

Anteproxecto TFG

GRAO EN CIENCIA E ENXEÑARÍA DE DATOS (GCED)

Datos da/o estudante:

Nome e apelidos	Eduardo Lucas Antolín
DNI	71959639E
Enderezo electrónico	eduardo.lucas.antolin@udc.es
Teléfono	+34 674 969 274

Título (galego):

Detección e clasificación morfolóxica de nebulosas planetarias en imaxes astrofísicas mediante aprendizaxe automática

Título (castelán):

Detección y clasificación morfológica de nebulosas planetarias en imágenes astrofísicas mediante aprendizaje automático

Título (inglés):

Automated detection and morphological classification of planetary nebulae in astrophysical images using machine learning

Tipo de proxecto:

☐ Clásico

☒ Investigación

Dirección:

— Titor(a) da FIC:

José Carlos Dafonte Vázquez

Diego Cantorna Berdullas

Breve descripción:

Uno de los mayores misterios que presenta nuestro universo son las Nebulosas Planetarias un tipo de nebulosa (nube de polvo y gas) de emisión que consiste en la envoltura estelar, expansiva y brillante, desprendida por una estrella gigante (de la misma masa que el Sol o incluso varias masas solares) en su última etapa de vida, cuando se encuentra enormemente expandida y fría, como consecuencia de las inestabilidades que sufre la estrella.

Durante muchos años, en los estudios teórico-interpretativos sobre nebulosas planetarias se asumía que las nebulosas tenían simetría esférica, aunque en la actualidad resulta cada vez más evidente que no es así y que tanto la morfología y la cinemática de las nebulosas es un amplio y prometedor campo de estudio que contiene gran cantidad de información sobre la estrella anterior a la nebulosa, la formación de esta y su posterior evolución.

La realización de un catálogo morfológico de nebulosas planetarias puede ser un proceso lento y meticuloso, además de ser propenso a una cierta variabilidad ya que dependiendo del observador la clasificación puede discernir. Estos catálogos son de gran importancia ya que sirven como una base esencial para futuras investigaciones, permitiendo la realización de cálculos hidrodinámicos y estudios sobre los factores que determinan la forma de las nebulosas, por lo que sería conveniente obtener tantas imágenes de nebulosas planetarias como fuera posible y tenerlas catalogadas según ciertos criterios objetivos.

En la era del *Big Data*, la automatización se ha convertido en una herramienta crucial para agilizar procesos que tradicionalmente requerían de mucho tiempo y esfuerzo. En este contexto, las técnicas de *Machine Learning* pueden desempeñar un papel fundamental. Al automatizar el proceso de segmentación y clasificación de nebulosas planetarias, no solo se puede acelerar la creación de estos valiosos catálogos, sino que también se puede mejorar su precisión.

Objetivos concretos:

Objetivos principales:

El objetivo principal de este proyecto es desarrollar una herramienta que, gracias a un modelo de Machine Learning, sea capaz de detectar con alta sensibilidad la morfología de nebulosas planetarias en imágenes espaciales tomadas por telescopios profesionales (como el telescopio español IAC80, situado en el Observatorio del Teide, Tenerife), pueda automatizar el proceso de segmentar (o al menos que sirva como herramienta auxiliar), y si fuera posible clasificar de manera robusta, siguiendo un estándar de clasificación no demasiado complejo.

Objetivos específicos:

1. Entrenamiento de modelos para segmentación

- 1.1. Extraer máscaras binarias (ground truth) manualmente de las imágenes astrofísicas y verificarlas junto a expertos para formar un conjunto de datos robusto.
- 1.2. Diseñar y refinar diferentes algoritmos y modelos de aprendizaje automático, tanto supervisados como no supervisados, para la segmentación de nebulosas planetarias en imágenes astrofísicas.
- 1.3. Evaluar el rendimiento de los modelos con diferentes métricas y afinarlos iterativamente hasta conseguir versiones optimizadas de estos.

2. Entrenamiento de modelos para clasificación

- 2.1. Definir una clasificación estándar a partir de las ya definidas por expertos.



- 2.2. Diseñar y refinar diferentes algoritmos y modelos de aprendizaje automático para la clasificación de las nebulosas planetarias según su morfología.
- 2.3. Conseguir un modelo que cumpla nuestro objetivo con cierta precisión.

