



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

**PREDICCIÓN DE REDSHIFTS FOTOMÉTRICOS USANDO IMÁGENES
MULTI BANDA Y MULTI RESOLUCIÓN**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL ELÉCTRICO

ALAN GERSON CONTRERAS MONTANARES

PROFESOR GUÍA:
FRANCISCO FÖRSTER BURÓN

MIEMBROS DE LA COMISIÓN:
PABLO ESTÉVEZ VALENCIA
FRANZ BAUER

SANTIAGO DE CHILE
2024

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE INGENIERO CIVIL ELÉCTRICO
POR: ALAN GERSON CONTRERAS MONTANARES
FECHA: 2024
PROF. GUÍA: FRANCISCO FÖRSTER BURÓN

PREDICCIÓN DE REDSHIFTS FOTOMÉTRICOS USANDO IMÁGENES MULTI BANDA Y MULTI RESOLUCIÓN

En esta memoria se evalúa el impacto que tiene añadir imágenes en multi-resolución en la tarea de predecir el *redshift* fotométrico de galaxias con imágenes en multi-banda mediante el uso de una red neuronal convolucional. Para ello, se comparó directamente con el uso de imágenes de resolución simple utilizadas en Pasquet et al. [1], las cuales tienen un campo de visión de 25.244 arcos de segundo (["]) y provienen del Sloan Digital Sky Server (SDSS). Al usar este tipo de imágenes se obtuvo un residuo promedio $\langle \Delta z \rangle = 0.00012$, una dispersión de los residuos $\sigma_{\text{MAD}} = 0.00906$ y un porcentaje de errores catastróficos $\eta = 0.31\%$. Por otro lado, al añadir imágenes en multi-resolución con 5 niveles y compartiendo los pesos del modelo se obtuvo un $\langle \Delta z \rangle = 0.00013$, $\sigma_{\text{MAD}} = 0.00914$ y $\eta = 0.34\%$. Sin embargo, para galaxias con un radio Petrosiano del 90% mayor a 70 [" se obtuvo un $\sigma_{\text{MAD}} = 0.0075$ al usar imágenes con resolución simple y $\sigma_{\text{MAD}} = 0.0037$ al usar imágenes en multi-resolución. Finalmente, se proponen nuevos enfoques para abordar este problema en trabajos futuros, tales como; la optimización de hiperparámetros, incluir un nuevo nivel y lograr una invarianza a escalas en el modelo.

*A mis padres, a mi hermano y baquita,
las estrellas más hermosas del firmamento.*

Agradecimientos

Primero que todo, gracias a Dios por sobre todas las cosas.

También quiero agradecer a mi familia, a mis padres y mi hermano por la constante preocupación y apoyo no solo en el pasar de estos años universitarios, sino en toda mi vida. Desde acompañarme para ir al colegio hasta dejarme al paradero en la madrugada incluso en días fríos. Por apoyarme y ayudarme siempre en momentos en los que ni yo sabía qué hacer. También por aguantar por tanto tiempo mi mal genio. Por todos los momentos que hemos pasado juntos, ya sean buenos o malos, y por todos los que pasaremos con el apoyo de Dios.

Quiero agradecer también a todos los hermanos y hermanas en la fe quienes han orado por mí durante todo este tiempo, cuyas oraciones me sirvieron en este periodo tanto física como espiritualmente. Dios les bendiga.

Adicionalmente quiero dar gracias al profesor Francisco por su buena disposición y guía a lo largo de esta memoria, y a los profesores Pablo y Franz por ser parte de la comisión y por todos los comentarios.

Tabla de Contenido

| | |
|--|-----------|
| 1. Introducción | 1 |
| 1.1. Objetivos | 2 |
| 1.1.1. Objetivo general | 2 |
| 1.1.2. Objetivos específicos | 2 |
| 2. Marco teórico y estado del arte | 3 |
| 2.1. Marco teórico | 3 |
| 2.1.1. Surveys astronómicos | 3 |
| 2.1.1.1. Sloan Digital Sky Survey (SDSS) | 3 |
| 2.1.1.2. Panoramic Survey Telescope and Rapid Response System (Pan-STARRS) | 5 |
| 2.1.2. Técnica de multi-resolución | 5 |
| 2.1.3. Redes neuronales profundas | 7 |
| 2.1.3.1. Redes neuronales artificiales | 7 |
| 2.1.3.2. Redes neuronales convolucionales | 9 |
| 2.1.3.3. Aprendizaje de redes neuronales | 11 |
| 2.2. Estado del arte | 14 |
| Beck et al. [32] | 14 |
| Pasquet et al. [1] | 15 |
| Hayat et al. [8] | 18 |
| Dey et al. [6] | 19 |
| Treyer et al. [9] | 19 |
| Ait Ouahmed et al. [10] | 20 |
| Förster et al. [19] | 21 |
| 3. Metodología | 22 |
| 3.1. Predicciones usando imágenes de resolución simple | 24 |
| 3.1.1. Usando imágenes de Pan-STARRS | 24 |
| 3.1.2. Usando imágenes del SDSS con HiPS2FITS | 25 |
| 3.2. Predicciones usando imágenes en multi-resolución | 25 |
| 3.2.1. Compartiendo pesos de los bloques convolucionales | 27 |
| 3.2.2. Sin compartir pesos de los bloques convolucionales | 28 |
| 4. Resultados | 29 |
| 4.1. Experimentos usando imágenes de resolución simple | 29 |
| 4.1.1. Usando imágenes de Pan-STARRS | 29 |
| 4.1.2. Usando imágenes del SDSS con HiPS2FITS | 31 |
| 4.2. Experimentos usando imágenes en multi-resolución | 33 |

| | | |
|-----------|--|-----------|
| 4.2.1. | Compartiendo pesos de los bloques convolucionales | 33 |
| 4.2.2. | Sin compartir pesos de los bloques convolucionales | 35 |
| 5. | Análisis | 37 |
| 5.1. | Análisis general | 39 |
| 6. | Conclusiones | 41 |
| 6.1. | Trabajo futuro | 42 |
| 6.1.1. | Mejorar la calidad de los experimentos | 42 |
| 6.1.2. | Minimizar la cantidad de información utilizada | 42 |
| 6.1.3. | Combinación con DELIGHT | 43 |
| | Bibliografía | 44 |
| | Anexos | 50 |
| | Anexo A. Antecedentes de términos astronómicos | 50 |
| | Anexo B. Comparación de colores y SNR entre SDSS y Pan-STARRS | 53 |
| | Anexo C. Experimentos adicionales | 57 |
| C.1. | Impacto de la banda u del SDSS en las predicciones | 57 |
| C.2. | Interpretando predicciones con SHAP | 59 |
| C.3. | Uso de niveles de forma independiente | 63 |
| | Anexo D. Resumen de resultados e impacto del SNR | 65 |

Índice de Tablas

| | | |
|------|--|----|
| 2.1. | Comparación de $\langle \Delta z \rangle$, σ_{MAD} y tasa de outliers (η) al usar o no usar imágenes del Stripe 82 en entrenamiento y/o test. Resultados extraídos de [1]. | 17 |
| 2.2. | Resumen de métricas de todos los trabajos. Resultados extraídos de [10]. | 20 |
| 4.1. | Resumen de métricas promedio de los 5 <i>folds</i> de test dadas distintas combinaciones de niveles utilizados de imágenes en multi-resolución, añadiendo los resultados de usar las imágenes de HiPS2FITS en resolución simple. | 34 |
| A.1. | Tablas de la base de datos del SDSS utilizadas en el servicio de CasJobs. | 51 |
| A.2. | Tabla de la base de datos de Pan-STARRS utilizada en el servicio de CasJobs. | 52 |
| C.1. | Resumen de métricas promedio de los 5 <i>folds</i> de test obtenidos al usar los niveles de forma independiente. | 63 |
| D.1. | Resumen de métricas promedio de los 5 <i>folds</i> de test obtenidos en todos los experimentos realizados. | 65 |

Índice de Ilustraciones

| | | |
|------|---|----|
| 2.1. | Respuestas finales obtenidas de los filtros <i>ugriz</i> del telescopio del SDSS. La ilustración fue obtenida de la librería <code>astroML</code> en Python (https://www.astroml.org/index.html). | 4 |
| 2.2. | Comparación entre una imagen de una galaxia en resolución simple (imagen de la izquierda) con respecto a usar imágenes en multi resolución en 5 niveles (imagen del centro). Adicionalmente se demarcan los 5 niveles para una mejor diferenciación entre ellos (imagen de la derecha). | 7 |
| 2.3. | Ilustración del modelo de una neurona artificial, donde x_n es la componente n-ésima del vector de entrada, w_n es el peso asociado a la componente x_n , b es el sesgo (o <i>bias</i>) y $f(\cdot)$ es una función de activación. Ilustración extraída de https://learnopencv.com/understanding-activation-functions-in-deep-learning/ | 8 |
| 2.4. | Ejemplo de arquitectura de una red neuronal convolucional correspondiente al modelo DELIGHT. Ilustración obtenida de [19]. | 11 |
| 2.5. | Resultados de la regresión lineal local del trabajo de Beck et al. [32]. En la figura de la izquierda se presenta la comparación del <i>redshift</i> predicho (<i>zphot</i>) con el espectroscópico (<i>zspec</i>), mientras que la figura de la derecha contiene los residuos no normalizados de las predicciones. Las líneas rojas punteadas corresponden a intervalos de confianza. Ilustración obtenida de [32]. | 15 |
| 2.6. | Resultados promedio de los cinco <i>folds</i> de test usando la CNN propuesta por Pasquet et al. (figura de la izquierda), comparado con el trabajo de Beck et al. sobre el mismo conjunto de datos (figura de la derecha). Ilustración extraída de [1]. | 17 |
| 2.7. | Resultados de realizar <i>fine-tuning</i> al modelo pre-entrenado (figura de la izquierda) comparado con entrenar el modelo desde cero (figura de la derecha). Ilustración extraída de [8]. | 18 |
| 2.8. | Resultados de la predicción del <i>redshift</i> en el modelo de Dey et al. sobre el conjunto de test. Ilustración extraída de [6]. | 19 |
| 2.9. | Ejemplo de imágenes en multi-resolución, donde A) es la imagen original, B) es la imagen con el fondo extraído y detección de objetos y C) contiene las cinco imágenes respectivas (D-H) en multi-resolución. Ilustración extraída de [19]. | 21 |
| 3.1. | Ilustración del procedimiento de realizar dos <i>stratified cross-validation</i> . El primero es para encontrar la mejor época promedio en donde se encuentra la menor pérdida en los 5 conjuntos de test, para ello se utiliza el conjunto de entrenamiento del primer <i>fold</i> del segundo <i>stratified cross-validation</i> , en este último se entregan las métricas promedio en la época encontrada anteriormente. | 23 |
| 3.2. | Ejemplo de una imagen en multi-resolución en cinco niveles, en donde se usó un campo de visión original de 202.752 ["]. | 26 |

| | | |
|------|---|----|
| 3.3. | Ejemplo de una imagen en multi-resolución en cinco niveles, en color verde se puede ver el campo de visión utilizado por [1], el cual equivale al nivel 4. . . . | 27 |
| 4.1. | Resultados de la regresión con imágenes de Pan-STARRS. Los resultados de los 5 <i>folds</i> de test fueron promediados, tanto en la regresión como en las métricas presentes. En la ilustración de la izquierda se usó la compresión con la función arcoseno hiperbólico, mientras que la de la derecha corresponde a la compresión realizada con la Ec. 3.1. Las líneas rojas punteadas corresponden a los límites que definen los errores catastróficos. | 30 |
| 4.2. | Promedio de la dispersión de los errores (σ_{MAD}) en los 5 <i>folds</i> de test y su respectiva desviación estándar en distintas características, usando las imágenes de Pan-STARRS con distintas compresiones. La ilustración superior es a lo largo del <i>redshift</i> , la inferior izquierda a lo largo del radio de la galaxia, mientras que la inferior derecha es a lo largo de su magnitud en la banda <i>r</i> . La Compresión 1 hace referencia al arcoseno hiperbólico, mientras que la Compresión 2 hace referencia a la Ec. 3.1. Los histogramas corresponden a la distribución de cada característica en todo del conjunto de datos. | 31 |
| 4.3. | Resultados de la regresión con imágenes del SDSS. Los resultados de los 5 <i>folds</i> de test fueron promediados, tanto en la regresión como en las métricas presentes. En la ilustración de la izquierda se usaron las imágenes de Pasquet et al. [1], mientras que en la de la derecha se usaron las de HiPS2FITS. Las líneas rojas punteadas corresponden a los límites que definen los errores catastróficos. . . . | 32 |
| 4.4. | Promedio de la dispersión de los errores (σ_{MAD}) en los 5 <i>folds</i> de test y su respectiva desviación estándar en distintas características usando las imágenes del SDSS procesadas en Pasquet et al. [1] y las de HiPS2FITS. La ilustración superior es a lo largo del <i>redshift</i> , la inferior izquierda a lo largo del radio de la galaxia, mientras que la inferior derecha es a lo largo de su magnitud en la banda <i>r</i> . Los histogramas corresponden a la distribución de cada característica en todo del conjunto de datos. | 33 |
| 4.5. | Promedio de la dispersión de los errores (σ_{MAD}) en los 5 <i>folds</i> de test y su respectiva desviación estándar en distintas características usando las imágenes del SDSS provenientes de HiPS2FITS en distintos niveles de multi-resolución. La ilustración superior es a lo largo del <i>redshift</i> , la inferior izquierda a lo largo del radio de la galaxia, mientras que la inferior derecha es a lo largo de su magnitud en la banda <i>r</i> . Los histogramas corresponden a la distribución de cada característica en todo del conjunto de datos. | 34 |
| 4.6. | Resultados de la regresión con 5 imágenes en multi-resolución del SDSS provenientes de HiPS2FITS. Los resultados de los 5 <i>folds</i> de test fueron promediados, tanto en la regresión como en las métricas presentes. En la ilustración de la izquierda se compartieron los pesos de los bloques convolucionales, mientras que la de la derecha se utilizaron bloques convolucionales independientes para cada sub-imagen. Las líneas rojas punteadas corresponden a los límites que definen los errores catastróficos. | 35 |

| | | |
|------|---|----|
| 4.7. | Promedio de la dispersión de los errores (σ_{MAD}) en los 5 <i>folds</i> de test y su respectiva desviación estándar en distintas características usando imágenes en multi-resolución del SDSS provenientes de HiPS2FITS tanto al compartir pesos como sin compartir pesos. La ilustración superior es a lo largo del <i>redshift</i> , la inferior izquierda a lo largo del radio de la galaxia, mientras que la inferior derecha es a lo largo de su magnitud en la banda <i>r</i> . Los histogramas corresponden a la distribución de cada característica en todo del conjunto de datos. | 36 |
| B.1. | Histogramas del SNR de las imágenes utilizadas de los <i>surveys</i> Pan-STARRS (color naranja) y SDSS (color morado). | 54 |
| B.2. | Comparación del color ($g - r$) en función del <i>redshift</i> utilizando magnitudes Petrosianas del SDSS (figura de la izquierda), magnitudes Kron de Pan-STARRS (figura del centro) y magnitudes con una apertura corregida también de Pan-STARRS (figura de la derecha). Las magnitudes fueron previamente corregidas por sus valores de extinción. | 55 |
| B.3. | Comparación del color ($r - i$) en función del <i>redshift</i> utilizando magnitudes Petrosianas del SDSS (figura de la izquierda), magnitudes Kron de Pan-STARRS (figura del centro) y magnitudes con una apertura corregida también de Pan-STARRS (figura de la derecha). Las magnitudes fueron previamente corregidas por sus valores de extinción. | 55 |
| B.4. | Comparación del color ($i - z$) en función del <i>redshift</i> utilizando magnitudes Petrosianas del SDSS (figura de la izquierda), magnitudes Kron de Pan-STARRS (figura del centro) y magnitudes con una apertura corregida también de Pan-STARRS (figura de la derecha). Las magnitudes fueron previamente corregidas por sus valores de extinción. | 56 |
| B.5. | Imágenes tipo <i>stack</i> extraídas de Pan-STARRS de galaxias. En la parte superior se encuentran las imágenes de la galaxia M99 en los filtros <i>grizy</i> (imágenes de izquierda a derecha), mientras que en la parte inferior se encuentran las de la galaxia M101. Ambos casos presentan artefactos en las mismas y/o distintas bandas. | 56 |
| C.1. | Resultados de las regresiones al usar imágenes en resolución simple del SDSS provenientes de HiPS2FITS sobre el conjunto de test. En la figura de la izquierda se utilizaron las bandas <i>ugriz</i> mientras que en la derecha se usaron las bandas <i>griz</i> . Las líneas rojas punteadas corresponden a los límites que definen los errores catastróficos. | 58 |
| C.2. | Dispersión de los errores (σ_{MAD}) sobre el conjunto de test en distintas características usando las imágenes de resolución simple del SDSS provenientes de HiPS2FITS al usar las bandas <i>ugriz</i> y <i>griz</i> . La ilustración superior es a lo largo del <i>redshift</i> , la inferior izquierda a lo largo del radio de la galaxia, mientras que la inferior derecha es a lo largo de su magnitud en la banda <i>r</i> . Los histogramas corresponden a la distribución de cada característica en el conjunto de test. . . | 59 |
| C.3. | Mean(shap values) obtenidos al usar las características extraídas de los bloques convolucionales en cada nivel con el clasificador del modelo, usando 9 galaxias con un $\text{PetroRad90} > 60$ ["]. La figura superior corresponde al caso de compartir pesos mientras que la inferior al caso de no compartir pesos. | 61 |

| | | |
|------|--|----|
| C.4. | Mean(shap values) obtenidos al usar las características extraídas de los bloques convolucionales en cada nivel con el clasificador del modelo, usando 10 galaxias con un $\text{PetroRad90} < 15$ ["]. La figura superior corresponde al caso de compartir pesos mientras que la inferior al caso de no compartir pesos. | 62 |
| C.5. | Promedio de la dispersión de los errores (σ_{MAD}) en los 5 <i> folds </i> de test y su respectiva desviación estándar en distintas características usando los niveles de forma independiente. La ilustración superior es a lo largo del <i>redshift</i> , la inferior izquierda a lo largo del radio de la galaxia, mientras que la inferior derecha es a lo largo de su magnitud en la banda <i>r</i> . Los histogramas corresponden a la distribución de cada característica en todo del conjunto de datos. | 64 |
| D.1. | $\langle \Delta z \rangle$ y σ_{MAD} promedio en función del SNR en cada banda del SDSS. Los resultados provienen de los experimentos de utilizar imágenes en resolución simple y de usar imágenes en multi-resolución con compartimiento de pesos. En ambos casos las imágenes provienen de HiPS2FITS. | 66 |

Capítulo 1

Introducción

La continua expansión del universo trae por consiguiente que los cuerpos celestes vayan interactuando entre ellos y también con el planeta Tierra, ya sea acercándose o alejándose. Esto trae por consiguiente que la información de flujo lumínico que se recopile de estos se vea perturbada por este fenómeno, ya que al alejarse dos cuerpos celestes las longitudes de onda de la luz observada se van desplazando hacia longitudes de onda mayores. Esto es lo que se conoce como *redshift*, el cual es un escalar que relaciona la longitud de onda observada con la emitida.

Conocer el *redshift* de los cuerpos celestes es fundamental si se quieren realizar estudios en estos, como por ejemplo, el cálculo de propiedades intrínsecas, tales como: luminosidades, masas, tasas de formación estelar, composición química, velocidades, distancias y cosmología en general. Ahora bien, para determinar el *redshift* con alta precisión se utilizan los espectros de los cuerpos celestes obtenidos de sondeos (*surveys*), sin embargo, este proceso requiere la observación de espectros en alta resolución y la SNR¹ para cada objeto [2], lo que es extremadamente laborioso y consume mucho tiempo. Es por eso que se introduce la idea de utilizar fotometría de banda ancha de imágenes de gran formato, las cuales poseen una resolución espectral menor pero pueden capturar una cantidad considerable de objetos más débiles, permitiendo a los astrónomos mapear estructuras a gran escala en el universo y estudiarlo en un contexto cosmológico más amplio.

Dado el surgimiento de las redes neuronales, en el año 2003 aparecieron modelos capaces de predecir el *redshift* de galaxias mediante características fotométricas como magnitudes y colores [3, 4]. Posteriormente, dada la popularidad de las redes neuronales convolucionales y del aprendizaje profundo, apareció por primera vez un modelo que utiliza las imágenes como tal de las galaxias en multi-bandas para la estimación del *redshift* [5]. Actualmente se siguen desarrollando modelos basados en redes neuronales convolucionales para esta tarea [1, 6–10], los cuales utilizan las imágenes centradas en galaxias en diversas bandas. Estas bandas corresponden a filtros que capturan el flujo lumínico en longitudes de ondas específicas. Ahora bien, para poder entrenar estos modelos con imágenes de igual tamaño, estos trabajos utilizan un campo de visión único ≈ 25.244 ["], no obstante, dado que las galaxias se encuentran a distintas distancias y tienen distintos tamaños, da por efecto que existan galaxias que no estén contenidas completamente dentro de ese campo de visión, por lo que la información como su morfología se pierde.

¹ De las siglas en Inglés *signal-to-noise ratio*.

Un problema similar ocurrió en una investigación en el *broker* ALeRCE [11], en donde buscaban crear un modelo que fuera capaz de encontrar la posición de la galaxia anfitriona dada una supernova. Pero como se mencionó anteriormente, dadas las distintas distancias y tamaños contener a una supernova con su galaxia anfitriona dentro de una misma imagen dependerá de cada caso. Si bien se podría utilizar un campo de visión suficientemente grande para poder capturar en todos los ejemplos a ambos cuerpos, los costos computacionales para almacenar estas imágenes aumentarán. Es por eso que utilizaron una técnica llamada multi-resolución, la cual consiste en tomar una imagen de cierta resolución y campo de visión, y a partir de esta, calcular sub-imágenes con distintas resoluciones y campos de visión a modo de representar la imagen original. La idea es que en cada sub-imagen (o también llamado nivel) el campo de visión se vaya reduciendo y a la vez aumentando proporcionalmente la resolución para que cada sub-imagen tenga las mismas dimensiones, reduciendo a su vez la memoria necesaria para almacenar estas sub-imágenes en comparación con la imagen original.

Considerando lo anterior, en esta memoria se busca evaluar la viabilidad de agregar imágenes en multi-resolución en la tarea de predecir el *redshift* de galaxias con imágenes en multi-banda. De esta forma, se lograría contener a todas las galaxias a utilizar dentro de las imágenes pero disminuyendo significativamente la cantidad de memoria computacional requerida para almacenarlas.

Para ello, se trabajará con las mismas galaxias que se utilizaron en la literatura, cuyas imágenes provienen del Sloan Digital Sky Survey (SDSS), pero añadiendo al estudio el uso de otro *survey* llamado Panoramic Survey Telescope and Rapid Response System (Pan-STARRS PS1). Como esta memoria busca ver el impacto de una técnica en la entrada del modelo y no en su arquitectura o configuración de hiperparámetros, se decidió utilizar el modelo de Pasquet et al. [1], ya que es un modelo base con el que se comparan los trabajos más recientes. Respecto a la forma en la que se realizará la comparación, lo que se hará será: (i) entrenar el modelo utilizando imágenes con la resolución y campo de visión utilizado en la literatura, (ii) entrenar el mismo modelo pero que reciba imágenes en multi-resolución en diferentes combinaciones de niveles, (iii) realizar la comparación de resultados en base a métricas de regresión que se han estado utilizando en estos trabajos.

1.1. Objetivos

1.1.1. Objetivo general

Evaluar el impacto que tiene añadir imágenes en multi-resolución en la tarea de predecir el *redshift* fotométrico de galaxias con imágenes en multi-banda, mediante el uso de una red neuronal convolucional.

1.1.2. Objetivos específicos

- O1 Utilizar y entrenar el modelo de Pasquet et al. [1] con imágenes de resolución simple.
- O2 Adaptar y entrenar el modelo de Pasquet et al. [1] para que sea capaz de recibir imágenes en multi-resolución.
- O3 Evaluar y comparar el desempeño de los modelos que usan imagen en multi-resolución con los que utilizan resolución simple.

Capítulo 2

Marco teórico y estado del arte

El presente capítulo se divide en dos secciones. En la primera se identifican los fundamentos teóricos y aspectos técnicos relacionados con el trabajo de la memoria, mientras que en la segunda se describe el estado del arte actual del problema.

2.1. Marco teórico

2.1.1. Surveys astronómicos

2.1.1.1. Sloan Digital Sky Survey (SDSS)

El Sloan Digital Sky Survey (SDSS) corresponde a un sondeo en donde se estudian dos aspectos; uno fotométrico y el otro espectroscópico. El objetivo de este estudio es crear una base de datos de imágenes digitales y espectroscopia del 25 % de la esfera celeste (aproximadamente 10.000 [deg]^2 o π esteridianos, centrado en el casquete polar galáctico norte) [12, 13].

El SDSS utiliza un telescopio dedicado de 2.5 metros en el Observatorio Apache Point en Nuevo México [14], el cual proporciona fotometría profunda de aproximadamente $r < 23$, con una resolución de 0.396 ["/pix] y utiliza una cámara CCD² compuesta de cinco bandas *ugriz* [15].

