



UNIVERSIDAD INTERNACIONAL DE VALENCIA

MÁSTER EN BIG DATA Y DATA SCIENCE

Trabajo Fin de Máster

DETECCIÓN DE FORMAS AISLADAS DE ASTEROIDES EN IMÁGENES ASTRONÓMICAS

Autor:

Laura Molinos Mayo

78811352N

Tutor:

René Duffard

Instituto de Astrofísica de Andalucía

Cotutor:

Rafael Morales

Instituto de Astrofísica de Andalucía

Octubre 2022

Índice

Resumen	6
1 Introducción y objetivos	8
2 Marco teórico	11
2.1 Asteroides	11
2.2 Fotometría de asteroides	16
2.3 Imágenes astronómicas	18
2.3.1 Detector CCD	21
2.3.2 Reducción de imágenes astronómicas	23
3 Detección de formas aisladas de asteroides	29
3.1 Descripción del proyecto	29
3.2 Datos de entrenamiento. Varuna	30
3.3 Preprocesamiento	34
3.4 El algoritmo. CNN	37
3.5 Compilado y entrenamiento	45
3.6 Comprobación y resultados	49
4 Conclusiones y trabajo futuro	64
5 Referencias	66
Notas	77
Apéndice	82

Índice de figuras

1	Las órbitas de los asteroides NEOs según su tipo (Observatoire Solaire, 2019)	12
2	Excentricidad frente al semieje mayor de la órbita de las principales familias del cinturón de asteroides (Elkins, 2010)	14
3	Curva de luz del asteroide del cinturón principal 201- Penelope (6/10/2006) y modelo tridimensional obtenido de varias curvas de luz (Observatorio Astronómico Los Molinos, s.f.; Astronomical Institute of the Charles University, s.f.)	17
4	Nebulosa de Cangrejo observada bajo diferentes longitudes del espectro (Caldú, s.f.)	19
5	Ejemplo de imagen astronómica de alta resolución a color. Los Pilares de la Creación en la nebulosa de Águila, tomada por Hubble en 1995 (National Geographic, 2022)	20
6	Ejemplo de imagen para fotometría de un campo denso de estrellas con un filtro en la región visible del espectro (filtro V) (Berry, s.f.)	20
7	Efecto de <i>blooming</i> en una imagen tomada con CCD (Astronomics, s.f.)	22
8	Ejemplo de fuente en una imagen astronómica (Doreste, 2002)	23
9	Ejemplo de reducción de imágenes astronómicas de la galaxia NGC 6118 (Ruiz, 2010)	27
10	Zonas de formación estelar y estrellas nuevas (Ruiz, 2010)	27
11	Escala de Pickering del <i>seeing</i> atmosférico (Bravo, 2019)	28
12	Ejemplo de <i>croppie</i> obtenido por el algoritmo <i>m2</i> . Se trata de un asteroide aislado (dentro del cuadro blanco) cercano a una fuente.	31
13	Ejemplos de <i>croppies</i> sin asteroide	32
14	Ejemplos de <i>croppies</i> con señales demasiado pequeñas o dispersas	32
15	Ejemplos de <i>croppies</i> de muy baja calidad	32

16	Ejemplos de <i>croppies</i> de baja calidad	33
17	Ejemplos de <i>croppies</i> contaminados	33
18	Ejemplos de <i>croppies</i> etiquetados.	34
19	<i>Croppies</i> de entrenamiento preprocesados correspondientes a la figura 18 . .	35
20	Operación de convolución de un <i>kernel</i> sobre un campo receptivo de la imagen (IBM, 2020)	38
21	Dimensión de la matriz de salida según el <i>stride</i> seleccionado (Colomer y Muñoz, s.f.)	39
22	Esquema de <i>zero padding</i> para conservar el tamaño de imagen (Dharmaraj, 2021)	40
23	Esquema de <i>maxpooling</i> 2x2 (Computer Science Wiki, 2018)	41
24	Código de la arquitectura de la CNN	44
25	Resumen de la arquitectura de la CNN con el número de parámetros	44
26	Esquema de los pasos de la CNN junto a las dimensiones	45
27	Compilado y entrenamiento de la CNN	48
28	Función de pérdida de los conjuntos de entrenamiento y validación durante el entrenamiento	49
29	<i>Accuracy</i> de los conjuntos de entrenamiento y validación durante el entre- namiento	49
30	Resultados del clasificador sobre el conjunto de <i>test</i> de Varuna	51
31	Falsos negativos de <i>m2</i> recuperados por el clasificador	52
32	Ejemplos <i>croppies</i> de <i>test</i> de 2002_ms4, originales y preprocesados	53
33	Falsos positivos del clasificador en el conjunto de <i>test</i> de 2002_ms4	54

34	Mezclas solapadas indistinguibles con fuentes aisladas del conjunto de Varuna. La mezcla 34a es muy similar a la aislada 34b y la mezcla 34c similar a la aislada 34d	54
35	Falsos negativos del clasificador en el 30 de junio de 2017	55
36	Verdaderos positivos del clasificador en el 30 de junio de 2017	56
37	Falso negativo y verdaderos positivos del clasificador en el 5 de julio de 2013	57
38	Falsos negativos de $m2$ y del clasificador	58
39	Falsos negativos de $m2$ recuperados por el clasificador	58
40	Ejemplos de las 70 mezclas contaminadas incluidas en el conjunto ampliado de <i>test</i> de 2002_ms4	60
41	Falsos positivos del clasificador en las 70 nuevas muestras	60
42	Secuencia del 3 de diciembre de 2013	82
43	Secuencia del 31 de enero de 2011	83

Índice de tablas

1	Métricas del clasificador sobre el conjunto de <i>test</i> de 2002_ms4	59
2	Métricas del clasificador sobre el conjunto de <i>test</i> de 2002_ms4 ampliado	61
3	Detecciones positivas y negativas del <i>dataset</i> de 2002_ms4 reales y obtenidas por ambos algoritmos	63

Resumen

En el grupo de ciencias planetarias del Instituto de Astrofísica de Andalucía (IAA) estudian las propiedades de los asteroides a partir de imágenes astronómicas. Para ello disponen de bases de datos de imágenes de asteroides y transneptunianos obtenidas durante las últimas décadas. Actualmente cuentan con la ayuda de un algoritmo que preprocesa las imágenes astronómicas al localizar al asteroide, recortar la región en la que se encuentra, segmentar las fuentes por colores y clasificar las imágenes según su validez. En muchas ocasiones el asteroide se encuentra superpuesto con otra fuente, normalmente una estrella, lo que impide la obtención de parámetros físicos del asteroide de manera independiente. El algoritmo vigente identifica las dos fuentes como elementos separados cuando se hallan cerca, pero tiene problemas en distinguirlas cuando ya están demasiado próximas o solapadas, considerando al conjunto de ambas fuentes como el asteroide y por tanto tomándola como una imagen válida. Esto conlleva la necesidad de discernir imágenes con el asteroide aislado y con superposición de fuentes de manera más efectiva. El propósito de este proyecto es crear un clasificador que detecte las fuentes aisladas y mezcladas para poder descartar imágenes que el algoritmo actual clasifica como válidas o recuperar aquellas que clasifica como erróneas. De esta manera se automatiza el procesamiento y se maximiza el aprovechamiento de las imágenes astronómicas.

Palabras clave: asteroides, imágenes astronómicas, clasificador, *machine learning*

Abstract

In the planetary science group of the Institute of Astrophysics of Andalusia they study the properties of asteroids from astronomical images. For that purpose they have databases of images of asteroids and trans-Neptunians obtained during the last decades. They currently have the help of an algorithm that preprocesses the astronomical images by locating the asteroid, cropping the region in which it is placed, segmenting the sources by color and classifying the images according to their validity. On many occasions, the asteroid is superimposed with another source, usually a star, which prevents obtaining physical parameters of the asteroid independently. The current algorithm identifies the two sources as separate elements when they are nearby, but has problems distinguishing them when they are already too close or overlapping, considering the set of both sources as the asteroid itself and therefore taking it as a valid image. This leads to the need to discern images with isolated asteroids and overlapping sources more effectively. The aim of this project is to create a classifier that detects isolated and mixed sources in order to discard images that the current algorithm classifies as valid or recover those that it classifies as erroneous. In this way, processing of astronomical images is automated and their use is maximized.

Keywords: asteroids, astronomical images, classifier, machine learning

1 Introducción y objetivos

La inmensa mayoría del conocimiento actual sobre los asteroides y otros astros en general se obtiene a través de imágenes astronómicas tomadas desde observatorios terrestres. Los astros se mueven sobre la bóveda celeste, es decir, sobre la esfera imaginaria concéntrica con la Tierra ([Sociedad Española de Astronomía, s.f.c](#)) por lo que en las imágenes pueden aparecer varios astros superpuestos, denominados fuentes en el campo de la fotometría. En el caso de los asteroides, lo habitual es que se superpongan con estrellas, mucho más alejadas de la Tierra que ellos. Es decir, los asteroides desde la perspectiva terrestre se mueven sobre un fondo de estrellas. Como interesa obtener las propiedades del asteroide, es conveniente que se encuentre aislado y se requiere por tanto la distinción de las imágenes con asteroides aislados y con superposición de fuentes. Dicha distinción es sencilla para el ojo humano, pero esto conlleva invertir demasiado tiempo y empeño, dados los tamaños de los *datasets* de imágenes. El propósito principal de este proyecto es el de crear un clasificador de aprendizaje automático supervisado que permita automatizar la tarea y por ende agilizar el procesamiento de las imágenes, previo a su análisis.

El clasificador está concebido para incorporarse en un futuro al conjunto de algoritmos actualmente en funcionamiento: *om*, *m2* y *henosis* ([Morales et. al, 2022](#)). Estos algoritmos permiten manejar el formato FITS, la detección de fuentes, el cálculo de centroides, la calibración de imágenes, astrometría de alta precisión y fotometría. En concreto *m2* (*Massive prOcessing Of aStronomical imagEs* version 2), localiza la posición del asteroide en la imagen, lo enmarca en un rectángulo blanco y recorta la región de la imagen donde se encuentra, convirtiéndolo al formato FITS o *png* (el recuadro blanco solo aparece en *png*). Además, asocia a cada una de las fuentes presentes en la imagen un color diferente, es decir, aquellos píxeles que han recibido señal presentan un color distinto al blanco y aquellos píxeles sin señal mantienen como es lógico el color negro. A los recortes *png* que devuelve el algoritmo se les llamará de ahora en adelante *croppies* y serán los utilizados para la creación del clasificador.

El algoritmo *m2* además realiza una clasificación inicial de los *croppies* en dos grupos: detecciones positivas y negativas. Las detecciones positivas son aquellas en las que el algoritmo identifica correctamente al asteroide y por tanto son las válidas en principio para el estudio posterior.

Las detecciones negativas, por el contrario, carecen de asteroide o este no es identificado por el algoritmo, por lo que no serían válidas para el análisis posterior. Sin embargo, como es lógico, el algoritmo no posee un rendimiento perfecto, por lo que existen detecciones válidas que terminan en el grupo de las negativas (falsos negativos) y detecciones erróneas en el grupo de las positivas (falsos positivos). Los falsos positivos pueden suprimirse en pasos posteriores pero los falsos negativos no, pues ya han sido descartados. Es ahí donde entra el clasificador del presente proyecto, por lo que los objetivos específicos, dentro del fin principal son los siguientes:

- Permitir descartar falsos positivos de $m2$ debidos a fuentes superpuestas, es decir, imágenes que son consideradas como detecciones positivas pero en realidad se tratan de mezclas.
- Permitir descartar falsos positivos de $m2$ debidos a poca calidad o contaminación. En los *datasets* existen multitud de imágenes de baja calidad y otras muchas de calidad totalmente mermada, debido a condiciones atmosféricas poco favorables durante la medición (nubes, turbulencias...) o al funcionamiento inefectivo de alguno de los componentes de medida (telescopio, filtros, detector...). Muchos *croppies* de mala calidad son descartados por $m2$ pero algunos logran pasar el filtro y se sitúan entre las detecciones positivas. Por ello es importante poder complementar al algoritmo actual también en este sentido.
- Permitir recuperar falsos negativos, es decir, imágenes que siendo válidas son descartadas por el algoritmo $m2$. En este grupo se incluyen asteroides aislados con una calidad aceptable de imagen, aptos para el análisis posterior.

De los objetivos anteriores se deduce que la meta final del proyecto es maximizar el número de muestras aprovechables de un *dataset* de imágenes astronómicas de asteroides para optimizar el estudio sobre estos cuerpos menores.

A continuación, se enumeran las secciones y subsecciones de este trabajo junto a una breve descripción:

- *Marco teórico.* En él se exponen de manera introductoria los conceptos teóricos relacionados con el proyecto.

- *Asteroides*. Se corresponde con una introducción a los tipos de asteroides y sus cualidades principales.
 - *Fotometría de asteroides*. Se comenta brevemente el estudio del brillo de los asteroides y otras propiedades a partir de imágenes astronómicas.
 - *Imágenes astronómicas*. Se sintetiza todo lo relativo a las imágenes astronómicas: equipo, obtención, procesamiento, etc.
- *Detección de formas aisladas de asteroides*. La sección relativa al proyecto propiamente dicho.
 - *Descripción del proyecto*. Como el nombre indica se trata de una breve síntesis del proyecto realizado.
 - *Datos de entrenamiento*. Varuna. Se explican las características del *dataset* del TNO Varuna y los tipos de *croppies* que deben descartarse.
 - *Preprocesamiento*. Se indica el preprocesamiento aplicado a los *croppies* para poder entrenar al clasificador.
 - *El algoritmo*. *CNN*. Se presentan los fundamentos teóricos de las CNN (*Convolutional Neural Networks*) y la arquitectura del clasificador.
 - *Compilado y entrenamiento*. Se describe lo relativo a las etapas de compilado y entrenamiento del algoritmo de clasificación.
 - *Comprobación y resultados*. Se exponen las métricas de rendimiento del clasificador y una valoración de sus aciertos y errores.
- *Conclusiones y trabajo futuro*. Se extraen los obstáculos y dificultades del proyecto y posibles mejoras.
- *Referencias*. Las fuentes externas consultadas para la redacción del trabajo.
- *Notas y apéndice*. Se tratan de anotaciones para aclarar conceptos e información complementaria al cuerpo principal de la memoria.

2 Marco teórico

2.1 Asteroides

Los asteroides son fundamentales para entender la composición de los planetas y obtener datos sobre la formación del Sistema Solar. Aunque su variedad permite la existencia de varias clasificaciones es habitual separarlos en los siguientes grupos: NEOs, cinturón principal, troyanos, centauros y transneptunianos ([Naves y Campàs, 2020](#)).

Los NEOs (*Near Earth Objects*) hacen referencia a los asteroides (y cometas) con perihelio¹ menor a 1,3 unidades astronómicas (UA)² y como su nombre indica son los más cercanos a la Tierra. El primer NEO descubierto fue 433-Eros, en 1898, avistado en la misma noche por Gustav Witt desde Berlín y Auguste Charlois desde Niza, con un diámetro de 16,84 km ([Mou y Webster, s.f.a](#)). Más tarde, en 1932, Karl Reinmuth descubrió desde Heidelberg el primer NEO que cruza la órbita terrestre, 1862-Apolo. Apolo tiene un diámetro más pequeño, 1,5 km, pero sigue siendo más grande que el 99 % de los asteroides conocidos ([Mou y Webster, s.f.b](#)).

Los NEOs a su vez se dividen en los siguientes grupos ([Burbine, 2016](#); [Observatoire Solaire, 2019](#)):

- Atira. Aquellos cuya órbita tiene afelio y semieje mayor menores que 1 UA y está totalmente contenida dentro de la órbita de la Tierra (y por lo tanto no la cruza). Deben su nombre al primer asteroide descubierto de este tipo, 163693-Atira.
- Amor. Aquellos con órbita contenida entre la Tierra y Marte que pueden cruzar la órbita marciana pero no la terrestre. Son los más fáciles de visualizar y deben su nombre a 1221-Amor. Eros pertenece también a este tipo.
- Atenas. Aquellos que poseen un semieje mayor más pequeño que el de la Tierra y pueden cruzar la órbita terrestre. Tienen un periodo menor a un año y deben su nombre a 2062-Atenas. Son solo visibles en los momentos del atardecer y amanecer.
- Apolo. Aquellos que poseen un semieje mayor más grande que el de la Tierra y pueden cruzar la órbita terrestre. Tienen periodo mayor a un año y deben su nombre a 1862-Apolo.

Otros tipos recientemente propuestos son los Vatira, cuya órbita estaría totalmente contenida en la de Venus y los Vulcanoides, cuya órbita estaría contenida en la de Mercurio.

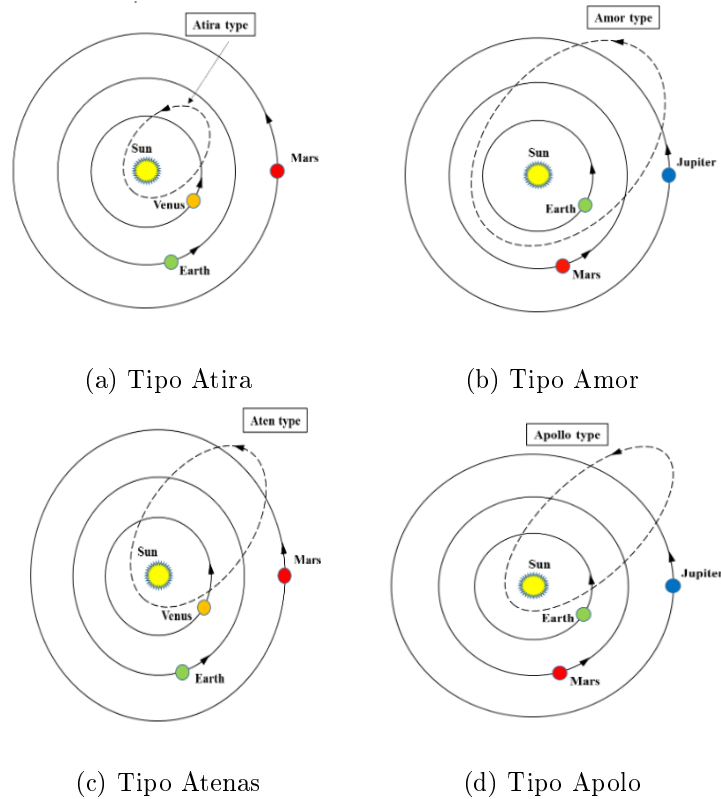


Figura 1: Las órbitas de los asteroides NEOs según su tipo ([Observatoire Solaire, 2019](#))

El cinturón principal de asteroides (*main belt*) se encuentra entre Marte y Júpiter y contiene la inmensa mayoría de los asteroides del Sistema Solar, estimándose su cantidad en millones. Su estudio es de gran importancia, pues según la teoría más extendida se trata de restos rocosos del Sistema Solar primigenio que no pudieron unirse a los planetas en formación debido a la gran influencia gravitatoria de Júpiter ([Swinburne University of Technology, s.f.c](#)). La mayor concentración de asteroides se encuentra entre las 2,12 y 3,3 UA del Sol y existen zonas con poca densidad de asteroides, conocidas como *Kirkwood Gaps*³. A pesar de su cantidad, si se fusionasen en un solo cuerpo, este no superaría el tamaño de la Luna por lo que la mayoría tienen tamaños pequeños o incluso diminutos, pero algunos alcanzan varios cientos de metros como Vesta, Pallas e Hygiea. El más grande es Ceres⁴, con 950 km de diámetro, que supone prácticamente un tercio de la masa del cinturón ([Tillman, 2017](#)).

Solo los masivos poseen formas esféricas mientras que los demás presentan formas bastante irregulares, como el caso de 216-Kleopatra, que recuerda a un hueso ([Astronomy Picture](#)

of the Day, 2000). Están compuestos principalmente de roca y metales (hierro y níquel los más abundantes). Los que más se alejan del Sol también pueden contener hielo dadas las bajas temperaturas y se convierten en cometas en su perihelio⁵.

El cinturón principal se clasifica a su vez en subgrupos en función de la composición (Tillman, 2017):

- Tipo C. Compuestos principalmente por carbono, con superficies de color casi negro. Suponen el grupo mayoritario en el cinturón y están más concentrados en la región externa.
- Tipo S. Contienen hierro, níquel y silicatos de magnesio. Más brillantes que los del tipo C, son el segundo grupo más abundante y se concentran en la región interna del cinturón.
- Tipo M. Compuestos en gran medida por hierro y níquel, son muy brillantes y se sitúan en tercer puesto en cuanto a cantidad. Pueblan la región media del cinturón.

Existen tipos menos abundantes como el tipo A, tipo D, tipo E, etc. También es habitual agruparlos según familias de asteroides que comparten características similares. A continuación se enumeran las más relevantes, aunque hay muchas más (Michel, Demeo y Bottke, 2016; Elkins, 2010):

- Hungaria. Formada por asteroides que se encuentran en la región más interna del cinturón (entre las 1,78 y 2 UA) con órbitas con baja excentricidad (casi circulares). Pueden mantener una resonancia⁶ con Júpiter de tipo 9:2 o con Marte de tipo 3:2. Deben su nombre a 434-Hungaria.
- Flora. Formada por asteroides de tipo S. Al igual que los Hungaria se encuentran en la región interna (de 2,1 a 2,3 UA) pero tienen excentricidades mucho más altas (órbitas elípticas). Los miembros de este grupo que están en resonancia con Saturno (llamada resonancia ν_6) suelen ser expulsados del cinturón principal al cruzar la órbita de Marte, alcanzando la zona interna del Sistema Solar, por lo que muchos NEOs se tratan de asteroides Flora eyectados.
- Koronis. Formada por asteroides en torno a 3 UA. Deben su nombre a 458-Koronis. Entre ellos destaca Ida⁷, de tipo S y de 58x23 *km* de tamaño, al tener satélite, Dactyl.

- Cybeles. Contiene asteroides que orbitan entre los 3,3 y 3,5 UA, con excentricidades relativamente bajas. Todos ellos proceden de un asteroide más masivo que ha sufrido fragmentación. Deben su nombre a 65-Cybele y sobresale 121-Hermione, que cuenta con un satélite al igual que Ida. En su mayoría se encuentran en resonancia 7:4 con Júpiter.
- Hilda. Formada por asteroides en la región más externa del cinturón (de 3,9 a 4,2 UA) en resonancia 3:2 con Júpiter. Deben su nombre a 153-Hilda.

Otras familias son Phocaea, Alinda, Pallas, Griqua, Thule, etc. Más allá de la familia Hilda existe una región (entre 4,2 y 5,05 UA) de menor concentración de asteroides, “limpiada” por la influencia de Júpiter, que los separa del siguiente gran grupo: los troyanos.

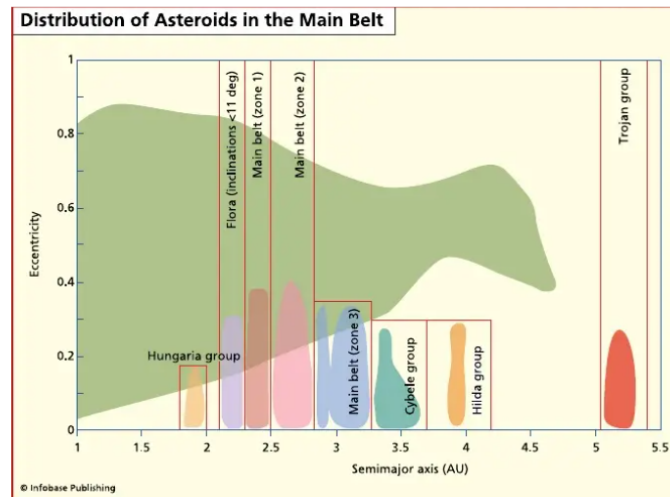


Figura 2: Excentricidad frente al semieje mayor de la órbita de las principales familias del cinturón de asteroides (Elkins, 2010)

Los troyanos comparten órbita con Júpiter, situándose a 60° por delante y por detrás del planeta, en los denominados puntos de Lagrange⁸. La mayoría de los asteroides cercanos a Júpiter son troyanos, aunque algunos de ellos tienen órbitas realmente inestables y son en ocasiones expulsados hacia el interior del Sistema Solar. El primer troyano en ser descubierto fue 624-Hektor y se trata del más grande (300x150 km).

Por su parte, los centauros son asteroides que se acercan o cruzan las órbitas de Júpiter y Saturno, algunos de ellos muy inestables. Sus órbitas pueden ser tan excéntricas como para llegar a la región interna del Sistema Solar o a la más externa.

El primer centauro descubierto fue 944-Hidalgo y el segundo 2060-Chiron, que cruza las órbitas de Saturno y Urano y es realmente caótico. Actualmente es considerado también un cometa al contar con coma gaseosa en su perihelio por lo que posee otra denominación: 95 P/Chiron. Un ejemplo de órbita muy excéntrica es el de 5335-Damocles, que se aproxima a Marte en su perihelio y llega más allá de Urano en su afelio.

Por último los transneptunianos (TNOs), que como su nombre indica incluye a los asteroides que orbitan más allá de Neptuno. Los TNOs son especialmente relevantes pues al encontrarse más alejados del Sol han sufrido menos procesos químicos y por lo tanto su composición es más fiel a la nebulosa primigenia que formó el Sistema Solar ([Fernández-Valenzuela et al, 2019](#)). Desde 1993 se descubren nuevos TNOs cada mes. Debido a su gran distancia al Sol es habitual que contengan hielo, aunque no de agua sino de otras sustancias (salvo casos puntuales). Dentro de ellos se encuentra el cinturón de Kuiper (entre los 30 y 49 UA), aunque existen otros TNOs independientes, normalmente con órbitas caóticas y fácilmente perturbadas, pudiendo convertirse en centauros con el tiempo. Se estima que la masa total del cinturón de Kuiper es el 20 % de la masa terrestre, cien veces más que el cinturón principal. Los asteroides del cinturón, también denominados KBOs (*Kuiper Belt Objects*) se pueden clasificar en subgrupos, en los que los más relevantes son plutinos, *cubewanos* y *twotinos* ([Daily Minor Planet, 2011](#); [Astrosigma, s.f.](#)).

