90º
 - Desplazamiento del centro de la galaxia
- La transformación del desplazamiento es la que mejor precisión obtuvo
- Error medio se redujo a 4.04



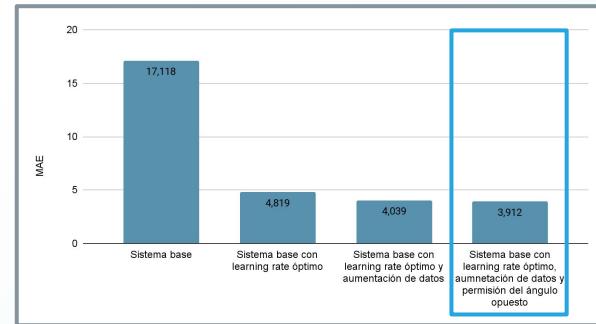
Diseño de la experimentación

4. Permitir la predicción del **ángulo opuesto**

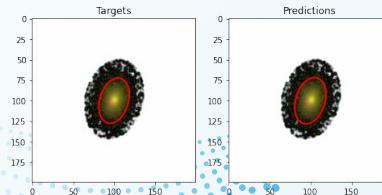
- Peor resultado:



- La red predice el ángulo opuesto al dado ($\text{ángulo} \pm 180^\circ$)
- Reducción del error medio a **3.91**



MAE = 0.431334



Diseño de la experimentación

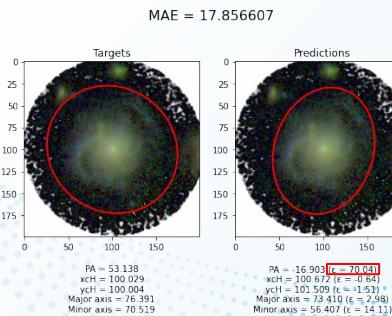
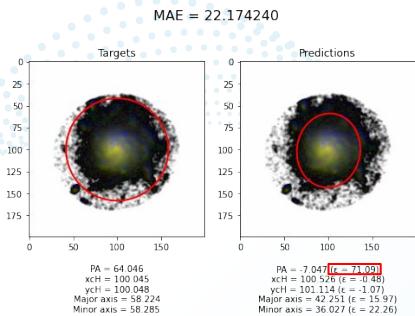
5. Construcción de Redes Neuronales Convolucionales propias

- Objetivo: aprender a construir Redes Neuronales Convolucionales propias
- Conclusión: una mayor profundidad implica una mayor capacidad de aprendizaje
- La mejor arquitectura es peor que la ResNet menos profunda



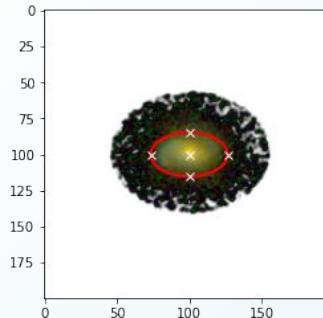
Posibles mejoras de la solución

- Penalizar la predicción del ángulo en función de la relación entre ejes
- A mayor semejanza de ejes, menor peso tendría el error del ángulo sobre el coste de la red



Posibles mejoras de la solución

- Predecir varios puntos en la imagen que puedan representar el borde de la galaxia
- No existiría el problema del ángulo de las galaxias con forma más redondeada



Índice

1. Objetivos
2. Contexto
3. Tecnologías
4. Clasificación morfológica de galaxias
5. Detección de bordes de las galaxias
6. Demostración práctica
7. Conclusiones

Índice

1. Objetivos
2. Contexto
3. Tecnologías
4. Clasificación morfológica de galaxias
5. Detección de bordes de las galaxias
6. Demostración práctica
7. Conclusiones

Conclusiones finales

- Aprendizaje de los **conocimientos** claves del Deep Learning y la astronomía
- Aprendizaje del uso de las diversas **tecnologías** de Deep Learning y Astroinformática
- Se ha conseguido un modelo de regresión y de clasificación capaz de automatizar la clasificación morfológica de galaxias
- Se ha obtenido un modelo de regresión que permite detectar automáticamente los bordes de las galaxias con una buena precisión

Trabajo Fin de Grado

Grado en Ingeniería Informática
Mención en Ingeniería del Software

Análisis y Experimentación Práctica de Frameworks Deep Learning Aplicados a la Astronomía

Autor: **Jorge Barrio Conde**

Tutor: **Benjamín Sahelices Fernández**

Cotutor: **Fernando Buitrago Alonso**