Las bandas corresponden a filtros tipo pasa-banda que dividen el rango de longitudes de onda de la luz, partiendo desde el corte ultravioleta atmosférico a 3000 [Å] hasta el límite de sensibilidad de los CCDs de silicio, el cual es de 11.000 [Å] . A continuación se muestran características de estos filtros:

- **Filtro *u*:** Su máxima sensibilidad se encuentra en 3500 [Å] , con una anchura efectiva (FWHM³) de 600 [Å] .
- **Filtro *g*:** Corresponde a una banda azul-verde con máxima sensibilidad en 4800 [Å] y anchura efectiva de 1400 [Å] .
- **Filtro *r*:** Corresponde a una banda de paso rojo con máxima sensibilidad en 6250 [Å] y anchura efectiva de 1400 [Å]

² *Charge-coupled device.*

³ *Full Width at Half Maximum.* Es una medida de la anchura del filtro en la que su eficiencia es al menos la mitad de su máximo.

- **Filtro *i***: Corresponde a una banda de infrarrojo lejano con máxima sensibilidad en 7700 [Å] y anchura efectiva de 1500 [Å].
- **Filtro *z***: Corresponde a una banda de paso en el infrarrojo cercano con máxima sensibilidad en 9100 [Å] y anchura efectiva de 1200 [Å].

En la Figura 2.1 se puede ver en detalle las respuestas de los distintos filtros descritos anteriormente, con la diferencia que los anteriores eran las respuestas efectivas esperadas, mientras que los de la figura corresponden a las respuestas finales obtenidas, las cuales dependen de factores externos tales como; respuesta del espejo, respuesta de la óptica, la eficiencia cuántica del CCD y la masa de aire presente en el lugar.

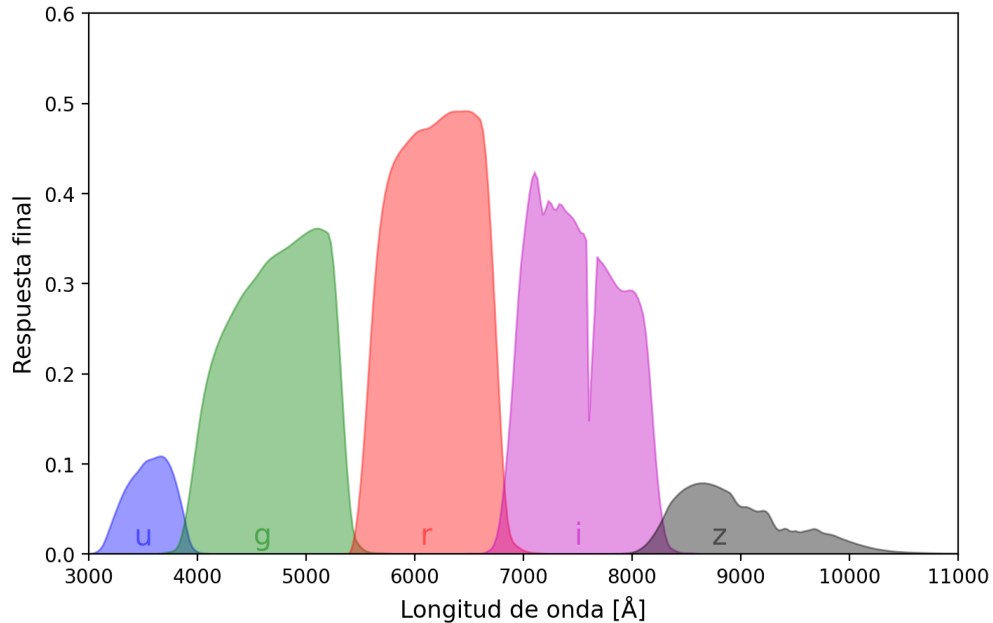


Figura 2.1: Respuestas finales obtenidas de los filtros *ugriz* del telescopio del SDSS. La ilustración fue obtenida de la librería `astroML` en Python (<https://www.astroml.org/index.html>).

La forma en la que se obtienen imágenes del cielo es mediante la cuenta de la cantidades de electrones que pasan en cada celda del CCD por el tiempo de exposición del telescopio, los cuales a su vez son determinados por los fotones recibidos. Luego de que haya terminado el tiempo de exposición se genera la imagen digital en base a estas cuentas. Los valores de cada pixel de esta imagen poseen una unidad particular llamada flujo, agregando de que cada pixel tiene un tamaño de 0.396 arcos de segundo ($''$), el cual es el ángulo utilizado para capturar un área del cielo en ese pixel.

Ahora bien, en este *survey* se realiza una transformación en las unidades de las imágenes, las cuales pasan de ser flujo a unidades de magnitud. Sin embargo, utilizan una forma particular de magnitud las cuales son informalmente llamadas *luptitudes* [16], en donde se aplica la transformación de la inversa del seno hiperbólico (arcoseno hiperbólico).

Adicionalmente, en base a las imágenes de cada cuerpo celeste se realizan diversos estudios, tales como: cálculos de magnitudes, errores, radios y fotometría en general, en donde todos los valores de magnitudes se encuentran en *luptitudes*.

2.1.1.2. Panoramic Survey Telescope and Rapid Response System (Pan-STARRS)

Corresponde a un observatorio de fotometría astronómica y procesamiento de datos desarrollado en el Instituto de Astronomía de la Universidad de Hawái. Su objetivo principal es realizar estudios en una gran cobertura del cielo y proporcionar respuestas rápidas a eventos astronómicos [17].

Pan-STARRS posee un telescopio principal llamado Pan-STARRS Telescope #1 (PS1), el cual es un telescopio de 1.8 metros equipado con una cámara de 1.4 gigapíxeles en multi bandas (*grizy_{P1}*), la cual posee una resolución de 0.258 ["/pix]. A diferencia del SDSS, Pan-STARRS presenta una cobertura mayor del área celeste, la cual es aproximadamente 30.000 [deg]² (o 3 π esteridianos) del casquete polar galáctico norte, a diferencia del SDSS que es tres veces menor.

Por otro lado la resolución es mayor, agregando que también existe una diferencia en los filtros que utiliza. En este caso se tiene la ausencia del filtro *u*, pero la presencia del filtro *y_{P1}*, dando por efecto una cobertura fotométrica más azulada en el SDSS y más cercanas al infrarrojo para PS1. Cabe destacar que la falta de la banda *u* en las estimaciones de *redshifts* (*z*) fotométricos grandes generalmente conduce a degeneraciones, ya que es importante para discriminar entre *redshifts* bajos ($z \lesssim 0.5$) y altos ($z \approx 2 - 3$). Sin embargo, es probable que esto no afecte demasiado este estudio, dado el rango de *redshift* investigado ($z < 0.4$).

Respecto a las imágenes que provee este *survey* hay dos tipos: *warp* y *stack*. El primer tipo son imágenes geoméricamente transformadas en una celda de cielo (*skycell* en ingles) en función del patrón de teselación del cielo de PS1, el cual es un conjunto de imágenes con una cuadrícula de píxeles común y proyecciones simples del cielo. Es decir, dada una imagen principal capturada por PS1 en una exposición, esta se divide en celdas del cielo conocidas determinadas por este patrón de teselación en una grilla de píxeles en común. Por otro lado, las imágenes *stack* son una combinación óptima de imágenes *warp* en la misma celda de cielo [18]. Con este tipo de proceso se obtienen imágenes más profundas y con mayor SNR.

Respecto a las unidades que se utilizan en las imágenes, éstas no están transformadas a magnitudes como en el SDSS, sino que están en unidades de flujo determinados por la cuenta de electrones por el tiempo de exposición de las imágenes.

2.1.2. Técnica de multi-resolución

En este trabajo se referirá a multi-resolución como una técnica de representación de una imagen en sub-imágenes de distintos campos de visión y resoluciones, tal como se describe en [19, 20].

La idea general de esta técnica o representación, es disminuir la cantidad de memoria necesaria para almacenar una imagen principal. Para ello se generan imágenes consecutivas las cuales van teniendo sucesivamente un menor campo de visión y mayor resolución. Si bien, esto se podría realizar de diferentes maneras, lo que hacen en la literatura es ir reduciendo

el campo de visión de la sub-imagen siguiente a la mitad pero aumentando su resolución al doble, entonces, dado $i \in 1, 2, 3, \dots, n$ el nivel o sub-imagen dentro de las n sub-imágenes a utilizar, c y r el campo de visión y resolución de la imagen principal respectivamente, las sub-imágenes tendrán un campo de visión y resolución igual a:

$$\text{FoV}_i = \frac{c}{2^{i-1}} [''], \quad (2.1)$$

$$\text{Resolucion}_i = r \cdot 2^{n-i} \left[\frac{''}{\text{pix}} \right]. \quad (2.2)$$

La forma en que se reduce el campo de visión es ir recortando la imagen a la mitad en cada nivel, mientras que para ir variando la resolución lo que se hace es submuestrear la imagen (*downsampling*) de cada nivel respectivo. Existen muchas formas de submuestrear una imagen, como por ejemplo; eliminar columnas y filas pares/impares, aplicar algún tipo de *pooling* de cierto tamaño como promedio, máximo, etc. En el caso de imágenes astronómicas, existen cambios muy abruptos de valores en los pixeles debido a la gran presencia de ruido, es por eso que se aplican ventanas cuadradas de la mediana debido a la robustez que tiene frente a valores anómalos extremos. El tamaño de esta ventana dependerá del nivel de la sub-imagen que se este trabajando y a la cantidad total de niveles a utilizar, dado por $2^{n-i} \times 2^{n-i}$.

Respecto a la cantidad de píxeles que tendrá una imagen cuadrada en una banda, esta será de $\frac{c^2}{r^2} [\text{pix}]^2$, en cambio, al utilizar multi-resolución con n sub-imágenes se tendrían:

$$N_{\text{pix}} = \sum_{i=1}^n \frac{\left[\frac{c}{2^{i-1}}\right]^2}{\left[r \cdot 2^{n-i}\right]^2} = \sum_{i=1}^n \frac{c^2}{r^2} \cdot \frac{1}{2^{2n-2}} = \frac{c^2}{r^2} \cdot \frac{n}{2^{2n-2}} [\text{pix}]^2, \quad (2.3)$$

de esto se pueden ver dos aspectos importantes; el primero es que cada nivel tiene la misma cantidad de pixeles, y el segundo es que la cantidad de pixeles total a utilizar en multi-resolución es proporcional a la cantidad de pixeles de la imagen principal en un factor de $\frac{n}{2^{2n-2}}$. Entonces dada una imagen cuadrada cualquiera, tal que sus lados sean divisibles por 2^n , la cantidad de pixeles que se utilizarán en multi-resolución para $n = 5$ será $\approx 2\%$ de la imagen original, es decir, se reduciría un 98% de la memoria necesaria para almacenar la imagen original.

Cabe destacar que esta técnica también trata de imitar el funcionamiento visual animal/humano. Al estar observando un punto, solo un área pequeña centrada en ese punto se puede ver con detalles, mientras que todo lo periférico a esa área no se puede observar con el mismo nivel de detalle. Esto se debe principalmente a la distribución de los conos en la retina, los cuales son responsables de la visión en alta resolución y la percepción del color, sin embargo, estos se encuentran concentrados en una pequeña área de la retina llamada fovea, la cual tiene aproximadamente 1-2 grados de campo visual. En la Figura 2.2 se muestra la comparación entre una imagen de resolución simple frente a usar multi-resolución, donde se puede ver el fenómeno descrito anteriormente.

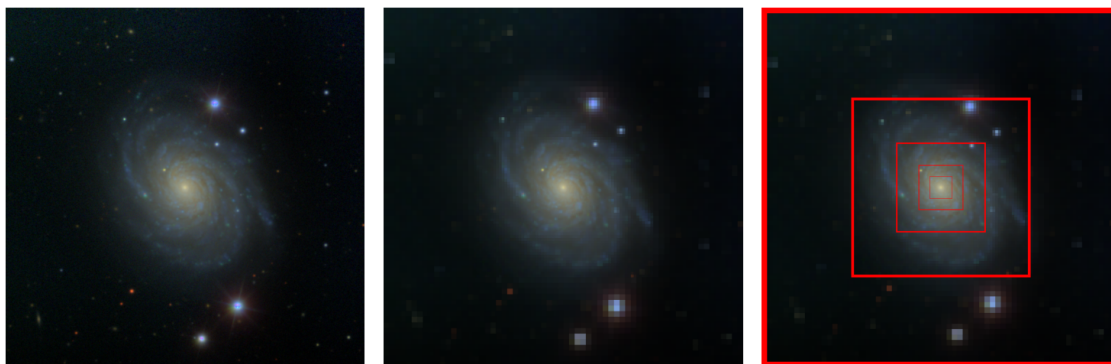


Figura 2.2: Comparación entre una imagen de una galaxia en resolución simple (imagen de la izquierda) con respecto a usar imágenes en multi resolución en 5 niveles (imagen del centro). Adicionalmente se demarcan los 5 niveles para una mejor diferenciación entre ellos (imagen de la derecha).

Se puede observar que esta técnica se enfoca en el objeto centrado en la imagen, reduciendo gradualmente los detalles desde el nivel con mayor campo de visión y concentrado los detalles en el objeto celeste como tal. Esta es la forma en la que se puede visualizar un objeto de gran tamaño pero sin darle mucha atención al contorno o contexto donde se encuentra, reduciendo el detalle de esa información y por ende a la cantidad de memoria RAM necesaria para almacenar su imagen.

2.1.3. Redes neuronales profundas

Cuando se habla de aprendizaje profundo no se tiene un consenso único sobre su definición. Lo más habitual es cuando se modifica la arquitectura de la red y se añaden varias capas ocultas, a este tipo de modelo se le puede llamar como red neuronal profunda. Por otro lado, cuando a una red artificial se le entrega una entrada cruda (o tipo *raw*) y es la misma red la que extrae patrones y características útiles que le sirven para resolver el problema en que se esta entrenando, a este tipo de técnica se le conoce como *end-to-end* en la que se utiliza una red neuronal profunda (como por ejemplo [1, 6–10]).

Si bien el aprendizaje profundo es un subconjunto de lo que son las redes neuronales y el aprendizaje de máquinas (*Machine Learning*), se mencionarán en orden de generalidad, partiendo desde lo que son las redes neuronales artificiales, luego de un tipo de arquitectura específico de estas redes que se utilizó en el trabajo de esta memoria, y por último del funcionamiento del aprendizaje de estas redes.

2.1.3.1. Redes neuronales artificiales

Las redes neuronales artificiales (o ANNs por su acrónimo en inglés) [21] son modelos matemáticos que son utilizados para resolver algún problema (o *task* en inglés) que se tenga, como por ejemplo clasificación o regresión (entre los más comunes). La idea es que estos modelos reciban una entrada específica y que de forma interna pueda entregar una salida deseada, ya sea un vector de probabilidades de las clases en un problema de clasificación o un valor específico en un problema de regresión.

Antes de ver cómo estas redes procesan la entrada, es fundamental mencionar que las ANNs están compuestas por unidades computacionales elementales interconectadas llamadas neuronas (o perceptrón [22]). Al igual que en la biología, donde una neurona natural es una célula que al excitarse eléctricamente procesa y transmite información a través de sinapsis, una neurona artificial fue definida de tal forma de imitar este proceso desde un punto de vista matemático.

Dentro del concepto anterior hay cuatro elementos básicos que componen a una neurona artificial; los pesos sinápticos que ponderan las entradas, un sumador o combinador lineal, un sesgo y una función de activación (generalmente no-lineal) que limita la amplitud de la salida. De esto, se tienen ecuaciones del procesamiento de una entrada en una neurona artificial, las cuales se detallan a continuación:

$$u_k = \sum_{j=1}^n w_{kj} \cdot x_j, \quad (2.4)$$

$$y_k = \varphi(u_k + b_k), \quad (2.5)$$

en estas se tiene que x_j es la entrada j -ésima, w_{kj} es el peso sináptico de una neurona k asociado a la entrada j -ésima, $\varphi(\cdot)$ es la función de activación no-lineal, u_k es la salida del combinador lineal, b_k es el sesgo o *bias*, y por ultimo, y_k es la salida de la neurona. Una forma visual de ver el modelo de una neurona se presenta en la Fig. 2.3.

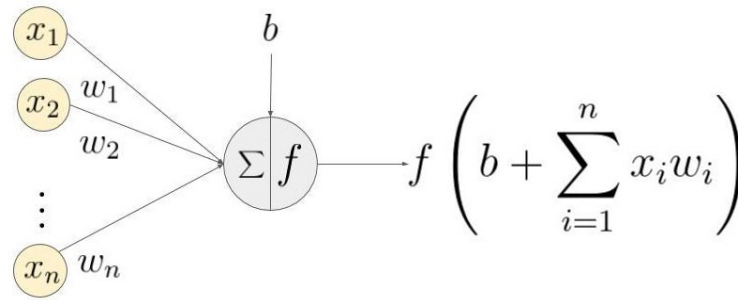


Figura 2.3: Ilustración del modelo de una neurona artificial, donde x_n es la componente n -ésima del vector de entrada, w_n es el peso asociado a la componente x_n , b es el sesgo (o *bias*) y $f(\cdot)$ es una función de activación. Ilustración extraída de <https://learnopencv.com/understanding-activation-functions-in-deep-learning/>.

Ahora bien, las ANN's están compuestas por capas, las cuales son agrupaciones de neuronas. A este tipo de capas se les llama capas totalmente conectadas o capas *fully connected* en inglés, debido a que cada una de las neuronas de la capa están totalmente conectadas a la entrada que se les este dando. De forma general, la salida de una capa totalmente conectada compuesta de m neuronas esta dada por:

$$\mathbf{y} = \phi(\mathbf{W}\mathbf{x} + \mathbf{b}), \quad (2.6)$$

donde $\mathbf{y} \in \mathbb{R}^m$ es la salida de la capa, $\mathbf{x} \in \mathbb{R}^l$ es la entrada de la red, $\mathbf{W} \in \mathbb{R}^{m \times l}$ es la matriz de pesos, $\mathbf{b} \in \mathbb{R}^m$ es el vector de sesgos y $\phi(\cdot)$ es la función de activación no-lineal. La configuración del número de capas o de neuronas define la arquitectura de la red neuronal.

Existen varios tipos de funciones de activación, dentro de las más utilizadas están: la función sigmoide, tangente hiperbólica, la unidad lineal rectificadora (ReLU por sus siglas en inglés) [23], etc. Las ecuaciones de las funciones mencionadas se muestran a continuación:

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}}, \quad (2.7)$$

$$\text{tanh}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \quad (2.8)$$

$$\text{ReLU}(x) = \max\{0, x\}. \quad (2.9)$$

El propósito de las funciones de activación no-lineales es que se puedan obtener, valga la redundancia, salidas no-lineales. La ventaja de esto es que pueden existir problemas los cuales no pueden ser resueltos con transformaciones lineales, por lo que estas funciones de activación ayudan a tener una mejor aproximación de solución en estos casos, dado que se le permite a la red aprender relaciones y patrones más complejos en los datos.

Ahora bien, para que estas redes puedan entregar salidas esperadas y no aleatorias, se ajustan los parámetros de la red (pesos y sesgos) de tal forma de minimizar una función de costo en un problema de optimización sobre un conjunto de datos. A este conjunto de datos se le conoce como conjunto de entrenamiento, y al proceso de optimización se le conoce como proceso de entrenamiento del modelo. Esto se verá a mayor detalle adelante.

2.1.3.2. Redes neuronales convolucionales

Un problema de las redes neuronales totalmente conectadas es que la entrada esta limitada exclusivamente a un vector con cierta dimensionalidad, agregando que no se aprovecha la información que puede ser extraída de entradas que tengan dimensiones adyacentes correlacionadas (como por ejemplo imágenes, series temporales, espectros, etc.). Es por estas limitaciones que aparecen las redes neuronales convolucionales (o CNNs por sus siglas en inglés) [24].

Una CNN es análoga a una ANN con capas *fully connected*, pero con diferencias claves [25]. Una de ellas es que la entrada ya no esta limitada a vectores, ahora la entrada puede ser una matriz de cierto tamaño y profundidad. Por otro lado, en vez de tener capas totalmente conectadas, las CNN poseen capas convolucionales, las cuales están compuestas por filtros (o *kernels* en inglés) que corresponden a una matriz de pesos de cierta profundidad y tamaño. Esta profundidad se le conoce como el número de canales (o *channels* en inglés) que tiene la capa convolucional. Dicho esto, la salida de una capa convolucional esta dada por:

$$\mathbf{y} = \phi(\mathbf{W} * \mathbf{x} + \mathbf{b}), \quad (2.10)$$

donde \mathbf{y} es la salida de la capa convolucional, \mathbf{x} es la entrada, \mathbf{W} es la matriz de filtros, \mathbf{b} es el vector de sesgos, $\phi(\cdot)$ es la función de activación no-lineal y $*$ corresponde al operador de convolución.

En este caso, cada componente de la entrada no tiene un peso único asociado (como en el caso de una capa totalmente conectada), sino que los filtros se aplican sobre toda la entrada,

aplicando la convolución en un área delimitada por el tamaño de los filtros y desplazándose a lo largo (y ancho en el caso de que la entrada fuera por ejemplo una imagen) de la entrada. Al compartir pesos, se logra extraer el mismo tipo de características que puede llegar a tener una entrada. Por ejemplo, en el caso de una imagen se podrían detectar bordes con un filtro Laplaciano de cierto tamaño, reduciendo la cantidad de parámetros entrenables del modelo y también teniendo la capacidad de poder extraer características locales de la entrada.

Como este trabajo se basa en utilizar imágenes, se dará el ejemplo del proceso de convolución entre una imagen y un filtro. Dada $\mathbf{x} \in \mathbb{R}^{w \times h \times c}$ una imagen de ancho w , alto h y c canales, $\mathbf{W} \in \mathbb{R}^{m \times n \times c \times k}$ una matriz de k filtros de tamaño $m \times n$ con c canales, el resultado de la convolución será un mapa de características de k canales de tamaño $a \times b$ ($\mathbb{R}^{a \times b \times k}$). El tamaño de este mapa de características dependerá del tamaño de la imagen, del tamaño del filtro y también de cómo el filtro se desplaza a lo largo del ancho y alto de la entrada, este tamaño de desplazamiento se le conoce como *stride*.

El *stride* es un hiperparámetro configurable por el usuario el cual es un valor entero estrictamente positivo, y dado este valor, se puede conocer el tamaño (ancho y alto) del mapa de características resultantes, el cual esta dado por:

$$U = \frac{E - D}{S} + 1, \quad (2.11)$$

donde U corresponde al tamaño de uno de los lados del mapa de características, E corresponde al tamaño de la entrada en el mismo lado, D es el tamaño del filtro en el mismo lado y S corresponde al valor de *stride*. Por ejemplo, el resultado de la convolución de una imagen de tamaño 8×8 pixeles con un filtro cualquiera de tamaño 2×2 y usando un *stride* igual a 2, dará por resultado un mapa de características de tamaño 4×4 .

Ahora bien, si se quiere que la salida tenga las mismas dimensiones de la entrada, no existe ningún valor de *stride* que provea esa condición. Para ello se utiliza un método llamado *padding*, el cual consiste en rellenar los bordes de la entrada con ceros. Así, se podría agrandar el tamaño de la entrada de tal forma de que existan varias posibles opciones de *stride* para que la salida tenga las mismas dimensiones que la entrada sin *padding*.

También existen otros elementos relevantes en una CNN, como las capas de *pooling*. Estas capas tienen un propósito similar a las funciones de activación no-lineales, agregando que tienen la capacidad de poder disminuir la dimensionalidad (tamaño) a la entrada. Los dos más usados son:

- **Avgpooling:** Entrega el promedio dentro de un vecindario de cierto tamaño en cada uno de los canales. Al igual que un *kernel*, este también se va desplazando a lo largo de la entrada con un cierto *stride*. Es análogo a realizar la convolución utilizando un *kernel* normalizado tal que sus pesos sean iguales y que su suma sea igual a 1.
- **Maxpooling:** Análogo al *pooling* anterior, solo que en este caso retorna el máximo dentro del vecindario.

Cuando se trabaja con imágenes, estas capas de *pooling* suelen ser cuadradas y se aplican comúnmente con un *stride* de 2. Lo mismo ocurre con los *kernels* de las capas convolucionales,

sin embargo estos son hiperparámetros configurables por el usuario, y no todos los problemas son iguales, por lo que pueden variar en cada caso.

Respecto a la arquitectura de este tipo de redes neuronales, usualmente está compuesta en primer lugar de bloques convolucionales, los cuales corresponden a un conjunto de tres elementos; capa convolucional, función de activación, capa de *pooling*. Generalmente se utilizan varios bloques convolucionales con diversos hiperparámetros (tamaños de *kernels*, cantidad de canales, tamaño de *stride*, etc.), los cuales sirven para ir extrayendo características cada vez más complejas de la entrada principal, en donde luego de extraer estas características, se suele aplanar este vector para ser introducido en capas totalmente conectadas en conjunto de funciones de activación, en donde finalmente se define una capa de salida del tamaño deseado. En la Figura 2.4 se puede ver un ejemplo de arquitectura de una CNN, la cual corresponde al modelo DELIGHT [19].