- Plutinos. Deben su nombre al TNO más famoso: Plutón⁹. Durante muchos años fue considerado un planeta hasta que en 2005 se descubrió un cuerpo más grande que él, también en el cinturón: Eris¹⁰. La IAU (*International Astronomical Union*) decidió en 2006 que Plutón y Eris son planetas enanos¹¹. Los plutinos mantienen órbitas estables en resonancia 2:3 con Neptuno. Otros miembros de este grupo son 90482-Orcus y 28978-Ixion.
- Twotinos. Poseen resonancias 1:2, también con Neptuno. Se conocen sobre una docena entre los que se encuentran 1966 *TR*₆₆ y 2002 *WC*₁₉.
- Cubewanos. También llamados KBOs clásicos, se tratan de asteroides que no mantienen ningún tipo de resonancia con Neptuno y nunca cruzan su órbita. Poseen órbitas estables casi circulares. En este grupo se incluyen, entre otros: Makemake¹² (planeta enano), 15760-Albión, 50000-Quaoar, 58534-Logos y 20000-Varuna.

Los TNOs con órbitas muy inclinadas y excéntricas y semieje mayor de más de 50 UA se incluyen en el disco disperso (*scattered-disk*). Son de los más distantes respecto al Sol por lo que contienen altos porcentajes de hielo. Se considera que sus órbitas inestables pueden ser perturbadas por los planetas gaseosos provocando su expulsión cara el Sistema Solar interno, donde se transforman en cometas periódicos. Eris es un miembro de este grupo.

La manera idónea de estudiar los asteroides, y los cuerpos en general, es a través de sondas espaciales que alcancen sus cercanías, como es el caso de la sonda Lucy de la NASA, pero esto solo ocurre para un conjunto pequeño de cuerpos menores. La mayoría son estudiados por vías más indirectas, desde la Tierra, a través de meteoritos que llegan a la superficie o mediante imágenes astronómicas desde los observatorios. Si el propósito es medir la localización del asteroide y caracterizar su movimiento en en plano celeste, es decir, conocer su órbita, el análisis entra dentro de la astrometría (Endl, 2014; Synider y Marschall, 2021). Si por el contrario, interesan más las propiedades de magnitud¹³ o brillo y color se habla de fotometría (AstroMía, s.f.). En el siguiente apartado se introduce de manera preliminar la fotometría de asteroides, donde es fundamental extraer las propiedades del asteroide de manera independiente y por tanto se requiere identificar cuando está superpuesto con una estrella, el objetivo del presente proyecto.

2.2 Fotometría de asteroides

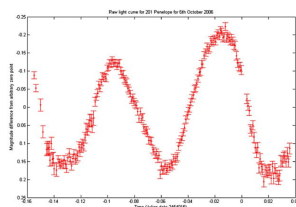
Mientras algunos cuerpos no presentan alteraciones de brillo desde la Tierra (estrellas no variables) los asteroides, en general, manifiestan complejas variaciones de magnitud con el tiempo. Esto se debe tanto a su forma irregular (solo los más masivos alcanzan formas más esféricas), a su rotación y a la composición en muchos casos heterogénea, que implica zonas de la superficie con diferentes albedos. Si se mide el brillo de un asteroide a lo largo de varias horas, si su periodo de rotación es corto, o a lo largo de muchos días o meses, si su periodo es largo, se puede conocer dicho periodo. La representación del brillo (eje Y) respecto al tiempo (eje X) se conoce como curva de luz. El periodo de rotación se deduce de la distancia entre puntos análogos de la curva (máximos y mínimos) (Casado y Serra, 2018). Pero la curva de luz no solo permite el cálculo del periodo de rotación del asteroide, sino también conocer su eje de rotación y forma tridimensional y la existencia de satélites que orbitan a su alrededor: los satélites al ocultar al cuerpo mayor en su movimiento de traslación generan periodicidades más sutiles en la curva de luz (García, 2012).

Es importante entender que mediante fotometría se puede obtener la forma tridimensional de un asteroide pero no su tamaño absoluto, dado que se recibe su brillo aparente y no absoluto. Si se quiere conocer el tamaño real se debe recurrir a ocultaciones estelares, una técnica muy utilizada en el estudio de asteroides.

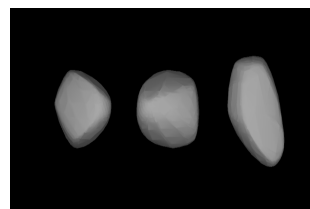
Para poder estudiar fotométricamente a un asteroide se necesita primero localizarlo. Para ello se puede recurrir a los datos astrométricos del MPC (*Minor Planet Center*)¹⁴ e introducirlos en algún programa de astrometría. Además, como se necesita seguir al asteroide durante un gran intervalo de tiempo es casi imprescindible contar con un sistema GoTo, es decir un telescopio que apunte en todo momento al asteroide gracias a un *software* vinculado.

Una vez localizado el asteroide se obtiene una secuencia de imágenes, en las que deben aparecer además dos estrellas de referencia (siempre las mismas). Dichas estrellas de referencia no pueden ser variables, es decir, no pueden manifestar variaciones de brillo con el tiempo y además deben tener brillo y color parecidos al asteroide. Para encontrar estrellas de referencia adecuadas se puede acudir a catálogos estelares¹⁵. El motivo de necesitar estrellas de referencia es que si se grafican también sus curvas de luz, las variaciones observadas serán producidas por las diferentes condiciones en cada noche de observación y los errores de instrumentación, que serán comunes a la curva del asteroide. Es decir, con la ayuda de estas estrellas se puede conocer el error asociado a cada punto de la curva de luz del asteroide.

Algunos asteroides presentan variaciones de magnitud muy elevadas y por tanto fáciles de medir con un equipo medianamente sencillo; sin embargo, otros asteroides manifiestan variaciones muy tenues y solo serán medibles desde observatorios con telescopios mayores.



(a) Curva de luz



(b) Modelo tridimensional

Figura 3: Curva de luz del asteroide del cinturón principal 201- Penelope (6/10/2006) y modelo tridimensional obtenido de varias curvas de luz ([Observatorio Astronómico Los Molinos, s.f.](#); [Astronomical Institute of the Charles University, s.f.](#))

2.3 Imágenes astronómicas

Las imágenes astronómicas son la principal fuente de conocimiento de todos los cuerpos astrofísicos en general. Proviene de observatorios terrestres o de sofisticados telescopios espaciales, aunque cada vez su obtención es asequible para más personas. Según el campo de estudio (astrometría, fotometría...) y el tipo de cuerpo estudiado (asteroides, planetas, galaxias, nebulosas...) sus características y requisitos variarán, pero en general, el equipo básico para su obtención se conforma de un telescopio, un detector o cámara (normalmente CCD) y un filtro o conjunto de filtros. Además, es prácticamente imprescindible una montura ecuatorial para el telescopio para localizar los cuerpos según sus coordenadas ecuatoriales¹⁶ y un sistema GoTo, para el correcto seguimiento de los astros en movimiento. El telescopio y detector apropiados dependerán de varios factores como son el campo de visión necesario o la resolución deseada en las imágenes.

El campo de visión o FOV (*Field of View*) como el nombre indica se corresponde con el área angular de cielo que se puede obtener con el equipo. Este depende de la distancia focal del telescopio (f) y el tamaño del detector. Con que menor sea la focal del telescopio mayor será el campo de visión. De igual manera con que más grande sea el detector más área de cielo contendrá la imagen obtenida. Un equipo con un telescopio de focal corta estará desaprovechado si se le acopla un detector pequeño y de la misma forma un detector de gran tamaño en un telescopio de focal larga dará lugar a imágenes defectuosas ([The American Association of Variable Star Observers \[AAVSO\], 2022](#)).

$$FOV \text{ (grados de arco}^2\text{)} = \frac{\text{ancho}_{\text{detector}} \text{ (mm)} \cdot \text{alto}_{\text{detector}} \text{ (mm)}}{f \text{ (mm)}} \cdot \left(\frac{180}{\pi}\right)^2 \quad (1)$$

Será necesario un FOV amplio en el estudio de cuerpos muy extensos como galaxias o nebulosas o cuerpos que se mueven a gran velocidad y pueden salirse del campo de visión antes de obtener la imagen. Para el estudio fotométrico de las estrellas, probablemente se precise de un FOV extenso para las brillantes por la dificultad de encontrar estrellas de referencia de su misma magnitud en la misma imagen. En el caso del estudio fotométrico de los asteroides, como son mucho más débiles en cuanto a brillo, no se requerirá de un campo de visión demasiado vasto a no ser que se mueva a elevada velocidad.

Por su parte la resolución también se trata de una variable angular, concretamente el campo de visión correspondiente a cada píxel individualmente.

Depende de la dimensión de píxel consiguiéndose las mejores resoluciones para dimensiones pequeñas (la resolución será mejor con que menor sea su valor). Así, para dos detectores del mismo tamaño alcanzará mejor resolución el de mayor cantidad de píxeles.

$$Resolucion \text{ (segundos de arco)} = \frac{ancho_{pixel} \text{ (mm)}}{f \text{ (mm)}} \cdot \frac{180}{\pi} \cdot 3600 \quad (2)$$

Se precisará de una resolución excelente en finalidades que requieran de gran detalle, como puede ser en la astrofotografía de planetas. Para otros fines resoluciones más modestas serán suficientes, como en general en la fotometría de asteroides. Como se deduce de las fórmulas, una focal corta ampliará el FOV pero perjudicará la resolución, por lo que habrá que encontrar el equilibrio adecuado para la finalidad.

Respecto a los filtros, colocados en el objetivo del telescopio o antes del detector, restringen la luz incidente a ciertas longitudes de onda. Solo en excepciones no se usarán filtros como el caso de un objeto extremadamente débil en brillo o fenómenos que emiten en todas las longitudes de onda con gran intensidad (por ejemplo supernovas). En todos los demás casos siempre se usará filtro. Esto permite limitar la información recibida y obtener por tanto conocimientos más valiosos del objeto emisor. Dependiendo del rango del espectro electromagnético de interés (rayos gamma, rayos X, ultravioleta, luz visible, infrarrojos...) se emplearán unos filtros u otros, siendo recomendable que todos estén estandarizados y homologados, para poder comparar con los trabajos de otros observadores. En la figura 4 se muestra el mismo objeto visto bajo diferentes longitudes de onda, es decir, con diferentes filtros.

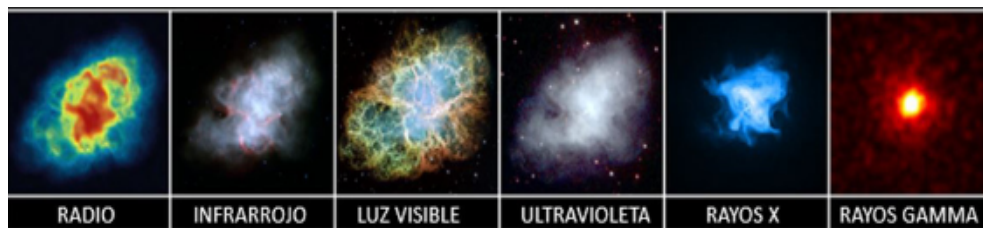


Figura 4: Nebulosa de Cangrejo observada bajo diferentes longitudes del espectro (Caldú, s.f.)

Las imágenes astronómicas producidas por el equipo son siempre en blanco y negro. Si se requieren colores, tendrán que obtenerse varias imágenes del mismo cuerpo con diferentes filtros y se atribuirá un color a cada filtro (ver figura 5).



Figura 5: Ejemplo de imagen astronómica de alta resolución a color. Los Pilares de la Creación en la nebulosa de Águila, tomada por Hubble en 1995 ([National Geographic, 2022](#))

La adición de colores se realiza en las etapas finales del procesamiento de imágenes, antes habrá que llevar a cabo un proceso común a todas las imágenes astronómicas, denominado reducción o calibrado, detallado en la subsección 2.3.2. Para la fotometría se trabaja con las imágenes calibradas en blanco y negro, obtenidas con los filtros adecuados para resaltar las emisiones en las longitudes de onda de interés (ver imagen 6).



Figura 6: Ejemplo de imagen para fotometría de un campo denso de estrellas con un filtro en la región visible del espectro (filtro V) ([Berry, s.f.](#))

2.3.1 Detector CCD

El sensor o detector CCD (*Charge-Coupled Device*), también conocido como cámara CCD, es el detector más utilizado actualmente para obtener imágenes astronómicas y se basa en el efecto fotoeléctrico¹⁷, pues los fotones recibidos por cada celda del sensor (píxel) formada por semiconductores (silicio) generan una cantidad de electrones proporcional. El sensor CCD se sitúa en el punto focal del telescopio (donde se forma la imagen) y está acoplado a una célula Peltier, encargada de enfriarlo y a un amplificador, que aumenta proporcionalmente la carga almacenada por cada celda en una tensión de magnitud medible, que posteriormente es transformada en una señal digital.

Entonces la imagen digital obtenida consistirá en una matriz de números (píxeles) con valores proporcionales a los fotones recibidos por la celda asociada del CCD: la celda con más fotones incidentes se corresponderá con el píxel de mayor valor y la que ha recibido menos fotones se corresponderá con el píxel de menor valor. Pero el humano procesa más fácilmente diferencias de color que una matriz numérica, por lo que al píxel de mayor valor se le asocia el color blanco, al de menor valor el negro y a todos los demás tonos grises intermedios. Dependiendo de la escala de grises seleccionada en el *software* de visualización habrá más o menos tonos diferentes, de manera que los píxeles con valores muy cercanos adquirirán el mismo tono. Por tanto una imagen astronómica antes de cualquier procesamiento se trata de un mapa de casillas grises, donde las fuentes se verán como conglomerados de casillas cercanas al blanco en torno a su centro (donde se reciben más fotones) y de casillas de tonalidades claras en las zonas colindantes, mientras que las regiones sin fuentes adquirirán tonos de gris oscuro o negro. En la figura 8 se muestra un ejemplo de la apariencia de una fuente en una imagen astronómica obtenida con una CCD, con una escala de grises de 64 tonos.

Hoy en día también está vigente otra tecnología: los dispositivos CMOS (*Complementary Metal-Oxide-Semiconductor*). Véase la comparativa entre ambos dispositivos ([Roca, 2014](#); [Espacio Profundo, 2018](#)):

- Economía. El CMOS es más sencillo de construir y por lo tanto más económico. Además consume menos energía.

- *Blooming*. El CMOS no sufre de *blooming* mientras que el CCD sí. El *blooming* ocurre cuando alguna de las celdas o píxeles recibe demasiados fotones, superando el número de electrones que puede almacenar y desbordándolos en los píxeles adyacentes. Existen cámaras CCD con sistema *antiblooming* que evitan este efecto pero pierden sensibilidad, por lo que no son favorables en astronomía. La solución en este ámbito es regular los tiempos de exposición según la magnitud de brillo del cuerpo medido, es decir, mayores tiempos de exposición para cuerpos débiles y tiempos de exposición más pequeños para estrellas brillantes. En la figura 7 se muestra un ejemplo de *blooming* extremo.



Figura 7: Efecto de *blooming* en una imagen tomada con CCD ([Astronomics, s.f.](#))

- Velocidad. El CCD contiene un único amplificador externo al detector que recibe y aumenta la señal de todos los píxeles, mientras que el CMOS posee un amplificador propio en cada una de los píxeles. Esto permite que la velocidad de obtención de la imagen sea mucho más rápida en el caso de CMOS pues los amplificadores trabajan simultáneamente.
- Uniformidad. Como el CMOS tiene un amplificador por cada píxel no existirá plena uniformidad. Esto implica que para el mismo objeto observado y mismo tiempo de exposición se puedan obtener resultados ligeramente distintos. En cambio, el CCD, al tener un único amplificador para todos los píxeles consigue una uniformidad asombrosa.
- Ruido. Como parte de la electrónica es externa al sensor, el ruido del dispositivo CCD es menor, aunque sigue siendo importante como se verá a continuación.

- Rango dinámico. Se corresponde con el rango que existe entre la mínima señal lumínica que pueden detectar y la señal que produce saturación o *blooming*. El CCD gana con creces en esta cualidad por lo que puede detectar un amplio rango de magnitudes.
- Linealidad. Un comportamiento lineal implica que el número de electrones generados es proporcional al de fotones recibidos y la constante de proporcionalidad no cambia con dicho número. Es decir, un píxel que recibe $2 \cdot n$ fotones generará el doble de electrones que uno que recibe n . El CCD es más lineal¹⁸ que el CMOS, con diferencia.

De todo lo anterior, parece que ambos están muy igualados, de hecho el CMOS domina en las cámaras de vídeo, *webcams*, etc, pero como en astronomía son fundamentales la linealidad, uniformidad y rango dinámico elevados y el menor ruido posible, el CCD es el más adecuado. Aunque algunos aficionados de la astronomía optan por la tecnología CMOS, en los observatorios terrestres y telescopios espaciales se emplean detectores CCD.

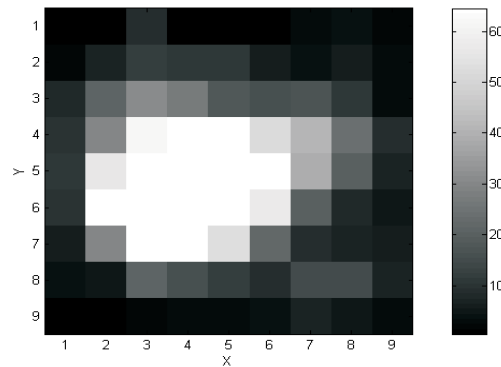


Figura 8: Ejemplo de fuente en una imagen astronómica (Doreste, 2002)

2.3.2 Reducción de imágenes astronómicas

Para obtener las imágenes finales que serán utilizadas en el estudio de los cuerpos (imágenes reducidas) se realizan una serie de operaciones con las imágenes originales (imágenes objeto) y un conjunto de imágenes detalladas a continuación (*bias*, *dark* y *flatfield*) para corregir los distintos defectos y ruidos introducidos por el equipo de medida: las lentes del telescopio o los filtros pueden contener suciedades en la fecha de obtención de las imágenes y la utilización de la cámara CCD puede producir:

- Ruido de lectura. Para contabilizar los electrones almacenados en cada celda y poder amplificarlos estos deben transitar hacia el registro de su columna de celdas (lectura). Tras este proceso el valor contabilizado puede variar ligeramente del real. Esto producirá pequeñas variaciones de los valores de los píxeles en la imagen digital.
- Ruido de *bias* u *offset*. Se corresponde con los electrones introducidos por la propia CCD para evitar que el ruido de lectura de lugar a valores negativos de carga (corriente de polarización). Estos electrones se suman a los demás como si fuesen procedentes de las fuentes astronómicas y por tanto alteran la imagen digital. El ruido de *offset* es prácticamente invariable con el tiempo y homogéneo a todas las celdas.
- Ruido térmico. Electrones generados por excitación térmica en los componentes electrónicos. Afecta de manera diferente y aleatoria a las diferentes celdas y empeora con el tiempo de exposición al aumentar la temperatura, por eso es fundamental el enfriamiento de la CCD con la célula de Peltier. De la misma forma alteran la imagen digital.

Para corregir estos ruidos y otros defectos se deben tomar las siguientes imágenes ([The American Association of Variable Star Observers \[AAVSO\], 2022](#)):

- Imágenes de *bias* u *offset*. Se obtienen con el objetivo del telescopio tapado o con el obturador de la CCD cerrado, con tiempos de exposición mínimos (milésima de segundo o similar), de manera que la carga almacenada por las celdas procede exclusivamente del *offset* de la CCD. Permiten por tanto sustraer el ruido de *offset* de las imágenes objeto. Normalmente es suficiente con obtener cuatro o cinco imágenes de este tipo en la fecha de obtención.
- Imágenes *dark* o tomas oscuras. Se obtienen nuevamente con el objetivo tapado o el obturador cerrado pero con tiempos de exposición más largos e iguales, del mismo orden que los de las imágenes objeto. Permiten corregir el ruido térmico de la CCD. Se suele tomar una antes y otra después de cada imagen objeto para tener en cuenta la variación de temperatura del equipo durante la medida. Si el equipo cuenta con una buena regulación de temperatura pueden tomarse cada cierto intervalo de tiempo.

- Imágenes *flatfield* o tomas planas. Se logran apuntando a fuentes lumínicas uniformes como puede ser una pantalla iluminada por luz artificial o zonas brillantes del cielo durante el amanecer o el atardecer. Permiten encontrar diferencias de sensibilidad entre las celdas del CCD o incluso celdas inoperativas (píxeles “rotos”) pero también suciedad en las lentes (si aparecen para todos los filtros) o en los filtros (si aparecen para un filtro en concreto). Deben tomarse las máximas posibles con cada uno de los filtros a utilizar en las imágenes objeto (en torno a 15 por cada filtro) y si se aplican al cielo es recomendable mover ligeramente el telescopio antes de una nueva imagen para evitar la contribución de las estrellas que puedan aparecer sobre el fondo uniforme. Además, se deben usar tiempos de exposición de manera que el número de electrones generados en cada celda se sitúe en torno a la mitad del rango dinámico de la cámara (con que más brillante esté el cielo menor será el tiempo apropiado). De esta forma el error de lectura de las imágenes será despreciable respecto a los demás (*offset*, térmico...). Dar con los tiempos adecuados se consigue con experiencia y esmero. De lo anterior se deduce que las *flatfield* son las imágenes más engorrosas de obtener pero si se desea una calidad refinada de las imágenes finales son necesarias. Si no se requiriese tanta calidad podría prescindirse de las *flatfield*. En ese caso las *bias* tampoco serían necesarias, pues bastaría con las *dark* para corregir tanto el ruido de *offset* como el térmico.

Lo primero es obtener la *super-bias*, que consiste en la imagen media de todas las *bias* tomadas en esa fecha. Después, se obtiene la *super-dark* realizando nuevamente la media de todas las *dark* y restándole la *super-bias*. De esta manera, la *super-dark* solo incluye la señal causada por el ruido térmico. A continuación se resta a las *flatfield* tanto el *super-bias* como el *super-dark*, en esa orden, se realiza la media (una para cada filtro) y se normaliza dividiendo entre el valor medio de píxel, consiguiendo las *super-flatfield*.

Una vez se tiene una única imagen *super-bias*, una *super-dark* y las *super-flatfield* de los diferentes filtros se corrigen las imágenes objeto. Para ello se les resta la *super-bias* y la *super-dark* y se dividen entre la *super-flatfield* correspondiente a su filtro. Al terminar estos pasos se aísla la contribución causada por la fuente o fuentes a estudiar y se eliminan todas las demás. Una vez se tienen las imágenes objeto corregidas, es habitual también combinar varias de ellas (para el mismo filtro por supuesto) para obtener una sola imagen objeto de mayor calidad.

Imagínese que el objeto a estudiar es una galaxia y en una de las imágenes objeto hay una zona de píxeles negros debidos a un rayo cósmico¹⁹, coincidiendo en uno de los brazos. Si se superpone dicha imagen con otras que han registrado correctamente esa zona, alinéandolas mediante estrellas de referencia, se puede compensar el defecto y contar con información de esa zona. Por ello, las imágenes objeto deben estar descentradas entre si, modificando el telescopio ligeramente en caso de tratarse de cuerpos fijos. Además, en ocasiones, si se desean resaltar solo cierto tipo de estrellas (como las recién formadas), puede resultar conveniente restar la imagen final de uno de los filtros a la de otro, para restringirse solo a ciertas longitudes de onda.

A *grosso modo*, este es el proceso de reducción o calibración para las imágenes astronómicas. Como es lógico hoy en día, todas esas operaciones sobre las imágenes son llevadas a cabo por un *software*, a elegir entre una gran diversidad de programas: IRIS, AstroArt, IRAF, StarTools, AIPS, PixInsight, etc.

En la figura 9 se muestra un ejemplo de reducción de imágenes astronómicas de la galaxia NGC 6118²⁰ realizado por el *software* IRAF (*Image Reducing and Analysis Facility*). Las imágenes 9a y 9b se tratan de una de las cuatro imágenes objeto con el filtro circular $R - Gunn$ y el filtro cuadrado $H - \alpha$ respectivamente. Las figuras 9c y 9d se tratan de las *super-flatfield* correspondientes en donde se observa como las celdas presentan sensibilidades diferentes (zonas más claras y oscuras) y píxeles no operativos. Además se aprecia una suciedad en alguna lente del telescopio pues es común a ambos filtros (defecto dentro del círculo blanco). Las figuras 9e y 9f son las imágenes objeto finales resultantes de combinar las cuatro imágenes objeto calibradas de cada uno de los filtros, que ya no poseen píxeles negros ni otros defectos.



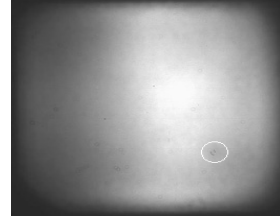
(a) Imagen objeto original con filtro $R - Gunn$



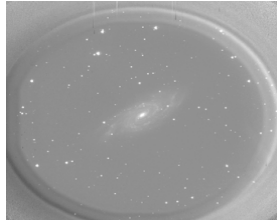
(b) Imagen objeto original con filtro $H - \alpha$



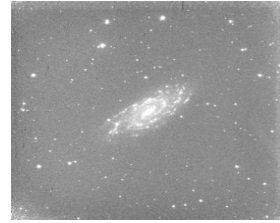
(c) *Super-flatfield* para el filtro $R - Gunn$



(d) *Super-flatfield* para el filtro $H - \alpha$



(e) Imagen objeto final con filtro $R - Gunn$



(f) Imagen objeto final con filtro $H - \alpha$

Figura 9: Ejemplo de reducción de imágenes astronómicas de la galaxia NGC 6118 (Ruiz, 2010)

En la figura 10 se muestra la imagen resultante de restar la contribución del filtro $R - Gunn$ al filtro $H - \alpha$, es decir la figura 9f menos la 9e. De esta manera se observan solo las estrellas nuevas y las zonas de formación estelares.