3. Desarrollo de una herramienta de implementación

- 3.1. Desarrollar una herramienta que integre los modelos de segmentación y clasificación para facilitar su uso en futuras investigaciones. Esta herramienta deberá de ser diseñada de manera que pueda ser fácilmente utilizada por otros investigadores, sin requerir conocimientos especializados en aprendizaje automático.

Cumplir con estos objetivos podría proporcionar una ayuda significativa a todos aquellos investigadores en el campo de la astrofísica, proporcionando una herramienta para la segmentación y clasificación de nebulosas planetarias. Al hacerlo, no solo se espera reducir la carga de trabajo manual para las tareas comentadas, sino que también se espera que, con el tiempo y la ampliación de datos, pueda resultar en una herramienta que mejore tanto la eficiencia como la precisión actual.

Método de trabajo:

Para el desarrollo de este Trabajo de Fin de Grado (TFG), se empleará la metodología Scrum, una de las técnicas más populares dentro del marco de trabajo Ágil.

La elección de este método de trabajo se debe a su enfoque en la creación y entrega de incrementos de trabajo que aportan valor al proyecto en ciclos cortos de tiempo, conocidos como Sprints. Cada Sprint en este proyecto tendrá una duración de entrega de dos semanas, aunque, gracias a la flexibilidad de esta metodología, se podría acortar la duración de estos según las necesidades del proyecto, siempre teniendo en cuenta que esto podría suponer un aumento del estrés y con ello una disminución de la efectividad del equipo, por lo que trataremos de evitarlo.

Ya que, este proyecto se trata de un trabajo individual con supervisión de uno o varios tutores, adaptaremos esta metodología con la siguiente organización:

1. **Reuniones de actualización:** se realizará una reunión con el tutor (o tutores) del TFG cada dos semanas, coincidiendo así con la finalización de cada sprint, para actualizar el estado del proyecto y analizar el avance de este.
2. **Análisis de los Sprints:** en cada reunión, es decir al final de cada sprint, se analizará detalladamente si los objetivos propuestos del Sprint han sido cumplidos y reflexionar sobre posibles mejoras para el desempeño del siguiente Sprint.
3. **Definición de Sprints:** al final de cada reunión, se realizará una fase en la que definiremos los objetivos a conseguir en el siguiente Sprint y haremos una breve lluvia de ideas sobre como poder llevar a cabo las tareas que nos harían conseguir nuestros objetivos específicos.

Fases principais do traballo:

Para la realización de este Trabajo de Fin de Grado, hemos dividido el desarrollo en varias fases esenciales, detalladas a continuación:

1. **Fase de investigación y contextualización del proyecto:** en esta fase inicial, se llevará a cabo un trabajo de investigación sobre el trabajo del proyecto. Esto se realizará leyendo literatura relevante, comprendiendo, en la máxima medida de lo posible, las técnicas actuales de segmentación y clasificación de nebulosas planetarias y familiarizándose con los datos y herramientas que serán utilizadas en el proyecto.



2. **Fase de preparación de datos:** una vez se esté familiarizado con los datos, se comenzará extrayendo máscaras binarias (ground truth) manualmente de las imágenes astrofísicas y se verificarán junto a expertos para formar un conjunto de datos robustos. Esta fase inicialmente tendrá el objetivo de obtener un conjunto de datos lo suficientemente amplio como para poder empezar a trabajar en la automatización de la segmentación, más adelante se irá ampliando el conjunto de datos creando más máscaras binarias en paralelo mientras trabajamos en otras fases.
3. **Fase de diseño y entrenamiento de modelos para segmentación:** en este punto, se diseñarán y refinarán diferentes modelos y algoritmos de aprendizaje automático para la segmentación de nebulosas planetarias en imágenes astrofísicas. Se evaluará el rendimiento de los modelos con diferentes métricas y serán afinados iterativamente hasta conseguir una versión lo más optimizada posible de los mismos.
4. **Fase de validación y experimentación para segmentación:** se validarán la precisión de los modelos de segmentación contra el ground truth y se realizarán experimentos para evaluar su rendimiento en diferentes condiciones.
5. **Fase de diseño y entrenamiento de modelos de clasificación:** en este punto, habrá que definir una clasificación morfológica estándar de las nebulosas planetarias a partir de las que ya están definidas por los expertos. Más adelante, se seguirá el mismo procedimiento que fue seguido para los modelos de segmentación.
6. **Fase de validación para clasificación:** se validarán la precisión de los modelos de clasificación contra el estándar definido y se escogerá el modelo óptimo para resolver el problema.
7. **Fase de desarrollo de herramientas:** Finalmente, se desarrollará una herramienta que integre los modelos de segmentación y clasificación con mejor desempeño.