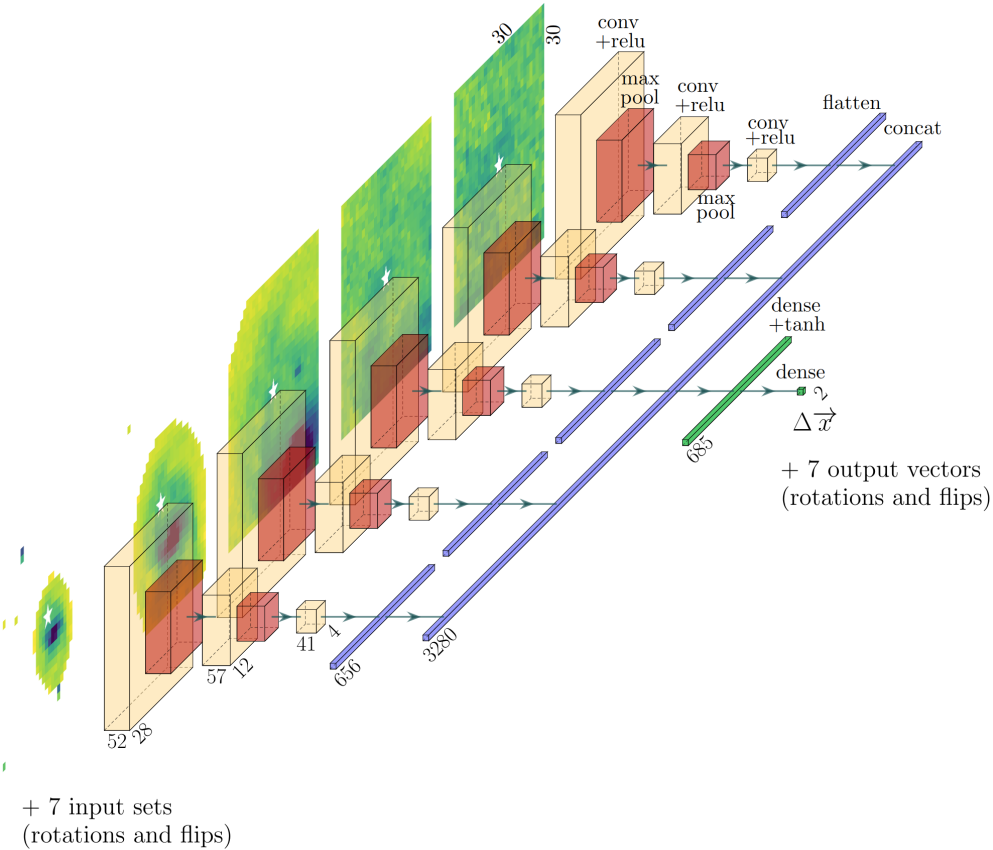


Figura 2.4: Ejemplo de arquitectura de una red neuronal convolucional correspondiente al modelo DELIGHT. Ilustración obtenida de [19].

2.1.3.3. Aprendizaje de redes neuronales

Como se mencionó anteriormente, el propósito del aprendizaje (o entrenamiento) de estas redes es ajustar iterativamente los pesos sinápticos y sesgos hasta que el modelo se comporte como se pretende, utilizando un set de entrenamiento durante el proceso de entrenamiento y el tipo de problema que se este abordando, como por ejemplo: clasificación, regresión, etc.

En otras palabras, dada una red neuronal artificial f_θ , lo que se busca es encontrar o aproximar una función $f_{\theta^*}(x) = y$, donde θ^* son los parámetros óptimos de la red que minimizan una función de costo dada durante el entrenamiento (y sobre el conjunto de entrenamiento). La forma de encontrar la optimalidad, es resolviendo el siguiente problema de optimización:

$$\theta^* = \arg \min_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^N \mathcal{L}(y^{(i)}, f_\theta(x^{(i)})), \quad (2.12)$$

donde \mathcal{L} es una función (diferenciable y convexa) de costo utilizada, $y^{(i)}$ es la variable objetivo del ejemplo i -ésimo dentro del conjunto de entrenamiento y $f_\theta(x^{(i)})$ es la predicción del modelo dada una entrada $x^{(i)}$ del conjunto de entrenamiento de N muestras.

Ahora bien, la forma de abordar este problema de optimización se hace mediante técnicas basadas en el método de gradiente descendente (Ecuación 2.13), en el cual, dada una instancia específica, lo que hace es actualizar los parámetros de la red (pesos y sesgos) en proporción y dirección opuesta al gradiente, ponderado por un escalar llamado tasa de aprendizaje (η), de tal forma de aproximarse idealmente al mínimo global de la función de costo:

$$\theta_{t+1} = \theta_t - \eta \nabla_{\theta_t} \left(\sum_{i=1}^N \mathcal{L}(y^{(i)}, f_\theta(x^{(i)})) \right). \quad (2.13)$$

La razón de utilizar el opuesto del gradiente se debe a que el gradiente de una función es un vector que apunta en dirección de máximo crecimiento de esa función evaluada en una instancia particular, por lo que el opuesto apunta en dirección de máximo decaimiento, lo cual es útil si se busca minimizar dicha función.

La forma de calcular el gradiente es mediante la técnica de *backpropagation*. Al ser la predicción del modelo una función compuesta por las distintas capas de la red neuronal, lo que se hace esta técnica es calcular los gradientes de cada uno de los parámetros de la red desde la salida hacia la misma entrada, esto lo hace mediante el uso de la regla de la cadena de la derivada. Una vez calculados los gradientes y evaluados en la instancia actual, se actualizan los parámetros de la red, a esta actualización se le conoce como iteración. Para buscar la convergencia lo que se hace es entrenar iterativamente la red, al utilizar todo el conjunto de entrenamiento por una vez en el proceso de entrenamiento se le conoce como época, y se utilizan cuantas épocas sean necesarias para llegar a una convergencia adecuada.

Sin embargo, usar todo el conjunto de entrenamiento es altamente costoso computacionalmente, especialmente si se tiene que calcular el gradiente de millones de parámetros en un conjunto de entrenamiento grande. Es por eso que nace *Stochastic Gradient Descent* (SGD por sus siglas en inglés) [26, 27], en donde se va tomando de forma individual cada muestra del conjunto de entrenamiento y actualizando los parámetros en función de cada una de ellas de forma independiente. No obstante, la convergencia puede ser muy ruidosa, debido a que los datos pueden ser distintos y no representativos, causando por ejemplo, gradientes que apunten en direcciones opuestas y que el aprendizaje no sea adecuado, en donde incluso puede ocurrir que la red no logre aprender nada.

Es por eso que posteriormente nace la idea de dividir el conjunto de entrenamiento en un subconjunto de muestras aleatorias (sin repetición) de cierto tamaño ($|\mathcal{B}|$), el cual es llamado *batch* o *mini-batch*. Esta es una forma no sesgada de calcular el gradiente y de utilizar el optimizador de SGD, la cual esta dada por la siguiente ecuación:

$$\theta_k = \theta_{k-1} - \eta \frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} \nabla_{\theta_t} \mathcal{L}(y^{(i)}, f_{\theta}(x^{(i)})), \quad (2.14)$$

cabe destacar que si el tamaño de *mini-batch* es igual a 1, es equivalente a utilizar SGD y si se utiliza un tamaño igual al tamaño del set de entrenamiento es lo mismo que utilizar el método de descenso de gradiente clásico.

Existen otros tipos de optimizadores, como por ejemplo el optimizador Adam [28], el cual es un algoritmo de optimización con tasa de aprendizaje adaptativa que ajusta η de forma automática. Este método utiliza los gradientes elevados al cuadrado para escalar la tasa de aprendizaje e incluye el promedio móvil de los gradientes en su formulación, esta estrategia es conocida como momentum, la cual ayuda a evitar la convergencia a un mínimo local durante la optimización. Los principales hiperparámetros de Adam son $\beta_1 = 0.9$ y $\beta_2 = 0.999$, que se relacionan al promedio móvil de los gradientes y los gradientes al cuadrado, respectivamente, regulando la tasa a la que se ajusta la tasa de aprendizaje.

También es posible regular la tasa de aprendizaje durante el entrenamiento con un *schedule*, ya sea usándolo con SGD como también con Adam, aunque con este último es menos común, ya que como se habló el mismo algoritmo se encarga de ajustar este hiperparámetro.

El aprendizaje también se basa en muchos aspectos fundamentales. Entre los más importantes está la selección de hiperparámetros, tales como; la función de costo utilizada, el optimizador, la tasa de aprendizaje, el tamaño del *mini-batch*, número de capas de la red, el número de épocas de entrenamiento, entre otros. Por ejemplo, utilizar una tasa de aprendizaje elevada puede provocar que la proporción con la que se aproxima al mínimo sea demasiado grande, dando por efecto que incluso en la actualización de parámetros se aleje aun más del mínimo, específicamente cuando se esta cerca de este. Por su contraparte, escoger una tasa de aprendizaje demasiado pequeña provoca una convergencia demasiado lenta, e incluso, durante el entrenamiento se podría llegar a un mínimo local, siendo demasiado difícil poder salir de este. Lo mismo ocurre con el resto de hiperparámetros.

El término del entrenamiento de estas redes (aparte de la convergencia en la optimización) depende exclusivamente del usuario, debido a que si bien se puede llegar a un valor muy bajo de costo sobre el conjunto de entrenamiento, esto no significa que frente a otro conjunto con datos no utilizados ocurra lo mismo. Es por eso que se utiliza un conjunto de validación, el cual sirve únicamente para monitorear (no para entrenar) el aprendizaje de tal forma que la red no se ajuste demasiado al conjunto de entrenamiento.

Dicho esto, una opción para que el entrenamiento se detenga es cuando la función de costo sobre el conjunto de validación no mejora en un número específico de épocas, pero el de entrenamiento sigue mejorando. A esto se le conoce como sobreajuste (*overfitting* en inglés), y es cuando la red se ajusta tanto al conjunto de entrenamiento que no le es posible generalizar frente a datos no vistos. A la técnica de detención se le conoce como *Early-Stopping* y a

la cantidad de épocas o iteraciones de monitoreo se le conoce como paciencia (*patience* en inglés), el cual es un hiperparámetro definido por el usuario.

Al igual que este criterio, también existen otros métodos que sirven para evitar el sobreajuste de la red, tales como; el uso de *dropout* [29], la implementación de *data augmentation* [30], la simplificación de la red [31], etc.

Una vez finalizado el entrenamiento de la red, lo que se hace es seleccionar el mejor modelo, o en otras palabras, seleccionar la mejor combinación de parámetros θ^* . Para ello se selecciona la combinación de parámetros cuando la función de costos sobre el conjunto de validación fue mínima (ya sea en una iteración o época). Esto se debe a que en esa instancia la red logró la mejor generalización durante el proceso de entrenamiento, es decir, logró funcionar de la mejor forma frente a un conjunto que no utilizó para actualizar sus parámetros.

Adicionalmente se suele utilizar un tercer conjunto llamado conjunto de test. Este es usado para obtener métricas finales del mejor modelo encontrado con datos que no se han usado ni para entrenar ni para monitorear su generabilidad.

2.2. Estado del arte

La estimación de los *redshifts* de galaxias en función de su fotometría es algo que ha estado en continua investigación. Es por eso que en esta sección se presentará un subconjunto de estudios actuales que comparten cosas en común, como por ejemplo; el *survey* utilizado para obtener datos y usos de técnicas de *Machine Learning* y redes neuronales convolucionales.

En particular, existen dos investigaciones principales de las cuales investigadores se motivaron a encontrar mejores soluciones a la estimación del *redshift*. La primera es de Beck et al. [32], en donde se utilizan características fotométricas en modelos de Machine Learning para este problema, mientras que el segundo es el de Pasquet et al. [1], quienes utilizan una red neuronal convolucional para solucionar el mismo problema usando directamente las imágenes de estas galaxias, y de esta forma, comparar su metodología con la de Beck et al. [32].

Posteriormente se han realizado más investigaciones que toman de base el trabajo de Pasquet et al. a modo de *baseline*, pero modificando la arquitectura del modelo, el tipo de aprendizaje, cambios en la entrada y otros aspectos. Sin embargo, en ninguno utilizan imágenes en multi resolución, es por eso que se describirá el trabajo de Förster et al. [19], donde se introduce esta técnica en imágenes astronómicas, agregando que es de esa investigación que surge la idea de usar esta técnica en el trabajo de esta memoria.

A continuación se presentarán todos los estudios mencionados en orden temporal.

Beck et al. [32]

En 2016, Beck et al. utilizaron una regresión lineal local para estimar el *redshift* de galaxias usando características fotométricas de estas. Para ello se utilizó datos del Data Release 12 (DR12 [12]) del SDSS, obteniendo 1.976.978 ejemplos de galaxias con un $z < 0.8$ y $r < 21.5$. Respecto a las características fotométricas, se utilizaron cuatro colores y la magni-

tud corregida en la banda r , teniendo así un vector de características de cinco dimensiones ($[r, (u - g), (g - r), (r - i), (i - z)]$). La magnitud que utilizaron fue la **cModelMag**, mientras que para los colores utilizaron magnitudes **modelMag**. La corrección de las magnitudes fue respecto a su extinción galáctica [33].

El entrenamiento se realizó con *cross-validation* utilizando un $k = 2$, es decir, se dividió el conjunto de datos en dos partes iguales, las cuales se utilizaron para entrenar, validar y viceversa. De esto, se obtuvo un residuo promedio de $\langle \Delta z \rangle = 5.84 \cdot 10^{-5}$, una desviación estándar igual a $\sigma = 0.0205$, y para calcular la tasa de errores catastróficos utilizaron un valor de 3σ , dándoles así un $\eta = 4.11\%$. En la Figura 2.5 se pueden ver los resultados de la regresión en función del *redshift* predicho y también de los residuos no normalizados ($z_{phot} - z_{spec}$).

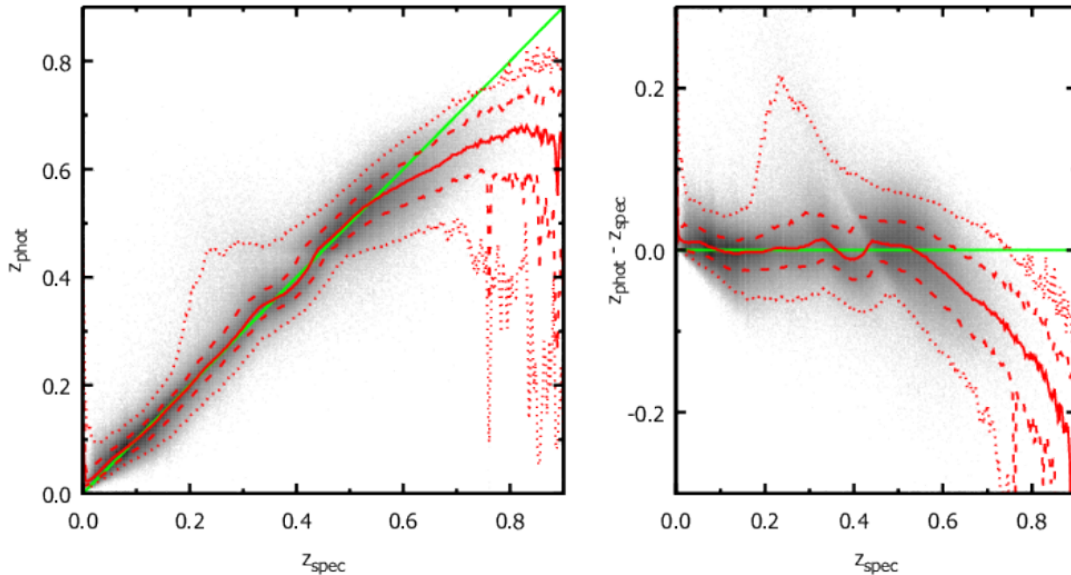


Figura 2.5: Resultados de la regresión lineal local del trabajo de Beck et al. [32]. En la figura de la izquierda se presenta la comparación del *redshift* predicho (z_{phot}) con el espectroscópico (z_{spec}), mientras que la figura de la derecha contiene los residuos no normalizados de las predicciones. Las líneas rojas punteadas corresponden a intervalos de confianza. Ilustración obtenida de [32].

Pasquet et al. [1]

Unos años más tarde, Paquet et al. [1] propusieron una red neuronal convolucional (CNN), la cual se basa en módulos de la red Inception [34] en conjunto de capas *fully connected*, donde la entrada de la red son imágenes centradas en galaxias utilizando las bandas *ugriz*. El campo de visión utilizado para cada imagen fue de $\approx 25.344''$ con una resolución de $0.396''/pix$. La cobertura fue de $z < 0.4$ y una magnitud $r < 17.8$. De esto, se obtuvieron 516,525 imágenes de galaxias de 64×64 píxeles en 5 canales (*ugriz*). La base de datos que utilizaron fue el DR12 del SDSS.

Algo fuera de lo común que se implementó, es que convirtieron este problema de regresión en uno de clasificación. Para ello dividieron linealmente el rango de *redshift* en 180 *bins* de tamaño $\delta z = 2.2 \times 10^{-3}$, asignándole a cada valor de *redshift* su número de *bin* respectivo. De

esta forma, en vez de solo obtener el valor de *redshift* predicho, se lograba obtener la función de densidad de probabilidad (PDF), lo cual entregaba una mejor visualización estadística de cada una de las predicciones. Luego, para poder obtener el valor de *redshift* final, lo que se hizo fue utilizar la esperanza matemática de la función de densidad de probabilidad en función del punto medio de cada *bin* asociado, tal como se muestra a continuación:

$$z_{phot} = \sum_{i=1}^{180} z_k P(z_k), \quad (2.15)$$

donde z_{phot} es el valor escalar final del *redshift* predicho, z_k es el punto medio del *bin* k -ésimo y $P(z_k)$ es la probabilidad obtenida del *bin* k -ésimo.

Adicionalmente incluyeron el valor de extinción galáctica de las galaxias (E(B-V)) [33], el cual fue concatenado en el vector de características final extraído de los bloques convolucionales. La razón de esto fue porque el polvo de las galaxias atenúa las longitudes de onda más pequeñas, dando por efecto que se vean más rojizas.

La forma en la que se realizó el entrenamiento fue utilizar cinco modelos con inicializaciones de pesos distintas, en donde se promediaron los cinco *redshifts* predichos como valor final para cada ejemplo. Además de eso, para evaluar de forma robusta los resultados se utilizó *cross-validation* con $k = 5$, en donde para cada *fold* se utilizaron los cinco modelos mencionados anteriormente.

Este trabajo se posicionó en el estado del arte del problema de la estimación del *redshift* fotométrico al obtener mejores resultados que los de Beck et al. [32] (Figura 2.6). Para comparar la *performance* de ambos métodos se promediaron los resultados de métricas de los *folds* de test. Las métricas utilizadas fueron las siguientes:

$$\langle \Delta z \rangle = \frac{1}{N} \sum_{i=1}^N \frac{z_{phot_i} - z_{spec_i}}{1 + z_{spec_i}}, \quad (2.16)$$

$$\sigma_{MAD} = 1.4826 \times MAD = 1.4826 \times \text{Mediana}(|\Delta z - \text{Mediana}(\Delta z)|), \quad (2.17)$$

$$\eta = \left(\frac{N_{\Delta z > 0.05}}{N} \right) \times 100, \quad (2.18)$$

donde $\langle \Delta z \rangle$ es el promedio de los residuos normalizados Δz el cual esta dado por $\Delta z = \frac{z_{phot} - z_{spec}}{1 + z_{spec}}$, $\sigma_{MAD} = 1.4826 \times MAD$, donde MAD (Median Absolute Deviation, por sus siglas en inglés) es la mediana de $|\Delta z - \text{Mediana}(\Delta z)|$ y η es la cantidad normalizada de residuos mayores a 0.05 ($|\Delta z| > 0.05$), también considerados como errores catastróficos.

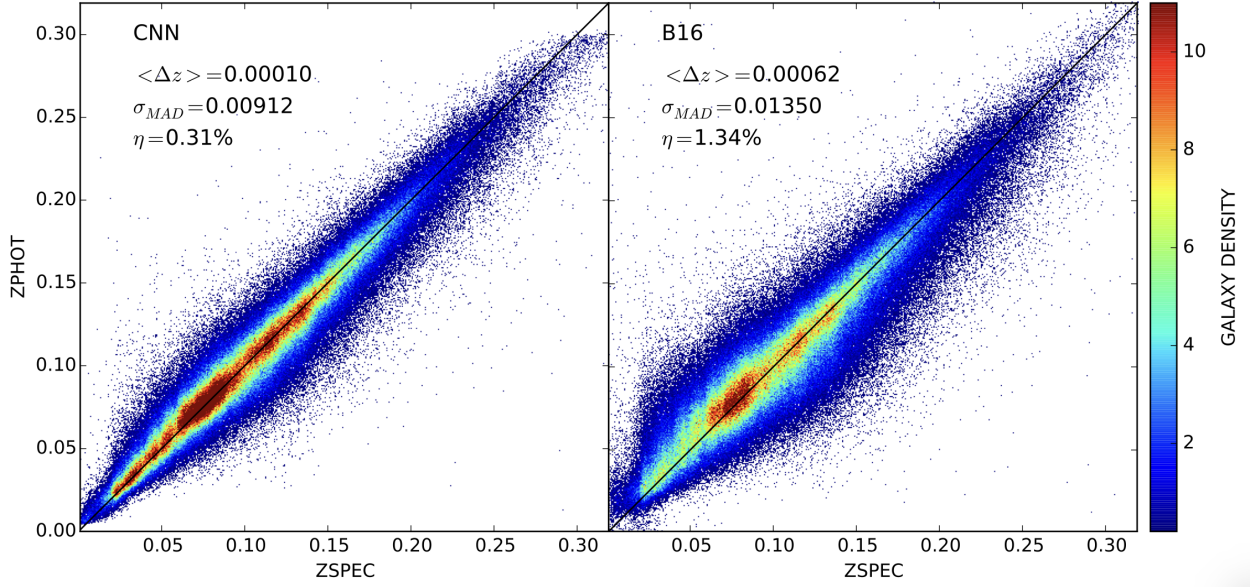


Figura 2.6: Resultados promedio de los cinco *folds* de test usando la CNN propuesta por Pasquet et al. (figura de la izquierda), comparado con el trabajo de Beck et al. sobre el mismo conjunto de datos (figura de la derecha). Ilustración extraída de [1].

Se puede apreciar que con este nuevo método se redujo seis veces el promedio de los residuos normalizados, la dispersión de los residuos se redujo en un 66.5% y la tasa de *outliers* se redujo en un 23% comparado con los resultados de [32].

Otro resultado de este trabajo fue que el aspecto más importante que determina la calidad de los resultados es el SNR de las imágenes. Para ello entrenaron el modelo utilizando solo imágenes del Stripe 82, el cual es un área del cielo donde se pueden obtener imágenes bastante profundas (SNR muy alto), entregando un resultado de dispersión de los residuos considerablemente menor con respecto a utilizar imágenes con SNR menor. Así también, se realizó un test con imágenes del Stripe 82 en el modelo entrenado con los datos originales, en donde se obtuvo el mismo resultado. Esto se puede ver en mayor detalle en la Tabla 2.1.

Tabla 2.1: Comparación de $\langle \Delta z \rangle$, σ_{MAD} y tasa de outliers (η) al usar o no usar imágenes del Stripe 82 en entrenamiento y/o test. Resultados extraídos de [1].

| Experimento | Tamaño set entrenamiento | Tamaño 1 set de test | $\langle \Delta z \rangle$ | σ_{MAD} | η |
|---|--------------------------|----------------------|----------------------------|----------------|--------|
| Entrenamiento con un 80% del dataset | 393,219 | | | | |
| Set de test completo | | 103,306 | 0.0001 | 0.00912 | 0.31 |
| Usando solo Stripe 82 | | 3,943 | -0.00009 | 0.00727 | 0.34 |
| Entrenamiento con Stripe 82 | 15,771 | | | | |
| Stripe 82 removido | | 478,274 | 0.00194 | 0.01341 | 1.15 |
| Usando solo Stripe 82 | | 3,942 | -0.00002 | 0.00795 | 0.38 |
| Entrenamiento sin Stripe 82 | 486,560 | | | | |
| Stripe 82 removido | | 97,607 | 0.00000 | 0.00914 | 0.33 |
| Usando solo Stripe 82 | | 19,714 | -0.00077 | 0.00077 | 0.41 |

Hayat et al. [8]

Luego de unos años, en [8] propusieron un modelo basado en un aprendizaje auto-supervisado, en el cual se aprenden representaciones (*embeddings*) de estas galaxias usando un *encoder* basado en la arquitectura de la red ResNet-50 [35] y a una función de costos de contraste llamada *NT-Xent* (*normalized temperature-scaled cross entropy loss* por sus siglas en inglés). A esta configuración se le conoce como *SimCLR* (*Simple Framework for Contrastive Learning of Visual Representations*) [36].

A modo general, el modelo trata de representar las imágenes en un espacio \mathbb{R}^n (con n el tamaño de los *embeddings*) de tal forma que imágenes semánticamente similares estén más juntas, alejando aquellas que no. Luego, se usaron estas representaciones para diversas tareas (también llamadas *downstream tasks*), una de ellas fue la predicción del *redshift*. Para ello, propusieron dos métodos distintos; el primero consistió en realizar *fine-tuning* al modelo pre-entrenado mientras que el segundo fue entrenar el modelo original sin utilizar los parámetros entrenados. En la Figura 2.7 se muestran los resultados de ambos métodos.