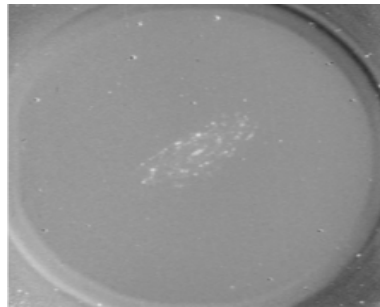


Figura 10: Zonas de formación estelar y estrellas nuevas (Ruiz, 2010)

Por último aclarar que con el proceso de reducción o calibrado tan solo se corrigen los principales efectos introducidos por el equipo de medida (telescopios, CCD y filtros) pero existen muchos otros factores que afectan a la calidad de las imágenes astronómicas. El más importante es la atmósfera, que puede presentar multitud de fenómenos, que además, varían en cada fecha. Lo primero a destacar es que la atmósfera absorbe gran cantidad de la luz que llega desde las fuentes y lo hace de manera desigual según la longitud de onda, lo que además está cambiando según avanza el efecto invernadero. Por otra parte, los fenómenos meteorológicos como las nubes y la lluvia pueden aplazar irremediamente sesiones de obtención de imágenes. Pero no es necesario toparse con condiciones tan desfavorables para obtener mala calidad de imagen; existen noches con meteorología en principio oportuna que dan lugar aún así a malas condiciones para la observación. Esto es lo que se conoce como *seeing* y se corresponde con el efecto de turbulencia o distorsión atmosférica, producido por corrientes de aire de diferentes densidades. Un mal *seeing* provocará que las fuentes temblequeen y aunque puede apreciarse a simple vista para estrellas brillantes empeora con el telescopio, especialmente si se emplean aumentos. Por ello cuando se requiera de imágenes nítidas de alta resolución habrá que seleccionar noches con muy buen *seeing*. Los observatorios situados a gran altitud pueden alcanzar valores de *seeing* muy buenos (0,4 segundos de arco) mientras que los situados a menor altura nunca bajarán del segundo de arco. En la figura 11 se muestra la escala de Pickering para el *seeing* desde el 1 (perfecto) al 10 (pésimo).

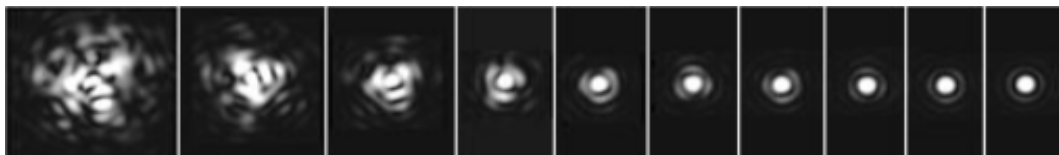


Figura 11: Escala de Pickering del *seeing* atmosférico (Bravo, 2019)

En ocasiones, aún habiendo buen *seeing* hay poca transparencia, debida a contaminación, polvo en suspensión, etc. En este caso se dificultará la observación de nebulosas y galaxias debido a su débil magnitud absoluta pero en cambio podrán obtenerse imágenes de alta resolución de por ejemplo planetas o estrellas cercanas.

Por ello conocer las condiciones atmosféricas es fundamental para la obtención de imágenes astronómicas. Otro factor a tener en cuenta es la contaminación lumínica aunque en general los observatorios están localizados en sitios estratégicos en este sentido.

3 Detección de formas aisladas de asteroides

3.1 Descripción del proyecto

Como se ha comentado en la introducción el objetivo del proyecto es la de crear un clasificador basado en aprendizaje automático (*machine learning*) supervisado, que permita discernir imágenes con asteroides aislados, es decir, no mezclados con otra fuente, de las que contienen superposición de fuentes. Para ello se trabajará sobre los *croppies* obtenidos por el algoritmo *m2* de dos asteroides diferentes, extraídos de la base de datos del IAA: con los del transneptuniano Varuna se entrenará el algoritmo y con los del también transneptuniano 2002_ms4 se comprobará su correcto rendimiento, lo que se conoce como testear en el ámbito del *machine learning*.

Al tratarse de un algoritmo supervisado no aprenderá de manera autónoma de las imágenes de entrenamiento, se le debe indicar a que tipo (aislado o mezcla de fuentes) pertenece cada una de ellas, para que posteriormente pueda decidir por si solo ante imágenes nuevas. Dicho de otra forma, habrá que pasarle los *croppies* etiquetados. Por lo tanto se debe realizar un etiquetado manual de los *croppies* con la ayuda del ojo humano para poder “enseñarle” al algoritmo. Este etiquetado se agilizará mediante un código que descarte las imágenes de poca calidad, como se verá en próximos apartados. Además, para facilitarle el aprendizaje al clasificador, habrá que preprocesar y estandarizar los *croppies* antes de introducirlos, lo que es prácticamente obligatorio en todos los algoritmos de *machine learning* para obtener resultados razonables.

Una vez creado el algoritmo y verificado su funcionamiento ante *croppies* de Varuna distintos a los de entrenamiento (*croppies* de *test*) se comprobará con otro conjunto de *croppies* totalmente nuevos para el clasificador, los del asteroide 2002_ms4, para observar su poder de generalización y saber si va a poder obtener un rendimiento decente en el procesamiento de otros asteroides.

El código correspondiente a todo el proyecto, no solo a la creación del clasificador, sino también del preprocesamiento de los *croppies* y otros pasos, se realiza en el lenguaje de programación Python. El entorno usado es Google Colab puesto que permite la inclusión de secciones y anotaciones de manera cómoda. El código en formato de cuaderno *.ipynb* puede encontrarse en el enlace de [GitHub](#) al final de la memoria, en el [apéndice](#).

3.2 Datos de entrenamiento. Varuna

Se cuenta con 3651 *croppies* del TNO Varuna producidos por el algoritmo *m2* y separados inicialmente en dos carpetas, correspondientes con las detecciones positivas y negativas clasificadas por dicho algoritmo. Se cuenta también con las imágenes astronómicas originales en formato FITS, el formato más utilizado para almacenar imágenes astronómicas, cuya función es la de acudir a ellas si se tienen dudas en algún *croppie*, puesto que la adición de colores puede suponer en ocasiones un entorpecimiento en la búsqueda de mezclas de fuentes sutiles. Del total de *croppies*, 2814 se tratan de detecciones positivas según el algoritmo *m2* y 837 de detecciones negativas.

Las imágenes astronómicas de Varuna de las que proceden los *croppies* han sido tomadas entre febrero de 2002 y el mismo mes del año actual (2022), por lo que abarcan dos décadas. En algunas fechas se han tomado multitud de imágenes, separadas por pocos minutos; en varias, una única imagen y en otras ninguna.

Es importante aclarar que los *croppies* de Varuna implicados en este proyecto no van a ser utilizados para la obtención de propiedades físicas del asteroide en un futuro, sino que su cometido se limita a crear el clasificador y favorecer el procesamiento de otras imágenes de este asteroide, u otros, que sí van a ser utilizadas para mediciones físicas y posibles hallazgos sobre los cuerpos menores.

Los *croppies* mantienen el nombre original de la imagen FITS asociada y como es habitual en este formato se trata de un nombre muy largo y complejo, al contener mucha información. Por consiguiente es conveniente cambiar los nombres originales por uno más sencillo para poder identificar cada imagen de manera más rápida. Lo más lógico es que este nuevo nombre se trate de la fecha y hora de la toma de la imagen, pues esto permite que al ordenar en cuanto a nombre en el directorio se ordenen también por fecha y se puedan ver secuencias temporales de *croppies*. Dichas secuencias posibilitan la apreciación de fenómenos, como son en este caso las superposiciones de fuentes.

La modificación del nombre se lleva a cabo a través de código, buscando la fecha y hora del nombre original y transformándola al formato *aaaa - mm - dd_hh : mm : ss*. Así, los *croppies* están ordenados por año y dentro de cada año se ordenan a su vez por mes. Algunas de las imágenes astronómicas contienen la fecha en el nombre de manera muy diferente a las demás o directamente no la contienen.

En el primer caso, se modifican los nombres manualmente para obtener la fecha en el formato nombrado; en el segundo, al no contar con fecha, se mantiene el nombre original, sin posibilidad de situarlo temporalmente respecto a los demás.

Por ejemplo, el nombre del *croppie* de la imagen 12 pasa del original *science_Varuna – 0002Cle_OSN_1_5_2022_02_02T00_33_08_170_open_400s_1024x1024_andor_Varuna* a *2022 – 02 – 02_00 : 33 : 08*.

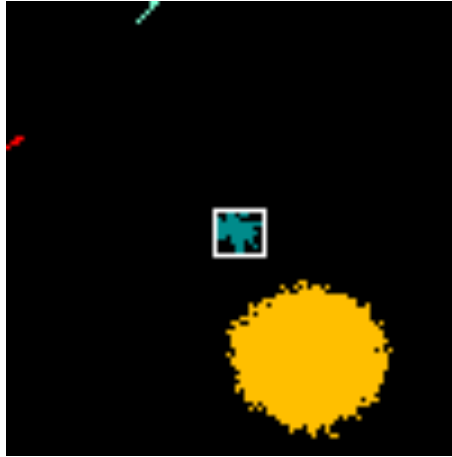


Figura 12: Ejemplo de *croppie* obtenido por el algoritmo *m2*. Se trata de un asteroide aislado (dentro del cuadro blanco) cercano a una fuente.

Una vez modificados los nombres, es hora de descartar los *croppies* que no poseen la calidad suficiente para el entrenamiento del clasificador. Esto podría realizarse también de manera manual pues son muy fáciles de identificar visualmente, pero de nuevo, la finalidad del proyecto es agilizar el procesamiento de las imágenes, por lo que es necesario crear un código que ejecute dicha tarea. Dentro de los *croppies* a descartar existen diversas variantes:

- *Croppies* sin asteroide. Son los más fáciles de identificar, pues dentro del marco blanco donde debería encontrarse el asteroide no existe ninguna fuente. Para descartarlos basta con recortar el interior del cuadro blanco y contar los píxeles ocupados, si no hay ninguno el *croppie* se desecha. En este tipo de *croppies* el algoritmo *m2* funciona de forma eficiente, descartando la totalidad de imágenes sin asteroide, por lo que no se presenta ninguna mejora al respecto con el presente código.

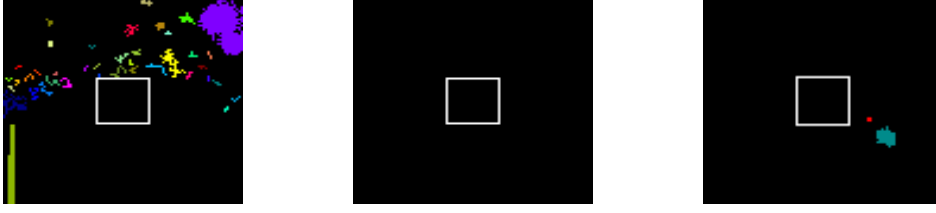


Figura 13: Ejemplos de *croppies* sin asteroide

- *Croppies* con una fuente pequeña o dispersa. Si el número de píxeles asociados a la fuente no es suficiente o tiene una forma poco compacta no se tratará del asteroide, sino de otras causas como los rayos cósmicos. Por lo tanto la fuente debe tener un tamaño mínimo y una forma medianamente compacta, lo que puede detectarse buscando píxeles ocupados consecutivos horizontal y verticalmente en el interior del cuadro blanco. El algoritmo *m2* deja pasar la inmensa mayoría de estos *croppies* a las detecciones positivas.



Figura 14: Ejemplos de *croppies* con señales demasiado pequeñas o dispersas

- *Croppies* de muy baja calidad. También sencillos de identificar visualmente, se presentan como una amálgama de colores en la que no se diferencia el asteroide respecto al fondo. Para descartarlos se busca un porcentaje de ocupación de píxeles igual o mayor al 30 % del *croppie*. El algoritmo *m2* filtra una gran cantidad de estos *croppies* pero no la totalidad.

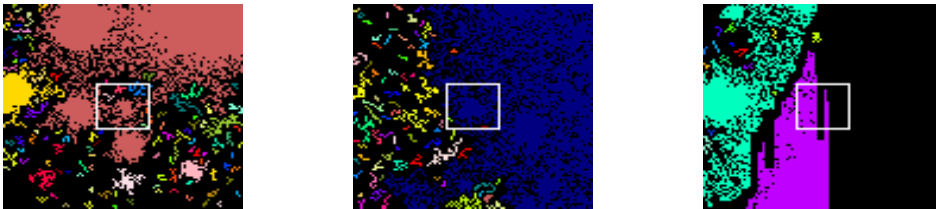


Figura 15: Ejemplos de *croppies* de muy baja calidad

- *Croppies* de baja calidad. En ellos el asteroide es en ocasiones apreciable pero se encuentra mezclado con defectos de la imagen o en una nube de píxeles dispersos.

Estos *croppies* se descartan en su totalidad al desechar los del siguiente grupo. El algoritmo *m2* deja pasar a las detecciones positivas bastantes *croppies* de este tipo.

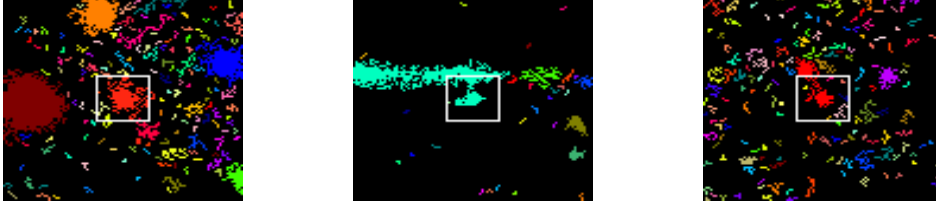


Figura 16: Ejemplos de *croppies* de baja calidad

- *Croppies* contaminados. En muchas ocasiones los *croppies* aparentan ser de buena calidad, el asteroide se diferencia bien y no existen defectos apreciables en la imagen, pero en realidad no son válidos al existir contaminación. Cuando una o varias fuentes están demasiado cercanas al asteroide, a pesar de no estar mezcladas, pueden contaminarlo lumínicamente, es decir, los píxeles de la región del asteroide pueden estar registrando señal que procede realmente de las fuentes colindantes y por lo tanto alterar las posteriores mediciones fotométricas. Es por eso que deben eliminarse todos los *croppies* que contengan píxeles ocupados en un radio de 4 píxeles por fuera del cuadro blanco. Al eliminar los contaminados se descartan también los *croppies* de baja calidad, pues siempre presentan píxeles ocupados en el entorno del asteroide. El algoritmo *m2* no descarta los *croppies* contaminados.

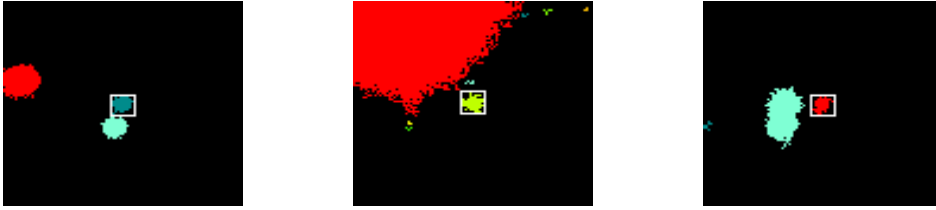


Figura 17: Ejemplos de *croppies* contaminados

Tras descartar todos estos *croppies* de Varuna, tanto de las detecciones positivas de *m2* como de las negativas, es hora de clasificar manualmente los *croppies* que han pasado el filtro de calidad en las dos clases, fuentes aisladas y mezcla de fuentes, para poder etiquetarlos y entrenar al clasificador.

Algunas de las mezclas son evidentes visualmente al observarse dos cuerpos esféricos fusionados parcialmente, pero otras son más sutiles al encontrarse las fuentes muy solapadas.

Estos casos, difíciles de distinguir de una fuente aislada por el ojo humano, fueron identificados gracias a contar con las imágenes anteriores y posteriores temporalmente, pudiendo observarse la fusión o separación de las fuentes. En el [apéndice](#) se adjuntan dos ejemplos de secuencias de *croppies* que manifiestan mezcla de fuentes (figuras 42 y 43).

Entonces, una vez etiquetados, se cuenta con 80 *croppies* con superposición o mezcla de fuentes y 2005 con fuentes aisladas. Debido a la gran diferencia de tamaño de las dos clases, habrá que llevar a cabo un equilibrado para poder entrenar al algoritmo, como se verá en el apartado siguiente.

En la figura 18 se muestran ejemplos de los *croppies* etiquetados manualmente. Los *croppies* 18a, 18b y 18c se corresponden con fuentes mezcladas y 18d, 18e y 18f con fuentes aisladas.

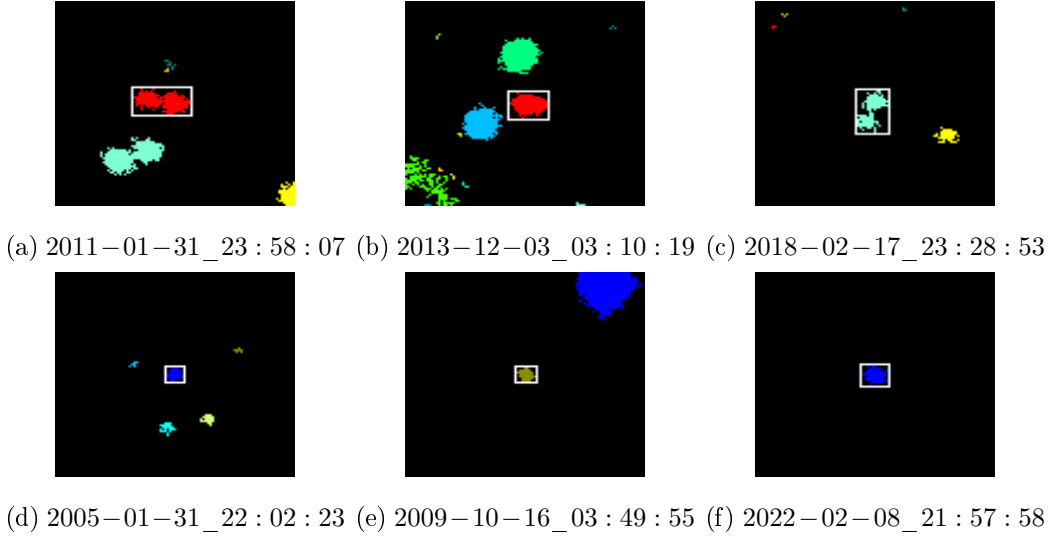


Figura 18: Ejemplos de *croppies* etiquetados.

3.3 Preprocesamiento

Para poder introducir datos de entrenamiento en un algoritmo de *machine learning*, no importa el tipo, siempre hay que aplicar un preprocesamiento y estandarización previos para que pueda aprender las características de interés y no otras secundarias. Esto es especialmente importante cuando se trabaja con imágenes, al contar con multitud de características. En este caso, interesa que el algoritmo analice la región donde se encuentra el asteroide o la mezcla de fuentes y aprenda las diferencias de forma entre ambos tipos.

No importa el resto del *croppie* original, por lo que es lógico centrarse en el interior del cuadro blanco; tampoco interesa de que color pueda ser la fuente o mezcla de fuentes, ya que no tiene ningún tipo de correlación con la tarea, simplemente es una forma de segmentación de las fuentes existentes que aplica *m2*.

En consecuencia, el preprocesamiento consiste en detectar el marco blanco en cada *croppie*, recortar el interior y pasar de una imagen a color (tres canales) a blanco y negro (un único canal), donde la fuente o fuentes mezcladas se correspondan con píxeles blancos y los píxeles sin fuentes sigan siendo negros. Además, como la región interior del cuadro blanco no posee el mismo tamaño en todos los *croppies* habrá que estandarizar también el tamaño del recorte resultante. En realidad, una CNN puede tomar entradas de tamaños diferentes siempre que se estandaricen antes de la capa de salida, pero es más sencilla la estandarización previa. Un tamaño estándar adecuado es el de 40x50 pues es mayor que la región interna del cuadro para todos los *croppies* y puede por tanto abarcar todas las fuentes aisladas o mezcladas, sin cortarlas.

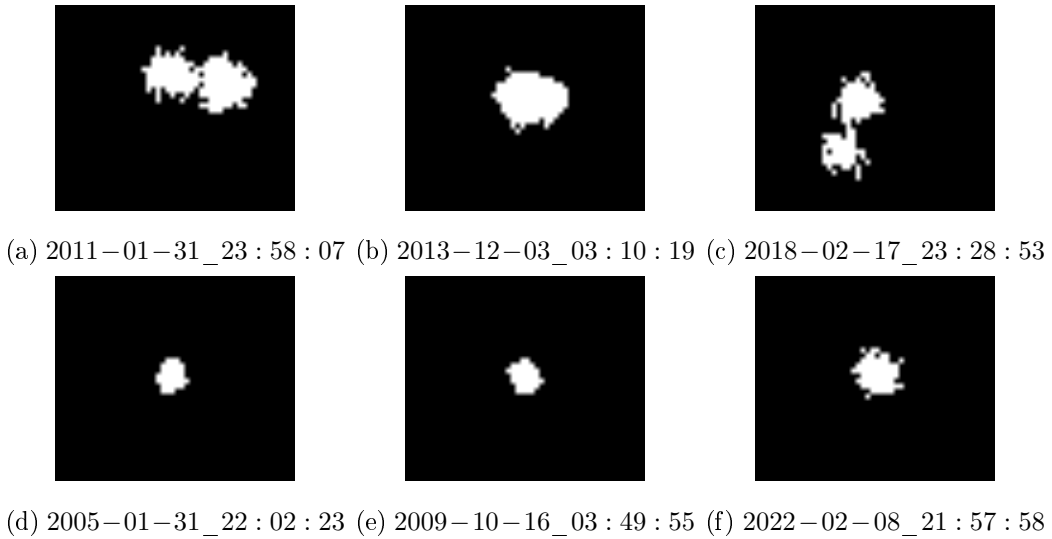


Figura 19: *Croppies* de entrenamiento preprocesados correspondientes a la figura 18

Después del preprocesamiento es momento de equilibrar las dos clases, pues como se ha comentado los *croppies* con mezcla son minoritarios. Si se introducen al algoritmo sin compensar las cantidades memorizará la clase mayoritaria (fuentes aisladas). Esto podría derivar en una falsa apariencia de buen rendimiento, pues el conjunto de *test* contendría también más fuentes aisladas y el clasificador acertaría en la mayoría de los casos. En cambio, en el momento en el que le llegasen nuevas imágenes con cantidades equitativas

de ambas clases, presentaría un gran porcentaje de error en las de fuentes mezcladas. Por consiguiente, es fundamental que las clases estén equilibradas. Para ello se utiliza *data augmentation*, una herramienta muy usada en *machine learning* que se aplica a la clase minoritaria y consiste en conseguir varias muestras a partir de una muestra original. Así pues, a cada *croppie* con mezcla preprocesada se le aplican traslaciones tanto en el eje vertical como en el horizontal, de manera que las imágenes modificadas (aumentadas) se comporten como imágenes “nuevas” para el algoritmo. Obviamente lo idóneo sería contar con más *croppies* con mezcla de buena calidad, pero al no ser posible, es la opción más apropiada.

Otra posibilidad sería reducir la cantidad de la clase mayoritaria, descartando imágenes de manera aleatoria, pero esto afectaría al rendimiento del clasificador puesto que habría aprendido de menos muestras. Es decir, a mayor número de imágenes de entrenamiento más eficiente será el algoritmo.

En razón de lo anterior, se opta por el *data augmentation* y se generan 25 *croppies* a partir de cada *croppie* con fuentes superpuestas. De esta manera se consiguen 2000 mezclas, que prácticamente igualan a las 2005 fuentes aisladas.

A continuación, solo queda almacenar los *croppies* preprocesados, tanto con fuentes aisladas como mezcladas en un mismo *array*. Dicho *array* tendrá una dimensión (4005,40,50), pues hay 4005 imágenes (2005 aisladas y 2000 mezcladas) y su dimensión es de 40x50. De igual manera, se almacenan las etiquetas asociadas a cada imagen en un array, también de longitud 4005 como es lógico. La etiqueta es “Sí” para las mezclas (*Sí* hay mezcla de fuentes) y “No” para las aisladas (*No* hay mezcla de fuentes). Sin embargo, para que el algoritmo pueda trabajar con ellas hay que transformarlas a entero: 1 para las mezcladas y 0 para las aisladas. Finalmente, por elección propia, se convierten a binario antes de pasarlas al algoritmo, cuyo motivo se indicará en el apartado siguiente. Por tanto la etiqueta de las mezclas será [0. 1.] y la de las aisladas [1. 0.].

Por último, se aplican dos pasos finales: se añade una cuarta dimensión al *array* de las imágenes, pues el algoritmo necesita conocer el número de canales. En este caso, al ser en blanco y negro, el número de canales es uno y el *array* tiene una dimensión final de (4005,40,50,1); si se tratasen de imágenes a color sería (4005,40,50,3). Finalmente, se dividen las imágenes y etiquetas asociadas en dos conjuntos: *train* y *test*.