Material, medios e recursos necesarios:

Las imágenes astrofísicas que vamos a utilizar como conjunto de datos pertenecen al Catálogo Morfológico de nebulosas planetarias Galácticas del hemisferio norte, confeccionado por investigadores del Instituto de Astrofísica de Canarias (IAC) y del Observatorio de Bolonia (Italia). Estas imágenes fueron tomadas utilizando dos telescopios de los Observatorios del IAC y fueron clasificadas según su morfología manualmente por los investigadores del centro. En este catálogo se pueden encontrar, en un principio, 589 imágenes, referentes a 243 objetos diferentes (nebulosas planetarias), en formato FITS. El formato FITS (Flexible Image Transport System) es un formato de archivo estándar muy utilizado en el campo de la astronomía para archivar datos astronómicos como imágenes de telescopios espaciales.

Para el desarrollo del proyecto se utilizará el lenguaje de programación Python el cual es un lenguaje fácil de entender y utilizar, al mismo tiempo que muy versátil y perfecto para este proyecto, debido a su gran variedad de bibliotecas en distintos ámbitos.

Entre las librerías que se utilizarán en este proyecto se encuentran algunas de código abierto las cuales están especializadas en los campos del aprendizaje automático y del aprendizaje profundo como pueden ser Tensorflow, Scikit-Learn o Pytorch. Además, se utilizarán otras librerías complementarias como ImageIO y AstroPy, para poder trabajar fácil y eficientemente con las imágenes de nuestro conjunto de datos.

Por último, dado el gran coste computacional que requiere trabajar con imágenes de formato FITS con excelente resolución espacial y el intensivo cálculo que requieren muchos modelos de aprendizaje automático y profundo, se va a utilizar una infraestructura computacional potente. En este caso, se

intentará obtener acceso a algún centro de supercomputación de alto rendimiento como puede ser el CESGA Finisterrae III o se realizará en algún equipo más sencillo que cuente con GPUs (Graphics Processing Units) que proporcione la capacidad de procesamiento necesaria para manejar la complejidad y el volumen de los datos.

Propiedade intelectual do traballo:

O Regulamento dos Traballos de Fin de Grao da Facultade de Informática da Coruña (aprobado pola Xunta de Centro o 18 de xullo de 2022) establece na sección 4, en relación aos dereitos derivados da propiedade intelectual dos traballos, o seguinte:

4.1. No caso dos traballos desenvolvidos en colaboración cunha entidade externa, a titularidade dos dereitos de propiedade e explotación dos resultados, se for o caso, rexerase polo establecido na relación contractual entre a/o estudante e a entidade externa. Neste caso, quen exerza a dirección académica non será titular dos dereitos de propiedade intelectual, salvo que se establecer doutra maneira nun documento asinado pola/o estudante, o profesorado encargado da dirección e un/ha representante da entidade externa.

4.2. No caso dos traballos desenvolvidos no ámbito do centro, a titularidade dos dereitos de propiedade intelectual, se for o caso, corresponderá á/ao estudante segundo queda recollido no apartado h) do artigo 8 do Real Decreto 1791/2010 do 30 de decembro, salvo que se establecer doutra maneira no anteproxecto asinado pola/o estudante e o profesorado encargado da dirección do TFG.

Os resultados xerados durante a realización do TFG poderán ser utilizados para a docencia na UDC, sempre con autorización e mención explícita de quen ostente a súa autoría.

Indique a continuación se o traballo se realiza en colaboración cunha entidade externa ou no ámbito do centro, e neste último caso, o acordo sobre os dereitos derivados da propiedade intelectual do traballo.

O traballo realízase en colaboración cunha entidade externa:

☐ Si; **Nome da empresa:**

☒ Non

Se o traballo non se realiza en colaboración cunha entidade externa, indique se os dereitos derivados da propiedade intelectual son compartidos entre a/o estudante e as/os directores:

☒ Si

☐ Non

En A Coruña, a 17 de abril de 2024

Asinado:

Estudiante

Titoras/es e Directoras/es