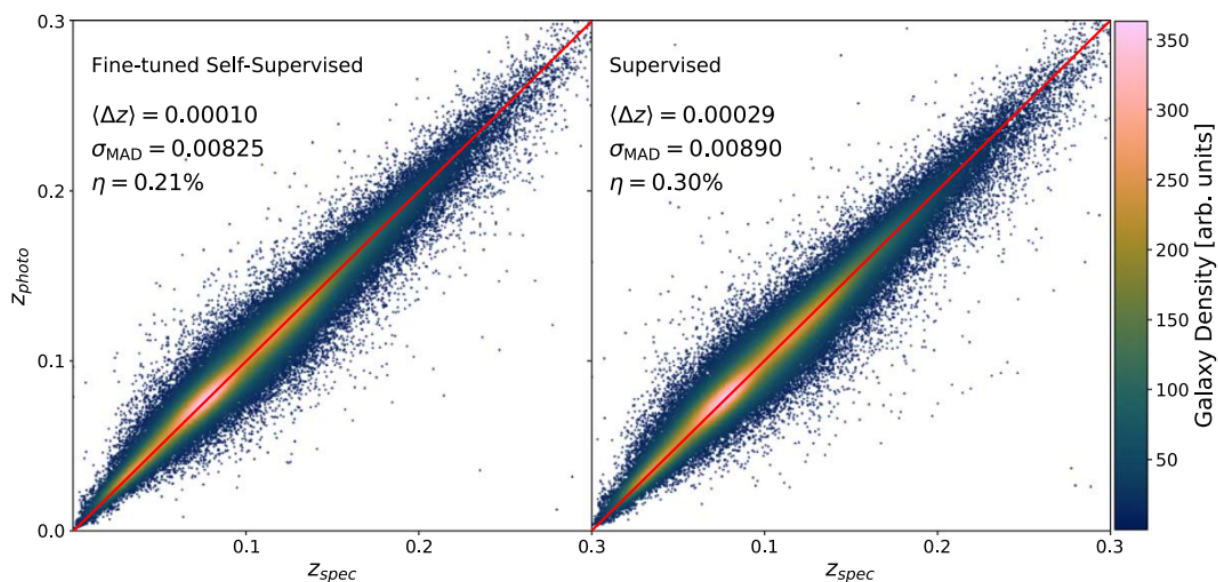


Figura 2.7: Resultados de realizar *fine-tuning* al modelo pre-entrenado (figura de la izquierda) comparado con entrenar el modelo desde cero (figura de la derecha). Ilustración extraída de [8].

Se puede observar que al no realizar *fine-tuning* se obtienen resultados similares que los de [1]. Sin embargo, al realizar *fine-tuning* se obtienen resultados significativamente mejores, infiriendo la gran capacidad de extracción de información relevante que tuvo el modelo en su primera fase de entrenamiento auto-supervisado.

Dey et al. [6]

En 2022, en [6] propusieron un modelo basado en cápsulas de redes convolucionales, las cuales se refieren a unidades específicas que tienen la capacidad de representar características o propiedades particulares de las imágenes de galaxias de forma más temprana. Las cápsulas se utilizan para codificar la morfología de las galaxias, y cada cápsula representa la probabilidad de que una imagen pertenezca a una de dos clases morfológicas: espiral o elíptica.

En su modelo utilizan estas cápsulas para lograr crear una arquitectura de *encoder*, donde se usan dos representaciones en su espacio latente. Uno de ellos para utilizarlo en un *decoder* para reconstruir la imagen y de esta forma darle interpretabilidad al modelo, mientras que la otra representación se conecta a capas totalmente conectadas para la predicción del *redshift*, en donde se obtuvieron los resultados de la Figura 2.8.

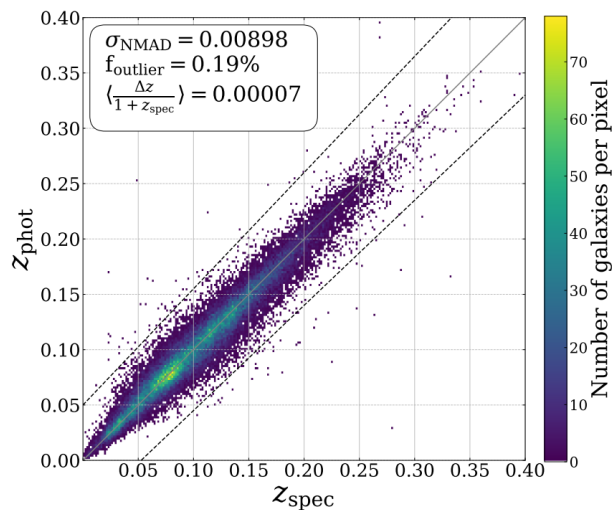


Figura 2.8: Resultados de la predicción del *redshift* en el modelo de Dey et al. sobre el conjunto de test. Ilustración extraída de [6].

Algo importante a destacar de este trabajo es que no se utilizó *cross-validation*. Lo que se hizo fue dividir el conjunto de datos en un 80 % para entrenar, 10 % para validar y 10 % para test, por lo que los resultados no pueden ser directamente comparables con los de Pasquet et al., ya que podrían existir sesgos introducidos por la forma en que se crearon los distintos conjuntos de datos, y no se puede ver el grado de generalización que tiene este modelo frente a datos nuevos.

Treyer et al. [9]

Posteriormente, en [9] se presentó una versión actualizada de la red introducida por Pasquet et al., en donde se redujo el tamaño del espacio latente (vector de características final obtenido de los bloques convolucionales) de 22,272 a 96, reduciendo considerablemente el número de parámetros entrenables de la red.

El objetivo principal de este trabajo fue extender la estimación del corrimiento al rojo en galaxias más tenues ($r \leq 20$). Para ello utilizaron otros *surveys* con fuentes espectroscópicas, tales como BOSS [37] y GAMA [38, 39], en donde la cobertura de predicción paso a $z \approx 0.7$.

Ahora bien, para compararse con los resultados de [1] lo que hicieron fue entrenar con los mismos datos y el mismo procedimiento pero con este nuevo modelo, en donde obtuvo un $\langle \Delta z \rangle = 0.0003$, $\sigma_{\text{MAD}} = 0.00808$ y $\eta = 0.17$, las cuales son mejores que todos los trabajos previos⁴.

Otro aspecto importante del modelo, es que en vez de tener una única capa final para la predicción se añade una segunda capa de una única salida, la cual correspondería al valor escalar del *redshift* predicho. Según los autores, se obtienen mejores resultados (no tan significativos) comparados a solo utilizar la capa de 180 neuronas de salida. Cabe resaltar que en este caso se utilizan dos funciones de costos, para la capa de 180 neuronas (clasificación) se utiliza *Cross Entropy Loss* mientras que para la capa de 1 neurona (regresión) se utilizó RMSE (*root mean square error* por sus siglas en inglés).

Ait Ouahmed et al. [10]

Finalmente, en [10] usaron como base el trabajo de Treyer et al. pero con un enfoque multimodal⁵, es decir, realizaron un procesamiento paralelo de varios subconjuntos de bandas de imágenes, cuyas salidas se fusionan luego para un procesamiento adicional a través de bloques convolucionales. De esta forma, tratan de encontrar correlaciones entre distintos pares y combinaciones de bandas que fueran útiles para predecir el *redshift*, lo cual según los autores se pierde al usar los cinco filtros de forma conjunta.

Al usar esta metodología se mejoraron los resultados previos (ver Tabla 2.2), posicionándose como el estado del arte más actual respecto a la predicción de *redshift* fotométrico.

Tabla 2.2: Resumen de métricas de todos los trabajos. Resultados extraídos de [10].

| Trabajo | σ [10^{-3}] | η [%] | $\langle \Delta z \rangle$ [10^{-3}] |
|------------------------------|---------------------------|---------------|---|
| Pasquet et al. 2019 [1] | 9.08 | 0.31 | 0.04 |
| Dey et al. 2022 [6] | 8.98 | 0.19 | 0.07 |
| Hayat et al. 2021 [8] | 8.25 | 0.21 | 0.1 |
| Treyer et al. 2023 [9] | 8.00 | 0.18 | -0.31 |
| Ait-Ouahmed et al. 2023 [10] | 7.85 | 0.16 | 0.31 |

⁴ En este caso el $\langle \Delta z \rangle$ no se considera una métrica tan relevante como el resto, de hecho, se esta considerando que la métrica más importante es la dispersión de los residuos (σ_{MAD}), debido a que es una métrica más generalizadora del error global del modelo.

⁵ Si bien la multimodalidad se conoce como una técnica de utilizar como entrada distintas fuentes de datos como audios, imágenes, texto, etc., en este caso se le muestra como el uso de distintas bandas, lo cual es distinto a lo que se le suele conocer.

Förster et al. [19]

Como se mencionó al inicio de este capítulo, no existen trabajos que utilicen imágenes en multi-resolución para estimar el *redshift* de las galaxias utilizando redes neuronales convolucionales. Sin embargo, en [19] propusieron un modelo basado en una red neuronal convolucional, la cual recibe una imagen en multi-resolución centrada en la posición de una supernova y calcula la distancia de este evento transiente hasta el centro de su galaxia anfitriona, tal como se muestra en la Figura 2.9.

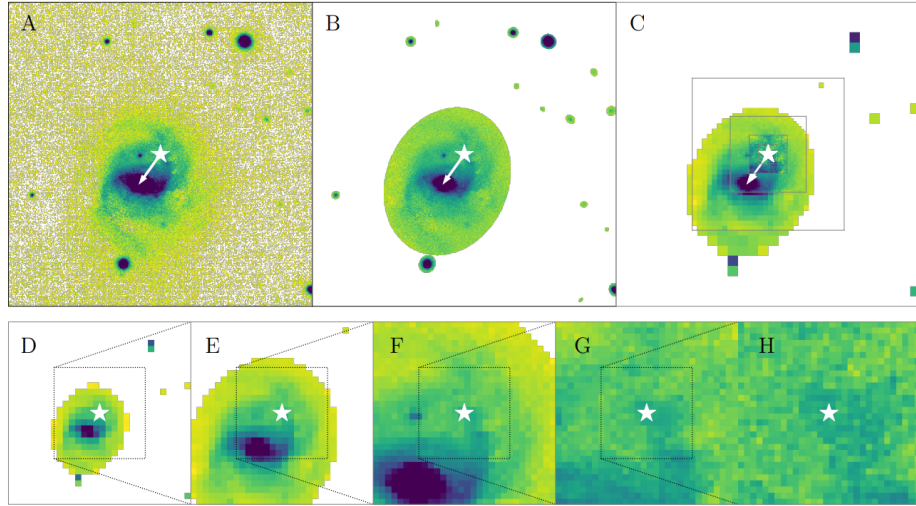


Figura 2.9: Ejemplo de imágenes en multi-resolución, donde A) es la imagen original, B) es la imagen con el fondo extraído y detección de objetos y C) contiene las cinco imágenes respectivas (D-H) en multi-resolución. Ilustración extraída de [19].

La forma en que lo abordaron fue utilizar imágenes del *survey* Pan-STARRS solamente en la banda *r*. Respecto al uso de multi-resolución, utilizaron cinco niveles calculadas a partir de una imagen principal con un FoV = 120 ["], las cuales son de 30×30 pixeles.

La arquitectura del modelo se muestra en la Figura 2.4, en donde se aplicaron rotaciones y *flips* a cada una de las cinco imágenes y se le pasaron a los mismos bloques convolucionales a modo de compartir pesos, esto con el objetivo de lograr una invarianza rotacional. Luego, los vectores de características obtenidos de estos bloques se concatenaron y se pasaron a capas totalmente conectadas, en donde la última de estas capas es de 2 neuronas de salidas, prediciendo así la distancia horizontal y vertical de la supernova hacia su galaxia anfitriona en unidades de la imagen de menor campo de visión (imagen H de la Figura 2.9).

Capítulo 3

Metodología

En este capítulo se describirán los experimentos realizados y por ende, de como se realizaron las predicciones de los *redshifts* de las galaxias. En primer lugar se mencionarán los pasos en común que se realizaron en cada uno, tales como; el entrenamiento de los modelos y la forma de evaluar los resultados, para posteriormente dar en detalle los pasos específicos efectuados en cada uno de ellos.

Como se mencionó en la Introducción, el trabajo de esta memoria se dividió en dos partes principales; (i) entrenar un modelo utilizando imágenes con la resolución y campo de visión utilizado en la literatura, (ii) entrenar el mismo modelo pero que reciba imágenes en multi-resolución en diferentes combinaciones de niveles. El modelo escogido fue el de Pasquet et al. [1], ya que corresponde a un modelo base del que se comparan trabajos actuales. Las predicciones también fueron en base a la esperanza matemática del PDF obtenido (Ec. 2.15) y se usó la misma base de datos de 516,525 galaxias.

Todo el código realizado se hizo en el lenguaje de programación **Python**, en donde los modelos y entrenamientos se llevaron a cabo con el *framework* **Pytorch Lightning** el cual se basa en la librería de **Pytorch** [40]. Como la arquitectura del modelo de Pasquet et al. se encuentra en el *framework* **Tensorflow** se tuvo que realizar el traspaso de la arquitectura hacia **Pytorch**.

Respecto a los experimentos lo que se hizo fue utilizar los mismos hiperparámetros usados en [1], dentro de estos se encuentran; el uso del optimizador Adam [28], un *batch size* de 128, una tasa de aprendizaje de 10^{-4} y la función de costo *Cross Entropy*. Adicionalmente se escogió un máximo de 40 épocas de entrenamiento y el *schedule Cosine Annealing* con un T_{max} igual al máximo de épocas. Durante el entrenamiento también se realizaron transformaciones aleatorias a las imágenes (*flips* horizontales, verticales y rotaciones en 90, 180 y 270 grados)

Para evaluar las predicciones de cada uno se utilizaron las métricas utilizadas en la literatura (Ecuaciones 2.16, 2.17 y 2.18). Para una mayor robustez se utilizó *stratified cross-validation* con $k = 5$, en donde la estratificación se realizó en base al *redshift*. Las métricas finales fueron en base al promedio de los *folds* de test, los cuales tienen un tamaño de 103.306 ejemplos.

Adicionalmente se evaluaron las predicciones en base a la dispersión de los errores (σ_{MAD}) a lo largo de tres características; el *redshift* (z), el radio Petrosiano del 90 % en la banda r (PetroR90) y la magnitud en la banda r (MAG).

Por temas de limitación de tiempo no se utilizaron 5 modelos por cada *fold* como lo hicieron en [1]. Aun así, se utilizó una única semilla (y por ende inicialización de pesos) en conjunto de las prácticas necesarias recomendadas por **Pytorch** para la reproducibilidad del código. Dicho esto, lo único que cambia en cada experimento es el tipo de entrada de cada uno y el tamaño de la primera capa totalmente conectada que sigue de los bloques convolucionales (debido a la diferencia de tamaño en las imágenes de entrada en algunos experimentos).

Como se está utilizando un modelo que ocupa métodos basados en descenso de gradiente, se tuvo que definir la época en la que se evaluarán cada uno de los 5 *fold*s. Esto se puede hacer de muchas maneras, por ejemplo, en [1] utilizaron $\approx 4\%$ del conjunto de entrenamiento en cada *fold* para evaluar y usaron el conjunto de test en la época de menor pérdida en validación. En este caso, se decidió por utilizar el primer *fold* de entrenamiento y realizar otro *stratified cross-validation* con $k = 5$ para definir la mejor época promedio en la que la función de pérdida de test es menor. Una vez definida la mejor época, se entrenó cada uno de los principales 5 *fold*s en donde se evaluó el modelo en la mejor época encontrada anteriormente. En la Figura 3.1 se puede apreciar de mejor manera lo descrito.

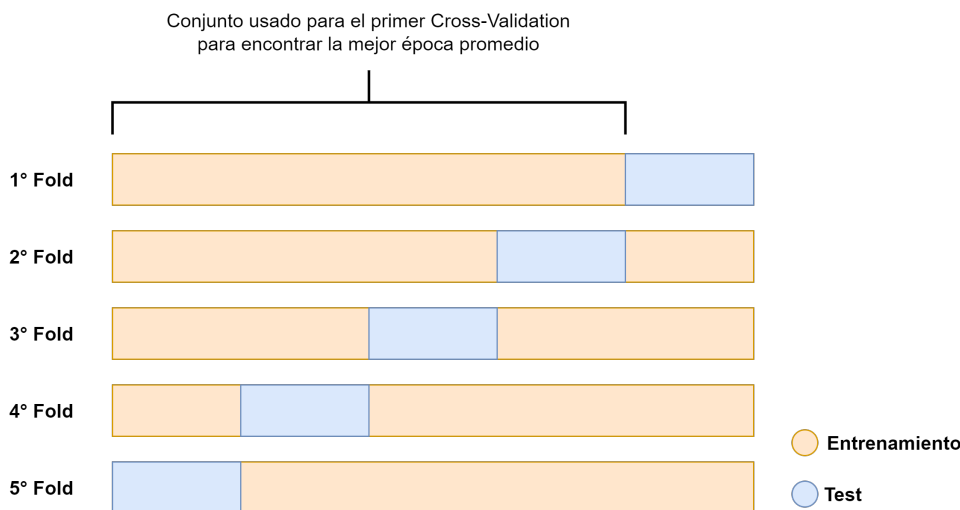


Figura 3.1: Ilustración del procedimiento de realizar dos *stratified cross-validation*. El primero es para encontrar la mejor época promedio en donde se encuentra la menor pérdida en los 5 conjuntos de test, para ello se utiliza el conjunto de entrenamiento del primer *fold* del segundo *stratified cross-validation*, en este último se entregan las métricas promedio en la época encontrada anteriormente.

Para que todos los experimentos sean comparables, se llevaron a cabo en el mismo computador con las mismas versiones de librerías, debido a que **Pytorch** no asegura reproducibilidad ni entre equipos distintos ni entre versiones distintas de su librería. Agregando que la estratificación al usar *cross-validation* se hizo con la misma semilla, por lo que en cada experimento se utilizan las mismas galaxias para entrenar y por ende también en el test.

Una vez definidos estos aspectos generales, se describirá en detalle cada experimento realizado en cada una de las dos partes principales.

3.1. Predicciones usando imágenes de resolución simple

En primer lugar se buscaba replicar los resultados de Pasquet et al. [1] con las mismas galaxias utilizadas por ellos. Sin embargo, se quiso probar utilizando imágenes del *survey* Pan-STARRS debido a que tiene una cobertura superior al *survey* utilizado por [1] (SDSS), por lo que se podría utilizar para realizar predicciones en galaxias que no están presentes en el SDSS.

Por otro lado, como en esta memoria se buscaba realizar una comparación con [1] se utilizarán imágenes del SDSS provenientes del servicio HiPS2FITS [41], por lo que se verificó si existen diferencias significativas en los resultados comparado con usar las imágenes procesadas en [1]. Esto se hizo principalmente para ver si es viable utilizar imágenes en multi-resolución provenientes de este mismo servicio.

A continuación, se mostrarán las metodologías realizadas en cada uno.

3.1.1. Usando imágenes de Pan-STARRS

Para adquirir las imágenes de Pan-STARRS se utilizó el servicio de **panstamps** el cual es un servidor que contiene la fotometría del sondeo, y mediante una API se puede solicitar una imagen FITS⁶ tipo *cutout* dada una coordenada celeste y campo de visión específico. En este caso se utilizaron imágenes tipo *stack* debido a su mayor profundidad comparado con las imágenes tipo *warp*, donde se usaron las coordenadas RA y DEC de las galaxias utilizadas en [1]⁷.

Para hacerlo comparable con el trabajo de [1] se utilizó un campo de visión lo más similar posible, el cual fue de 24 ["]. Sin embargo, como este *survey* tiene una mayor resolución que el SDSS, se usó una ventana cuadrada de tamaño 2×2 de la mediana, así se reduce el tamaño de las imágenes de 98×98 pixeles a 46×46 . Esto se hizo principalmente por limitaciones computacionales a la hora de entrenar.

Por último, como las imágenes se encuentran en unidades de flujo, se debía realizar una compresión de los valores de los pixeles de las imágenes, a esto se le conoce en la literatura como compresión del rango dinámico de los pixeles. Cabe recalcar que no se debe estandarizar o normalizar los valores de forma independiente en cada filtro como usualmente se hace con otro tipo de imágenes, debido a que se perdería toda relación entre las bandas, y por ende los colores y distancias, dando por efecto que el modelo solo aprenda a relacionar el *redshift* por el tamaño y morfología de las galaxias, lo cual es incorrecto.

Dicho esto, se probaron 2 compresiones, la primera fue utilizar el arcoseno hiperbólico y la otra compresión es propuesta por [42], en la cual se divide el valor de los pixeles en 1000 y luego se aplica la función dada por la siguiente ecuación:

⁶ *Flexible Image Transport System*.

⁷ El catálogo con los metadatos de cada ejemplo utilizado por [1] puede ser descargado de <https://d-scholarship.pitt.edu/42023/>.

$$x' = \text{sign}(x) \cdot \left(\sqrt{\text{sign}(x) \cdot x + 1} - 1 \right), \quad (3.1)$$

donde x' es el pixel escalado, x el pixel ya dividido por 1000 y $\text{sign}()$ la función signo.

Como el tamaño de las imágenes es menor comparado a lo que usaron en [1], el vector de características extraído de los bloques convolucionales es de menor tamaño (12,528) por lo que se tuvo que modificar la entrada de la primera capa totalmente conectada y que fuera de tamaño 12,529 (considerando el E(B-V)).

Finalmente se realizó el procedimiento descrito al principio de este capítulo para entrenar el modelo con estas imágenes.

3.1.2. Usando imágenes del SDSS con HiPS2FITS

El servicio de HiPS2FITS contiene datos de imágenes de distintos *surveys*, uno de ellos es el DR9 del SDSS. La ventaja de este servicio es que las imágenes de cada filtro ya vienen en una grilla común, a diferencia de las imágenes en el sitio del SDSS, las cuales necesitan este procesamiento. Adicionalmente, se puede utilizar una API en la cual mediante una coordenada celeste, un campo de visión e incluso una resolución, se puede descargar una imagen FITS con dichas características.

Dicho esto, para descargar cada imagen se utilizaron las coordenadas RA y DEC de cada galaxia, además de utilizar el mismo campo de visión y resolución utilizadas en [1], los cuales son 25,344 ["] y 0,396 ["/pix] respectivamente. Como a las imágenes del SDSS utilizadas por [1] no se les realizó un procesamiento adicional, tampoco se les realizó un procesamiento adicional a las descargadas en HiPS2FITS.

Para poder comparar estos resultados con los de [1] se tuvo que replicar esos resultados con los mismos *folds* y en el mismo computador para que no existan sesgos adicionales. Para esto se descargaron las imágenes utilizadas por ellos las cuales están compartidas de forma pública⁸. Como son las mismas imágenes en donde solo se diferencia con las de HiPS2FITS por el procesamiento que realizaron, la arquitectura se mantiene igual.

Finalmente se realizó el procedimiento descrito al principio de este capítulo para entrenar ambos modelos con sus imágenes respectivas.

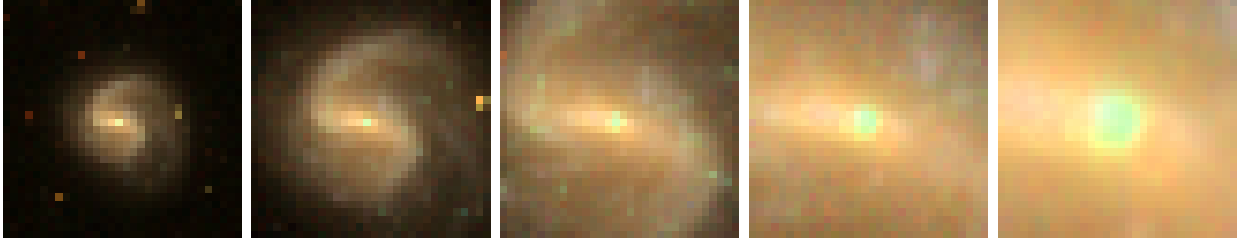
3.2. Predicciones usando imágenes en multi-resolución

Para entrenar el modelo de [1] usando imágenes en multi-resolución, primero se definió el campo de visión máximo de la imagen original y la cantidad de niveles que se utilizarán. En este caso se quería conservar el hecho de que las imágenes puedan ser divisibles por 2 al menos 3 veces (debido a la cantidad de capas de *pooling* en el modelo), agregando de que fueran de menor dimensión que las imágenes utilizadas originalmente. Para ello se definió el campo de visión máximo de 202.752 ["] y 5 niveles, por lo que cada imagen será de 32×32 pixeles con resoluciones de 6.336, 3.168, 1.584, 0.792 y 0.396 ["/pix] y campos de visión de

⁸ Las imágenes pueden ser descargadas de <https://deepdip.iap.fr/#item/60ef1e05be2b8ebb048d951d>.

202.752, 101.376, 50.688, 25.344 y 12.672 ["] respectivamente en cada nivel.

Para una mejor comparación, en la Figura 3.2 se puede apreciar visualmente los cinco niveles, en donde se aprovechó el hecho de que en HiPS2FITS se puede utilizar cualquier resolución, por lo que no es necesario utilizar una ventana de *pooling* de mediana para ir reduciendo la resolución.



(a) Nivel 1, (b) Nivel 2, (c) Nivel 3, (d) Nivel 4, (e) Nivel 5,
FoV = 202.752 ["], FoV = 101.376 ["], FoV = 50.688 ["], FoV = 25.344 ["], FoV = 12.672 ["],
Res. = 6.336 ["/pix]. Res. = 3.168 ["/pix]. Res. = 1.584 ["/pix]. Res. = 0.792 ["/pix]. Res. = 0.396 ["/pix].

Figura 3.2: Ejemplo de una imagen en multi-resolución en cinco niveles, en donde se usó un campo de visión original de 202.752 ["].

En la Figura 3.3 se puede ver el mismo ejemplo anterior, pero destacando el campo de visión utilizado por [1], el cual en este caso es igual al del nivel 4.