Las de *train* serán usadas para el entrenamiento del clasificador y las de *test* para comprobar los resultados. Es de extrema importancia separar de manera correcta los dos conjuntos pues las imágenes de *test* tienen que ser totalmente nuevas para el algoritmo, no puede ocurrir que algunas imágenes pertenezcan simultáneamente al conjunto de entrenamiento y de *test*. Si esto ocurriese, el algoritmo acertaría en las imágenes de *test* por el simple hecho de que ya las ha visto; en cambio obtendría muy malos resultados en imágenes realmente nunca vistas para él. Un símil sería el de un alumno que aprueba un examen porque conoce a priori las preguntas que habrá, no porque haya aprendido nada del temario; en el momento en el que se enfrente a preguntas inesperadas no las superará. No obstante, la separación eficaz de ambos conjuntos es muy sencilla pues las librerías de *machine learning* en Python cuentan con funciones que se encargan de ello.

Lo habitual es destinar el 80 % de los datos a entrenamiento y el 20 % restante a *test*, por lo que se trabaja con 3204 imágenes etiquetadas de entrenamiento y 801 de test. Al concluir todos estos pasos, es hora de entrenar al algoritmo de clasificación, cuyos fundamentos y arquitectura se explican a continuación.

3.4 El algoritmo. CNN

Aunque existen multitud de algoritmos de clasificación (árboles de decisión, *random forest*, *support vector machine*, etc) para esta tarea se selecciona una red neuronal, concretamente una red neuronal convolucional o CNN (*Convolutional Neural Network*). Tanto las redes neuronales convencionales como las convolucionales se incluyen en un subcampo dentro del *machine learning*, el aprendizaje profundo o *deep learning*. Esta denominación proviene de que las redes imitan al aprendizaje humano y por tanto permiten la resolución de problemas más complejos que el aprendizaje automático convencional. Mientras que las redes convencionales simulan el funcionamiento de neuronas humanas genéricas interconectadas, las CNN se asemejan más al funcionamiento del córtex visual, la zona del cerebro que se encarga de procesar la información visual. Es por ello que las CNN son ampliamente usadas en clasificación de imágenes, aunque también pueden aplicarse a secuencias temporales como señales de audio y otras entradas que puedan representarse de forma matricial.

Las redes convencionales también pueden clasificar imágenes, pero si estas son realmente sencillas, de lo contrario no darán resultado.

Esto es debido a que no trabajan con datos bidimensionales sino unidimensionales y al transformar una imagen a un *array* unidimensional (aplanamiento o *flatten*) se pierde gran cantidad de información. La complejidad de la presente tarea es ya demasiada para una red convencional pues las fuentes aisladas no tienen la misma forma ni el mismo tamaño en todas las imágenes y mucho menos las fuentes superpuestas. En el supuesto caso de que todas las fuentes aisladas tuviesen prácticamente la misma forma y tamaño salvo una variación mínima y las mezcladas también sí podría resolverse con una red convencional; es por eso que debe recurrirse a una red convolucional. Sin embargo, dentro de los problemas resolubles por una CNN, este no es de lo más complejos por lo que bastará con una arquitectura relativamente sencilla, mucho menos exigente computacionalmente que la que se tendría en la clásica clasificación de imágenes de perros y gatos por ejemplo.

El fundamento de la CNN es el de operar los valores de los píxeles de cada imagen contra varias matrices denominadas filtros o *kernels*. Cada *kernel* recorre la imagen de izquierda a derecha y de arriba abajo barriendo en cada operación un píxel diferente junto a su vecindario de píxeles, denominados campo de recepción (*receptive field*). La operación realizada, aunque se denomina convolución en este ámbito, se trata en realidad de una correlación cruzada, es decir, el producto de cada coeficiente del *kernel* con el píxel correspondiente del campo receptivo y la posterior suma. Tras cada operación se obtiene un píxel de la matriz resultante.

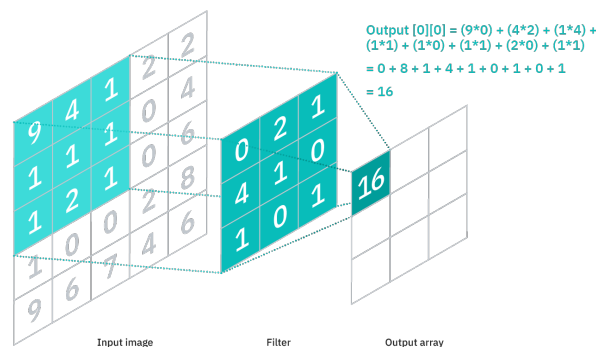


Figura 20: Operación de convolución de un *kernel* sobre un campo receptivo de la imagen (IBM, 2020)

El tamaño de la matriz resultante depende de varios factores:

- *Stride*. El *stride* es el paso que se desliza el *kernel* entre una operación y la siguiente.

Por defecto es de valor unidad por lo que se desplaza un píxel horizontalmente en cada operación desde la esquina izquierda superior hasta alcanzar la esquina derecha y a continuación se desliza hacia abajo un píxel empezando nuevamente a barrer desde el límite izquierdo de la imagen. Si se selecciona un *stride* de 2 o mayor el *kernel* barrerá la imagen en menor tiempo y la matriz de salida tendrá menor dimensión, con la consecuente menor carga computacional, a expensas de pérdida de información de la imagen. Incluso es posible seleccionar un *stride* diferente horizontal y verticalmente si las características de interés se restringen a una de las direcciones. Para el proyecto actual, no se requiere de un gran coste computacional puesto que las imágenes no son de gran dimensión y se puede seleccionar un *stride* de 1.

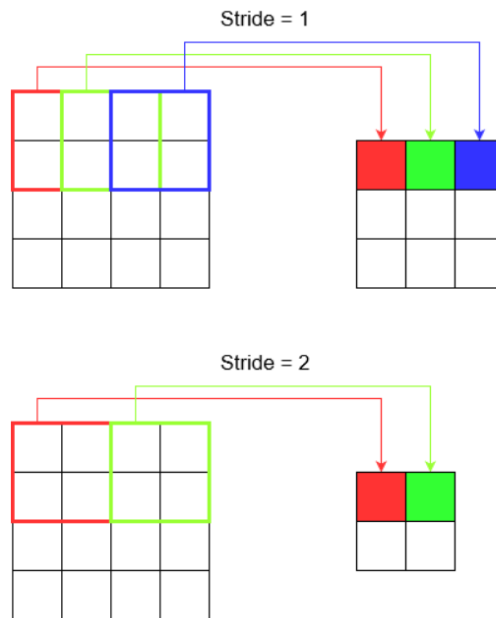


Figura 21: Dimensión de la matriz de salida según el *stride* seleccionado (Colomer y Muñoz, s.f.)

- **Tamaño del *kernel*.** Con que mayor sea el tamaño del *kernel* menor será la dimensión de la matriz resultante, puesto que la imagen se barre con menos operaciones al ocupar el filtro campos receptivos más grandes. Lo normal es que el *kernel* sea cuadrado salvo casos específicos y el tamaño habitual es 3x3. Por tanto ese es el *kernel* usado en el proyecto.
- ***Padding*.** De lo anterior se deduce que las matrices de salida tienen siempre menor tamaño que las imágenes de entrada, aún seleccionando *stride* unidad.

Para el tamaño de imagen utilizado, 40x50, las matrices resultantes serían de 39x49 a no ser que se aplique *padding*, en cuyo caso el tamaño de la matriz sería el mismo que el de entrada. El *padding* consiste en rellenar con píxeles los alrededores de la imagen para que los *kernels* barran también esas regiones. Lo habitual es rellenar con ceros (píxeles negros), lo que se conoce como *zero padding* y es el que se utiliza en el proyecto. Lógicamente, los ceros no afectan al funcionamiento del algoritmo, solo permiten que no se pierda información de la imagen de entrada. El inconveniente del *padding* es que los píxeles de los bordes de la imagen tienen menos peso en las matrices de salida y por tanto también están menos representados para el clasificador. Esto no es un problema para el caso de estudio puesto que las fuentes tras el preprocesamiento no ocupan los píxeles de los bordes (a pesar de aplicar traslaciones no tocan los bordes gracias al tamaño estándar escogido) por lo que estos son menos relevantes que los píxeles centrales.

0	0	0	0	0	0	0
0	60	113	56	139	85	0
0	73	121	54	84	128	0
0	131	99	70	129	127	0
0	80	57	115	69	134	0
0	104	126	123	95	130	0
0	0	0	0	0	0	0

Kernel

0	-1	0
-1	5	-1
0	-1	0

114	328	-26	470	158
53	266	-61	-30	344
403	116	-47	295	244
108	-135	256	-128	344

Figura 22: Esquema de *zero padding* para conservar el tamaño de imagen (Dharmaraj, 2021)

- *Pooling*. El *pooling* es una herramienta que reduce en gran medida la dimensión de las matrices de salida. Consiste en aplicar una operación sobre cada vecindario de píxeles (tamaño a seleccionar) de la matriz de salida para obtener otra de menor dimensión. La operación más utilizada, al haber demostrado mayor rendimiento, es tomar el máximo valor del conjunto de píxeles (*maxpooling*). Esta técnica permite reducir el coste computacional y quedarse con las características más relevantes pero no es recomendable si se desean buscar dichas características en regiones concretas de la imagen. Por ejemplo, si en una tarea se necesitase buscar una curva concreta en la esquina superior derecha de la imagen no podría aplicarse *pooling*.

El algoritmo detectaría que existe la curva pero no sabría localizarla espacialmente con precisión, puesto que esa información espacial se pierde en el *pooling*. En la tarea actual se debe identificar la forma de las fuentes aisladas o solapadas pero no importa si se encuentran más o menos desplazadas horizontal o verticalmente, dado que en el *data augmentation* se aplican traslaciones, por lo que el *pooling* puede implementarse. El tamaño del *maxpooling* utilizado es el usual, 2x2, por lo que se busca el máximo en conjuntos de 4 píxeles; es decir, para la entrada de 40x50 la matriz de salida es de 20x25.

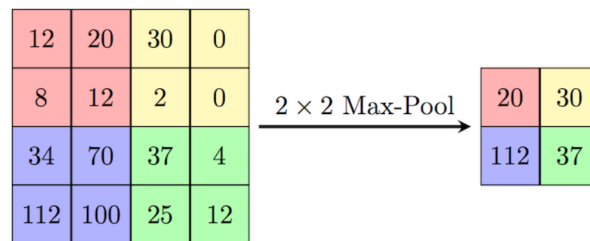


Figura 23: Esquema de *maxpooling* 2x2 ([Computer Science Wiki, 2018](#))

La CNN está formada por varias capas convolucionales que a su vez se conforman de varios *kernels*, cada uno con unos coeficientes arbitrarios que se van optimizando a medida que la CNN entrena. Las capas se agrupan en etapas convolucionales. En la misma etapa todas las capas poseen el mismo número de *kernels* y en cada etapa consecutiva el número de *kernels* aumenta. De esta manera cada etapa encuentra características cada vez más complicadas. En el ejemplo de clasificación de perros y gatos la primera etapa detectaría líneas, la segunda curvas y así consecutivamente hasta poder detectar agrupaciones de elementos complejas como hocicos, colas y patas.

Como al operar con cada filtro se genera una matriz de salida, tras una capa convolucional se generarán n matrices, siendo n el número de filtros. Si las imágenes fueran a color el *kernel* operaría sobre el valor de píxel de los tres canales (rojo, verde, azul) obteniéndose tres matrices de salida, que se combinarían para formar una única matriz. El conjunto de matrices resultantes de cada capa se hacen llamar mapa de activación o características. Para una imagen de 40x50, tras una capa de 32 filtros, se obtendría un mapa de tamaño 32x40x50 (con *zero padding* y sin *pooling*). En la siguiente capa de 32 filtros cada uno de ellos operaría paralelamente sobre las 32 matrices de salida de la anterior generando una sola matriz nueva y se obtendría de nuevo un mapa de 32x40x50.

En la CNN del proyecto se incluye una sola etapa convolucional, suficiente para la complejidad de la tarea, formada por dos capas convolucionales, ambas de 32 filtros. Es importante mencionar que antes de pasar de una capa convolucional a otra se aplica otra operación sobre cada matriz de salida, lo que se conoce como función de activación. La función más ampliamente usada por su eficiencia es la ReLu, que básicamente convierte los valores negativos de la matriz a cero y deja los positivos intactos. En caso de utilizar *pooling* este es el último paso de todos y se aplica sobre la matriz resultante de la función de activación.

Por tanto las dos capas convolucionales extraen las características diferenciadoras de las imágenes de las dos clases (fuentes aisladas y mezcladas) pero para poder llevar a cabo la clasificación deben acoplarse a una red neuronal convencional tras aplanar la información de salida (convertir a una dimensión, *flatten*). Entonces, la última fase de la CNN contiene dos capas convencionales (capas densas): una de 64 neuronas y la final, la de salida, de dos neuronas, que coincide con el número de clases a clasificar. Si el problema fuese más complejo serían necesarias más capas densas, cuyo número de neuronas debería ir reduciéndose, pero en este caso basta con una sola capa densa previa a la de salida.

Con esto ya se podría construir una arquitectura básica de CNN para clasificar, pero existen herramientas adicionales, conocidas como técnicas de regularización, que previenen el sobreaprendizaje u *overfitting*. El sobreaprendizaje ocurre cuando el algoritmo se especializa demasiado en las imágenes sobre las que aprende, es decir las de entrenamiento, pero no generaliza bien a nuevas imágenes (imágenes de *test*). Esto es debido a que la arquitectura de la CNN cuenta con muchos parámetros a entrenar (los coeficientes de todos los filtros) lo que equivale a un modelo muy complejo que logra ajustarse con gran precisión a los datos que “ve”. Siguiendo con el símil del alumno, el sobreaprendizaje equivale a que invierta todo su tiempo de estudio para el examen de biología a especializarse en las plantas: acertará con la máxima puntuación las preguntas sobre las plantas pero no cubrirá las preguntas sobre el reino animal. Lo idóneo es que adquiera conocimientos más básicos en el reino animal y vegetal y pueda superar con una puntuación razonable cualquier pregunta a la que se enfrente. Volviendo al *machine learning*, la solución ante el *overfitting* es por tanto reducir la complejidad del modelo, de manera que no se ajuste con tanta fidelidad a las muestras de entrenamiento pero obtenga mejor rendimiento ante muestras de *test*.

Una vía para conseguirlo es usar *dropout*, que de manera simplificada, consiste en desactivar o ignorar los coeficientes de algunos *kernels* durante el proceso de entrenamiento.

De esta forma se impide que el modelo alcance su máxima potencia. El porcentaje de coeficientes a desactivar puede regularse aunque no es recomendable superar el 50 % puesto que podría derivar en un modelo demasiado pobre. Otra técnica de regularización es el *batch normalization* que normaliza los valores de las matrices de salida para que no alcancen valores demasiado altos, factor que puede influir en el aprendizaje del algoritmo: como se ha comentado, la función ReLu limita el crecimiento de los números negativos pero no modifica los positivos y estos pueden aumentar rápidamente al realizarse los productos de la convolución.

Tanto *dropout* como *batch normalization* puede aplicarse tras las capas convolucionales o en las capas densas de la parte convencional de la arquitectura. Se emplea *dropout* al final de la etapa convolucional con un porcentaje del 25 % y tras la capa densa de 64 neuronas, con un porcentaje del 50 %. El *batch normalization* se introduce después de las dos capas convolucionales y también después de la de 64 neuronas.

En la imagen 24 se muestra la arquitectura de la CNN donde se refleja todo lo explicado, creada usando Keras (biblioteca de *deep learning* en Python) ejecutada sobre TensorFlow (biblioteca de código abierto de *machine learning* desarrollada por Google). El modelo se crea indicando cual será el tamaño de entrada (imágenes de 40x50 y un canal) y cual será la salida: capa densa de 2 neuronas (número de clases) con función de activación *softmax*, la adecuada para problemas de clasificación, que devolverá la probabilidad de pertenencia a cada clase entre 0 y 1. Al haber solo dos clases (clasificación binaria) también podría usarse la función de activación *sigmoid*.

En la imagen 25 puede verse el resumen de la arquitectura aportado por Keras, con el número de parámetros a entrenar (coeficientes de los filtros) en cada una de las capas y el total. Es decir el modelo durante el entrenamiento encontrará los valores óptimos de más de un millón de parámetros, derivados de las imágenes de entrada. Pueden parecer muchos, pero de trabajar con una red neuronal convencional en vez de una CNN el número de parámetros sería aún más ingente.

Finalmente en la figura 26 se muestra un esquema de los pasos de la arquitectura junto a las dimensiones de salida de cada una de las etapas. Se aprecia como en el *flatten* la información pasa de ser multidimensional a unidimensional y como la salida del modelo es un vector de dos elementos con la probabilidad de pertenecer a cada una de las dos clases.

La predicción del clasificador se corresponderá con la de mayor probabilidad.

```

▶ input = Input(shape=(40, 50, 1))

x1 = Conv2D(32, (3, 3), padding = 'same', activation = 'relu')(input)
x1 = BatchNormalization()(x1)
x1 = Conv2D(32, (3, 3), padding = 'same', activation = 'relu')(x1)
x1 = BatchNormalization()(x1)
x1 = MaxPooling2D((2, 2))(x1)
x1 = Dropout(0.25)(x1)

xtp = Flatten()(x1)
xtp = Dense(64, activation = 'relu')(xtp)
xtp = BatchNormalization()(xtp)
xtp = Dropout(0.5)(xtp)

prediccion = Dense(2, activation='softmax')(xtp)

model = Model([input], prediccion)

```

Figura 24: Código de la arquitectura de la CNN

Model: "model"

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 40, 50, 1)]	0
conv2d (Conv2D)	(None, 40, 50, 32)	320
batch_normalization (BatchNormalization)	(None, 40, 50, 32)	128
conv2d_1 (Conv2D)	(None, 40, 50, 32)	9248
batch_normalization_1 (BatchNormalization)	(None, 40, 50, 32)	128
max_pooling2d (MaxPooling2D)	(None, 20, 25, 32)	0
dropout (Dropout)	(None, 20, 25, 32)	0
flatten (Flatten)	(None, 16000)	0
dense (Dense)	(None, 64)	1024064
batch_normalization_2 (BatchNormalization)	(None, 64)	256
dropout_1 (Dropout)	(None, 64)	0
dense_1 (Dense)	(None, 2)	130

=====
 Total params: 1,034,274
 Trainable params: 1,034,018
 Non-trainable params: 256

Figura 25: Resumen de la arquitectura de la CNN con el número de parámetros

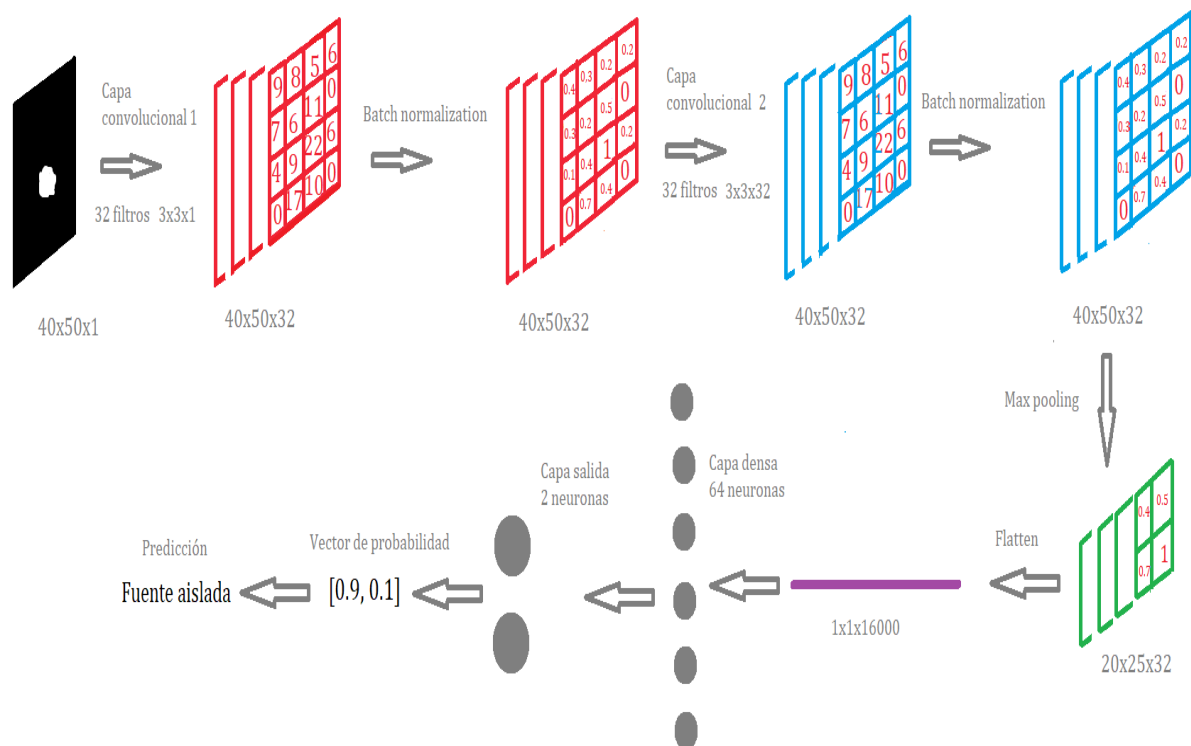


Figura 26: Esquema de los pasos de la CNN junto a las dimensiones

3.5 Compilado y entrenamiento

Una vez creado el código para el modelo es hora de compilarlo y entrenarlo. El compilado es una preparación previa al entrenamiento que básicamente transforma las capas definidas en matrices eficientes para el cálculo en la GPU y requiere de la definición de tres parámetros: la función de pérdida (*loss function*), el algoritmo de optimización y la métrica a registrar. La función de pérdida es aquella que ha de minimizarse para hallar los valores óptimos de los parámetros del modelo. Cuando se selecciona *softmax* como función de activación de la capa de salida la función de pérdida apropiada es *categorical cross entropy*, que requiere pasar las etiquetas a binario, como se comentó anteriormente. Si se quisiese trabajar con las etiquetas enteras (0 y 1) habría que escoger como función de pérdida *sparse categorical cross entropy*. La *categorical cross entropy* calcula la diferencia entre las etiquetas predichas para las imágenes de entrenamiento y las etiquetas reales. El objetivo es por tanto que dicha diferencia disminuya durante el entrenamiento hasta encontrar el mínimo: el valor de los parámetros en ese instante será el óptimo.

El algoritmo de optimización es el encargado de buscar el mínimo de la función de pérdida. El usado en la CNN es el más habitual, el Adam (*Adaptive Moment Estimation*) que es bastante eficaz buscando el mínimo absoluto de la función y no se estanca en mínimos locales. Al Adam solo hay que pasarle la tasa de aprendizaje o *learning rate* con la que debe trabajar, que indica con qué rapidez ha de aproximarse al mínimo: con que más pequeño sea más cautelosos serán los pasos del algoritmo, reduciendo el riesgo de saltarse el mínimo a expensas de un mayor coste computacional; si la tasa es demasiado grande, seguramente no encuentre al mínimo. Lo ideal es fijar un valor intermedio que encuentre el mínimo en un tiempo de búsqueda razonable. Por último, la métrica se trata de la magnitud que se registrará durante el entrenamiento para poder visualizarla al finalizar y será un indicativo de si ha sido fructífero o no. La métrica más adecuada para una clasificación es la exactitud o *accuracy*, que se trata del porcentaje de aciertos respecto al total de predicciones.

Para poder entrenar al modelo hay que tomar tres decisiones más: considerar o no conjunto de validación y fijar el tamaño de lote (*batch size*) y el número de épocas (*epochs*). El conjunto de validación se trata de un grupo de imágenes de entrenamiento que se separan de las demás para que el algoritmo no aprenda sobre ellas y sean nuevas para él. Es por tanto un conjunto de *test* pero que se utiliza durante el entrenamiento del modelo y no a posteriori y permite hacerse una idea de cual será el poder de generalización del algoritmo. Si los resultados sobre las imágenes de validación son insatisfactorios también lo serán sobre el conjunto de *test*, por lo que podrán realizarse las modificaciones pertinentes hasta obtener resultados decentes sobre el conjunto de validación; en el momento que esto ocurra y nunca antes, podrá testearse el modelo con imágenes de *test*. Entonces, lo natural es reservar un porcentaje de las imágenes de entrenamiento para validación, aunque esto implique contar con menos muestras para el aprendizaje; de lo contrario el entrenamiento es a ciegas y no se tendrá una idea de la calidad del clasificador hasta aplicarlo sobre las imágenes de *test*. El porcentaje destinado a validación es del 20 % por lo que finalmente, de las 3204 imágenes de entrenamiento iniciales, 2563 serán propiamente de entrenamiento y 641 de validación. La separación de ambos conjuntos puede realizarse nuevamente con Keras, introduciendo *validation split=0,2* en los argumentos de *model.fit* (figura 27).

Además, durante el entrenamiento los datos de *train* se separan en subconjuntos (lotes o *batches*) que van pasando a la CNN de uno en uno; al terminar con el primero de los lotes se actualizarían los parámetros del modelo y se introduciría el siguiente lote.

Una vez que todos los lotes (totalidad de los datos de *train*) han pasado una vez por el algoritmo se completa una época. Si se escoge un tamaño de lote demasiado pequeño lo más probable es que los valores de los parámetros oscilen de manera más brusca y el entrenamiento sea más caótico; si por el contrario se toman lotes demasiado grandes, habrá menos actualizaciones de parámetros y el entrenamiento será más lento, teniendo que pasar el conjunto de entrenamiento más veces al algoritmo (más épocas). Después de probar con varios valores, todos potencias de dos como es usual (8,16,32...), el que brinda mejor comportamiento es el de 64. Por consiguiente las 3204 imágenes de entrenamiento se dividen en 41 grupos: 40 de 64 y otro de tan solo 3, al no ser 3204 divisible entre 64. En cada época se llevan a cabo entonces 41 actualizaciones de los parámetros. Si indicamos el argumento *shuffle=True*, Keras permuta aleatoriamente las imágenes de entrenamiento al inicio de cada época, de manera que la distribución en los lotes cambia y el entrenamiento es más eficiente.

El número de épocas óptimo no puede conocerse a priori, también habría que probar con varios valores: si fijamos uno demasiado pequeño puede que el algoritmo no haya tenido tiempo a alcanzar el mínimo de la función de pérdida y no devuelva los valores óptimos; si es demasiado grande el algoritmo puede alcanzar sobreaprendizaje, lo que tampoco es bueno como ya se ha comentado. Pero existe un mecanismo muy útil para evitarse multitud de pruebas: la parada temprana (*early stopping*). En el momento en el que la métrica seleccionada, en este caso la función de pérdida sobre el conjunto de validación (*val_loss*), no mejore tras tres épocas (u otro valor deseado), el entrenamiento se parará y se guardarán los parámetros correspondientes al mejor valor de la pérdida. La parada temprana, al impedir el *overfitting* es también considerada una técnica de regularización.

En resumen, durante el entrenamiento se van actualizando los valores de los parámetros tras cada lote o *batch* con el objetivo de minimizar la función de pérdida sobre los datos de entrenamiento. Al final de cada época se registra y muestra por pantalla la pérdida para el conjunto de entrenamiento y para el de validación junto con la métrica escogida de ambos conjuntos, en este caso la *accuracy*. Cuando la función de pérdida en la validación deja de mejorar se para el entrenamiento al haber alcanzado el máximo poder de generalización del modelo y se guardan los parámetros óptimos.

En la imagen [27](#) puede verse la pérdida y la *accuracy* de entrenamiento y validación. Estos valores son de los mejores que se pueden obtener durante un entrenamiento.

Las pérdidas parten de valores bastante bajos desde el primer lote y las *accuracy* de valores muy cercanos a la unidad (valor óptimo), lo que indica que el algoritmo aprende muy fácilmente de las imágenes de entrenamiento. A lo largo de las épocas se refinan los resultados hasta que se alcanzan tres épocas sin mejora alguna, cuando la parada temprana actúa.

Se ha lidiado con el sobreaprendizaje gracias a la regularización. En caso de presentarse, las pérdidas de entrenamiento disminuirían mucho más bruscamente que las de validación y la *accuracy* aumentaría también más bruscamente. Además, el modelo arroja un porcentaje de acierto muy bueno en el conjunto de validación; un modelo muy pobre se acercaría al comportamiento estocástico, es decir, un acierto del 50 % y uno modesto no sobrepasaría el 80 %.

En las figuras 28 y 29 se grafican las pérdidas y *accuracy* de ambos conjuntos respecto a la época. Se confirma que el comportamiento es el deseado, puesto que la curva de validación converge a la de entrenamiento en ambas gráficas, salvo oscilaciones razonables. Si existiese *overfitting* la pérdida de validación se encontraría por encima de la de entrenamiento, a una distancia mayor y la *accuracy* de validación estaría por debajo de la de entrenamiento, también a mayor distancia.

```
model.compile(Adam(learning_rate=0.001), loss=['categorical_crossentropy'], metrics=['accuracy'])
MyCallbacks = [EarlyStopping(monitor="val_loss",patience=3,restore_best_weights=True)]
M=model.fit(x_train, y_train, batch_size=64, epochs=25, validation_split=0.2,shuffle=True,callbacks=MyCallbacks)
```

```
Epoch 1/25
41/41 [=====] - 20s 445ms/step - loss: 0.1045 - accuracy: 0.9637 - val_loss: 0.1468 - val_accuracy: 0.9719
Epoch 2/25
41/41 [=====] - 20s 490ms/step - loss: 0.0382 - accuracy: 0.9867 - val_loss: 0.0473 - val_accuracy: 0.9844
Epoch 3/25
41/41 [=====] - 18s 438ms/step - loss: 0.0330 - accuracy: 0.9902 - val_loss: 0.0450 - val_accuracy: 0.9860
Epoch 4/25
41/41 [=====] - 18s 437ms/step - loss: 0.0246 - accuracy: 0.9922 - val_loss: 0.0279 - val_accuracy: 0.9891
Epoch 5/25
41/41 [=====] - 18s 436ms/step - loss: 0.0169 - accuracy: 0.9957 - val_loss: 0.0357 - val_accuracy: 0.9860
Epoch 6/25
41/41 [=====] - 18s 430ms/step - loss: 0.0162 - accuracy: 0.9957 - val_loss: 0.0333 - val_accuracy: 0.9875
Epoch 7/25
41/41 [=====] - 17s 426ms/step - loss: 0.0117 - accuracy: 0.9969 - val_loss: 0.0167 - val_accuracy: 0.9953
Epoch 8/25
41/41 [=====] - 18s 437ms/step - loss: 0.0085 - accuracy: 0.9984 - val_loss: 0.0121 - val_accuracy: 0.9969
Epoch 9/25
41/41 [=====] - 19s 472ms/step - loss: 0.0066 - accuracy: 0.9992 - val_loss: 0.0157 - val_accuracy: 0.9922
Epoch 10/25
41/41 [=====] - 18s 441ms/step - loss: 0.0075 - accuracy: 0.9980 - val_loss: 0.0065 - val_accuracy: 0.9969
Epoch 11/25
41/41 [=====] - 18s 428ms/step - loss: 0.0066 - accuracy: 0.9969 - val_loss: 0.0115 - val_accuracy: 0.9969
Epoch 12/25
41/41 [=====] - 17s 424ms/step - loss: 0.0072 - accuracy: 0.9984 - val_loss: 0.0173 - val_accuracy: 0.9938
Epoch 13/25
41/41 [=====] - 18s 428ms/step - loss: 0.0085 - accuracy: 0.9973 - val_loss: 0.0095 - val_accuracy: 0.9984
```

Figura 27: Compilado y entrenamiento de la CNN

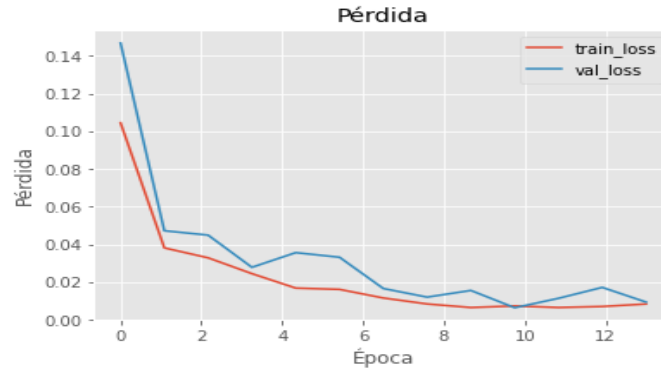


Figura 28: Función de pérdida de los conjuntos de entrenamiento y validación durante el entrenamiento

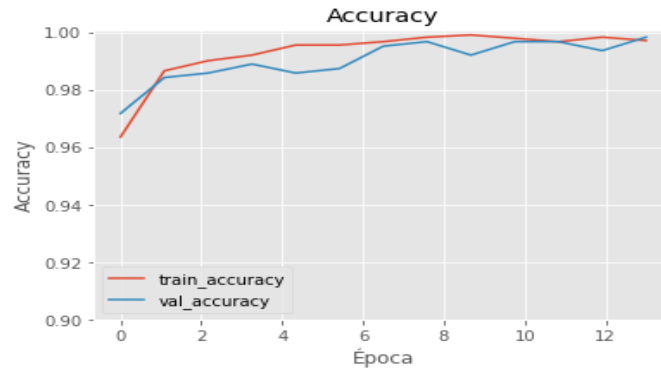


Figura 29: *Accuracy* de los conjuntos de entrenamiento y validación durante el entrenamiento

Los valores óptimos de los parámetros suponen una pérdida de validación por debajo de las 2 centésimas y una *accuracy* por encima de 0,98. Estas métricas son lo suficientemente favorables para pasar a la siguiente etapa: comprobación o testeo del modelo.

3.6 Comprobación y resultados

Primero se comprueba el modelo sobre el conjunto de *test* de Varuna: las 801 imágenes reservadas en su momento para el testeo. Estas contienen 408 fuentes aisladas y 393 muestras mezcladas. Para ello se obtienen las etiquetas predichas por el modelo para esas imágenes con *model.predict* de Keras y se comparan con las etiquetas reales mediante el reporte de *classification_report* de la librería de *machine learning* en Python *scikit-learn*.

El reporte contiene cuatro de las métricas más usadas para conocer el rendimiento de un modelo: *accuracy* o exactitud, *precision* o precisión, *recall* o exhaustividad y *f1-score* o *valor-F*.

Como ya se ha comentado, la *accuracy* se corresponde con el número de aciertos, teniendo en cuenta los de todas las clases, entre el total de muestras.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

TP= verdaderos positivos. Fuentes aisladas clasificadas por el modelo como aisladas.

TN= verdaderos negativos. Fuentes mezcladas clasificadas por el modelo como mezcladas.

FP= falsos positivos. Fuentes mezcladas clasificadas por el modelo como fuentes aisladas.

FN= falsos negativos. Fuentes aisladas clasificadas por el modelo como mezcladas.

La precisión, el *recall* y el *f1-score* se calculan para cada clase, pero el reporte también contiene su media global (*macro avg*) y la media ponderada (*weighted avg*) que tiene en cuenta el número de muestras de cada clase:

$$macro\ avg = 0,5 \cdot magnitud_{aisladas} + 0,5 \cdot magnitud_{mezclas} \quad (4)$$

$$weighted\ avg = \frac{n^\circ\ muestras\ aisladas}{n^\circ\ muestras\ total} \cdot magnitud_{aisladas} + \frac{n^\circ\ muestras\ mezcladas}{n^\circ\ muestras\ total} \cdot magnitud_{mezclas} \quad (5)$$

La precisión indica el porcentaje de muestras clasificadas en una clase que realmente pertenecen a esa clase.

$$precision_{aisladas} = \frac{TP}{TP + FP} \quad (6)$$

$$precision_{mezclas} = \frac{TN}{TN + FN} \quad (7)$$

El *recall* se corresponde con el porcentaje de muestras que han sido clasificadas en una clase correctamente del total de muestras que pertenecen a esa clase.

$$recall_{aisladas} = \frac{TP}{TP + FN} \quad (8)$$

$$recall_{mezclas} = \frac{TN}{TN + FP} \quad (9)$$

Por último el *f1-score* es un cálculo en el que interviene la precisión y el *recall*:

$$f1_{aisladas} = 2 \cdot \frac{precision_{aisladas} \cdot recall_{aisladas}}{precision_{aisladas} + recall_{aisladas}} \quad (10)$$

$$f1_{mezclas} = 2 \cdot \frac{precision_{mezclas} \cdot recall_{mezclas}}{precision_{mezclas} + recall_{mezclas}} \quad (11)$$