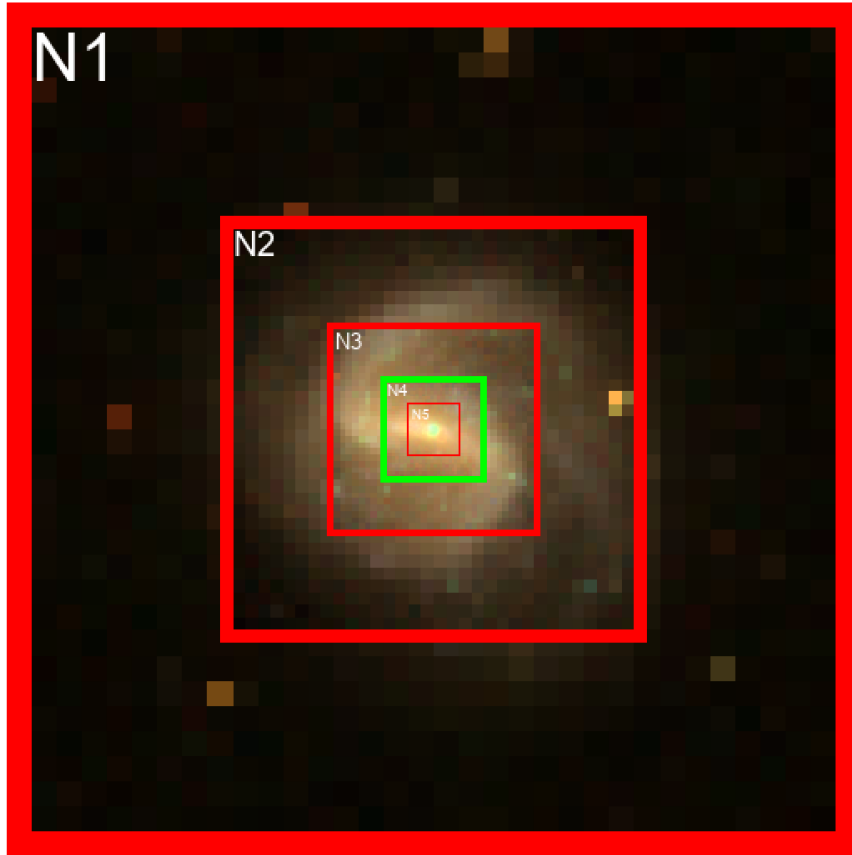


Figura 3.3: Ejemplo de una imagen en multi-resolución en cinco niveles, en color verde se puede ver el campo de visión utilizado por [1], el cual equivale al nivel 4.

Como en Förster et al. [19] la mejor solución fue compartir los pesos de los bloques convolucionales, se quiso ver si en este caso también lo era, ya que son problemas distintos y podría ser que en este caso la mejor solución sea no compartir estos pesos. Las metodologías de ambos casos y sus experimentos respectivos se detallan a continuación.

3.2.1. Compartiendo pesos de los bloques convolucionales

Para descargar las imágenes se usó nuevamente el servicio HiPS2FITS con los campos de visión y resoluciones mencionados anteriormente para cada ejemplo. Tampoco se realizó un procesamiento previo.

Ahora bien, se tuvo que modificar la arquitectura del modelo original el cual recibía imágenes de tamaño 64×64 píxeles en 5 canales (filtros *ugriz*). En este caso ahora las imágenes son de tamaño 32×32 píxeles en los mismos 5 canales, por lo que el vector de características obtenido de los bloques convolucionales paso de tener un largo de 22,272 a 5,568 características, pero como ahora son 5 niveles se utilizó la misma concatenación efectuada en [19] (ver Figura 2.4). Esta concatenación da como resultado un vector de 27,840 características, por lo que la entrada a la primera capa totalmente conectada sería de largo 27,841 (ya que se le suma el E(B-V)).

Como se quiere compartir pesos, durante el entrenamiento cada nivel se pasa por los mismos bloques convolucionales, por lo que no hay otra modificación adicional en la arquitectura. Sin embargo, para ver la utilidad de cada nivel se realizaron experimentos en los cuales se fue eliminando sucesivamente desde el primer nivel en adelante, es decir, primero se utilizaron los 5 niveles, luego los últimos 4, luego los últimos 3, y así hasta llegar a usar solo el quinto nivel. En cada caso se fue cambiando el tamaño de la capa totalmente conectada siguiendo la misma lógica anterior.

En cada combinación de niveles se realizó el procedimiento descrito al principio de este capítulo.

3.2.2. Sin compartir pesos de los bloques convolucionales

Para este caso se utilizaron en su totalidad los cinco niveles del experimento anterior, esto es únicamente por limitaciones de tiempo. Como las imágenes son las mismas, la descarga de las imágenes es exactamente igual al caso anterior, y tampoco se hizo un procesamiento posterior en éstas.

Ahora bien, lo único distinto en este experimento es la configuración de la red. Como no se quiere compartir pesos se replicaron 5 veces los bloques convolucionales, uno para cada nivel respectivamente, por lo que el vector de característica extraído de estos bloques es del mismo tamaño que el caso anterior el cual es de 5,568 características por cada nivel. Se realizó la misma concatenación de estos 5 niveles, dando un resultado de 27,841 características (considerando el E(B-V)) y por ende la modificación en la entrada de la primera capa totalmente conectada.

Al igual que todos los experimentos, se realizó el procedimiento descrito al principio de este capítulo.

Capítulo 4

Resultados

En este capítulo se presentan los resultados de los cuatro principales experimentos mencionados. Estos se mostrarán en base a las dos partes principales; (i) uso de imágenes de resolución simple y (ii) uso de imágenes en multi-resolución.

Tal como se mencionó en la metodología, las métricas utilizadas para evaluar los modelos fueron promediadas en base a los cinco *folds* de test.

4.1. Experimentos usando imágenes de resolución simple

4.1.1. Usando imágenes de Pan-STARRS

En la Figura 4.1 se muestran los resultados de las regresiones usando las dos compresiones de pixeles, donde Compresión 1 hace alusión al arcoseno hiperbólico y Compresión 2 hace alusión a la Ecuación 3.1. Adicionalmente se muestra la Figura 4.2 en donde se puede ver en detalle el σ_{MAD} en atributos de interés, tales como; el *redshift* (z), el radio Petrosiano del 90% en la banda r (PetroR90) y la magnitud en la banda r (MAG).

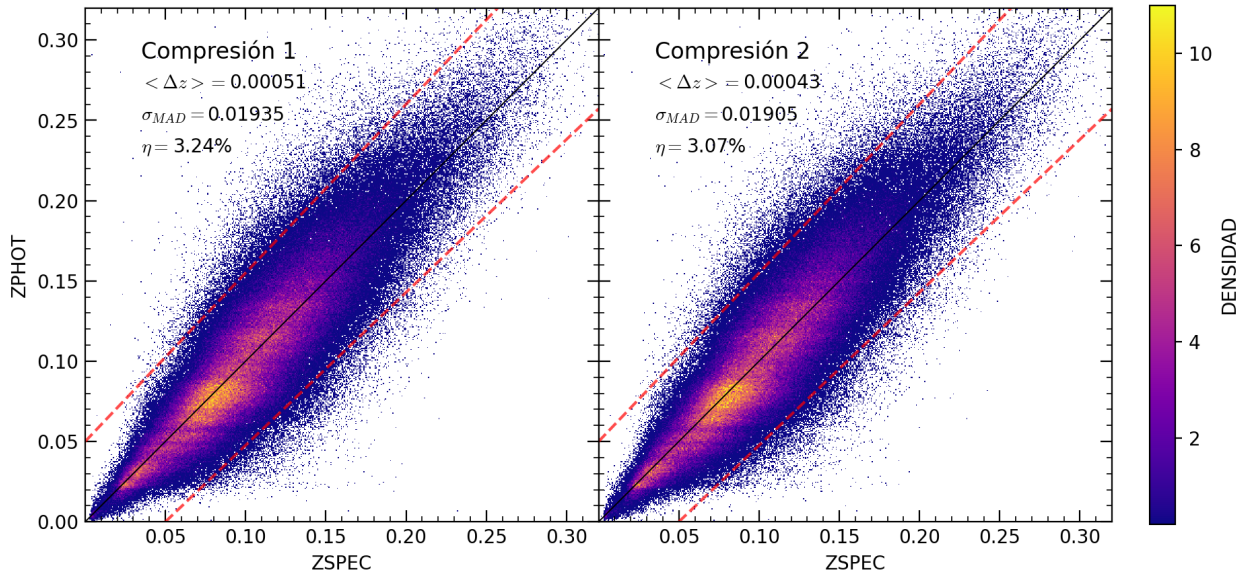


Figura 4.1: Resultados de la regresión con imágenes de Pan-STARRS. Los resultados de los 5 *folds* de test fueron promediados, tanto en la regresión como en las métricas presentes. En la ilustración de la izquierda se usó la compresión con la función arcoseno hiperbólico, mientras que la de la derecha corresponde a la compresión realizada con la Ec. 3.1. Las líneas rojas punteadas corresponden a los límites que definen los errores catastróficos.

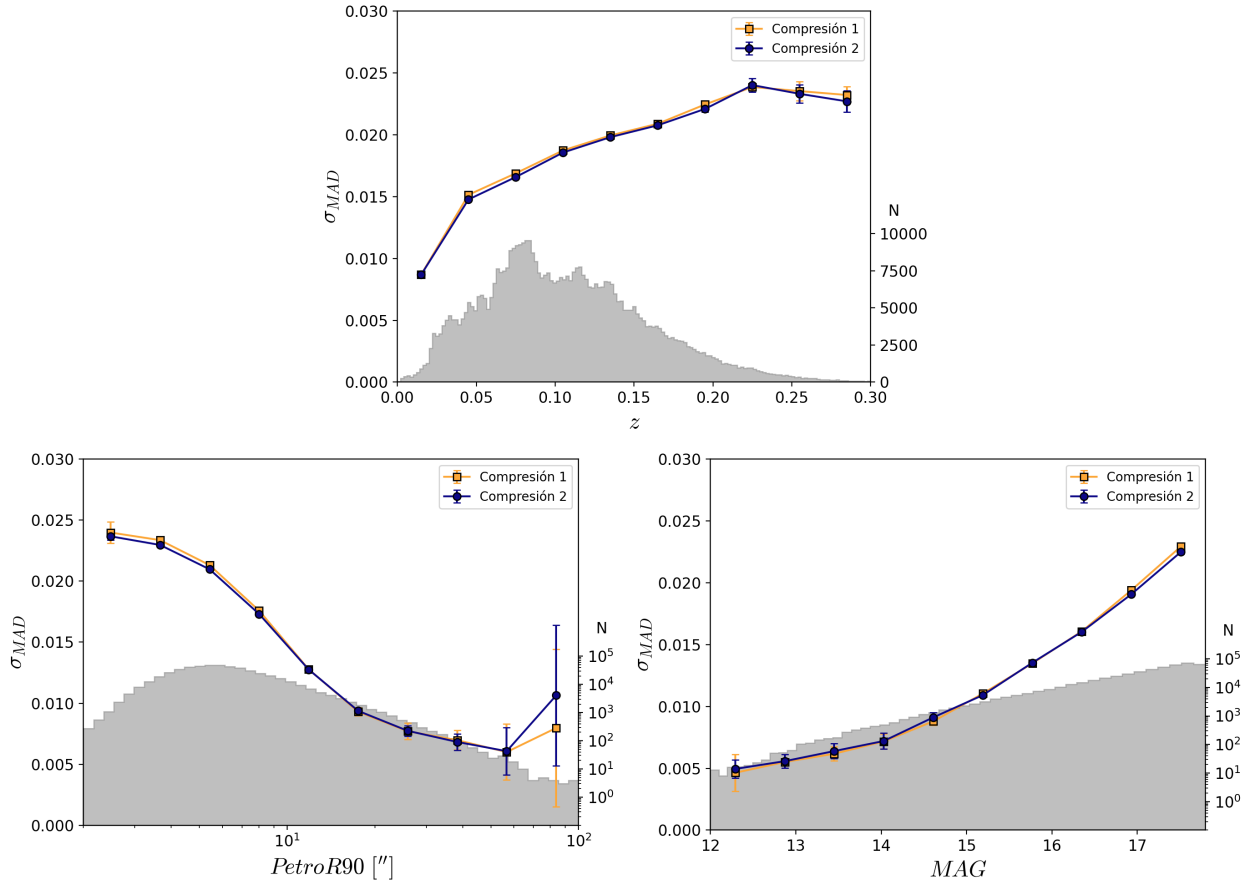


Figura 4.2: Promedio de la dispersión de los errores (σ_{MAD}) en los 5 *fold*s de test y su respectiva desviación estándar en distintas características, usando las imágenes de Pan-STARRS con distintas compresiones. La ilustración superior es a lo largo del *redshift*, la inferior izquierda a lo largo del radio de la galaxia, mientras que la inferior derecha es a lo largo de su magnitud en la banda *r*. La Compresión 1 hace referencia al arcoseno hiperbólico, mientras que la Compresión 2 hace referencia a la Ec. 3.1. Los histogramas corresponden a la distribución de cada característica en todo del conjunto de datos.

4.1.2. Usando imágenes del SDSS con HiPS2FITS

En la Figura 4.3 se muestran los resultados de las regresiones usando las imágenes de [1] con las obtenidas en HiPS2FITS provenientes del mismo *survey*. Por otro lado, en la Figura 4.4 se puede ver en detalle el σ_{MAD} en atributos de interés, tales como; el *redshift* (z), el radio Petrosiano del 90 % en la banda *r* (PetroR90) y la magnitud en la banda *r* (MAG).

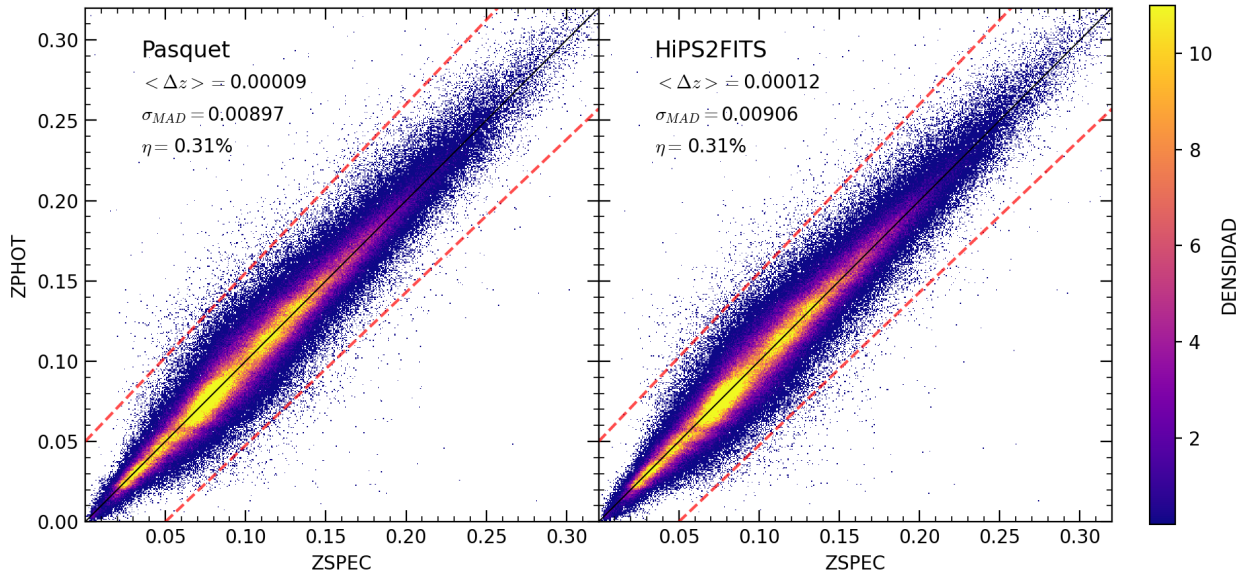


Figura 4.3: Resultados de la regresión con imágenes del SDSS. Los resultados de los 5 *folds* de test fueron promediados, tanto en la regresión como en las métricas presentes. En la ilustración de la izquierda se usaron las imágenes de Pasquet et al. [1], mientras que en la de la derecha se usaron las de HiPS2FITS. Las líneas rojas punteadas corresponden a los límites que definen los errores catastróficos.

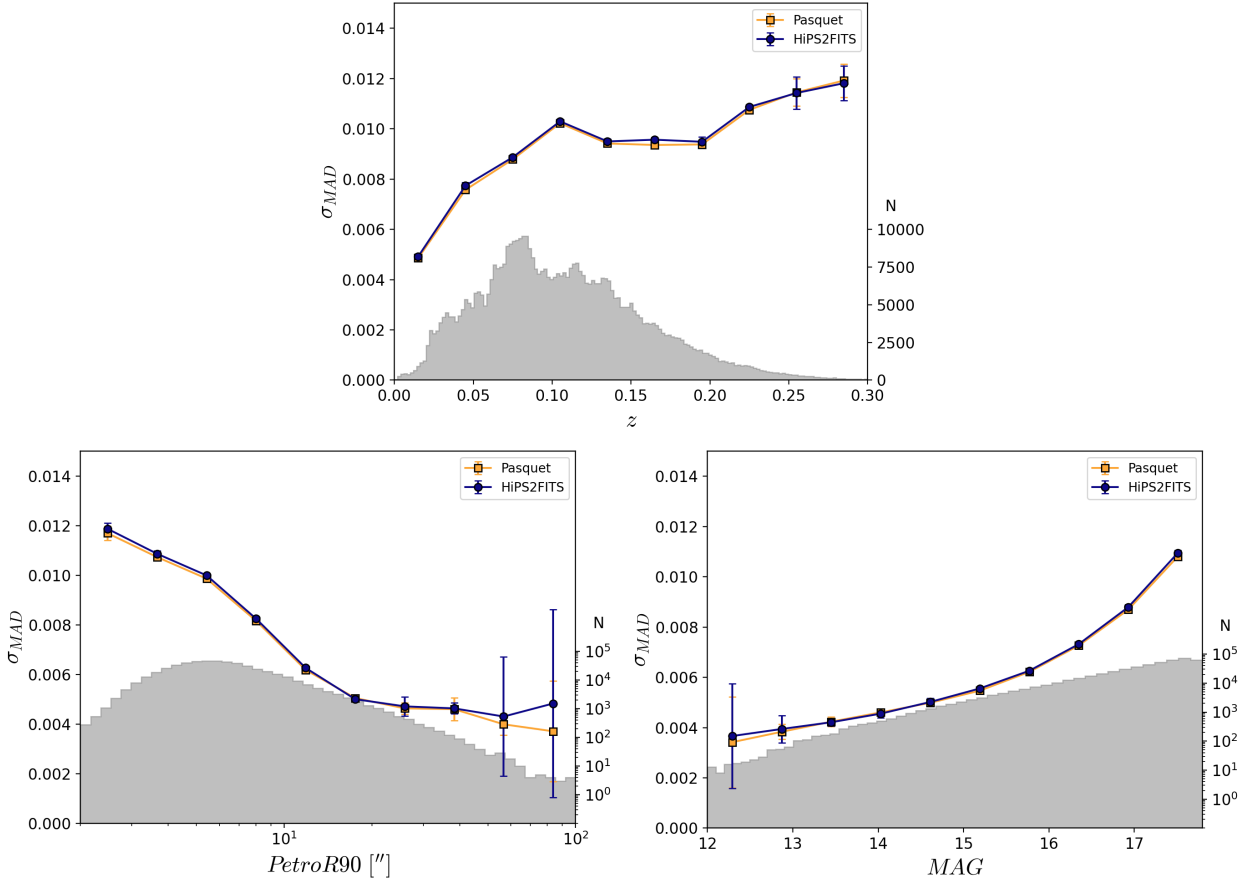


Figura 4.4: Promedio de la dispersión de los errores (σ_{MAD}) en los 5 *folds* de test y su respectiva desviación estándar en distintas características usando las imágenes del SDSS procesadas en Pasquet et al. [1] y las de HiPS2FITS. La ilustración superior es a lo largo del *redshift*, la inferior izquierda a lo largo del radio de la galaxia, mientras que la inferior derecha es a lo largo de su magnitud en la banda *r*. Los histogramas corresponden a la distribución de cada característica en todo del conjunto de datos.

4.2. Experimentos usando imágenes en multi-resolución

4.2.1. Compartiendo pesos de los bloques convolucionales

Para poder evaluar el impacto que tiene el uso de las imágenes en distintos niveles de resolución y campos de visión, se realizaron diversos experimentos en los cuales se iba eliminando sucesivamente desde el primer nivel, es decir, se entrenó en primer lugar utilizando los 5 niveles, luego los últimos 4, luego los últimos 3, y así hasta llegar a usar solo el quinto nivel (el de menor campo de visión). Adicionalmente se incorporaron las métricas de utilizar las imágenes de resolución simple del SDSS también extraídas de HiPS2FITS.

En la Tabla 4.1 se presentan las métricas promedio de los 5 *folds* de test para cada experimento agregando el experimento de usar imágenes de resolución simple (se le adjuntó con el nombre “Original”), adicionalmente se muestra el σ_{MAD} dados distintos radios Petrosianos de las galaxias (R_P). Por otro lado, en la Figura 4.5 se muestra el σ_{MAD} en atributos de

interés, tales como; el *redshift* (z), el radio Petrosiano del 90% en la banda r (PetroR90) y la magnitud en la banda r (MAG).

Tabla 4.1: Resumen de métricas promedio de los 5 *folds* de test dadas distintas combinaciones de niveles utilizados de imágenes en multi-resolución, añadiendo los resultados de usar las imágenes de HiPS2FITS en resolución simple.

| Niveles utilizados | $\langle \Delta z \rangle [10^{-4}]$ | $\eta [\%]$ | $\sigma_{\text{MAD}} [10^{-3}]$ | $\sigma_{\text{MAD}}(R_p > 30) [10^{-3}]$ | $\sigma_{\text{MAD}}(R_p > 50) [10^{-3}]$ | $\sigma_{\text{MAD}}(R_p > 70) [10^{-3}]$ |
|--------------------|--------------------------------------|--------------|---------------------------------|---|---|---|
| Original | 1.215 | 0.307 | 9.057 | 4.6 | 4.553 | 7.454 |
| 1-2-3-4-5 | 1.282 | 0.343 | 9.141 | 4.206 | 4.262 | 3.694 |
| 2-3-4-5 | 1.258 | 0.347 | 9.174 | 4.325 | 4.241 | 3.043 |
| 3-4-5 | 1.338 | 0.342 | 9.154 | 4.468 | 3.884 | 3.866 |
| 4-5 | 1.404 | 0.354 | 9.240 | 4.887 | 5.595 | 3.486 |
| 5 | 1.166 | 0.332 | 9.295 | 5.474 | 7.3361 | 7.377 |

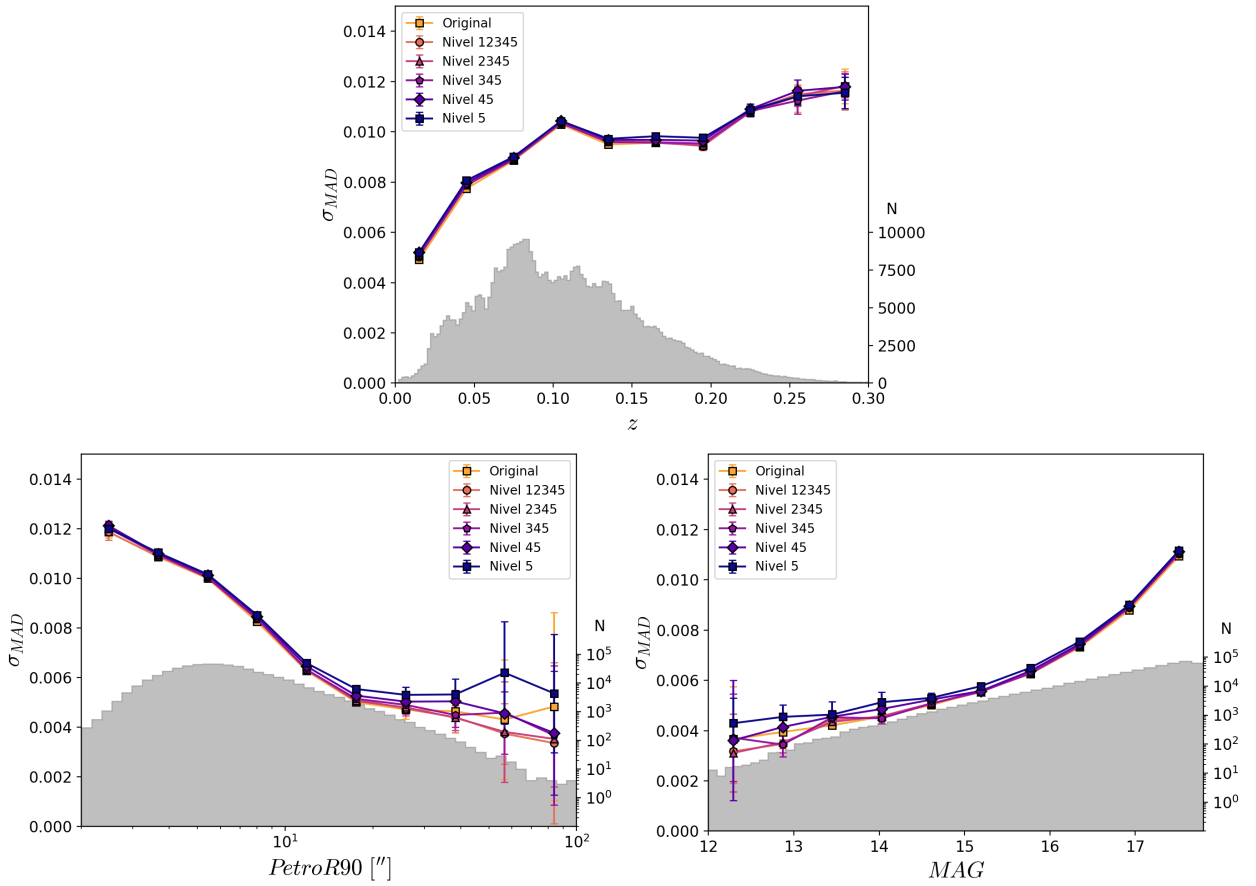


Figura 4.5: Promedio de la dispersión de los errores (σ_{MAD}) en los 5 *folds* de test y su respectiva desviación estándar en distintas características usando las imágenes del SDSS provenientes de HiPS2FITS en distintos niveles de multi-resolución. La ilustración superior es a lo largo del *redshift*, la inferior izquierda a lo largo del radio de la galaxia, mientras que la inferior derecha es a lo largo de su magnitud en la banda r . Los histogramas corresponden a la distribución de cada característica en todo del conjunto de datos.