En la figura 30 puede verse el reporte. El valor unidad de todas las métricas se traduce en que el clasificador no ha fallado en ninguna de las 801 imágenes de *test* de Varuna, es decir, las ha clasificado todas correctamente.

```
[ ] predictions = model.predict(x_test, batch_size=64)
    print(classification_report(y_test.argmax(axis=1), predictions.argmax(axis=1)))
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	408
1	1.00	1.00	1.00	393
accuracy			1.00	801
macro avg	1.00	1.00	1.00	801
weighted avg	1.00	1.00	1.00	801

Figura 30: Resultados del clasificador sobre el conjunto de *test* de Varuna

En el [apéndice](#) se adjunta un [enlace](#) al directorio de Google Drive donde se encuentran los 408 *croppies* con fuentes aisladas del conjunto de *test* de Varuna clasificados por el algoritmo. De ellos, 10 formaban parte de las detecciones negativas de *m2*, es decir habrían sido descartadas por ese algoritmo (ver figura 31). Entonces el clasificador permite recuperar falsos negativos de *m2*, una de sus meta principales.

Pero esto no es suficiente para comprobar el buen funcionamiento del clasificador; es necesario testarlo con imágenes de otro asteroide para ver si tiene un poder de generalización aceptable o por el contrario solo es aplicable a Varuna. Es ahí donde entra en juego el *dataset* de 2002_ms4, mucho más exigente, al contar con campos de fuentes muy densos y de peor calidad. Al igual que el de Varuna está separado en las detecciones positivas de *m2* (553 *croppies*) y negativas (745 *croppies*). Las imágenes astronómicas de 2002_ms4 han sido tomadas desde febrero de 2002 al mismo mes de 2022, como en el caso de Varuna.

De igual forma que con los de Varuna, se cambian los nombres originales de los *croppies* por su fecha, para poder identificar mezclas más fácilmente. A continuación se descartan todos los *croppies* citados en el apartado 3.2 con el mismo código aplicado para Varuna. Los *croppies* que pasan el filtro de calidad se corresponden con el conjunto de *test* sobre el que se comprobará el clasificador.

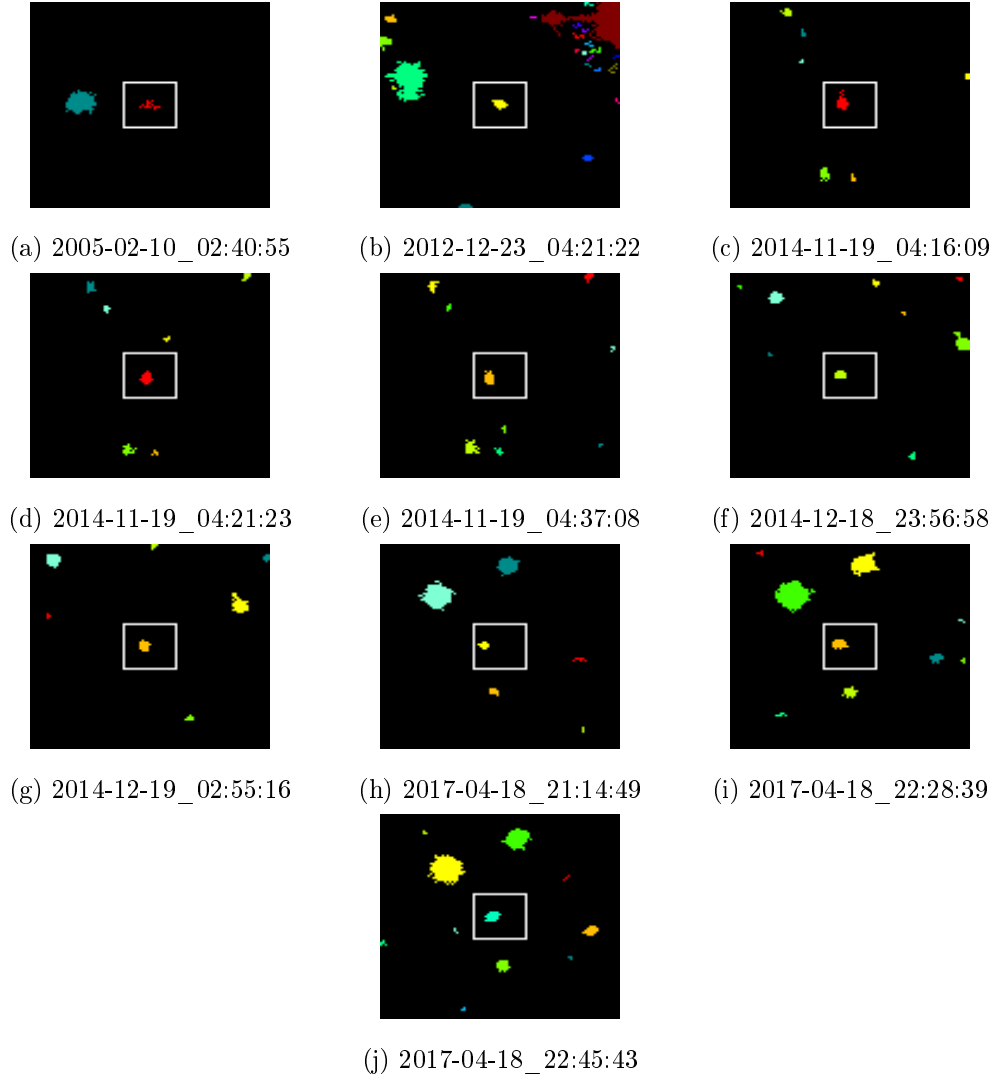


Figura 31: Falsos negativos de $m2$ recuperados por el clasificador

De las detecciones positivas de $m2$, 248 son descartadas mediante el código y 305 pasan el filtro; en cambio de las negativas solo 4 logran pasar el filtro. Esto es lógico puesto que los *croppies* de 2002_ms4 son de lo peor que puede encontrarse en cuanto a calidad y hay muy pocos no contaminados.

Para poder clasificar el conjunto de *test* hay que aplicarle previamente el mismo preprocesamiento que el conjunto de entrenamiento sobre el que ha entrenado la CNN. Se recuerda que el conjunto de *test* de Varuna ya estaba preprocesado, al ser extraído de los datos iniciales de entrenamiento. En la figura 32 se presentan ejemplos de *croppies* de *test* de 2002_ms4 junto a las imágenes preprocesadas asociadas que pasarán al clasificador.

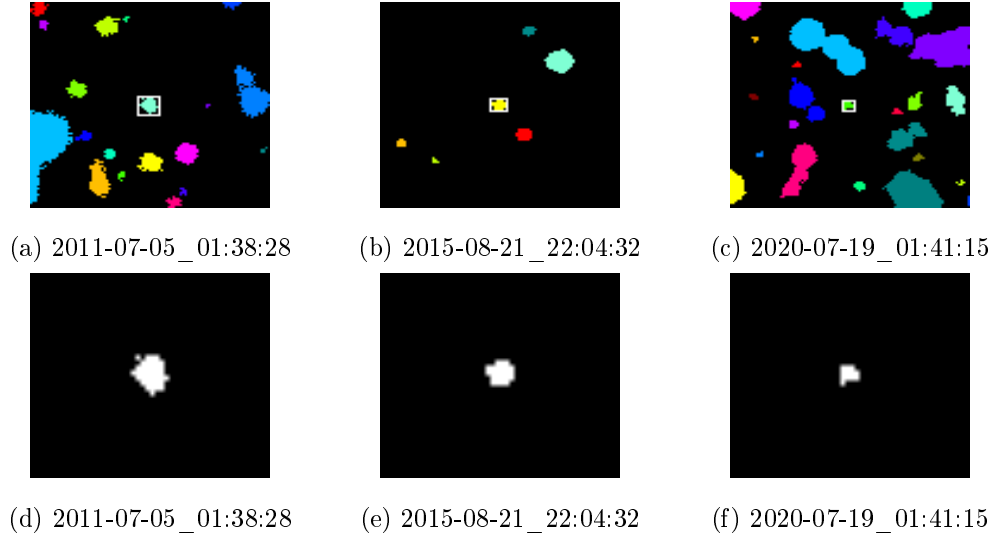


Figura 32: Ejemplos *croppies* de *test* de 2002_ms4, originales y preprocesados

En este caso, en vez de etiquetar manualmente los *croppies* antes de clasificar para poder comprobar los resultados del clasificador, se hará el proceso inverso: se clasifican los *croppies* y después se comprueba manualmente cuales ha acertado y cuales ha fallado.

La CNN clasifica, de los 309 *croppies* de *test*, 298 como fuentes aisladas y 11 como fuentes mezcladas. En el [apéndice](#) se adjunta un [enlace](#) al directorio de Google Drive donde se encuentran las 298 *croppies* clasificados como fuentes aisladas. Examinándolos se encuentran 5 falsos positivos, es decir, fuentes mezcladas que la CNN ha clasificado como aisladas, mostrados en la figura 33. Todos ellos pertenecen a la misma fecha (6 de junio de 2010). Probablemente que el clasificador no haya acertado se deba a que estas mezclas poseen formas diferentes a las contenidas en el conjunto de entrenamiento de Varuna y tamaños más reducidos.

Además los 11 *croppies* clasificados como fuentes mezcladas son falsos negativos, es decir, se tratan en realidad de fuentes aisladas. Ocho de ellos son de la misma fecha (30 de junio de 2017), son casi idénticos y se corresponden con los casos problemáticos para el clasificador: fuentes aisladas de gran tamaño y forma aproximadamente esférica que confunde con fuentes mezcladas muy solapadas. Esto es normal, pues la CNN ha entrenado sobre fuentes aisladas y mezclas muy superpuestas con formas prácticamente indistinguibles (ver figura 34).

De hecho, dichas mezclas fueron identificadas visualmente gracias a contar con las secuencias de *croppies* anteriores y posteriores; si es por tanto complicado distinguir esos casos para el ojo humano también lo será para la CNN y no pueden presentarse mejoras en este sentido. Por ende, en *croppies* de ese estilo la CNN “dudará”, atribuyendo probabilidades más equitativas a ambas clases y como consecuencia en algunos casos acertará con poca certeza y en otros fallará también con poca certeza.

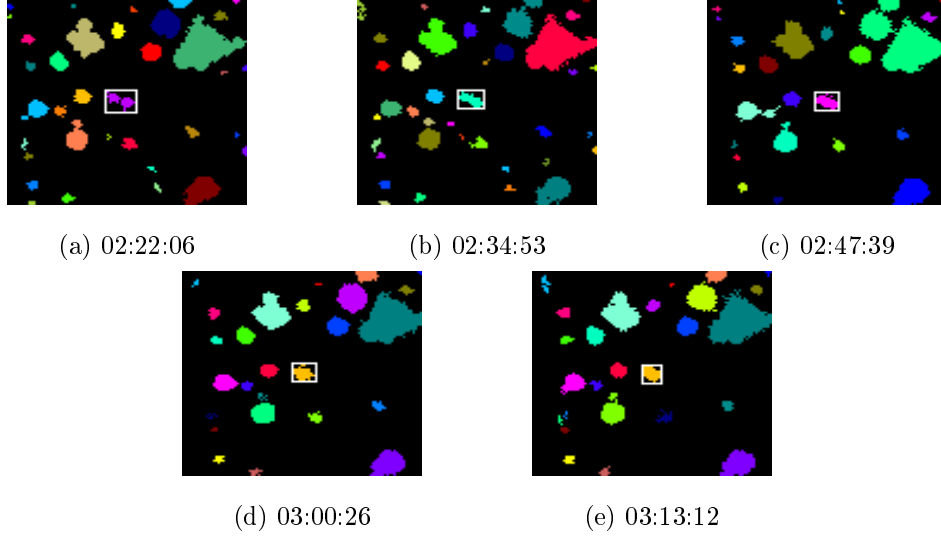


Figura 33: Falsos positivos del clasificador en el conjunto de *test* de 2002_ms4

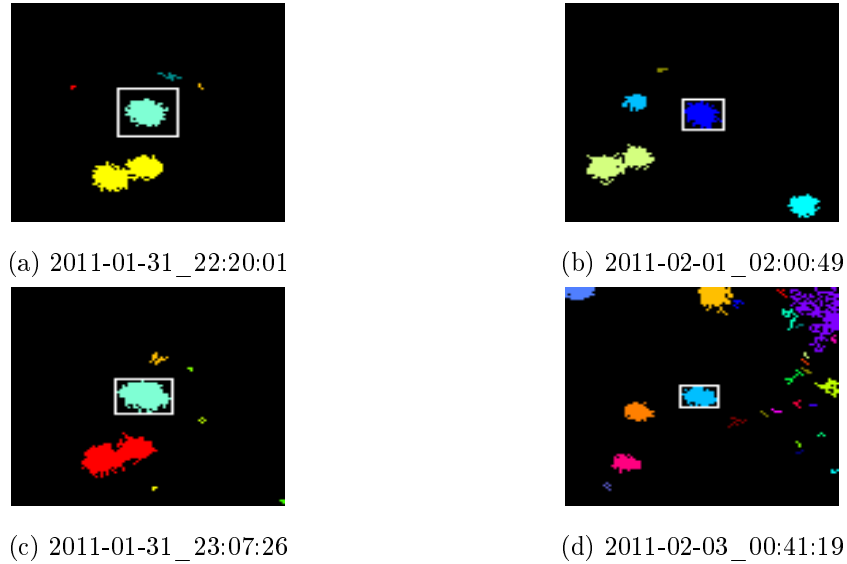


Figura 34: Mezclas solapadas indistinguibles con fuentes aisladas del conjunto de Varuna. La mezcla 34a es muy similar a la aislada 34b y la mezcla 34c similar a la aislada 34d

En la figura 35 se muestran los 8 falsos negativos del 30 de junio de 2017. En la figura 36 se muestran los 8 verdaderos positivos de la misma fecha, es decir los *croppies* correctamente clasificados como fuentes aisladas.

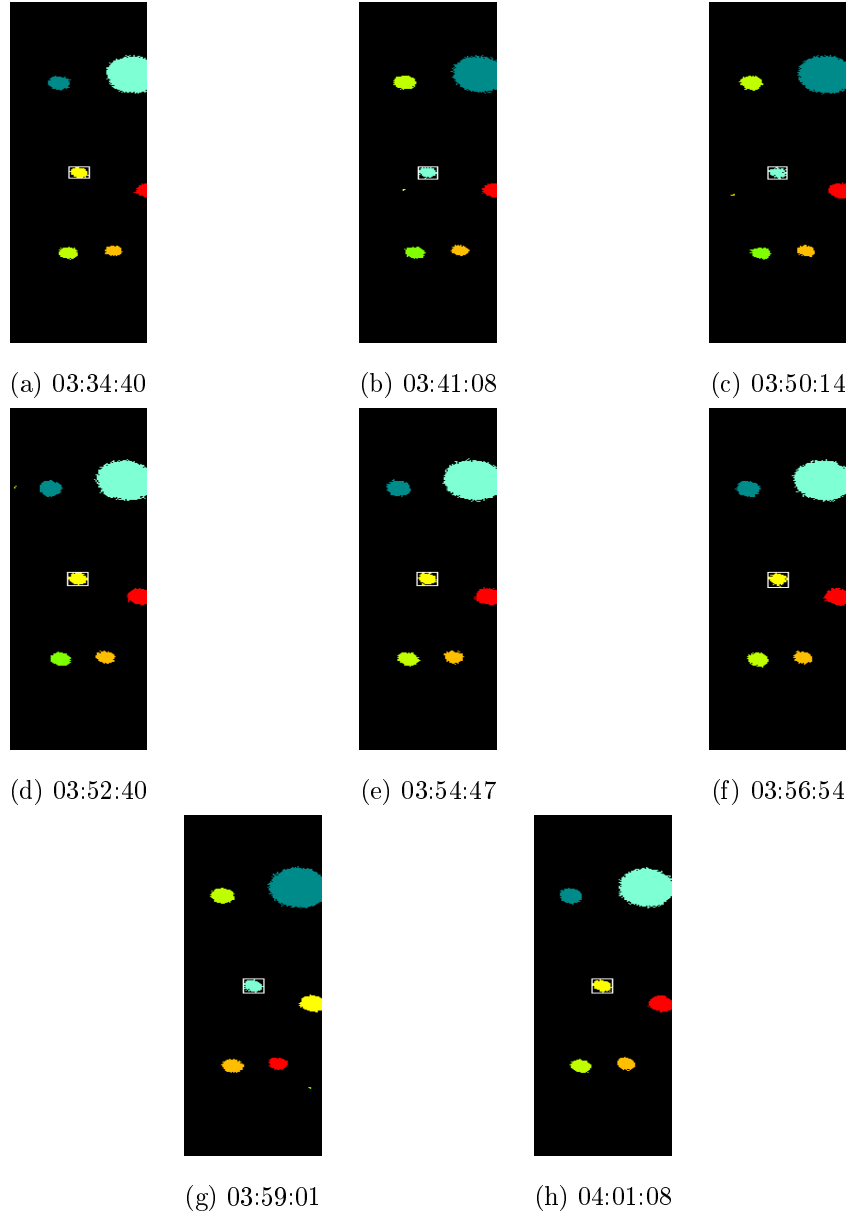


Figura 35: Falsos negativos del clasificador en el 30 de junio de 2017

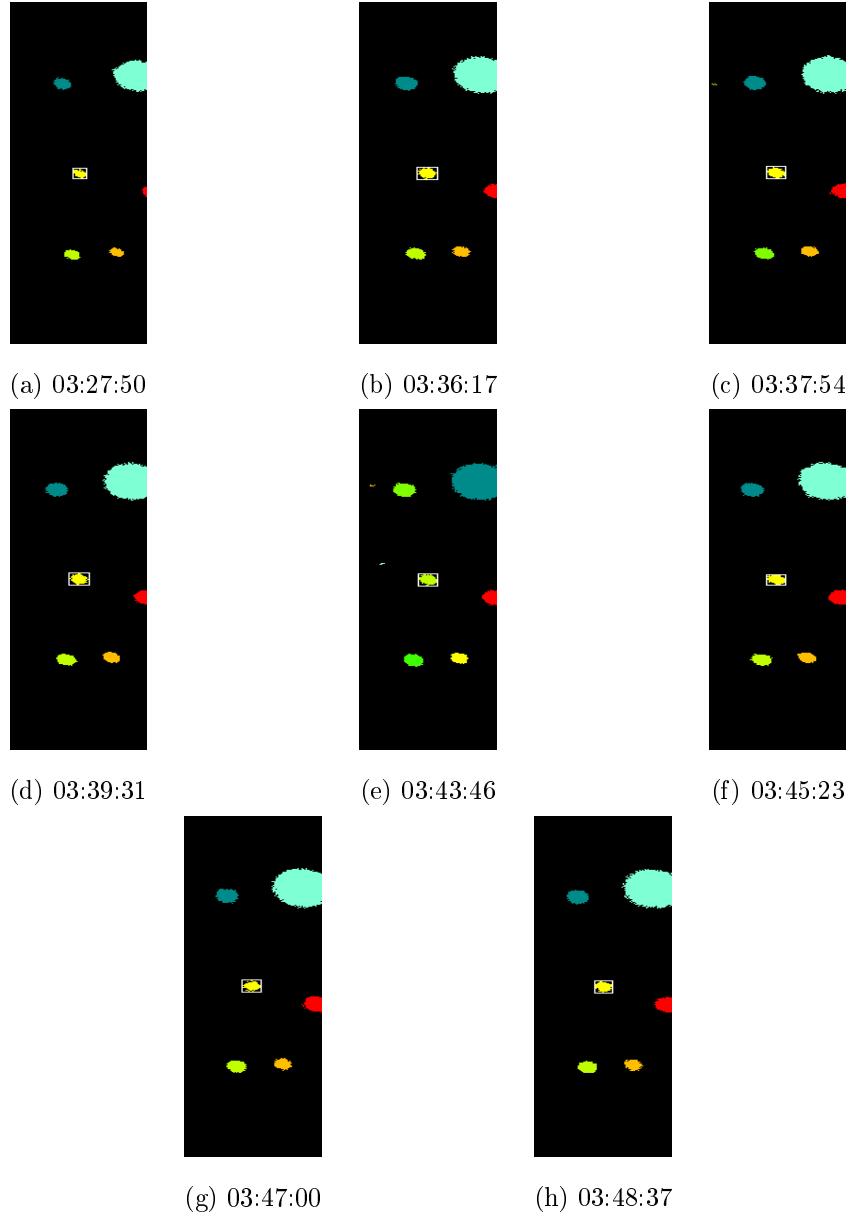


Figura 36: Verdaderos positivos del clasificador en el 30 de junio de 2017

De los tres falsos negativos restantes, el de 5 de julio de 2013 puede incluirse también entre los casos problemáticos al ser la forma de la fuente aislada similar a una mezcla solapada. En la figura 37 se presenta dicho falso negativo (37d) junto a los 8 verdaderos positivos de la misma fecha.

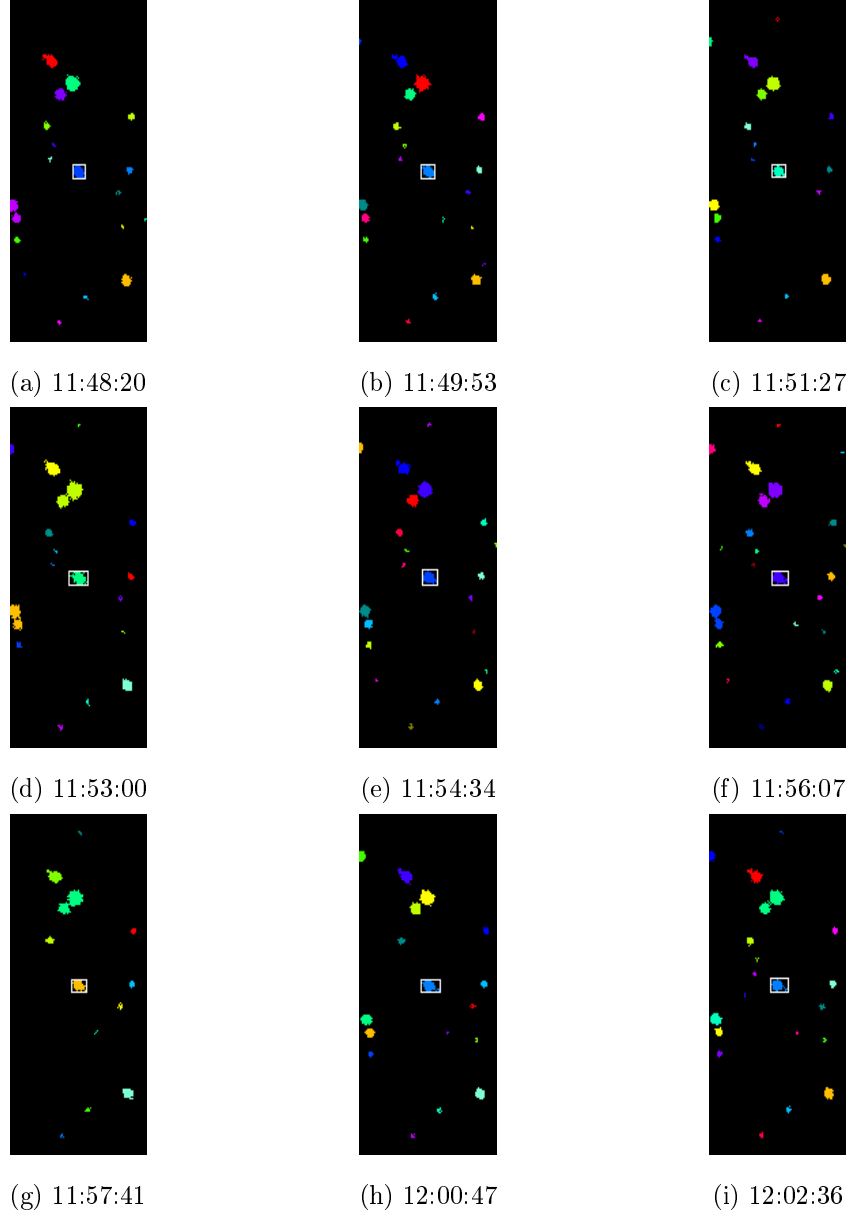
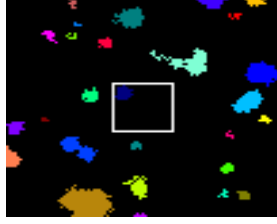


Figura 37: Falso negativo y verdaderos positivos del clasificador en el 5 de julio de 2013

Finalmente, los dos falsos negativos sobrantes no se encuentran dentro de los casos problemáticos y son simplemente fallos casuísticos del clasificador, puede que incentivados por la poco habitual localización del asteroide muy cercana a los bordes, aunque en general clasifica bien ese tipo de fuentes aisladas. Ambos falsos negativos formaban parte de las detecciones negativas de $m2$ por lo que también serían descartados por dicho algoritmo (figura 38). En cambio, los otros dos falsos negativos de $m2$ sí fueron recuperados por el clasificador (figura 39).

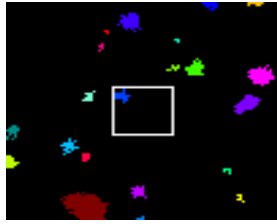


(a) 2013-07-10_00:42:14



(b) 2020-10-12_19:30:26

Figura 38: Falsos negativos de $m2$ y del clasificador



(a) 2013-07-10_00:50:45



(b) 2019-07-27_02:25:48

Figura 39: Falsos negativos de $m2$ recuperados por el clasificador

Entonces teniendo en cuenta los falsos positivos y negativos del clasificador las métricas sobre el conjunto de *test* de 2002_ms4 serían las siguientes:

$$TP = 293$$

$$TN = 0$$

$$FP = 5$$

$$FN = 11$$

$$precision_{aisladas} = \frac{293}{293 + 5} \approx 0,98$$

$$precision_{mezclas} = \frac{0}{11} = 0$$

$$recall_{aisladas} = \frac{293}{293 + 11} \approx 0,96$$

$$recall_{mezclas} = \frac{0}{5} = 0$$

$$f1_{aisladas} = 2 \cdot \frac{0,98 \cdot 0,96}{0,98 + 0,96} \approx 0,97$$

$$f1_{mezclas} = 2 \cdot \frac{0 \cdot 0}{0 + 0} = ind$$

$$precision_{media} = 0,5 \cdot 0,98 + 0,5 \cdot 0 \approx 0,49$$

$$precision_{media,ponderada} = \frac{304}{309} \cdot 0,98 + \frac{5}{309} \cdot 0 \approx 0,96$$

$$recall_{media} = 0,5 \cdot 0,96 + 0,5 \cdot 0 \approx 0,48$$

$$recall_{media,ponderada} = \frac{304}{309} \cdot 0,96 + \frac{5}{309} \cdot 0 \approx 0,94$$

$$f1_{media} = 0,5 \cdot 0,97 + 0,5 \cdot ind = ind$$

$$f1_{media,ponderada} = \frac{304}{309} \cdot 0,97 + \frac{5}{309} \cdot ind = ind$$

$$accuracy = \frac{293 + 0}{309} \approx 0,95 \rightarrow 95\%$$

	precision	recall	f1-score	número muestras
aisladas	0,98	0,96	0,97	304
mezcladas	0	0	ind	5
media	0,49	0,48	ind	309
media ponderada	0,96	0,94	ind	309
accuracy	0,95			309

Tabla 1: Métricas del clasificador sobre el conjunto de *test* de 2002_ms4

La precisión y *recall* del conjunto de mezcladas son nulas y el *f1-score* de las mezclas y las medias son indeterminados ya que ninguna de las 5 mezclas ha sido clasificada como tal. Esto es consecuencia del desequilibrio de las clases en el conjunto de *test*, debido a que la inmensa mayoría de mezclas del *dataset* sufren de contaminación y por lo tanto son eliminadas con el código de descarte. La media normal de las métricas no es representativa cuando existe desequilibrio pero sí la media ponderada, pues como se ha dicho anteriormente, considera el número de muestras de cada clase. Para evitar las indeterminaciones y conocer métricas representativas para las mezclas es necesario introducir más fuentes mezcladas en el conjunto de *test*.

Una opción es tomar las muestras mezcladas contaminadas que han sido descartadas a priori de la clasificación. Se incluyen entonces las 70 mezclas descartadas en el conjunto de *test*, se clasifican nuevamente y se calculan las métricas con el conjunto ampliado. Las clases siguen estando desequilibradas pero no tan desmesuradamente de forma que las métricas para las mezclas serán más representativas.

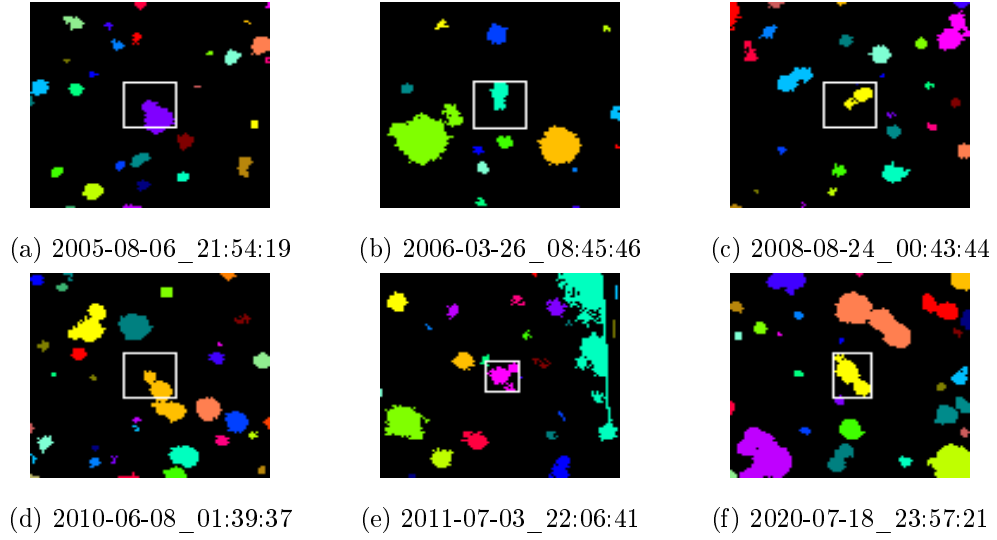


Figura 40: Ejemplos de las 70 mezclas contaminadas incluidas en el conjunto ampliado de *test* de 2002_ms4

De las 70 mezclas incluidas 62 son clasificadas como mezclas y 8 como aisladas (falsos positivos).

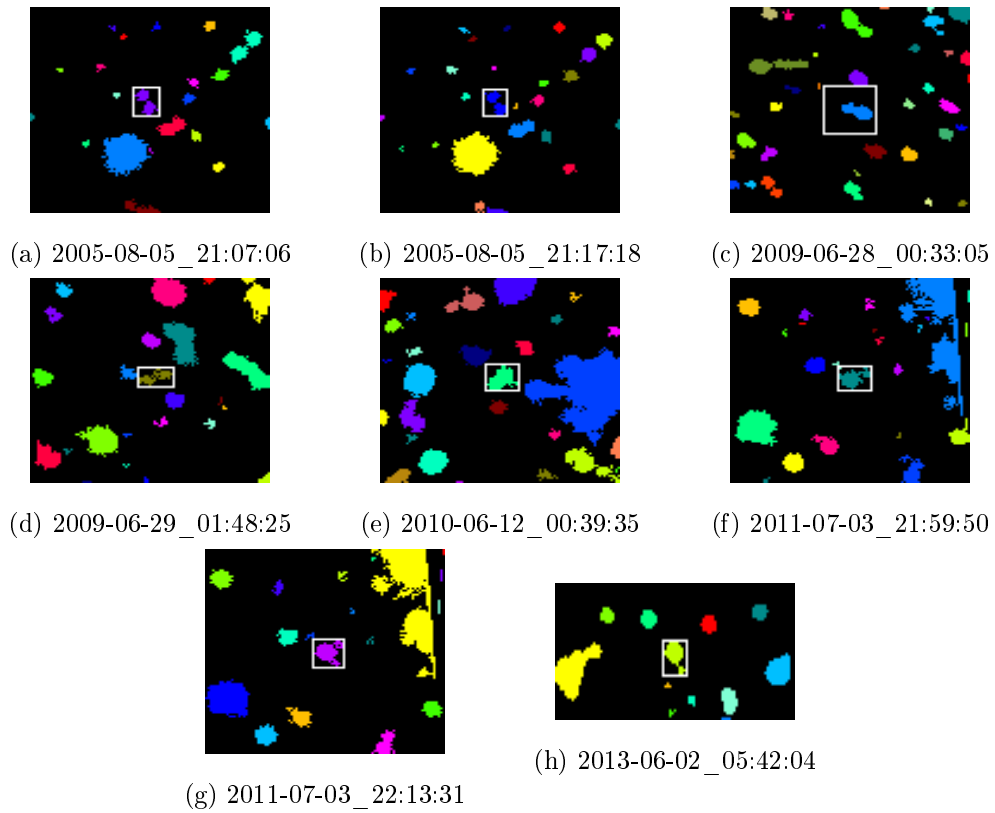


Figura 41: Falsos positivos del clasificador en las 70 nuevas muestras

Por lo tanto las métricas para el conjunto ampliado son las siguientes:

$$\begin{aligned}
TP &= 293 & TN &= 62 \\
FP &= 13 & FN &= 11 \\
precision_{aisladas} &= \frac{293}{293 + 13} \approx 0,96 & precision_{mezclas} &= \frac{62}{62 + 11} \approx 0,85 \\
recall_{aisladas} &= \frac{293}{293 + 11} \approx 0,96 & recall_{mezclas} &\approx \frac{62}{62 + 13} \approx 0,83 \\
f1_{aisladas} &= 2 \cdot \frac{0,96 \cdot 0,96}{0,96 + 0,96} \approx 0,96 & f1_{mezclas} &= 2 \cdot \frac{0,85 \cdot 0,83}{0,85 + 0,83} \approx 0,84 \\
precision_{media} &= 0,5 \cdot 0,96 + 0,5 \cdot 0,85 \approx 0,91 \\
precision_{media,ponderada} &= \frac{304}{379} \cdot 0,96 + \frac{75}{379} \cdot 0,85 \approx 0,94 \\
recall_{media} &= 0,5 \cdot 0,96 + 0,5 \cdot 0,83 \approx 0,90 \\
recall_{media,ponderada} &= \frac{304}{379} \cdot 0,96 + \frac{75}{379} \cdot 0,83 \approx 0,93 \\
f1_{media} &= 0,5 \cdot 0,96 + 0,5 \cdot 0,84 \approx 0,90 \\
f1_{media,ponderada} &= \frac{304}{379} \cdot 0,96 + \frac{75}{379} \cdot 0,84 \approx 0,94 \\
accuracy &= \frac{293 + 62}{379} \approx 0,94 \rightarrow 94\%
\end{aligned}$$

	precision	recall	f1-score	número muestras
aisladas	0.96	0.96	0.96	304
mezcladas	0.85	0.83	0.84	75
media	0.91	0.90	0.90	379
media ponderada	0.94	0.93	0.94	379
accuracy	0.94			379

Tabla 2: Métricas del clasificador sobre el conjunto de *test* de 2002_ms4 ampliado

Como era de esperar, no son tan buenas como las obtenidas para Varuna pero son aún así favorables. Las métricas de la clase mezcladas son más bajas que las de la clase aisladas.

Esto podría indicar que el clasificador se comporta mejor con las aisladas o ser consecuencia del aún existente desequilibrio entre ambas clases. Para conocer la opción correcta sería necesario contar con más muestras con mezclas.

Además, es más importante reducir el número de falsos negativos que el de falsos positivos, debido a que los negativos se descartarán en el procesamiento de otros *datasets* (en principio no se podrán recuperar) pero los falsos positivos podrán suprimirse en pasos posteriores. En el conjunto de *test* ampliado hay 11 falsos negativos y 13 falsos positivos, es decir, el número de falsos negativos es ligeramente menor, pero ambos son del mismo orden. Esto se traduce en que la precisión y el *recall* son iguales para las aisladas y la precisión de las mezcladas es ligeramente mayor que su *recall*. Como la mayoría de falsos negativos, como se ha visto, se tratan de los casos problemáticos para el clasificador (fuentes “indistinguibles”) no es posible una mejora de cara a reducir su número.

Finalmente se adjuntan en la tabla 3 las detecciones positivas y negativas reales del *dataset* de 2002_ms4 fruto de examinar todos los *croppies* con detenimiento (inspección humana) y la comparativa con los resultados del algoritmo vigente (*m2*) y la combinación del código de descarte y el clasificador creados en el proyecto.

En el *dataset* existen por tanto 304 detecciones positivas, es decir *croppies* con fuentes aisladas no contaminadas y 994 detecciones negativas, de las cuales 989 poseen alguno de los defectos tratados en la sección 3.2 (mala calidad, ausencia o tamaño pequeño del asteroide y/o contaminación) y solo 5 se tratan de *croppies* no contaminados pero que presentan mezcla (ver figura 33).

El algoritmo *m2*, como ya se ha visto, clasifica el *dataset* en 553 detecciones positivas y 745 negativas. De las 553 que clasifica como positivas, 253 son falsos positivos: 248 presentan alguno de los defectos de calidad y 5 se corresponden con las mezclas no contaminadas. De las 745 negativas, 4 son falsos negativos.

El código de descarte y clasificador del proyecto consiguen por su parte una clasificación muy similar a la real del *dataset*, salvo por los 5 falsos positivos y los 11 falsos negativos.

Inspección humana				
Detecciones positivas		Detecciones negativas		
Fuentes aisladas no contaminadas	304	-Mala o muy mala calidad -Ausencia de asteroide -Asteroide pequeño o disperso -Contaminados	989	994
		Fuentes mezcladas no contaminadas	5	
Algoritmo m2				
Detecciones positivas		Detecciones negativas		
553		745		
Código de descarte + clasificador				
Detecciones positivas		Detecciones negativas		
Fuentes aisladas no contaminadas	298	-Mala o muy mala calidad -Ausencia de asteroide -Asteroide pequeño o disperso -Contaminados	989	1000
		Fuentes mezcladas no contaminadas	11	

Tabla 3: Detecciones positivas y negativas del *dataset* de 2002_ms4 reales y obtenidas por ambos algoritmos

4 Conclusiones y trabajo futuro

La mayor dificultad de este proyecto ha sido la meticulosa inspección de los *croppies*, que conlleva tiempo y perseverancia. Existe una gran variedad de ellos y mientras algunos son de fácil interpretación otros requieren de una observación más cuidadosa. Esto explica la necesidad de automatizar esta tarea puesto que en las bases de datos no existen pocos miles sino cientos de miles de imágenes, que suponen un cometido impracticable para el humano.

Además de la diversidad de *croppies* en un solo *dataset* debe sumarse la variación de características de un *dataset* a otro (los de Varuna y 2002_ms4 muestran claras diferencias) dependiendo del equipo o equipos utilizados y las cualidades intrínsecas de cada asteroide. Esto implica que los *croppies* de baja calidad presenten patrones dispares en cada conjunto de datos y supone la dificultad de crear un código que descarte la totalidad de *croppies* defectuosos para cualquier *dataset*. Es por ello que quizás el código propuesto en este proyecto, a pesar de ser aplicable a los dos *datasets* utilizados, no obtenga un filtrado pleno de los *croppies* deficientes y requiera de modificaciones en nuevos conjuntos. Aún así, proporcionará un mayor porcentaje de descarte que el algoritmo *m2*, que es efectivo ante algún tipo de defecto pero no ante otros, como se comenta en la sección 3.2. El código de descarte, que en principio era complejo y requirió varias ejecuciones de prueba para refinarlo, se simplificó en gran medida al percatarse de que al suprimir los *croppies* contaminados ya se descartaban otros defectos.

En cuanto al clasificador, el diseño de la arquitectura no ha sido desafiante debido a su sencillez, a la existencia de las bibliotecas Keras y TensorFlow que simplifican enormemente el proceso y a los conceptos adquiridos durante el presente máster, en la asignatura obligatoria *07MBID Machine Learning* y la optativa *12MBID Redes Neuronales y Deep Learning*. Por su parte, el preprocesamiento de los *croppies* previo al clasificador ha sido favorecido por el manejo en el lenguaje Python y la multitud de librerías de procesamiento de imágenes existentes para dicho lenguaje: *OpenCV*, *scipy.ndimage*, *scikit-image*, *PIL*, etc.

De lo visto en los resultados, el clasificador ha logrado un 100 % de acierto en el conjunto de *test* de Varuna y un 94 % para 2002_ms4, lo que supone en principio un poder de generalización razonable. Además, ha superado con éxito una de las metas: la de descartar falsos positivos de *m2* debidos a cuestiones de calidad o defectos.

También ha demostrado poder recuperar fuentes aisladas no contaminadas de los falsos negativos de $m2$ (figuras 31 y 39), otra de las metas. Respecto al objetivo restante, la supresión de falsos positivos de $m2$ debidos a mezclas, se cumplirá también salvo en los casos problemáticos ya comentados o formas muy diferentes a las de entrenamiento como las de la figura 33. El grado de satisfacción real de cada uno de los objetivos y la consecuente utilidad del clasificador más allá del fin académico no serán conocidos hasta aplicarlo a varios conjuntos nuevos de otros asteroides.

Uno de los inconvenientes en esta tarea de clasificación ha sido el desequilibrio de clases, comentado en varias ocasiones, lo que ha obligado a recurrir a *data augmentation* y en un menor rendimiento respecto al caso idóneo en el que se contasen con más muestras mezcladas. El otro gran obstáculo de la tarea es la también comentada presencia de fuentes mezcladas muy solapadas, de gran similitud con algunas fuentes aisladas. Esto implica inevitables falsos positivos y también los más preocupantes falsos negativos. Aunque la mejora en este sentido no es viable con la CNN, sí es posible procurar nuevas vías, más innovadoras y como es lógico de mayor complejidad. El IAA plantea el futuro diseño de otro clasificador, que enlazado con el de este proyecto, permita lidiar con esos casos problemáticos. Para ello habrá que adentrarse en la geometría computacional: el futuro clasificador, además de las imágenes, contará con nueva información de la fuente (o fuentes mezcladas) como su excentricidad, es decir, su desviación respecto a la forma circular; las parábolas envolvente e interna máxima y otros parámetros geométricos.