4.2.2. Sin compartir pesos de los bloques convolucionales

Como se mencionó, este problema es distinto al que intentan solucionar en [19], por lo que no necesariamente compartir pesos sea también la mejor solución en este caso. Dicho esto, se ajustó la arquitectura del modelo añadiendo 4 bloques convolucionales idénticos, por lo que ahora cada una de las 5 sub-imágenes pasa por un bloque convolucional independiente. Este experimento solo se realizó utilizando los 5 niveles debido a limitaciones de tiempo.

En la Figura 4.6 se presenta la comparación de regresión entre compartir o no los pesos de los bloques convolucionales, mientras que en la Figura 4.7 se muestra el σ_{MAD} de ambos métodos sobre atributos de interés, tales como; el *redshift* (z), el radio Petrosiano del 90% en la banda r (PetroR90) y la magnitud en la banda r (MAG). El termino “With SW” hace referencia a *With Sharing Weights* o “Compartiendo pesos” en español, mientras que el termino “W/O SW” hace referencia a *Without Sharing Weights* o “Sin compartir pesos” en español.

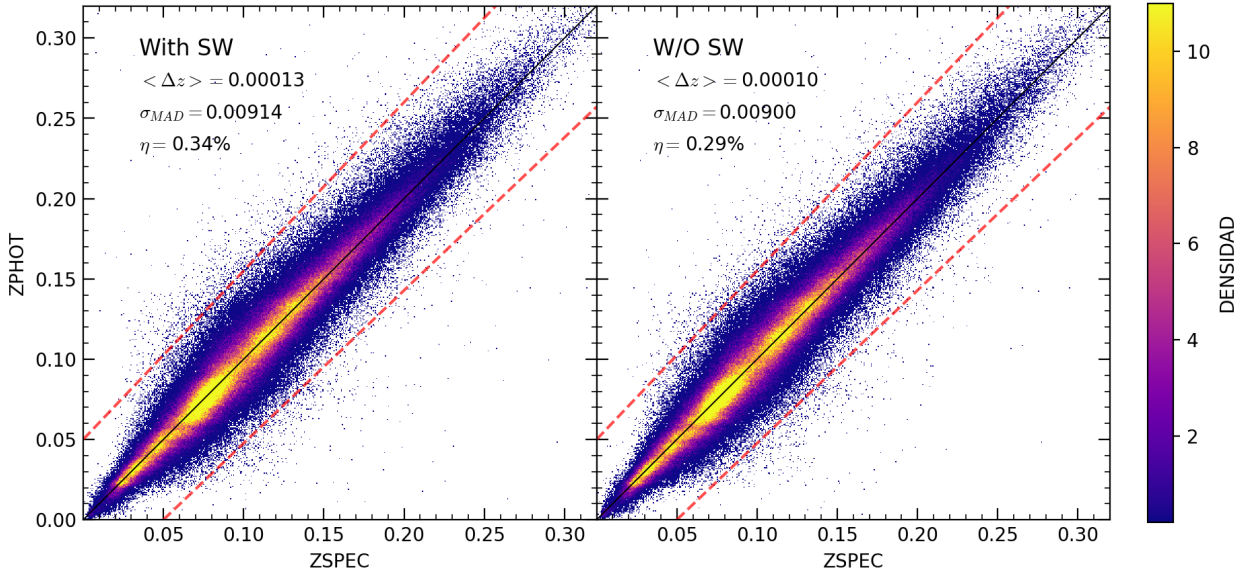


Figura 4.6: Resultados de la regresión con 5 imágenes en multi-resolución del SDSS provenientes de HiPS2FITS. Los resultados de los 5 *folds* de test fueron promediados, tanto en la regresión como en las métricas presentes. En la ilustración de la izquierda se compartieron los pesos de los bloques convolucionales, mientras que la de la derecha se utilizaron bloques convolucionales independientes para cada sub-imagen. Las líneas rojas punteadas corresponden a los límites que definen los errores catastróficos.

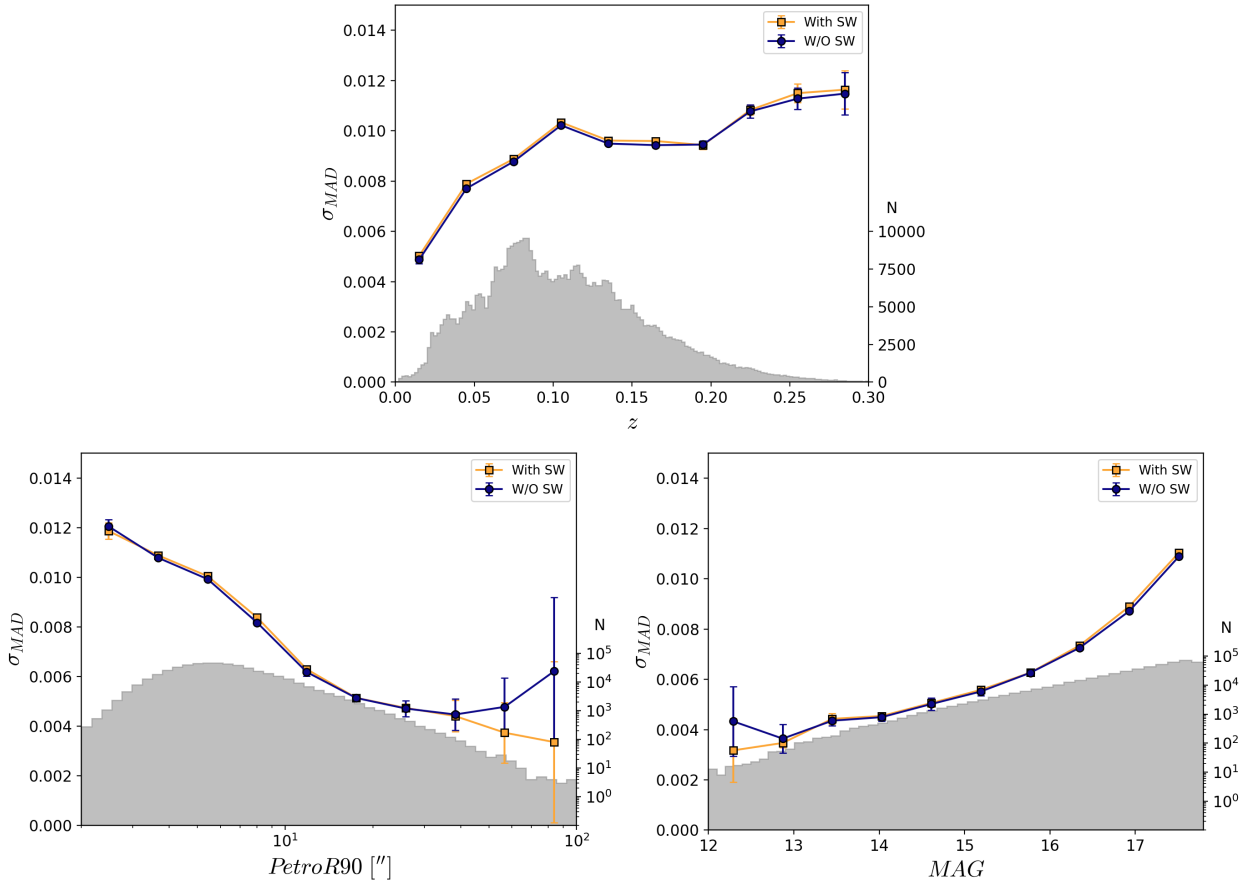


Figura 4.7: Promedio de la dispersión de los errores (σ_{MAD}) en los 5 *folds* de test y su respectiva desviación estándar en distintas características usando imágenes en multi-resolución del SDSS provenientes de HiPS2FITS tanto al compartir pesos como sin compartir pesos. La ilustración superior es a lo largo del *redshift*, la inferior izquierda a lo largo del radio de la galaxia, mientras que la inferior derecha es a lo largo de su magnitud en la banda *r*. Los histogramas corresponden a la distribución de cada característica en todo del conjunto de datos.

Capítulo 5

Análisis

Como se mencionó, en primera instancia se buscaba utilizar imágenes de Pan-STARRS debido a su mayor cobertura con respecto al SDSS, sin embargo, los resultados no fueron similares a los obtenidos en Pasquet et al. [1]. Si bien al usar la compresión de la Ec. 3.1 se obtienen mejores métricas que al usar la compresión del arcoseno hiperbólico, estas siguen siendo significativamente peores que las obtenidas por Pasquet et al. [1], en donde el σ_{MAD} de Pan-STARRS es más del doble que el obtenido con imágenes del SDSS.

Hay varias posibles causas de este resultado, por ejemplo las distintas bandas que poseen estos *surveys*. En experimentos previos se verificó que realmente la banda u del SDSS es muy importante y que al eliminarla la dispersión de los errores aumenta $\approx 6.6\%$ y la tasa de errores catastróficos aumenta casi un 131% (ver Anexo C.1), sin embargo, sigue siendo insignificante con los valores obtenidos con Pan-STARRS, en donde cada métrica empeora un 210% y 1000% respectivamente. Otra posible hipótesis es por errores de calibración en el telescopio PS1, agregando de que en algunas imágenes es posible visualizar artefactos, tales como líneas o manchas. No obstante, se tiene como hipótesis que la principal causa de esto es la calidad de las imágenes de Pan-STARRS respecto a la señal obtenida de las galaxias.

En [43] se realizó el mismo experimento de Beck et al. [32] para estimar el *redshift* de galaxias mediante el uso de colores y magnitudes pero usando fotometría de Pan-STARRS. En dicho trabajo se realiza una comparación de correlación entre el color ($r - i$) de galaxias de los *surveys* SDSS y Pan-STARRS con respecto a sus *redshifts* espectroscópicos, en donde se obtuvo una mayor correlación del primero y una mayor dispersión en el segundo. Adicionalmente, se podría esperar que la SNR de las imágenes del SDSS sea mayor que las de Pan-STARRS, lo cual se relacionaría directamente con los resultados de [1] (Tabla 2.1). Un análisis más detallado de esto se encuentra en el Anexo B

Dada que una de las prioridades era comparar los resultados al usar imágenes en multi-resolución con los trabajos que utilizaban imágenes del SDSS con resolución simple, se optó por utilizar imágenes de este *survey*. El problema de esto es que las imágenes de cada filtro no vienen en la misma grilla de pixeles (alineadas), por lo que se tiene que realizar un procesamiento previo que toma un tiempo considerable. Sin embargo, se dio cuenta del servicio HiPS2FITS [41] en el cual las imágenes ya vienen en una misma grilla, por lo que no se tiene que realizar un procesamiento adicional, pero se debía comparar si existía alguna diferencia con las imágenes procesadas por [1].

De esta comparación se obtuvieron los resultados de las Figuras 4.3 y 4.4, en donde no existen diferencias significativas en las métricas generales ($\langle \Delta z \rangle$, σ_{MAD} y η). Tampoco se observan diferencias en el σ_{MAD} a lo largo del *redshift* y magnitud, pero sí en galaxias con radios grandes. Esto en parte puede deberse a que existen muy pocos datos en el último rango de radios que se usó para calcular la dispersión (≈ 10), por lo que hubieron *folds* en los que se tomaron incluso menos datos para entrenar o testear (debido a que la estratificación se hizo respecto al *redshift*), dando por efecto una gran varianza de la dispersión en ese rango, y por ende también del promedio.

Una vez verificado que no existen diferencias significativas con usar imágenes del SDSS provenientes de HiPS2FITS se pasó a descargar las imágenes en multi-resolución en este mismo servicio. Obteniendo los resultados de la Tabla 4.1 y Fig. 4.5, en donde se compartieron los pesos de los bloques convolucionales y se probaron distintas combinaciones de niveles. Se puede apreciar que ninguna combinación logró superar los resultados de utilizar imágenes con resolución simple (excepto el $\langle \Delta z \rangle$ al usar solo el nivel 5). No obstante, si se observa en detalle el σ_{MAD} a lo largo del tamaño de las galaxias (PetroR90) se tiene que al usar imágenes en multi-resolución se logra reducir el σ_{MAD} en galaxias grandes, lo cual comprueba la hipótesis que se tenía originalmente.

Si bien al usar imágenes en multi-resolución se logra predecir de mejor manera el *redshift* de galaxias grandes, a modo general las predicciones no son mejores (respecto a las métricas obtenidas), esto puede deberse a que en parte se pierde resolución/información de las imágenes en resolución simple con el uso de la multi-resolución, información que al parecer le es útil al modelo para realizar mejores predicciones de forma general.

El último experimento que se realizó fue con el propósito de verificar si compartir pesos es lo mejor o no en este problema, debido a que al compartir pesos en los bloques convolucionales se utilizan los mismos *kernels* en todas las imágenes en multi-resolución, por lo que quizás no son capaces de extraer características mucho más informativas en comparación a que cada sub-imagen utilice sus propios *kernels*.

La comparación se llevo a cabo únicamente en los casos de utilizar los 5 niveles, de esto se obtuvieron los resultados de las Figuras 4.6 y 4.5. Se puede ver que a modo general se mejoran los resultados al no compartir pesos y que no existen diferencias significativas del σ_{MAD} a lo largo del *redshift* ni de la magnitud, pero sí respecto al tamaño, en donde esta vez se obtienen peores predicciones para galaxias grandes. Esto puede deberse a que la red puede priorizar los niveles con mejor resolución (últimos niveles), teniendo por efecto que en los niveles inferiores no se extraiga información importante, y es precisamente en estos niveles donde están contenidas las galaxias grandes.

Aunque es de esperar que al no compartir pesos se logren extraer mejores características en cada nivel que al compartirlos, eso no significa que la red utilizará todas esas características para realizar las predicciones. La CNN utilizada se compone de dos partes principales; una convolucional y la de clasificación, en la parte convolucional es donde se extraen patrones y características que la red considera relevantes para solucionar el problema por el que se le esta entrenando, por otro lado en la parte de la clasificación (parte con las capas *fully connected*) es donde se utilizan estas características para realizar la predicción.

Es por esto que aunque se extraigan mejores características al no compartir pesos, puede que en la parte de clasificación el modelo solo priorice las características extraídas de los últimos niveles, debido a que esas características pueden ser significativamente más importantes que el resto de niveles a modo general. Esto tiene sentido, ya que el conjunto de datos de galaxias esta dominado por galaxias pequeñas, por lo que la red podría centrarse en niveles superiores en los cuales es capaz de extraer una mayor diversidad de patrones comparado a los niveles inferiores, en los cuales estas galaxias están representadas en casi 1 pixel.

Por otro lado, el hecho de compartir pesos se podría ver como un tipo de regularización para que el modelo no priorice niveles particulares, ya que durante el entrenamiento se esta forzando a que con los mismos *kernels* se puedan extraer características igual de importantes en cada nivel. Si bien para galaxias pequeñas es esperable que el modelo no pueda extraer características importantes en todos los niveles (por la misma razón que se comentó en el párrafo anterior), en galaxias grandes el modelo si lo logra hacer, y al tener características relevantes de diferentes niveles puede realizar mejores predicciones de estas. En el Anexo C.2 se pudo verificar esto.

5.1. Análisis general

A modo general, se pudo corroborar que utilizar imágenes en multi-resolución es útil para poder extraer una mayor información de galaxias que no estaban contenidas en el campo de visión utilizado en la literatura, pero esto se da únicamente al compartir pesos. También se pudo verificar que a modo general, se obtienen mejores resultados al no compartir pesos ($\sigma_{\text{MAD}} = 0.009$, $\eta = 0.29\%$) comparado a compartir pesos ($\sigma_{\text{MAD}} = 0.00914$, $\eta = 0.34\%$) e incluso a utilizar las imágenes de la literatura en resolución simple ($\sigma_{\text{MAD}} = 0.00906$, $\eta = 0.31\%$), por otro lado, las predicciones para galaxias de gran tamaño son mejores al compartir pesos, en donde galaxias con un $\text{PetroR90} > 70$ ["] se obtuvo un $\sigma_{\text{MAD}} = 0.0037$ al compartir pesos, $\sigma_{\text{MAD}} = 0.0061$ al no compartir pesos y $\sigma_{\text{MAD}} = 0.0075$ usando resolución simple.

Si bien se esperaba que al usar imágenes con mayor campo de visión se mejoraran las predicciones de galaxias grandes, se obtuvo un resultado similar al utilizar los dos últimos niveles de las imágenes en multi-resolución, en donde la imagen con mayor campo de visión es igual al que se utilizó en la literatura (25.244 ["]). Se logró que las predicciones de galaxias con un $\text{PetroR90} > 70$ ["] fueran mejores al usar los dos últimos niveles ($\sigma_{\text{MAD}} = 0.0035$) que al usar una única imagen con el mismo campo de visión y el doble de resolución ($\sigma_{\text{MAD}} = 0.0075$). Sin embargo, esto no ocurre con galaxias con un $\text{PetroR90} > 30$ ["] y $\text{PetroR90} > 50$ ["], por lo que habría que ver en detalle la razón de esto. Una de ellas podría ser el cálculo del radio de la galaxia, en donde quizás existen galaxias en el cual el método para estimarlo no sea el adecuado, generando sesgos y errores posteriores en los análisis.

Por otro lado, cabe destacar que mejorar las predicciones de *redshifts* de galaxias grandes no impacta de forma global las métricas de las predicciones, ya que estas galaxias corresponden a un subconjunto minoritario, agregando que tienden a tener un valor de magnitud menor (galaxias más brillantes) y por ende un SNR alto. Esto facilita que el modelo realice una buena predicción de *redshifts* de estas galaxias. Por el contrario, las galaxias pequeñas tienden a tener un flujo menor (alta magnitud) debido a que es más difícil de capturar su

señal (entre otras cosas), dando por efecto mayores variaciones en sus mediciones y a su vez un SNR menor. La multi-resolución podría utilizarse para hacer que la red sea equivariante o invariante a escala como lo hacen en [44], realizando un enfoque principalmente en galaxias pequeñas que son las que dominan el conjunto de datos. Si bien al compartir pesos se logra en parte que el modelo pueda extraer características a distintas escalas, no se logra una invarianza como tal.

Siguiendo este hilo, se podría haber obtenido una mejora considerable a modo general en las métricas si se hubiera añadido una cantidad mayor de galaxias de gran tamaño. Sin embargo, estas galaxias estarían fuera de los rangos de estudio ($z < 0.4$ y $r \leq 17.8$), por lo que se podrían introducir otro tipo de problemas, como sesgos en predicciones de objetos con *redshifts* elevados o sesgos en predicciones de galaxias muy tenues.

Otro aspecto relevante, es que se pudo obtener métricas similares a la literatura incluso al utilizar una cantidad menor de datos (ver Tabla 4.1), en donde al utilizar un 75% menos de pixeles que en la literatura (Nivel 5) el σ_{MAD} aumentó en un 2.6% y la tasa de errores catastróficos en un 8% con respecto a utilizar las imágenes de la literatura. Si bien los resultados empeoran, la cantidad necesaria para almacenar las imágenes de estas galaxias disminuye considerablemente al igual que el tiempo de inferencia, lo cual es útil hoy en día que se transmiten millones de datos cada noche y optimizar la forma de utilizar memoria computacional es fundamental. También se podría probar aumentar aún más la cantidad de niveles, por ejemplo a 6, en donde cada nivel sería de 16×16 pixeles (un 93.75% menos de pixeles que en la literatura) y ver el grado en el que difieren las métricas con respecto a la literatura.

En cuanto al experimento como tal, puede ser que no se haya realizado adecuadamente, debido a que existen muchas formas en las que se pueden definir las imágenes en multi-resolución, lo cual depende del campo de visión de la imagen principal y del número de niveles utilizados, en donde quizás se hubieran descubierto otros hallazgos en los resultados al escoger una cantidad mayor de niveles. Por otro lado, como se quería comparar los resultados de [1] con respecto a añadir imágenes en multi-resolución, lo más adecuado hubiese sido que el campo de visión utilizado en [1] hubiese estado en el nivel del medio y no en un extremo. En base a esto último, quizás otra forma de abordar este problema es tratar de utilizar las imágenes de [1], y con respecto a esta calcular imágenes en multi-resolución con el propósito de minimizar la cantidad de información utilizada, pero tratando de que las predicciones sean lo más similar o incluso mejores a las de [1].

Otro aspecto crítico es que en todos los experimentos se utilizaron los mismos hiperparámetros, por lo que no se explotó a fondo el potencial que pueden tener las redes en cada experimento. Esta búsqueda de hiperparámetros no se pudo realizar debido a limitaciones de tiempo, ya que a que realizar todos los experimentos mostrados tomó alrededor de 40 días continuos de entrenamiento y variar solo un hiperparámetro podría duplicar este tiempo.

Capítulo 6

Conclusiones

Se logró analizar de forma introductoria el impacto que tiene el uso de imágenes en multi-resolución para predecir el *redshift* de galaxias, comparado con usar imágenes en resolución simple usando una red neuronal convolucional. Se pudo verificar que al usar esta técnica se logran obtener mejores predicciones en galaxias grandes, lo cual era una de las hipótesis principales de este trabajo.

También se pudo verificar que al usar imágenes del *survey* Pan-STARRS se obtienen predicciones considerablemente peores que al utilizar imágenes del SDSS. Se discutió que las razones principales se debían a artefactos y/o a problemas de calibración presentes en el *survey* Pan-STARRS. Sin embargo, esto podría ser útil para la estimación de *redshift* de galaxias fuera de la cobertura del SDSS, aunque dichas predicciones tendrían una mayor incertidumbre en comparación con las predicciones de los modelos entrenados con imágenes del SDSS.

Adicionalmente, se obtuvo que utilizar imágenes provenientes del servicio HiPS2HITS los resultados no difieren significativamente comparado con el procesamiento llevado a cabo por [1] sobre las mismas imágenes, por lo que puede ser usado para realizar predicciones sobre nuevas galaxias sin la necesidad de realizar un procesamiento adicional en las imágenes, ahorrando tiempo considerable en todo el proceso.

Respecto al uso de imágenes de multi-resolución, se verificó que la dispersión de los residuos normalizados de galaxias grandes disminuye considerablemente en comparación a usar imágenes en resolución simple, lo cual da cuenta que efectivamente al darle a la CNN un mayor campo de visión de galaxias grandes esta puede extraer información que es usada para mejores predicciones. Sin embargo, las predicciones generales no mejoraban considerablemente, debido a que en el conjunto de datos utilizados predominaba la presencia de galaxias pequeñas (de igual o menor tamaño que el campo de visión utilizado en la literatura).

También se pudo observar que la red es, en cierta medida, capaz de extraer información a distintas escalas, lo cual se comprobó con el experimento realizado de no compartir pesos, en el cual las predicciones mejoraban únicamente en galaxias de menor tamaño y empeoraban para galaxias grandes. Sin embargo, no se logra una invarianza a escala.

Este trabajo muestra una introducción de las ventajas de utilizar imágenes en multi-resolución en este problema, en donde los resultados obtenidos dan indicios de que la meto-

dología o los experimentos realizados no fueron los más adecuados. Se piensa que una mejor manera de aprovechar las imágenes en multi-resolución es lograr que la red sea equivariante o invariante a escala, incluso sin la necesidad de que las predicciones se hagan con varios niveles.

De todos los resultados, análisis y conclusiones, se proponen recomendaciones en iteraciones futuras de este trabajo, las cuales se detallan en la siguiente sección.

6.1. Trabajo futuro

Dentro del análisis se pudo dar cuenta de dos aspectos relevantes: el primero fue que los experimentos no fueron configurados para ser los mejores en cuanto a los hiperparámetros, cantidad de niveles y campos de visión utilizados, mientras que el segundo aspecto es que se pueden obtener buenos resultados usando menos cantidad de información de las imágenes. De estos surgen recomendaciones para posibles iteraciones futuras, agregando un tercer trabajo futuro que relaciona este trabajo con DELIGHT [19], lo cual se muestra a continuación.

6.1.1. Mejorar la calidad de los experimentos

Para poder asegurar que cada experimento sea el más óptimo, se podrían realizar diversas correcciones al trabajo realizado. La primera es la búsqueda de los mejores hiperparámetros, lo cual no se hizo porque entrenar todos los modelos tomó un tiempo aproximado de 40 días consecutivos, sin embargo, de esta forma se podría ver realmente la máxima (o lo más cercano a lo mejor) *performance* de cada experimento respecto a las métricas utilizadas. Quizás se podría obtener el mismo o un mejor resultado en cada experimento de multi-resolución con respecto a los resultados de la literatura, lo cual no fue en este caso al usar los mismos hiperparámetros que se utilizaron en [1].

Adicionalmente se podría probar usando 6 niveles en vez de 5, y ver el impacto que tiene añadir este nivel adicional (usando el mismo campo de visión de la imagen original). Por otro lado, se podría cambiar el enfoque del entrenamiento de modo que la red pueda ser invariante a escala y ver efectivamente si los resultados mejoran. Como se usó la misma arquitectura que [1], probablemente los resultados al usar imágenes en multi-resolución están limitados a la arquitectura, la cual no fue creada para lograr esta invarianza.

Estos aspectos se mencionan con el propósito de poder sacar el mejor provecho del uso de la multi-resolución, y así, determinar finalmente su impacto de forma general en esta tarea específica.

6.1.2. Minimizar la cantidad de información utilizada

Por otro lado, como se pudieron obtener resultados similares a la literatura usando un 75 % menos de información comparado con [1], y dado que hoy en día se transmiten cientos de miles de datos por noche, se podría realizar un enfoque en el que se obtengan resultados similares pero minimizando la cantidad de información utilizada en las imágenes.

Esto se puede hacer de muchas maneras, y como todos los problemas son distintos se tendrían que realizar diversos experimentos que puedan resolver el problema con este enfoque. Uno de ellos es el que se comentó anteriormente, el cual es lograr una invarianza a escala, esto

se podría realizar de forma interna en la configuración del modelo como de forma externa en el entrenamiento, como por ejemplo mediante el uso de *data augmentation*, de esta forma se podría entrenar el modelo usando solo un nivel de forma aleatoria, esperando que el modelo pueda dar la misma predicción de *redshift* independiente del nivel dado, así solo se necesitaría 1 sola imagen en multi-banda para realizar una predicción.

Se trabajó con imágenes de tamaño 32×32 píxeles, pero si se añadiera un nivel adicional (manteniendo el mismo campo de visión de la imagen original) se tendrían imágenes de 16×16 píxeles por lo que se podría comprobar si utilizar este nuevo formato es útil, y así, se necesitaría mucha menos información, añadiendo de que quizás no sea necesario utilizar los 6 niveles.

También se podría tratar de replicar los resultados de [1], ya sea aplicando la multi-resolución en sus imágenes, como también usando solo el doble de su campo de visión y no el óctuple como se utilizó en este trabajo.

Otra opción es utilizar un *autoencoder* o una de sus variantes [45], en donde se podría utilizar como entrada imágenes en multi-resolución y como salida esperada la imagen original de resolución simple. Si bien al usar esta arquitectura se necesitaría una mayor cantidad de memoria para almacenar la entrada y salida, a la hora de predecir solo se utilizaría el encoder, donde se podría usar su espacio latente para entrenar otro clasificador, o se le podría hacer *fine-tuning* tal como se hizo en [8]. Esto se podría hacer con diversas combinaciones de niveles, reduciendo aun más la cantidad de información utilizada como entrada al modelo.

6.1.3. Combinación con DELIGHT

Finalmente, se podría integrar este trabajo con DELIGHT [19] de modo de tener un modelo multi-tarea, en el cual, dada una misma entrada, se pueda realizar la predicción de la distancia de una supernova a su galaxia anfitriona y además del *redshift* de la misma galaxia.

Para ello se tendría que tener en cuenta diversos factores, como por ejemplo que DELIGHT usa imágenes centradas en la supernova y no en la galaxia, también que DELIGHT utiliza solo la banda *r* a diferencia de este trabajo que se utilizaron las bandas *ugriz*, por otro lado DELIGHT utiliza imágenes de Pan-STARRS y el trabajo actual utiliza imágenes del SDSS. Adicionalmente también se tendría que tomar en cuenta la cantidad de ejemplos que se tienen en cada modelo, ya que en este caso se utilizaron 516.525 ejemplos distintos, mientras que en DELIGHT solo se utilizaron 16.791 ejemplos, en el cual podrían existir ejemplos que no estén dentro del primer conjunto de datos (debido a la menor cobertura del SDSS).

Bibliografía

- [1] Pasquet, Johanna, Bertin, E., Treyer, M., Arnouts, S., y Fouchez, D., “Photometric redshifts from sdss images using a convolutional neural network”, *AA*, vol. 621, p. A26, 2019, [doi:10.1051/0004-6361/201833617](https://doi.org/10.1051/0004-6361/201833617).
- [2] Newman, J. A. y Gruen, D., “Photometric redshifts for next-generation surveys”, *Annual Review of Astronomy and Astrophysics*, vol. 60, p. 363–414, 2022, [doi:10.1146/annurev-astro-032122-014611](https://doi.org/10.1146/annurev-astro-032122-014611).
- [3] Firth, A. E., Lahav, O., y Somerville, R. S., “Estimating photometric redshifts with artificial neural networks”, *Monthly Notices of the Royal Astronomical Society*, vol. 339, pp. 1195–1202, 2003, [doi:10.1046/j.1365-8711.2003.06271.x](https://doi.org/10.1046/j.1365-8711.2003.06271.x).
- [4] Collister, A. A. y Lahav, O., “ANNz: Estimating Photometric Redshifts Using Artificial Neural Networks”, *MNRAS*, vol. 116, no. 818, pp. 345–351, 2004, [doi:10.1086/383254](https://doi.org/10.1086/383254).
- [5] Hoyle, B., “Measuring photometric redshifts using galaxy images and Deep Neural Networks”, *Astronomy and Computing*, vol. 16, pp. 34–40, 2016, [doi:10.1016/j.ascom.2016.03.006](https://doi.org/10.1016/j.ascom.2016.03.006).
- [6] Dey, B., Andrews, B. H., Newman, J. A., Mao, Y.-Y., Rau, M. M., y Zhou, R., “Photometric redshifts from SDSS images with an interpretable deep capsule network”, *Monthly Notices of the Royal Astronomical Society*, vol. 515, pp. 5285–5305, 2022, [doi:10.1093/mnras/stac2105](https://doi.org/10.1093/mnras/stac2105).
- [7] Timalisina, U., Broll, B., Moore, K., Budavári, T., y Lédeczi, , “Deepforge for astronomy: Deep learning sdss redshifts from images”, *Astronomy and Computing*, vol. 40, p. 100601, 2022, [doi:https://doi.org/10.1016/j.ascom.2022.100601](https://doi.org/10.1016/j.ascom.2022.100601).
- [8] Hayat, M. A., Stein, G., Harrington, P., Lukić, Z., y Mustafa, M., “Self-supervised representation learning for astronomical images”, *The Astrophysical Journal Letters*, vol. 911, p. L33, 2021, [doi:10.3847/2041-8213/abf2c7](https://doi.org/10.3847/2041-8213/abf2c7).
- [9] Treyer, M., Ait Ouahmed, R., Pasquet, J., Arnouts, S., Bertin, E., y Fouchez, D., “CNN photometric redshifts in the SDSS at $r \leq 20$ ”, *Monthly Notices of the Royal Astronomical Society*, vol. 527, pp. 651–671, 2023, [doi:10.1093/mnras/stad3171](https://doi.org/10.1093/mnras/stad3171).
- [10] Ait Ouahmed, R., Arnouts, S., Pasquet, J., Treyer, M., y Bertin, E., “Multimodality for improved cnn photometric redshifts”, *AA*, vol. 683, p. A26, 2024, [doi:10.1051/0004-6361/202347395](https://doi.org/10.1051/0004-6361/202347395).
- [11] Förster, F., Cabrera-Vives, G., Castillo-Navarrete, E., Estévez, P. A., Sánchez-Sáez, P., Arredondo, J., Bauer, F. E., Carrasco-Davis, R., Catelan, M., Elorrieta, F., Eyheramendy, S., Huijse, P., Pignata, G., Reyes, E., Reyes, I., Rodríguez-Mancini, D., Ruz-Mieres, D., Valenzuela, C., Álvarez Maldonado, I., Astorga, N., Borissova, J., Clocchiat-

- ti, A., Cicco, D. D., Donoso-Oliva, C., Hernández-García, L., Graham, M. J., Jordán, A., Kurtev, R., Mahabal, A., Maureira, J. C., Muñoz-Arancibia, A., Molina-Ferreiro, R., Moya, A., Palma, W., Pérez-Carrasco, M., Protopapas, P., Romero, M., Sabatini-Gacitua, L., Sánchez, A., Martín, J. S., Sepúlveda-Cobo, C., Vera, E., y Vergara, J. R., “The automatic learning for the rapid classification of events (alerce) alert broker”, *The Astronomical Journal*, vol. 161, p. 242, 2021, [doi:10.3847/1538-3881/abe9bc](https://doi.org/10.3847/1538-3881/abe9bc).
- [12] York, D. G., Adelman, J., Anderson, John E., J., Anderson, S. F., Annis, J., Bahcall, N. A., Bakken, J. A., Barkhouser, R., Bastian, S., Berman, E., Boroski, W. N., Bracker, S., Briegel, C., Briggs, J. W., Brinkmann, J., Brunner, R., Burles, S., Carey, L., Carr, M. A., Castander, F. J., Chen, B., Colestock, P. L., Connolly, A. J., Crocker, J. H., Csabai, I., Czarapata, P. C., Davis, J. E., Doi, M., Dombeck, T., Eisenstein, D., Ellman, N., Elms, B. R., Evans, M. L., Fan, X., Federwitz, G. R., Fiscelli, L., Friedman, S., Frieman, J. A., Fukugita, M., Gillespie, B., Gunn, J. E., Gurbani, V. K., de Haas, E., Haldeman, M., Harris, F. H., Hayes, J., Heckman, T. M., Hennessy, G. S., Hindsley, R. B., Holm, S., Holmgren, D. J., Huang, C.-h., Hull, C., Husby, D., Ichikawa, S.-I., Ichikawa, T., Ivezić, Ž., Kent, S., Kim, R. S. J., Kinney, E., Klaene, M., Kleinman, A. N., Kleinman, S., Knapp, G. R., Korienek, J., Kron, R. G., Kunszt, P. Z., Lamb, D. Q., Lee, B., Leger, R. F., Limmongkol, S., Lindenmeyer, C., Long, D. C., Loomis, C., Loveday, J., Lucinio, R., Lupton, R. H., MacKinnon, B., Mannery, E. J., Mantsch, P. M., Margon, B., McGehee, P., McKay, T. A., Meiksin, A., Merelli, A., Monet, D. G., Munn, J. A., Narayanan, V. K., Nash, T., Neilsen, E., Neswold, R., Newberg, H. J., Nichol, R. C., Nicinski, T., Nonino, M., Okada, N., Okamura, S., Ostriker, J. P., Owen, R., Pauls, A. G., Peoples, J., Peterson, R. L., Petravick, D., Pier, J. R., Pope, A., Pordes, R., Prosapio, A., Rechenmacher, R., Quinn, T. R., Richards, G. T., Richmond, M. W., Rivetta, C. H., Rockosi, C. M., Ruthmansdorfer, K., Sandford, D., Schlegel, D. J., Schneider, D. P., Sekiguchi, M., Sergey, G., Shimasaku, K., Siegmund, W. A., Smee, S., Smith, J. A., Snedden, S., Stone, R., Stoughton, C., Strauss, M. A., Stubbs, C., SubbaRao, M., Szalay, A. S., Szapudi, I., Szokoly, G. P., Thakar, A. R., Tremonti, C., Tucker, D. L., Uomoto, A., Vanden Berk, D., Vogeley, M. S., Waddell, P., Wang, S.-i., Watanabe, M., Weinberg, D. H., Yanny, B., Yasuda, N., y SDSS Collaboration, “The Sloan Digital Sky Survey: Technical Summary”, , vol. 120, pp. 1579–1587, 2000, [doi:10.1086/301513](https://doi.org/10.1086/301513).
- [13] Margony, B., “The sloan digital sky survey”, *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 357, p. 93–103, 1999, [doi:10.1098/rsta.1999.0316](https://doi.org/10.1098/rsta.1999.0316).
- [14] Gunn, J. E., Siegmund, W. A., Mannery, E. J., Owen, R. E., Hull, C. L., Leger, R. F., Carey, L. N., Knapp, G. R., York, D. G., Boroski, W. N., Kent, S. M., Lupton, R. H., Rockosi, C. M., Evans, M. L., Waddell, P., Anderson, J. E., Annis, J., Barentine, J. C., Bartoszek, L. M., Bastian, S., Bracker, S. B., Brewington, H. J., Briegel, C. I., Brinkmann, J., Brown, Y. J., Carr, M. A., Czarapata, P. C., Drennan, C. C., Dombeck, T., Federwitz, G. R., Gillespie, B. A., Gonzales, C., Hansen, S. U., Harvanek, M., Hayes, J., Jordan, W., Kinney, E., Klaene, M., Kleinman, S. J., Kron, R. G., Kresinski, J., Lee, G., Limmongkol, S., Lindenmeyer, C. W., Long, D. C., Loomis, C. L., McGehee, P. M., Mantsch, P. M., Eric H. Neilsen, J., Neswold, R. M., Newman, P. R., Nitta, A., John Peoples, J., Pier, J. R., Prieto, P. S., Prosapio, A., Rivetta, C., Schneider, D. P., Snedden, S., y i Wang, S., “The 2.5 m telescope of the sloan digital sky survey”, *The*

- Astronomical Journal, vol. 131, p. 2332, 2006, [doi:10.1086/500975](https://doi.org/10.1086/500975).
- [15] Fukugita, M., Ichikawa, T., Gunn, J. E., Doi, M., Shimasaku, K., y Schneider, D. P., “The Sloan Digital Sky Survey Photometric System”, , vol. 111, p. 1748, 1996, [doi:10.1086/117915](https://doi.org/10.1086/117915).
- [16] Lupton, R. H., Gunn, J. E., y Szalay, A. S., “A modified magnitude system that produces well-behaved magnitudes, colors, and errors even for low signal-to-noise ratio measurements”, The Astronomical Journal, vol. 118, p. 1406–1410, 1999, [doi:10.1086/301004](https://doi.org/10.1086/301004).
- [17] Chambers, K. C., Magnier, E. A., Metcalfe, N., Flewelling, H. A., Huber, M. E., Waters, C. Z., Denneau, L., Draper, P. W., Farrow, D., Finkbeiner, D. P., Holmberg, C., Koppenhoefer, J., Price, P. A., Rest, A., Saglia, R. P., Schlafly, E. F., Smartt, S. J., Sweeney, W., Wainscoat, R. J., Burgett, W. S., Chastel, S., Grav, T., Heasley, J. N., Hodapp, K. W., Jedicke, R., Kaiser, N., Kudritzki, R. P., Luppino, G. A., Lupton, R. H., Monet, D. G., Morgan, J. S., Onaka, P. M., Shiao, B., Stubbs, C. W., Tonry, J. L., White, R., Bañados, E., Bell, E. F., Bender, R., Bernard, E. J., Boegner, M., Boffi, F., Botticella, M. T., Calamida, A., Casertano, S., Chen, W. P., Chen, X., Cole, S., Deacon, N., Frenk, C., Fitzsimmons, A., Gezari, S., Gibbs, V., Goessl, C., Goggia, T., Gourgue, R., Goldman, B., Grant, P., Grebel, E. K., Hambly, N. C., Hasinger, G., Heavens, A. F., Heckman, T. M., Henderson, R., Henning, T., Holman, M., Hopp, U., Ip, W. H., Isani, S., Jackson, M., Keyes, C. D., Koekemoer, A. M., Kotak, R., Le, D., Liska, D., Long, K. S., Lucey, J. R., Liu, M., Martin, N. F., Masci, G., McLean, B., Mindel, E., Misra, P., Morganson, E., Murphy, D. N. A., Obaika, A., Narayan, G., Nieto-Santisteban, M. A., Norberg, P., Peacock, J. A., Pier, E. A., Postman, M., Primak, N., Rae, C., Rai, A., Riess, A., Riffeser, A., Rix, H. W., Röser, S., Russel, R., Rutz, L., Schilbach, E., Schultz, A. S. B., Scolnic, D., Strolger, L., Szalay, A., Seitz, S., Small, E., Smith, K. W., Soderblom, D. R., Taylor, P., Thomson, R., Taylor, A. N., Thakar, A. R., Thiel, J., Thilker, D., Unger, D., Urata, Y., Valenti, J., Wagner, J., Walder, T., Walter, F., Watters, S. P., Werner, S., Wood-Vasey, W. M., y Wyse, R., “The pan-starrs1 surveys”, 2019.
- [18] Waters, C. Z., Magnier, E. A., Price, P. A., Chambers, K. C., Burgett, W. S., Draper, P. W., Flewelling, H. A., Hodapp, K. W., Huber, M. E., Jedicke, R., Kaiser, N., Kudritzki, R.-P., Lupton, R. H., Metcalfe, N., Rest, A., Sweeney, W. E., Tonry, J. L., Wainscoat, R. J., y Wood-Vasey, W. M., “Pan-starrs pixel processing: Detrending, warping, stacking”, The Astrophysical Journal Supplement Series, vol. 251, p. 4, 2020, [doi:10.3847/1538-4365/abb82b](https://doi.org/10.3847/1538-4365/abb82b).
- [19] Förster, F., Arancibia, A. M. M., Reyes-Jainaga, I., Gagliano, A., Britt, D., Cuellar-Carrillo, S., Figueroa-Tapia, F., Polzin, A., Yousef, Y., Arredondo, J., Rodríguez-Mancini, D., Correa-Orellana, J., Bayo, A., Bauer, F. E., Catelan, M., Cabrera-Vives, G., Dastidar, R., Estévez, P. A., Pignata, G., Hernández-García, L., Huijse, P., Reyes, E., Sánchez-Sáez, P., Ramírez, M., Grandón, D., Pineda-García, J., Chabour-Barra, F., y Silva-Farfán, J., “Delight: Deep learning identification of galaxy hosts of transients using multiresolution images”, The Astronomical Journal, vol. 164, p. 195, 2022, [doi:10.3847/1538-3881/ac912a](https://doi.org/10.3847/1538-3881/ac912a).
- [20] Reyes-Jainaga, I., Förster, F., Arancibia, A. M. M., Cabrera-Vives, G., Bayo, A., Bauer, F. E., Arredondo, J., Reyes, E., Pignata, G., Mourão, A. M., Silva-Farfán, J., Galbany, L., Álvarez, A., Astorga, N., Castellanos, P., Gallardo, P., Moya, A., y Rodríguez, D., “Multiscale stamps for real-time classification of alert streams”, The Astrophysical Jour-

- nal Letters, vol. 952, p. L43, 2023, [doi:10.3847/2041-8213/ace77e](https://doi.org/10.3847/2041-8213/ace77e).
- [21] Bailer-Jones, C. A. L., Gupta, R., y Singh, H. P., “An introduction to artificial neural networks”, 2001, <https://arxiv.org/abs/astro-ph/0102224>.
- [22] Rosenblatt, F., “The perceptron: a probabilistic model for information storage and organization in the brain.”, *Psychological review*, vol. 65 6, pp. 386–408, 1958, <https://api.semanticscholar.org/CorpusID:12781225>.
- [23] Nair, V. y Hinton, G. E., “Rectified linear units improve restricted boltzmann machines”, en *ICML 2010*, pp. 807–814, 2010.
- [24] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., y Jackel, L. D., “Backpropagation Applied to Handwritten Zip Code Recognition”, *Neural Computation*, vol. 1, pp. 541–551, 1989, [doi:10.1162/neco.1989.1.4.541](https://doi.org/10.1162/neco.1989.1.4.541).
- [25] O’Shea, K. y Nash, R., “An introduction to convolutional neural networks”, 2015, <https://arxiv.org/abs/1511.08458>.
- [26] Robbins, H. y Monro, S., “A Stochastic Approximation Method”, *The Annals of Mathematical Statistics*, vol. 22, no. 3, pp. 400 – 407, 1951, [doi:10.1214/aoms/1177729586](https://doi.org/10.1214/aoms/1177729586).
- [27] Kiefer, J. y Wolfowitz, J., “Stochastic Estimation of the Maximum of a Regression Function”, *The Annals of Mathematical Statistics*, vol. 23, no. 3, pp. 462 – 466, 1952, [doi:10.1214/aoms/1177729392](https://doi.org/10.1214/aoms/1177729392).
- [28] Kingma, D. P. y Ba, J., “Adam: A method for stochastic optimization”, 2017, <https://arxiv.org/abs/1412.6980>.
- [29] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., y Salakhutdinov, R., “Dropout: A simple way to prevent neural networks from overfitting”, *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014, <http://jmlr.org/papers/v15/srivastava14a.html>.
- [30] Shorten, C. y Khoshgoftaar, T., “A survey on image data augmentation for deep learning”, *Journal of Big Data*, vol. 6, 2019, [doi:10.1186/s40537-019-0197-0](https://doi.org/10.1186/s40537-019-0197-0).
- [31] Ying, X., “An overview of overfitting and its solutions”, *Journal of Physics: Conference Series*, vol. 1168, p. 022022, 2019, [doi:10.1088/1742-6596/1168/2/022022](https://doi.org/10.1088/1742-6596/1168/2/022022).
- [32] Beck, R., Dobos, L., Budavári, T., Szalay, A. S., y Csabai, I., “Photometric redshifts for the SDSS Data Release 12”, *Monthly Notices of the Royal Astronomical Society*, vol. 460, pp. 1371–1381, 2016, [doi:10.1093/mnras/stw1009](https://doi.org/10.1093/mnras/stw1009).
- [33] Schlegel, D. J., Finkbeiner, D. P., y Davis, M., “Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds”, *ApJ*, vol. 500, no. 2, pp. 525–553, 1998, [doi:10.1086/305772](https://doi.org/10.1086/305772).
- [34] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., y Rabinovich, A., “Going deeper with convolutions”, 2014, <https://arxiv.org/abs/1409.4842>.
- [35] He, K., Zhang, X., Ren, S., y Sun, J., “Deep residual learning for image recognition”, 2015, <https://arxiv.org/abs/1512.03385>.
- [36] Chen, T., Kornblith, S., Norouzi, M., y Hinton, G., “A simple framework for contrastive learning of visual representations”, 2020, <https://arxiv.org/abs/2002.05709>.