En caso de que esa opción no sea fructífera se podría probar con un algoritmo que en vez de clasificar *croppies* individualmente maneje secuencias temporales de *croppies* de la misma fecha. De esta manera, podría identificar mezclas muy solapadas gracias a los *croppies* anteriores donde la mezcla es más fácil de clasificar. Para ello el camino más viable sería probablemente la cooperación de la red convolucional del proyecto con una red neuronal recurrente o RNN (*Recurrent Neural Network*) que puede trabajar con la dimensión temporal y “recordar” los *croppies* anteriores mientras clasifica el actual.

Referencias

- [1] Astromía (s.f.). *Fotometría*. Recuperado el 22 de julio de 2022 de <https://www.astromia.com/glosario/fotometria.htm>
- [2] Arizona Skies Meteorites (s.f.). *Iron Meteorite Widmanstatten Patterns*. Recuperado el 19 de julio de 2022 de <https://www.arizonaskiesmeteorites.com/Widmanstatten/>
- [3] Astronomical Institute of Charles University, s.f.. *Asteroid (201) Penelope*. Database of Asteroid Models from Inversion Techniques (DAMIT). Recuperado el 22 de julio de 2022 de https://astro.troja.mff.cuni.cz/projects/damit/pages/about_damit
- [4] Astronomics (s.f.). *Blooming*. Recuperado el 26 de agosto de 2022 de <https://www.astronomics.com/info-library/astronomical-terms/blooming/>
- [5] Astrosigma (s.f.). *Objetos transneptunianos*. Recuperado el 14 de julio de 2022 de <https://astrosigma.com/objetos-transneptunianos/>
- [6] Barrera, L. (2022). *El astrolabio: qué es y como funciona*. Sobrehistoria.com. Recuperado el 10 de julio de 2022 de <https://sobrehistoria.com/que-es-el-astrolabio/>
- [7] Barucci y Fulchignoni (1982). The dependence of asteroid lightcurves on the orientation parameters and the shapes of asteroids. *Moon and the Planets*, 27, 47-57. Bibcode: [1982M&P....27...47B](#)
- [8] Belskaya et. al (2006). Low phase angle effects in photometry of trans-neptunian objects: 20000 Varuna and 19308 (1996 TO66). *Icarus*, 184, 277-284. doi:[10.1016/j.icarus.2006.04.015](#)
- [9] Bennett, S. (2020). *Dwarf planet Ceres has salty water and appears geologically active*. NOVA. <https://www.pbs.org/wgbh/nova/article/dwarf-planet-ceres-water-geologically-active/>
- [10] Bravo, R. (2018). *Tipos de meteoritos*. AstroAfición. <https://astroaficion.com/2018/08/13/tipos-de-meteoritos/#comments>
- [11] Bravo, R. (2019). *Calidad del cielo: Seeing y transparencia*. AstroAfición. <https://astroaficion.com/2019/08/19/calidad-cielo-seeing-transparencia/>

- [12] Berman, B. (2015). *Weird Object: Asteroid 90 Antiope*. Astronomy. <https://astronomy.com/magazine/weirdest-objects/2015/02/44-asteroid-90-antiope>
- [13] Berry, R. (s.f.). *Photometry of BL Camelopardalis with the QSI 532ws CCD Camera*. Digital Imaging with Richard Berry. <http://www.wvi.com/~rberry/index.html>
- [14] Brown, P. *et al.* (2002). The flux of small near-Earth objects colliding with the Earth. *Nature*, 420, 294-296
- [15] Burbine, T.H. (2016). Near-Earth Asteroids and the Impact Threat. En Burbine, T.H. *Asteroids: Astronomical and Geological Bodies*. Cambridge: Cambridge University Press
- [16] Caldú, A. (s.f.). *¿Qué es la radioastronomía?*. Ciencia y Desarrollo. <https://www.cyd.conacyt.gob.mx/?p=articulo&id=508>
- [17] Casado, J.C. y Serra-Ricart, M. (2018). *Cálculo de rotación asteroides*. Astroaula. <https://www.astroaula.net/rotacion-asteroides/>
- [18] Casado, J.C., Serra-Ricart, M. y Cuesta, L. (2004). *Unidad Didáctica Ocultaciones*. Instituto de Astrofísica de Canarias [IAC]. Material no publicado.
- [19] Center for Near Earth Object Studies (s.f.a). *NEO Basics. Target Earth*. Recuperado el 11 de julio de 2022 de https://cneos.jpl.nasa.gov/about/target_earth.html
- [20] Center for Near Earth Object Studies (s.f.b). *NEO Groups*. Recuperado el 12 de julio de 2022 de https://cneos.jpl.nasa.gov/about/neo_groups.html
- [21] Center for Near Earth Object Studies (s.f.c). *Sentry: Earth Impact Monitoring. Impact Risk Data*. Recuperado el 16 de julio de 2022 de <https://cneos.jpl.nasa.gov/sentry/>
- [22] Churyumov, K.I., Kruchynenko, V.G., Chubko, L.S. y Churyumova, T.K. (2009). The meteoroid above Mediterranean Sea on July, 6th 2002 was a fragment of a cometary nucleus? en *263th Icy Bodies of the Solar System Proceedings IAU Symposium, 2009, Rio de Janeiro*, 246-248
- [23] Ciencia de la NASA (2008). *El evento de Tunguska - Cien años después*. <https://ciencia.nasa.gov/>
- [24] Cobiella, N. (s.f.). *El astrolabio. La medida de las estrellas*. Terra Australis Incognita. Recuperado el 20 de julio de 2022 de <http://www.arauco.org/SAPEREAUDE/>

[terraaustralisincognita/historiasdeandalus/astrolabio.html#:~: text=Su%20origen%20se%20le%20atribuye, astr%C3%B3nomo%20griego%20Hiparco%20de%20Nicea.](http://terraaustralisincognita/historiasdeandalus/astrolabio.html#:~:text=Su%20origen%20se%20le%20atribuye, astr%C3%B3nomo%20griego%20Hiparco%20de%20Nicea.)

- [25] Collins, M.D. (2020). The YORP Effect Can Efficiently Destroy 100 Kilometer Planetesimals At The Inner Edge Of The Solar System en *235th American Astronomical Society Meeting Abstracts, 2020, Honolulu*, p. 427. Bibcode:[2020AAS...23527701C](#)
- [26] Colomer, A. y Muñoz, G.E. (s.f.). *Redes neuronales y aprendizaje profundo*. Universidad Internacional de Valencia [VIU]. Material no publicado.
- [27] Daily Minor Planet. Blog of the Minor Planet Center (2011). *Asteroid Classification I-Dynamics*. <https://web.archive.org/web/20110418072433/https://minorplanetcenter.net/blog/asteroid-classification-i-dynamics/>
- [28] Dharmaraj, D (2021). *Zero-Padding in Convolutional Neural Networks*. Medium. <https://medium.com/@draj0718/zero-padding-in-convolutional-neural-networks-bf1410438e99>
- [29] Dartnell (2019). *How do binary asteroids form?*. BBC Sky at Night Magazine. <https://www.skyatnightmagazine.com/space-science/how-do-binary-asteroids-form/>
- [30] De la Fuente, C. y De la Fuente, R. (2020). On the orbital evolution of 2020 AV2, the first asteroid ever observed to go around the Sun inside the orbit of Venus. *Monthly Notices of the Royal Astronomical Society*, 494, 6-10. doi:[10.1093/mnrasl/slaa027](#)
- [31] Department of Physics and Astronomy of Georgia State University. *Albedo*. Hyperphysics. Recuperado el 10 de julio de 2022 de <http://hyperphysics.phy-astr.gsu.edu/hbasees/phyopt/albedo.html>
- [32] Díaz, J.V. (2021). *Los catálogos estelares más importantes*. UNIVERSO Blog. <https://josevicentediaz.com/2021/08/20/los-catalogos-estelares-mas-importantes/>
- [33] Doreste, J.L. (2002). *Tratamiento de Imágenes CCD Astronómicas*. Material no publicado. https://www.researchgate.net/publication/268296055_Tratamiento_de_Imagenes_CCD_Astronomicas
- [34] *Los Enanos del Sistema Solar*. Duffard, R. [IAA-Comunicación]. (2016). [Video] Youtube. https://www.youtube.com/watch?v=N4ZqaGJZK0g&ab_channel=IAA-Comunicaci%C3%B3n

- [35] *Meteoritos y Meteoroides*. Duffard, R. [IAA-Comunicación]. (2017). [Video] Youtube. https://www.youtube.com/watch?v=IDZsmCDe8qg&ab_channel=IAA-Comunicaci%C3%B3n
- [36] *Ocultaciones estelares por asteroides y TNOs*. Duffard, R. [Federación de Asociaciones Astronómicas de España]. (2021). [Video] Youtube. https://www.youtube.com/watch?v=RpkNuyj87JI&ab_channel=FAAE
- [37] Elkins, L.T. (2010). *Asteroids, Meteorites and Comets (The Solar System)*, p.92. Nueva York: Facts On File.
- [38] Elliot, J.L., Person, M.J. y McDonalds, S.W. (2000). Prediction and Observation of the 1997 July 18 Stellar Occultation by Triton: More Evidence for Distortion and Increasing Pressure in Triton's Atmosphere. *Academic Press*, 148(2), 347-369. doi:[10.1006/icar.2000.6508](https://doi.org/10.1006/icar.2000.6508)
- [39] Emelyanenko, V.V., Asher, D.J. y Bailey, M. E. (2007). The fundamental role of the Oort Cloud in determining the flux of comets through the planetary system. *Monthly Notices of the Royal Astronomical Society*, 381 (2), 779-789. doi: [10.1111/j.1365-2966.2007.12269.x](https://doi.org/10.1111/j.1365-2966.2007.12269.x)
- [40] Endl, M.; Spohn, T., Breuer, D., Johnson, T.V. (Eds.). *Encyclopedia of the Solar System*. 3edⁿ. Elsevier.
- [41] European Space Agency [ESA](s.f.a). *GAIADATA RELEASE 3 (GAIA DR3)*. COSMOS. Recuperado el 14 de agosto de 2022 de <https://www.cosmos.esa.int/web/gaia/dr3>
- [42] ESA (s.f.b). *What are Lagrange points?*. Recuperado el 14 de agosto de 2022 de https://www.esa.int/Enabling_Support/Operations/What_are_Lagrange_points
- [43] ESA (2018). *Gaia crea el mapa estelar más completo de nuestra Galaxia... y más allá*. https://www.esa.int/Space_in_Member_States/Spain/Gaia_crea_el_mapa_estelar_mas_completo_de_nuestra_Galaxia_y_mas_alla
- [44] ESA (2019a). *A history of astrometry-part I mapping the sky from ancient to pre-modern times*. <https://sci.esa.int/web/gaia/-/53196-the-oldest-sky-maps#:~:text=The%20first%20documented%20records%20of,and%20recorded%20their%20periodic%20motions.>

- [45] ESA (2019b). *Gaia's biggest operation since launch*. https://www.esa.int/Enabling_Support/Operations/Gaia_s_biggest_operation_since_launch
- [46] Espacio Profundo (2018). *Tipos de sensores y sus características*. <https://www.espacioprofundo.com/camaras/c%C3%B3mo-elegir-una-c%C3%A1mara-para-astrofotograf%C3%ADa/introducci%C3%B3n/tipos-de-sensores-y-sus-caracter%C3%ADsticas-r2/>
- [47] Farnham, T.L. (2001). Rotation and Color Studies of Centaurs, KBOs and Comets en *33rd Division for Planetary Sciences [DPS] meeting, 2001, Nueva Orleans*
- [48] Fernández-Valenzuela, E. et. al (2019). The Changing Rotational Light-curve Amplitude of Varuna and Evidence for a Close-in Satellite. *The Astrophysical Journal Letters*, 883. doi:[10.3847/2041-8213/ab40c2](https://doi.org/10.3847/2041-8213/ab40c2)
- [49] Fernández, J.L (s.f.). *Leyes de Kepler*. Fisicalab. Recuperado el 10 de julio de 2022 de <https://www.fisicalab.com/apartado/leyes-kepler>
- [50] *Charla-taller: Astrometría de asteroides con Tycho Tracker*. Fornas, A. [A.V.A. Astronomía Valencia]. (2022). [Vídeo] Youtube.
- [51] García, F. (2021). *Asteroides, curvas, gráficas polares y astroturismo con Faustino García*. La Esfera Celeste: Podcast Astronomía Amateur. <https://laesferaceleste.com/>
- [52] García, J. (2017). *Subastado el astrolabio andalusí más antiguo que se conserva*. ABC. https://www.abc.es/cultura/abci-subastado-astrolabio-andalusi-mas-antiguo-conserva-201704270109_noticia.html?ref=https%3A%2F%2Fwww.abc.es%2Fcultura%2Fabci-subastado-astrolabio-andalusi-mas-antiguo-conserva-201704270109_noticia.html
- [53] Genebriera, J. (s.f.). *Medición del paralaje de un asteroide*. Observatorio de Tacande, La Palma. Material no publicado.
- [54] Gil, L. (2021). *La radiación cósmica: Por qué no debería ser motivo de preocupación*. Organismo Internacional de Energía Atómica [IAEA]. <https://www.iaea.org/es/newscenter/news/radiacion-cosmica-por-que-no-deberia-ser-motivo-de-preocupacion-en-ingles>

- [55] Granvik, M. *et al.* (2016). Super-catastrophic disruption of asteroids at small perihelion distances. *Nature*, 530, 303-306. doi: <https://doi.org/10.1038/nature16934>
- [56] Gregersen, E. (s.f.). *Light from the stars*. Britannica. Recuperado el 11 de julio de 2022 de <https://www.britannica.com/science/star-astronomy/Stellar-positions>
- [57] Herald, D., George, T., Timerson, B. y Maley, P.D. (2017). (113) Amalthea And Its Satellite S/2017 (113)1. *Journal for Occultation Astronomy*, 7(3)
- [58] Herrera, D. (2017). *La magnitud de las estrellas*. AstroAfición. <https://astroaficion.com/2017/01/19/magnitud/>
- [59] Hicks, M., Simonelli, D. y Buratti, B. (2005). Photometric behavior of 20000 Varuna at very small solar phase angles. *Icarus*, 176, 492-498. doi:10.1016/j.icarus.2005.01.013
- [60] High Energy Astrophysics Science Archive Research Center [HEASARC] (2012). *GSC - HST Guide Star Catalog*. <https://heasarc.gsfc.nasa.gov/W3Browse/all/gsc.html>
- [61] Hedo, J.M. et al. (2018). On the minimum orbital intersection distance computation: a new effective method. *Monthly Notices of the Royal Astronomical Society*, 479, 3288-3299. doi:10.1093/mnras/sty1598
- [62] Instituto de Astrofísica de Andalucía (2019). *Se hallan indicios de la existencia de un satélite alrededor de Varuna, un cuerpo menor situado más allá de Neptuno*. <https://www.iaa.csic.es/noticias/se-hallan-indicios-existencia-satelite-alrededor-varuna-cuerpo-menor-situado-mas-alla>
- [63] International Astronomical Union [IAU] (s.f.). *Minor Planet Center (MPC)*. Recuperado el 21 de julio de 2022 de https://www.iau.org/science/scientific_bodies/centres/MPC/
- [64] International Astronomical Union [IAU] (2006). *RESOLUTION B5. Definition of a Planet in the Solar System*. https://www.iau.org/administration/resolutions/general_assemblies/
- [65] International Union of Pure and Applied Chemistry [IUPAC] (1997). *Compendium of Chemical Terminology-The "Gold Book". 2edⁿ*. Oxford: Blackwell Scientific Publications. <http://goldbook.iupac.org>

- [66] International Business Machines Corporation [IBM] (2020). *Convolutional Neural Networks*. IBM Cloud Education. <https://www.ibm.com/cloud/learn/convolutional-neural-networks>
- [67] Jet Propulsion Laboratory [JPL] (s.f.a). *Small-Body Database Lookup. 53319 (1999 JM8)*. Recuperado el 11 de julio de 2022 de https://ssd.jpl.nasa.gov/tools/sbdb_lookup.htm
- [68] Jet Propulsion Laboratory [JPL] (s.f.b). *Small-Body Database Lookup. 87 Sylvia (A866 KA)*. Recuperado el 13 de julio de 2022 de https://ssd.jpl.nasa.gov/tools/sbdb_lookup.html#/?sstr=87&view=OSP
- [69] Jet Propulsion Laboratory [JPL] (s.f.c). *Small-Body Database Lookup. 20000 Varuna (2000 WR106)*. Recuperado el 13 de agosto de 2022 de https://ssd.jpl.nasa.gov/tools/sbdb_lookup.html#/?sstr=2020000&view=OPD
- [70] Jewitt, D. y Sheppard, S. (2002). Physical properties of trans-neptunian object (20000) Varuna. *The Astronomical Journal*, 123(4), 2110-2120
- [71] Lacerda, P. y Jewitt, D. (2007). Densities of Solar System Objects from Their Rotational Light Curves. *The Astronomical Journal*, 133(4). doi:[10.1086/511772](https://doi.org/10.1086/511772)
- [72] Obtención del Código MPC. *Tribuna de Astronomía y Universo*, 48, 74-80
- [73] Lellouch et. al (2013).“TNOs are Cool”: A survey of the trans-Neptunian region IX. Thermal properties of Kuiper belt objects and Centaurs from combined Herschel and Spitzer observations. *Astronomy & Astrophysics*, 557(A60). doi:[0.1051/0004-6361/201322047](https://doi.org/0.1051/0004-6361/201322047)
- [74] Li, M., Juang, Y. y Gong, S. (2019). Assessing the risk of potentially hazardous asteroids through mean motion resonances analyses. *Astrophysics and Space Science*, 364 (78). doi: [10.1007/s10509-019-3557-5](https://doi.org/10.1007/s10509-019-3557-5)
- [75] Li, M., Juang, Y. y Gong, S. (2019). Survey of asteroids in retrograde mean motion resonances with planets. *Astronomy & Astrophysics*, 630 (60). doi: [10.1051/0004-6361/201936117](https://doi.org/10.1051/0004-6361/201936117)
- [76] Li, M., Wang, Y. [Yirui], Wang, Y. [Youliang], Zhou, B. y Zheng, W. (2020). Enhanced Kinetic Impactor for Deflecting Large Potentially Hazardous Asteroids via Maneuvering Space Rocks. *Scientific Reports*, 10(8506). doi: [0.1038/s41598-020-65343-z](https://doi.org/0.1038/s41598-020-65343-z)

- [77] Licandro, J., Oliva, E. y Di Martino, M. (2001). NICS-TNG infrared spectroscopy of trans-neptunian objects 2000 EB173 and 2000 WR106. *Astronomy & Astrophysics*, 373. doi:[10.1051/0004-6361:20010758](https://doi.org/10.1051/0004-6361:20010758)
- [78] Lunar & Planetary Laboratory (s.f.). *Spacewatch Discovery of Minor Planet 2000 WR106*. Recuperado el 13 de agosto de 2022 de <https://spacewatch.lpl.arizona.edu/news/spacewatch-discovery-minor-planet-2000-wr106>
- [79] Madieto, J.M. (2015). *Tus ojos jamás verían esto: así se colorean las fotos astronómicas*. El País. https://verne.elpais.com/verne/2015/01/16/articulo/1421396454_992571.html
- [80] Computer Science Wiki (2018). *Max-pooling/Pooling*. https://computersciencewiki.org/index.php/Max-pooling/_Pooling
- [81] Marchis et. al (2014). The puzzling mutual orbit of the binary trojan asteroid (624) Hektor. *The Astrophysical Journal Letters*, 784. doi:[10.1088/2041-8205/783/2/L37](https://doi.org/10.1088/2041-8205/783/2/L37)
- [82] Margot et. al (2002). Binary Asteroids in the Near-Earth Object Population. *Science*, 296(5772), 1445-1448. doi:[10.1126/science.1072094](https://doi.org/10.1126/science.1072094)
- [83] Michel, P., DeMeo, F.E. y Bottkle, W.F (Eds.) (2016). *Asteroids IV*, p.332. Tucson: The University of Arizona Space Science.
- [84] Meza, E. y Pereyra, A. (s.f.). *Linealidad de un CCD para uso astronómico*. Material no publicado. https://www.academia.edu/21305688/LINEALIDAD_DE_UN_CCD_PARA_USO_ASTRON%C3%93MICO
- [85] Morbidelli, A. (2008). Origin and dynamical evolution of comets and their reservoirs en *35th Saas-Fee advanced course, Suíza*
- [86] Morales, R. et. al (2022). Cataloguing astronomical ground-based images of asteroids and TNOs en *Europlanet Science Congress 2022*, 2022, Granada
- [87] Morgado, B.E. et. al (2021). Refined physical parameters for Chariklo's body and rings from stellar occultations observed between 2013 and 2020. *Astronomy & Astrophysics*, 652(A141). doi:[10.1051/0004-6361/202141543](https://doi.org/10.1051/0004-6361/202141543)
- [88] Mou, J. y Webster, I. (s.f.). *Apollo*. Space Reference. Recuperado el 10 de julio de 2022 de <https://www.spacereference.org/asteroid/1862-apollo-1932-ha>

- [89] Mou, J. y Webster, I. (s.f.). *Eros*. Space Reference. Recuperado el 10 de julio de 2022 de <https://www.spacereference.org/asteroid/433-eros-a898-pa>
- [90] NASA (s.f.). *NASA Team Finds Riches in Meteorito Treasure Hunt*. Recuperado el 11 de julio de 2022 de https://www.nasa.gov/topics/solarsystem/features/asteroid_treasure_hunt.html
- [91] NASA/IPAC Extragalactic Database (s.f.). *NED results for object NGC 6118*. Recuperado el 27 de agosto de 2022 de http://ned.ipac.caltech.edu/cgi-bin/nph-objsearch?objname=ngc+6118&extend=no&hconst=73&omegam=0.27&omegav=0.73&corr_z=1&out_cs=Equatorial&out_equinox=J2000.0&obj_sort=RA+or+Longitude&of=pre_text&zv_breaker=30000.0&list_limit=5&img_stamp=YES
- [92] National Geographic (2022). *Imágenes desde Hubble*. <https://www.nationalgeographic.es/photography/2017/03/imagenes-desde-hubble?image=1190.600x450>
- [93] *Observación de cometas y asteroides. Astrometría y fotometría*. Naves, R. y Campàs, M. [Agrupación Astronómica de Madrid]. (2020). [Vídeo] Youtube. https://www.youtube.com/watch?v=MrbArhz7pTo&ab_channel=Agrupaci%C3%B3nAstron%C3%B3micadeMadrid
- [94] Nemiroff, R. y Bonnell, J. (Eds.) (2000). *Dog-Bone Shaped Asteroid 216 Kleopatra*. Astronomy Picture of the Day (NASA), 10-05-2000. <https://apod.nasa.gov/apod/ap000510.html>
- [95] Notkin, G. (s.f.). *Meteorite types and classification*. Geology.com. Recuperado el 16 de julio de 2022 de <https://geology.com/meteorites/meteorite-types-and-classification.shtml>
- [96] Observatoire Solaire (2019). *Near, but not too near!*. <http://www.observatoiresolaire.eu/astronomy-blog/previous/18>
- [97] Observatorio Astronómico Los Molinos (s.f.). *Curvas de Luz de Asteroides*. Recuperado el 22 de julio de 2022 de <https://www.oalm.gub.uy/curvas-de-luz>
- [98] Oliveira et. al (2022). Constraints on the structure and seasonal variations of Triton's atmosphere from the 5 October 2017 stellar occultation and pre-

- vious observations. *Astronomy & Astrophysics*, 659(136). doi:[10.1051/0004-6361/202141443](https://doi.org/10.1051/0004-6361/202141443)Publishedonline06April2022
- [99] Oliveira et. al (2018). Stellar Occultation of Triton on October 5th, 2017 en *European Planetary Science Congress 2018, Berlín*
- [100] Ortiz, J.L. et. al (2015). Possible ring material around centaur (2060) Chiron. *Astronomy & Astrophysics*, 576(A18). doi: [10.1051/0004-6361/201424461](https://doi.org/10.1051/0004-6361/201424461)
- [101] Ortiz, J.L. et. al (2017). The size, shape, density and ring of the dwarf planet Haumea from a stellar occultation. *Nature*, 550, 219-223. doi:[10.1038/nature24051](https://doi.org/10.1038/nature24051)
- [102] *A GPU-Accelerated Synthetic Tracker for Blind Detection of NEOs, MBAs, and other Moving Objects*. Material no publicado. <https://www.tycho-tracker.com/links>
- [103] Peterson, B. (s.f.). *Quasar*. Britannica. Recuperado el 14 de agosto de 2022 de <https://www.britannica.com/science/quasar>
- [104] Ribas et. al (2014). Stellar occultations by transneptunian and centaurs objects: results from more than 10 observed events en *XIV Latin American Regional IAU Meeting, 2013, Florianópolis*
- [105] Roca, F. (2014). *Diseño, construcción y uso de una cámara CCD para observaciones astronómicas*. Proyecto final de grado. Escola Politècnica Superior d'Enginyeria de Vilanova i la Geltrú, Vilanova i la Geltrú.
- [106] Rodríguez, F. (s.f.). *Coordenadas Celestes*. Astroilusión. Recuperado el 20 de julio de 2022 de <https://www.astroilusion.es/coordenadas-celestes/>
- [107] Rodríguez, H. (2022). *Los planetas enanos del sistema solar:Plutón, Eris, Haumea, Makemake y Ceres*. National Geographic España. https://www.nationalgeographic.com.es/ciencia/5-planetes-enanos-sistema-solar-que-les-diferencia-resto_18447
- [108] Rommel et. al (2020). Stellar occultations enable milliarcsecond astrometry for Trans-Neptunian objects and Centaurs. *Astronomy & Astrophysics*, 644(A40). doi:[10.1051/0004-6361/202039054](https://doi.org/10.1051/0004-6361/202039054)

- [109] Ruiz, T. (2010). *Imágenes en la astronomía profesional*. Asociación Astronómica Quarks. https://www.aquarks.com/web/Asociacion_Astronomica_Quarks/Portada/Entradas/2010/3/1_Imagenes_en_la_astronomia_profesional.html
- [110] Sánchez, C. y Cuevas, S. (2009). El telescopio y su historia. *Ciencias*, 95, 28-31. <https://www.revistacienciasunam.com/en/42-revistas/revista-ciencias-95/184-el-telescopio-y-su-historia.html>
- [111] Santos-sanz et. al (2013). Thermal light curve of 20000 Varuna with Herschel-PACS en *European Planetary Science Congress, 2013, London*
- [112] Schabel, C. (2021). *La increíble precisión del mundo de las ocultaciones estelares*. La Esfera Celeste: Podcast Astronomía Amateur. <https://laesferaceleste.com/>
- [113] Sheppard, S. (2022). In the glare of the Sun: Searches during twilight toward the Sun have found several asteroids near Venus orbit. *Science*, 377 (6604), 366-367. doi:[10.1126/science.abj9820](https://doi.org/10.1126/science.abj9820)
- [114] Serrano, A. (s.f.). *Observando el cielo a simple vista*. Dr. Ángel Serrano Sánchez de León. Recuperado el 20 de julio de 2022 de <https://aserrano.es/>
- [115] Sociedad Española de Astronomía [SEA] (s.f.a). *Año-luz*. Recuperado el 10 de julio de 2022 de <https://www.sea-astronomia.es/glosario/ano-luz>
- [116] Sociedad Española de Astronomía [SEA] (s.f.b). *Cometa*. Recuperado el 13 de julio de 2022 de <https://www.sea-astronomia.es/glosario/cometa>
- [117] Sociedad Española de Astronomía [SEA] (s.f.c). *Esfera Celeste*. Recuperado el 20 de julio de 2022 de <https://www.sea-astronomia.es/glosario/esfera-celeste>
- [118] Sociedad Española de Astronomía [SEA] (s.f.d). *Unidad Astronómica*. Recuperado el 10 de julio de 2022 de <https://www.sea-astronomia.es/glosario/unidad-astronomica>
- [119] Stansberry et. al (2007). Physical Properties of Kuiper Belt and Centaur Objects: Constraints from the Spitzer Space Telescope. doi:[10.48550/arXiv.astro-ph/0702538](https://doi.org/10.48550/arXiv.astro-ph/0702538)
- [120] Swinburne University of Technology (s.f.a). *Albedo*. COSMOS-The Swinburne Astronomy Online [SAO] Encyclopedia of Astronomy. Recuperado el 10 de julio de 2022 de <https://astronomy.swin.edu.au/cosmos/A/Albedo>

- [121] Swinburne University of Technology (s.f.b). *Kirkwood Gaps*. COSMOS-The Swinburne Astronomy Online [SAO] Encyclopedia of Astronomy. Recuperado el 12 de julio de 2022 de <https://astronomy.swin.edu.au/cosmos/K/Kirkwood+Gaps>
- [122] Swinburne University of Technology (s.f.c). *Main Asteroid Belt*. COSMOS-The Swinburne Astronomy Online [SAO] Encyclopedia of Astronomy. Recuperado el 12 de julio de 2022 de <https://astronomy.swin.edu.au/cosmos/m/main+asteroid+belt>
- [123] Synider, G. y Marschall, L. (2021). *Astrometry of Asteroids. Student Manual*. Gettysburg College. Material no publicado.
- [124] Teledyne Photometrics (s.f.). *Binning*. Recuperado el 21 de julio de 2022 de <https://www.photometrics.com/learn/imaging-topics/binning>
- [125] The American Association of Variable Star Observers [AAVSO] (2022). *CCD Photometry Guide*. https://www.aavso.org/sites/default/files/publications_files/ccd_photometry_guide/CCDPhotometryGuide.pdf
- [126] Universidad Diego Portales [UDP]: Facultad de Ingeniería y Ciencias. *Lo que no sabías sobre Plutón: 5-Sus 5 lunas*. Núcleo de Astronomía udp (2015). <https://astronomia.udp.cl/es/>
- [127] Tillman, N. T. (2017). *Asteroid Belt: Facts & Formation*. SPACE.com. <https://www.space.com/>
- [128] University of Southern Maine (s.f.). PLANETARIUM: Why is the vernal equinox called the "First Point of Aries" when the Sun is actually in Pisces on this date?. Recuperado el 20 de julio de 2022 de <https://usm.maine.edu/planet/why-vernal-equinox-called-first-point-aries-when-sun-actually-pisces-date>
- [129] Venditti, F.C.F., Marchi, L.O., Misra A.K., Sanchez, Diogo M. y Prado, A.F.B.A. (2020). Dynamics of tethered asteroid systems to support planetary defense. *The European Physical Journal Special Topics*, 229, 1463-1477. doi:[10.1140/epjst/e2020-900183-y](https://doi.org/10.1140/epjst/e2020-900183-y)
- [130] Vera C. Rubin Observatory (s.f.). *Potentially Hazardous Asteroids (PHAs)*. Recuperado el 16 de julio de 2022 de <https://www.lsst.org/science/solar-system/potentially-hazardous-asteroids>

Notas

- ¹ El perihelio se trata del punto más cercano al Sol por el que pasa un cuerpo al trazar su órbita. El punto más lejano se denomina afelio. Durante el perihelio el cuerpo alcanza su velocidad máxima y durante el afelio la mínima, como se deduce de la segunda ley de Kepler ([Fernández, s.f.](#)).
- ² Una unidad astronómica, de símbolo UA (AU en inglés), es una unidad de longitud ampliamente utilizada para distancias dentro del Sistema Solar y equivale a la distancia media entre la Tierra y el Sol, aproximadamente 150 millones de kilómetros ([SEA, s.f.d](#)). La otra unidad astronómica más usada es el año luz (a.l.), *light year* en inglés (*l.y.*), que se corresponde con la distancia que recorre la luz en el vacío en un año, es decir, algo menos de 9 billones y medio de kilómetros ([SEA, s.f.a](#)).
- ³ Los *Kirwood Gaps* ocurren por la influencia gravitatoria de Júpiter. Es poco probable que los asteroides se encuentren en ciertas órbitas resonantes con Júpiter (2:1, 3:1, 5:2, 7:3) puesto que esto conllevaría que se encontrasen cerca del planeta cada n vueltas y su fuerza gravitatoria los empujaría hacia otra órbita. Esto deriva en regiones en forma de anillo con poca concentración de asteroides ([Swinburne University of Technology, s.f.b](#)).
- ⁴ La nave Dawn de la NASA ha encontrado sales en un cráter de la superficie de Ceres (cráter Occator). Esto junto al hallazgo de minerales hidratados, indica la existencia (actual o no) de agua líquida ([Duffard, 2016](#)). Algunos investigadores sostienen la hipótesis de un posible océano subterráneo en el planeta enano ([Bennett, 2020](#)).
- ⁵ Los cometas hacen referencia a todos los cuerpos celestes pequeños que cuentan con volátiles (hielo y polvo) y al acercarse al Sol subliman formando una coma. Los asteroides que contienen hielos pueden transformarse en cometas al alcanzar su perihelio, por lo que muchos cuerpos cuentan con denominación de asteroide y cometa simultáneamente ([Sociedad Española de Astronomía \[SEA\], s.f.b](#)). En el presente trabajo no se tratan los cometas.
- ⁶ La resonancia orbital ocurre cuando el asteroide y un planeta tienen una relación entera de los periodos orbitales, es decir, cuando el asteroide da n vueltas al Sol el planeta da m , siendo n y m enteros. Esta resonancia puede ser inestable si la órbita del asteroide se ve perturbada con el tiempo o estable, si se mantiene largos intervalos de tiempo.

⁷ Se conoce la masa de Ida con bastante precisión, pues es posible obtener la densidad de los asteroides cuando tienen satélites, usando sus tamaños, el periodo del satélite y la distancia entre ambos. Dactyl orbita a tan solo 90 *km* de Ida y sorprendentemente es esférico aún teniendo un diámetro tan pequeño (1,5 *km*). Ida y Dactyl son considerados un asteroide binario a pesar de la diferencia de tamaño. Otros asteroides binarios son 2000 DP107, con un cuerpo de 800 *m* de diámetro y otro de 300 *m*, que curiosamente siempre apunta la misma zona hacia el cuerpo mayor, como ocurre con la Luna ([Casado, Serra-Ricart y Cuesta, 2004](#); [Margot et. al, 2002](#)) y el simétrico Antiope, formado por dos cuerpos prácticamente iguales en tamaño (84 y 88 *km*) ([Berman, 2015](#)). Incluso existen asteroides ternarios como 87-Sylvia, con un cuerpo mayor y dos satélites, Rómulo y Remo ([Jet Propulsion Laboratory \[JPL\], s.f.b](#)).

⁸ En total existen cinco puntos de Lagrange (L_1 , L_2 , L_3 , L_4 y L_5) en torno a Júpiter en los que los asteroides pueden orbitar de manera estable. Los troyanos se sitúan concretamente en L_4 y L_5 . Otros planetas como Urano y Neptuno también cuentan con puntos de Lagrange, incluso en la Tierra la basura espacial tiende a acumularse en esos puntos ([Naves y Campàs, 2020](#)).

⁹ Plutón cuenta con cinco satélites: Caronte, el más grande con 1186 *km* de diámetro, Nix, Hydra, Cerbero y Estigia ([Universidad Diego Portales \[UDP\], 2015](#)). Parte de su corteza presenta cráteres, por lo que es geológicamente inactiva mientras que otra parte sí presenta actividad. Además, las curvas encontradas en su superficie son indicios de vientos en la atmósfera. La presencia de montañas de gran altura también indica la existencia de hielos de agua ([Duffard, 2016](#)).

¹⁰ Eris es uno de los cuerpos del Sistema Solar con mayor albedo (90 %). El albedo se trata de la cantidad de luz reflejada por el asteroide (u otro cuerpo celeste) respecto a la luz que recibe de alguna fuente, en este caso el Sol ([Swinburne University of Technology, s.f.a](#)). El albedo de Bond se corresponde con el cociente de la luz reflejada (en todas las longitudes de onda y todas direcciones) entre la luz total recibida. El albedo geométrico en cambio, se trata del porcentaje de luz reflejada por el asteroide respecto a la que reflejaría un reflector perfecto (superficie de Lambert) ([Georgia State University, s.f.](#)). Los asteroides de bajo albedo (*low albedo*) reflejan menos del 10 % y son también llamados asteroides oscuros, mientras que los de alto albedo se conocen como asteroides brillantes. El alto albedo de

Eris se debe a la capa de hielo de metano depositada en su superficie. Es probable que esta capa se convierta a gas al llegar a su perihelio, pero dado que su periodo es de más de 500 años, pasará mucho tiempo hasta comprobarlo. Eris tiene un satélite, Disnomia, lo que permitió calcular su densidad con mucha precisión ([Duffard, 2016](#)).

¹¹ La definición de planeta enano y planeta coinciden en varios puntos: es un cuerpo celestial que tiene suficiente masa para que las fuerzas gravitatorias superen a las de sólido rígido y se alcance equilibrio hidroestático, es decir, forma esférica y no es satélite de otro cuerpo no estelar. La diferencia reside en que el planeta ha “limpiado” el entorno de su órbita y el planeta enano no; en otras palabras, el planeta puede evacuar o dominar los pequeños cuerpos de su vecindario y un planeta enano no ([International Astronomical Union \[IAU\], 2006](#)). Actualmente se reconocen cinco planetas enanos, ordenados de mayor a menor en tamaño: Eris (TNO), Plutón (TNO), Makemake (TNO), Haumea (TNO) y Ceres (cinturón principal). Son candidatos Orcus, Salacia, Quaoar, Sedna y Varuna ([Rodríguez, 2022](#)).

¹² Makemake cuenta con un alto albedo, en torno al 80 %. No tiene satélites ([Duffard, 2016](#)).

¹³ La magnitud aparente de un cuerpo celeste se corresponde con el brillo medido desde la Tierra, sin tener en cuenta a la distancia a la que se encuentra. La magnitud absoluta, sin embargo, se trata del brillo que tendría dicho cuerpo si se encontrase a 10 *parsecs* de la Tierra (1 *parsec* equivale a aproximadamente 3,26 años luz) y por lo tanto permite la comparación de varios cuerpos: el Sol se ve tan brillante porque está muy “cerca” de la Tierra (magnitud aparente de -26,75), pero si se encontrase a 10 *parsecs* sería mucho más difícil de ver a simple vista (magnitud absoluta de 4,82). Cuanto más grande es la magnitud más débil es el brillo por lo que los cuerpos más brillantes cuentan con magnitudes negativas como las estrellas Deneb, Rigel y Betelgeuse ([Gegersen, s.f.](#)).

¹⁴ El MPC es un organismo dependiente de la IAU encargado del seguimiento de los cuerpos pequeños del Sistema Solar (planetas enanos, asteroides, cometas y satélites), es decir, se ocupa del cálculo, recepción, verificación y difusión de sus observaciones astrométricas y sus órbitas, así como de su numeración y denominación oficial ([IAU, s.f.](#)).

¹⁵ Los catálogos estelares contienen datos astrométricos muy precisos de gran cantidad de estrellas. El más amplio y uno de los más utilizados es Gaia, obtenido gracias a la nave con el mismo nombre, lanzada en 2013 y vigente en la actualidad. Otros catálogos relevantes son

el GSC (*Guide Star Catalog*), creado en su momento para facilitar la labor del telescopio Hubble; SAO (*Smithsonian Astrophysical Observatory Star Catalog*), que data de 1966; Henry Draper; Hipparcos y BSC (*Bright Star Catalogue*) ([ESA \(s.f.a\)](#); [HEASARC,2012](#); [Díaz, J.V., 2021](#))..

- ¹⁶ Las coordenadas ecuatoriales definen la posición de los astros en la bóveda celeste y son la declinación (DEC o δ) y la ascensión recta (AR o α). La declinación es similar a la latitud de las coordenadas geográficas y se corresponde con el ángulo formado entre el ecuador celeste (prolongación del ecuador terrestre) y el paralelo que pasa por el astro a medir. La declinación se mide en unidades angulares: grados ($^{\circ}$), minutos ($'$) y segundos ($''$) y va desde $+90^{\circ}$ (polo norte de la esfera) a -90° (polo sur de la esfera). La ascensión recta, similar a la longitud geográfica, se corresponde con la distancia angular entre el meridiano cero y el meridiano que pasa por el astro. Se da en unidades de tiempo, horas (h), minutos (m) y segundos (s) y va de 0 a 24 h ([Serrano, s.f.](#)).
- ¹⁷ El efecto fotoeléctrico consiste en la emisión de electrones por parte de un metal cuando es iluminado por una fuente de luz de frecuencia suficiente. Lo que ocurre es que los fotones de dicha fuente portan la energía necesaria para vencer la débil fuerza que mantiene ligados los electrones semi-libres al metal, de forma que estos son expulsados con una velocidad que depende de la energía del fotón incidente. Con que mayor es la intensidad, es decir, el número de fotones que inciden por unidad de tiempo mayor será la cantidad de electrones emitidos. Si se ilumina un metal pero la energía de los fotones no supera la de ligadura de los electrones no se dará el efecto fotoeléctrico.
- ¹⁸ Aún así, los dispositivos CCD solo cuentan con linealidad real por debajo de un número de fotones, cuando se acercan al nivel de saturación (*blooming*) empiezan a presentar comportamiento no lineal ([Meza y Pereyra, s.f.](#)).
- ¹⁹ Los rayos cósmicos se tratan de partículas subatómicas, normalmente partículas α (núcleos de helio), γ (fotones) y protones de elevada energía que proceden del Sol o de eventos estelares fuera del Sistema Solar, como las supernovas ([Gil, 2021](#)).
- ²⁰ La galaxia NGC 6118 se encuentra a 83 millones de años luz y tiene un tamaño similar al de la Vía Láctea (110 mil años luz de diámetro). Es también una galaxia espiral pero con unos brazos menos definidos ([NASA/IPAC Extragalactic Database, s.f.](#)).

Apéndice

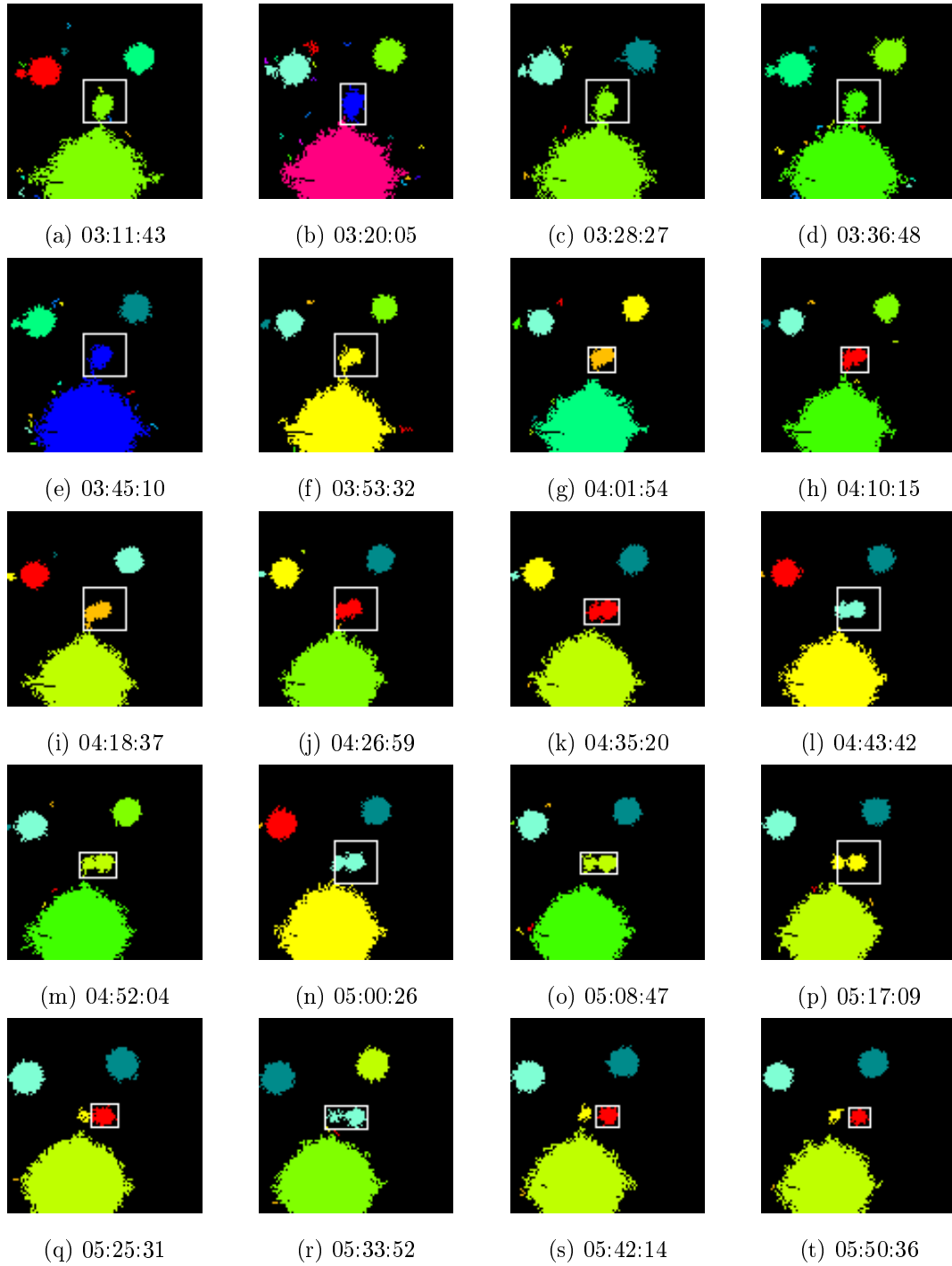


Figura 42: Secuencia del 3 de diciembre de 2013

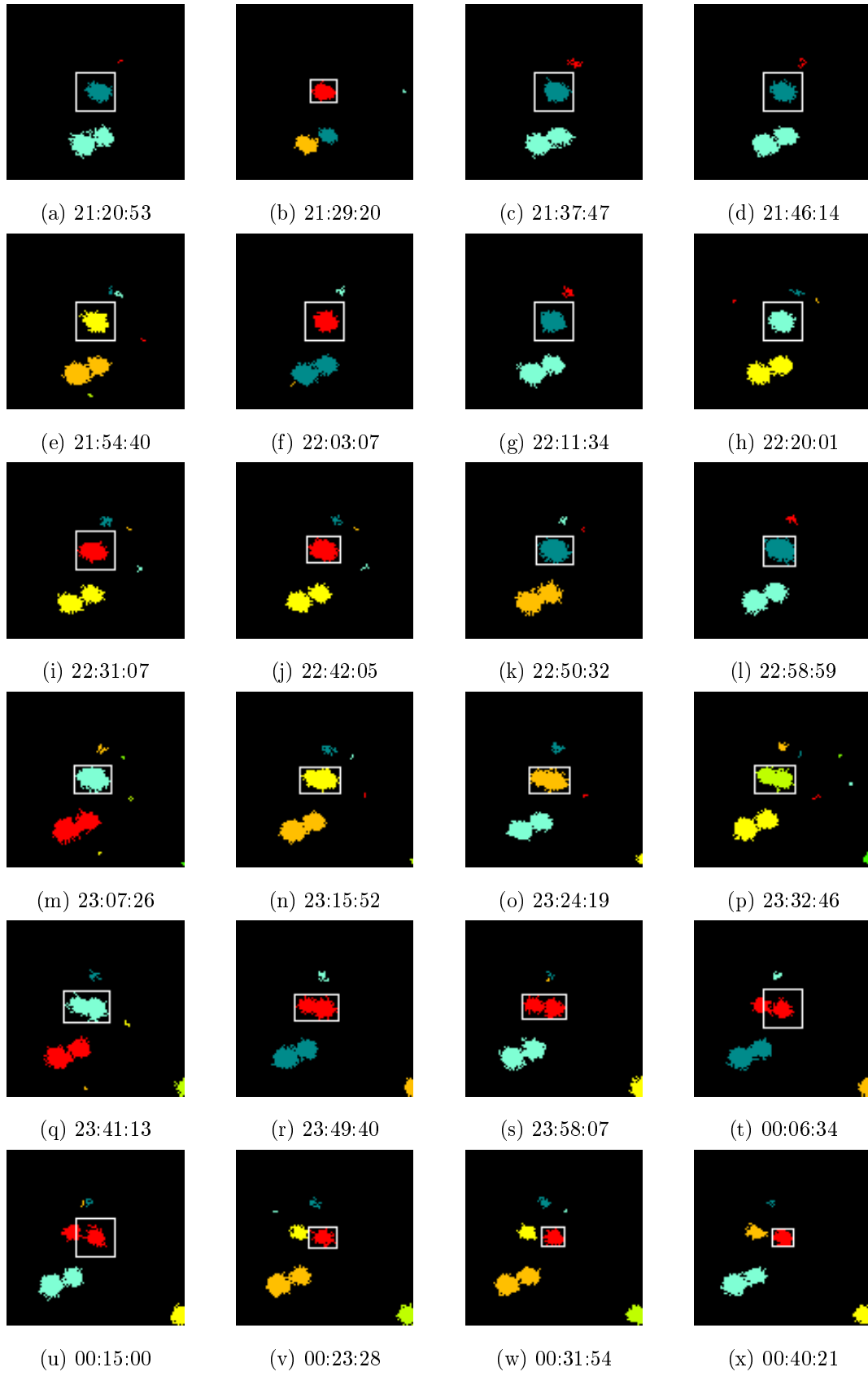


Figura 43: Secuencia del 31 de enero de 2011

- Enlace a [GitHub](#) con el código del proyecto en formato *ipynb*.
- Enlace a [Google Drive](#) con las fuentes aisladas del conjunto de *test* de Varuna.
- Enlace a [Google Drive](#) con las fuentes aisladas del conjunto de *test* de 2002_ms4.