- [37] Dawson, K. S., Schlegel, D. J., Ahn, C. P., Anderson, S. F., Aubourg, , Bailey, S., Barkhouser, R. H., Bautista, J. E., Beifiori, A., Berlind, A. A., Bhardwaj, V., Bizyaev, D., Blake, C. H., Blanton, M. R., Blomqvist, M., Bolton, A. S., Borde, A., Bovy, J., Brandt, W. N., Brewington, H., Brinkmann, J., Brown, P. J., Brownstein, J. R., Bundy, K., Busca, N. G., Carithers, W., Carnero, A. R., Carr, M. A., Chen, Y., Comparat, J., Connolly, N., Cope, F., Croft, R. A. C., Cuesta, A. J., da Costa, L. N., Davenport, J. R. A., Delubac, T., de Putter, R., Dhital, S., Ealet, A., Ebelke, G. L., Eisenstein, D. J., Escoffier, S., Fan, X., Filiz Ak, N., Finley, H., Font-Ribera, A., Génova-Santos, R., Gunn, J. E., Guo, H., Haggard, D., Hall, P. B., Hamilton, J.-C., Harris, B., Harris, D. W., Ho, S., Hogg, D. W., Holder, D., Honscheid, K., Huehnerhoff, J., Jordan, B., Jordan, W. P., Kauffmann, G., Kazin, E. A., Kirkby, D., Klaene, M. A., Kneib, J.-P., Le Goff, J.-M., Lee, K.-G., Long, D. C., Loomis, C. P., Lundgren, B., Lupton, R. H., Maia, M. A. G., Makler, M., Malanushenko, E., Malanushenko, V., Mandelbaum, R., Manera, M., Maraston, C., Margala, D., Masters, K. L., McBride, C. K., McDonald, P., McGreer, I. D., McMahon, R. G., Mena, O., Miralda-Escudé, J., Montero-Dorta, A. D., Montesano, F., Muna, D., Myers, A. D., Naugle, T., Nichol, R. C., Noterdaeme, P., Nuza, S. E., Olmstead, M. D., Oravetz, A., Oravetz, D. J., Owen, R., Padmanabhan, N., Palanque-Delabrouille, N., Pan, K., Parejko, J. K., Pâris, I., Percival, W. J., Pérez-Fournon, I., Pérez-Ràfols, I., Petitjean, P., Pfaffenberger, R., Pforr, J., Pieri, M. M., Prada, F., Price-Whelan, A. M., Raddick, M. J., Rebolo, R., Rich, J., Richards, G. T., Rockosi, C. M., Roe, N. A., Ross, A. J., Ross, N. P., Rossi, G., Rubiño-Martin, J. A., Samushia, L., Sánchez, A. G., Sayres, C., Schmidt, S. J., Schneider, D. P., Scóccola, C. G., Seo, H.-J., Shelden, A., Sheldon, E., Shen, Y., Shu, Y., Slosar, A., Smee, S. A., Snedden, S. A., Stauffer, F., Steele, O., Strauss, M. A., Streblyanska, A., Suzuki, N., Swanson, M. E. C., Tal, T., Tanaka, M., Thomas, D., Tinker, J. L., Tojeiro, R., Tremonti, C. A., Vargas Magaña, M., Verde, L., Viel, M., Wake, D. A., Watson, M., Weaver, B. A., Weinberg, D. H., Weiner, B. J., West, A. A., White, M., Wood-Vasey, W. M., Yèche, C., Zehavi, I., Zhao, G.-B., y Zheng, Z., “The baryon oscillation spectroscopic survey of sdss-iii”, *The Astronomical Journal*, vol. 145, p. 10, 2012, [doi:10.1088/0004-6256/145/1/10](https://doi.org/10.1088/0004-6256/145/1/10).
- [38] Driver, S. P., Norberg, P., Baldry, I. K., Bamford, S. P., Hopkins, A. M., Liske, J., Loveday, J., Peacock, J. A., y (listed below), G. T., “GAMA: towards a physical understanding of galaxy formation”, *Astronomy Geophysics*, vol. 50, pp. 5.12–5.19, 2009, [doi:10.1111/j.1468-4004.2009.50512.x](https://doi.org/10.1111/j.1468-4004.2009.50512.x).
- [39] Driver, S. P., Hill, D. T., Kelvin, L. S., Robotham, A. S. G., Liske, J., Norberg, P., Baldry, I. K., Bamford, S. P., Hopkins, A. M., Loveday, J., Peacock, J. A., Andrae, E., Bland-Hawthorn, J., Brough, S., Brown, M. J. I., Cameron, E., Ching, J. H. Y., Colless, M., Conselice, C. J., Croom, S. M., Cross, N. J. G., De Propriis, R., Dye, S., Drinkwater, M. J., Ellis, S., Graham, A. W., Grootes, M. W., Gunawardhana, M., Jones, D. H., van Kampen, E., Maraston, C., Nichol, R. C., Parkinson, H. R., Phillipps, S., Pimblet, K., Popescu, C. C., Prescott, M., Roseboom, I. G., Sadler, E. M., Sansom, A. E., Sharp, R. G., Smith, D. J. B., Taylor, E., Thomas, D., Tuffs, R. J., Wijesinghe, D., Dunne, L., Frenk, C. S., Jarvis, M. J., Madore, B. F., Meyer, M. J., Seibert, M., Staveley-Smith, L., Sutherland, W. J., y Warren, S. J., “Galaxy and mass assembly (gama): survey diagnostics and core data release: Gama”, *Monthly Notices of the Royal Astronomical Society*, vol. 413, p. 971–995, 2011, [doi:10.1111/j.1365-2966.2010.18188.x](https://doi.org/10.1111/j.1365-2966.2010.18188.x).
- [40] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin,

- Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., y Chintala, S., “Pytorch: An imperative style, high-performance deep learning library”, 2019, <https://arxiv.org/abs/1912.01703>.
- [41] Boch, T., Fernique, P., Bonnarel, F., Chaitra, C., Bot, C., Pineau, F. X., Baumann, M., y Michel, L., “HiPS2FITS: Fast Generation of FITS Cutouts From HiPS Image Datasets”, en *Astronomical Data Analysis Software and Systems XXIX* (Pizzo, R., Deul, E. R., Mol, J. D., de Plaa, J., y Verkouter, H., eds.), vol. 527 de *Astronomical Society of the Pacific Conference Series*, p. 121, 2020.
- [42] Lin, Q., Fouchez, D., Pasquet, J., Treyer, M., Ait Ouahmed, R., Arnouts, S., y Ilbert, O., “Photometric redshift estimation with convolutional neural networks and galaxy images: Case study of resolving biases in data-driven methods”, *Astronomy; Astrophysics*, vol. 662, p. A36, 2022, [doi:10.1051/0004-6361/202142751](https://doi.org/10.1051/0004-6361/202142751).
- [43] Tarrío, P. y Zarattini, S., “Photometric redshifts for the pan-starrs1 survey”, *Astronomy; Astrophysics*, vol. 642, p. A102, 2020, [doi:10.1051/0004-6361/202038415](https://doi.org/10.1051/0004-6361/202038415).
- [44] Sosnovik, I., Szmajda, M., y Smeulders, A., “Scale-equivariant steerable networks”, 2020, <https://arxiv.org/abs/1910.11093>.
- [45] Bank, D., Koenigstein, N., y Giryes, R., “Autoencoders”, 2021, <https://arxiv.org/abs/2003.05991>.
- [46] Blanton, M. R., Dalcanton, J., Eisenstein, D., Loveday, J., Strauss, M. A., SubbaRao, M., Weinberg, D. H., Anderson, John E., J., Annis, J., Bahcall, N. A., Bernardi, M., Brinkmann, J., Brunner, R. J., Burles, S., Carey, L., Castander, F. J., Connolly, A. J., Csabai, I., Doi, M., Finkbeiner, D., Friedman, S., Frieman, J. A., Fukugita, M., Gunn, J. E., Hennessy, G. S., Hindsley, R. B., Hogg, D. W., Ichikawa, T., Ivezić, Ž., Kent, S., Knapp, G. R., Lamb, D. Q., Leger, R. F., Long, D. C., Lupton, R. H., McKay, T. A., Meiksin, A., Merelli, A., Munn, J. A., Narayanan, V., Newcomb, M., Nichol, R. C., Okamura, S., Owen, R., Pier, J. R., Pope, A., Postman, M., Quinn, T., Rockosi, C. M., Schlegel, D. J., Schneider, D. P., Shimasaku, K., Siegmund, W. A., Smee, S., Snir, Y., Stoughton, C., Stubbs, C., Szalay, A. S., Szokoly, G. P., Thakar, A. R., Tremonti, C., Tucker, D. L., Uomoto, A., Vanden Berk, D., Vogeley, M. S., Waddell, P., Yanny, B., Yasuda, N., y York, D. G., “The Luminosity Function of Galaxies in SDSS Commissioning Data”, , vol. 121, pp. 2358–2380, 2001, [doi:10.1086/320405](https://doi.org/10.1086/320405).
- [47] Yasuda, N., Fukugita, M., Narayanan, V. K., Lupton, R. H., Strateva, I., Strauss, M. A., Ivezić, Ž., Kim, R. S. J., Hogg, D. W., Weinberg, D. H., Shimasaku, K., Loveday, J., Annis, J., Bahcall, N. A., Blanton, M., Brinkmann, J., Brunner, R. J., Connolly, A. J., Csabai, I., Doi, M., Hamabe, M., Ichikawa, S.-I., Ichikawa, T., Johnston, D. E., Knapp, G. R., Kunszt, P. Z., Lamb, D. Q., McKay, T. A., Munn, J. A., Nichol, R. C., Okamura, S., Schneider, D. P., Szokoly, G. P., Vogeley, M. S., Watanabe, M., y York, D. G., “Galaxy Number Counts from the Sloan Digital Sky Survey Commissioning Data”, , vol. 122, pp. 1104–1124, 2001, [doi:10.1086/322093](https://doi.org/10.1086/322093).
- [48] Lundberg, S. y Lee, S.-I., “A unified approach to interpreting model predictions”, 2017, <https://arxiv.org/abs/1705.07874>.

Anexo A

Antecedentes de términos astronómicos

A lo largo de esta memoria se mencionaron varias características de las galaxias, pero no se especificó cómo se calculaban o conseguían dichos valores. Uno de ellos es el radio Petrosiano, el cual sirvió para determinar el tamaño de las galaxias utilizadas, pero, ¿cómo calcular el radio Petrosiano?. Las galaxias están a distintas distancias de donde son observadas, esto hace que la comparación de sus tamaños sea complejo debido a este factor. Es por eso que en el SDSS ocupan un método que no se basa en la distancia del objeto [46, 47], el cual consiste en determinar el radio de un annulus (un círculo) en el cual el flujo promedio en ese annulus sea 0.2 veces el flujo promedio dentro del mismo. Formalmente la ecuación para encontrar esta proporción (\mathcal{R}_P) para determinar el radio Petrosiano es la siguiente:

$$\mathcal{R}_P(r) \equiv \frac{\int_{\alpha r}^{\beta r} dr' 2\pi r' I(r') / [\pi(\alpha^2 - \beta^2)r^2]}{\int_0^r dr' 2\pi r' I(r') / [\pi r^2]}, \quad (\text{A.1})$$

donde el numerador corresponde al flujo promedio dentro de anillo de cierto grosor determinado por el radio de dos annulus, donde el annulus de menor tamaño es de radio βr y el de mayor de tamaño es de radio αr , I es el flujo y r es el radio Petrosiano. Por otro lado, el denominador es el flujo promedio desde el centro de la galaxia hasta el radio r . Finalmente, para determinar el radio Petrosiano el SDSS utiliza la proporción $\mathcal{R}_P = 0.2$, un $\alpha = 1.25$ y un $\beta = 0.8$.

Otra característica de las galaxias que se utilizó fue la magnitud. Las magnitudes son una transformación del flujo (I), la cual por defecto es igual a:

$$\text{MAG}(I) = -2.5 \log_{10}(I) + C. \quad (\text{A.2})$$

Como se puede comprobar, cuerpos celestes con una mayor cantidad de flujo tienen una menor magnitud que aquellos con menor cantidad de flujo. Para determinar finalmente la magnitud de una galaxia lo que se hace es calcular el total del flujo (o magnitud) usando el radio Petrosiano de la misma. Esto se hace de la siguiente forma:

$$F_P = \int_0^{N \cdot r_p} dr' 2\pi r' I(r'), \quad (\text{A.3})$$

donde F_P es el flujo Petrosiano, r_p es el radio petrosiano, I es el flujo y N es un ponderador.

En el SDSS se ocupa un $N = 2$, en donde se asegura de capturar el total de flujo de una galaxia dentro de un annulus utilizando por ende, un radio igual al doble del radio Petrosiano. Otro aspecto importante es que para calcular la magnitud petrosiana en cada banda del SDSS, se utiliza el radio Petrosiano de la banda r .

En base al flujo Petrosiano se pueden calcular otro tipo de radios Petrosianos, los utilizados por el SDSS son el **PetroRad90** y **PetroRad50**. Estos simplemente son los radios en los cuales se captura el 90 % y 50 % del flujo Petrosiano respectivamente. En este trabajo se utilizó particularmente el **PetroRad90**.

Pero, ¿cómo obtener estos o más datos fotométricos?. Existen diversas formas, una de las que se utiliza en la literatura (y que también se usó en esta memoria) es utilizar el servicio de CasJobs, en el cual se pueden hacer consultas (*queries*) en lenguaje SQL⁹ donde se van extrayendo y conectando estos metadatos de diversas tablas de diversos *surveys*. Cada *survey* posee diversas tablas, al igual que diversas páginas en donde realizar estas consultas¹⁰.

En la Tabla A.1 se presentan las principales tablas de la base de datos del SDSS utilizadas en esta memoria y una breve descripción de los datos que contiene, mientras que en la Tabla A.2 se presenta lo mismo pero para el caso de Pan-STARRS.

Tabla A.1: Tablas de la base de datos del SDSS utilizadas en el servicio de CasJobs.

| Tabla | Tipo | Descripción |
|-------------|-----------------|---|
| SpecObjAll | Espectroscópico | Contiene los parámetros medidos para un espectro. Esta es una tabla base que contiene TODA la información espectroscópica, incluyendo muchos datos duplicados y erróneos. En su lugar, se recomienda usar SpecObj, que tiene los datos debidamente filtrados para asegurar su limpieza. Estas tablas contienen datos de los espectrógrafos BOSS y SDSS. |
| PhotoObjAll | Fotométrico | Contiene las cantidades completas del catálogo fotométrico para las imágenes SDSS. Esta tabla contiene una entrada por cada detección, con los parámetros fotométricos asociados medidos por PHOTO, y calibrados astrométrica y fotométricamente. |
| Photoz | Fotométrico | Los redshifts estimados fotométricamente para todas las galaxias en PhotoObjAll. |
| galSpecInfo | Espectroscópico | Información general para el reanálisis espectroscópico MPA-JHU. Esta tabla contiene una entrada por cada observación espectroscópica. Puede unirse con las otras tablas galSpec que contienen las mediciones, o con specObjAll, usando specObjId. |

⁹ *Structured Query Language* por sus siglas en inglés.

¹⁰ Para consultas de datos del SDSS esta la pagina <https://skyserver.sdss.org/casjobs/>, mientras que las consultas para datos de Pan-STARRS se hacen en <https://mastweb.stsci.edu/mcasjobs/>.

Tabla A.2: Tabla de la base de datos de Pan-STARRS utilizada en el servicio de CasJobs.

| Tabla | Tipo | Descripción |
|----------------|-------------|---|
| StackPetrosian | Fotométrico | Contiene las magnitudes y radios Petrosianos para fuentes extendidas. |

También se presenta el Código A.1, el cual fue la consulta realizada en Pasquet et al. [1] para obtener la metadata de las galaxias utilizadas. Las tablas a las que se le hicieron consultas se describen en la Tabla A.1. Adicionalmente se le realizó una leve modificación, ya que la original actualmente tiene problemas de sintaxis en el servicio.

Código A.1: Consulta SQL para obtener la metadata de las galaxias utilizadas en esta memoria. Los datos son del DR12 del SDSS mientras que la consulta es una modificación de [1].

```

1  SELECT
2  za.specObjID,za.bestObjID,za.class,za.subClass,za.z,za.zErr,
3  po.objID,po.type,po.flags, po.ra,po.dec,
4  (po.petroMag_r - po.extinction_r) as dered_petro_r,
5  zp.z as zphot, zp.zErr as dzphot,
6  zi.e_bv_sfd,zi.primtarget,zi.sectarget,zi.targettype,
7  zi.spectrotype, zi.subclass,
8  FROM SpecObjAll za
9  JOIN PhotoObjAll po ON (po.objID = za.bestObjID)
10 JOIN Photoz zp ON (zp.objID = za.bestObjID)
11 JOIN galSpecInfo zi ON (zi.SpecObjID = za.specObjID)
12 WHERE
13 za.z BETWEEN 0 and 0.4
14 and za.zWarning=0
15 and za.targetType ='SCIENCE' and za.survey='sdss'
16 and za.class='GALAXY' and zi.primtarget>=64
17 and po.clean=1 and po.insideMask=0
18 and (po.petroMag_r - po.extinction_r)<=17.8

```

Unas de las ventajas de utilizar directamente el servicio CasJobs en su página son los tiempos de consulta, los cuales son menores comparados con usar su API mediante el uso de Python. Otra de ellas es que se pueden realizar consultas de manera asíncrona, es decir, que se podría consultar sin la necesidad de estar presente durante todo el tiempo de ejecución (computador apagado), a diferencia de usar la API, en los cuales se tiene que estar con el computador encendido y con el código ejecutándose.

Anexo B

Comparación de colores y SNR entre SDSS y Pan-STARRS

Para poder determinar si realmente las imágenes de Pan-STARRS poseen un menor *signal-to-noise ratio*, lo que se hizo fue realizar una consulta a la base de datos mediante el servicio de CasJobs, en donde se consultó por los errores de las magnitudes Petrosianas de cada ejemplo de cada banda en la tabla `StackPetrosian`, debido a que las imágenes utilizadas son de tipo *stack* y también para realizar una comparación usando errores del mismo tipo de magnitud.

Sin embargo, como las imágenes tipo *stack* se sobreponen, existen objetos con más de una medición o incluso que tienen errores en las mismas. Para una comparación justa se utilizó únicamente datos sin errores de medición, por ejemplo, si había una galaxia con algún error de medición, se eliminaba tanto en la base de datos de Pan-STARRS como en la del SDSS.

Como en este caso solo se tenía metadata del SDSS, se descargaron los datos utilizados en [43] y luego se buscaron las galaxias que estuvieran en las coordenadas celestes de los datos del SDSS en un margen de error de 0.0001° , considerando aquellos ejemplos en los que se detectaba 1 solo elemento y que el *redshift* espectrográfico sea igual, logrando un cruce de aproximadamente 450.000 ejemplos.

Una vez realizado el cruce se utilizaron los `objID` de los datos de Pan-STARRS y se realizó la consulta por sus errores respectivos. Es importante destacar que el `objID` de un mismo objeto no es el mismo entre *surveys*, ya que cada uno tiene su forma para determinarlo.

Tras haber obtenido los errores de magnitud y eliminado los mismos objetos que tuvieran errores de medición en ambas bases de datos, se calculó el SNR obtenido en ambos *surveys* en las bandas comunes (*griz*), el cual está determinado por:

$$\text{SNR}_x = \frac{1.086}{\text{petroMagErr}_x}, \quad (\text{B.1})$$

en donde SNR_x es el *signal-to-noise ratio* en la banda x y petroMagErr_x es el error de la magnitud petrosiana en la banda x . En la Figura B.1 se presentan los resultados obtenidos.

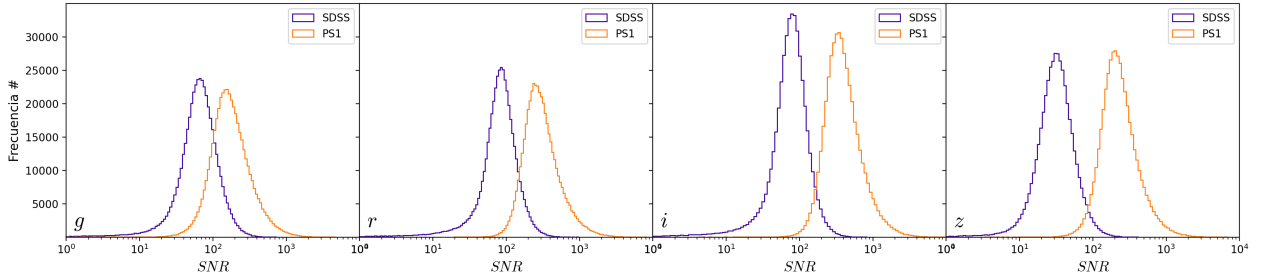


Figura B.1: Histogramas del SNR de las imágenes utilizadas de los *surveys* Pan-STARRS (color naranja) y SDSS (color morado).

Se puede observar que las imágenes de Pan-STARRS presentan un mayor SNR que los del SDSS, lo cual no era lo que se esperaba. Sin embargo, también se debía verificar si las imágenes presentan artefactos, debido a que podría haber un artefacto en una banda que afecte la relación con las demás.

Verificar esto para cada banda de cada galaxia es impracticable, es por eso que se calcularon los colores de cada objeto en cada *survey*. Los colores son diferencias entre las magnitudes de bandas colindantes, es decir, el color $(g - r)$ de una galaxia es la diferencia de su magnitud entre la banda g con la banda r . Por otro lado, como el polvo galáctico dispersa la luz generando que los objetos se vean más rojos, se le debe restar esta diferencia de color de cada banda a cada magnitud respectiva, este valor se le conoce como la extinción.

Pero, hay un problema con usar magnitudes de Pan-STARRS, debido a que lo mejor es utilizar una misma apertura en todas las bandas que considere en su totalidad al cuerpo celeste. Como se dijo, en el SDSS se utiliza el radio Petrosiano de la banda r para calcular las magnitudes en todas las bandas, esto se debe a que la banda r es la que mejor respuesta tiene (ver Figura 2.1), por lo que es capaz de capturar más flujo y a las galaxias como tal.

Según la documentación de Pan-STARRS la magnitudes Kron son las mejores para fuentes no puntuales como galaxias agregando que tiene el mejor *signal-to-noise ratio*, el problema es que en cada banda se utilizan aperturas distintas. Sin embargo, como se tiene acceso a los datos utilizados en [43], se utilizaron las mismas magnitudes, en las cuales se utilizó una apertura corregida, en donde seleccionaron (para cada galaxia) la apertura más cercana a la utilizada para calcular la magnitud Kron en la banda r .

A continuación se muestran los colores en común; $(g - r)$, $(r - i)$ e $(i - z)$ de cada *survey* en función del *redshift* espectrográfico, en donde las magnitudes fueron previamente corregidas por sus valores de extinción.

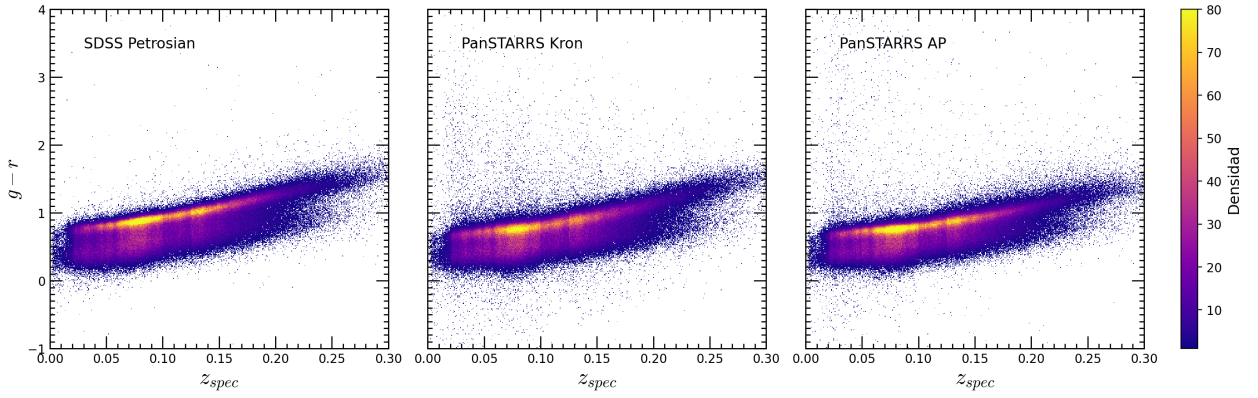


Figura B.2: Comparación del color ($g - r$) en función del *redshift* utilizando magnitudes Petrosianas del SDSS (figura de la izquierda), magnitudes Kron de Pan-STARRS (figura del centro) y magnitudes con una apertura corregida también de Pan-STARRS (figura de la derecha). Las magnitudes fueron previamente corregidas por sus valores de extinción.

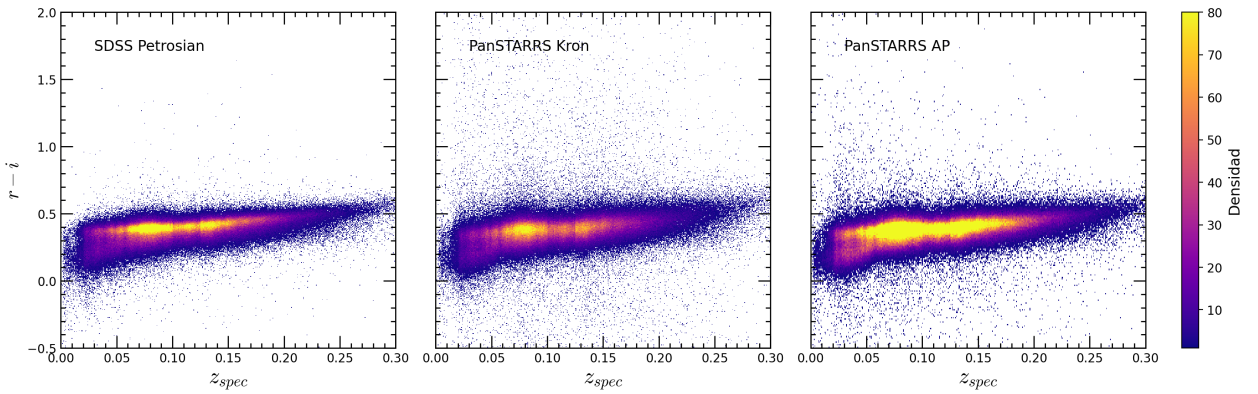


Figura B.3: Comparación del color ($r - i$) en función del *redshift* utilizando magnitudes Petrosianas del SDSS (figura de la izquierda), magnitudes Kron de Pan-STARRS (figura del centro) y magnitudes con una apertura corregida también de Pan-STARRS (figura de la derecha). Las magnitudes fueron previamente corregidas por sus valores de extinción.

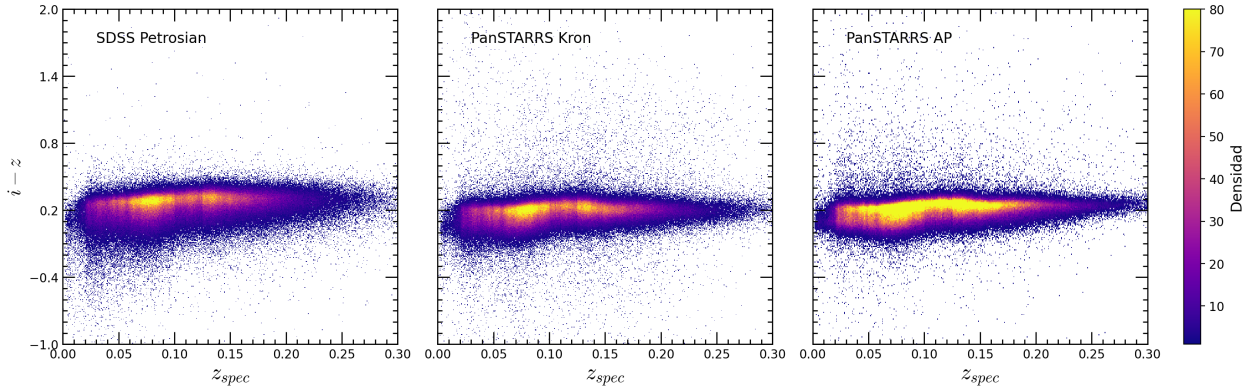


Figura B.4: Comparación del color ($i - z$) en función del *redshift* utilizando magnitudes Petrosianas del SDSS (figura de la izquierda), magnitudes Kron de Pan-STARSS (figura del centro) y magnitudes con una apertura corregida también de Pan-STARSS (figura de la derecha). Las magnitudes fueron previamente corregidas por sus valores de extinción.

Se observa que existe una mayor dispersión en los colores de Pan-STARSS comparados a los colores del SDSS. De esto se podría inferir que realmente podría existir un problema de artefactos en varios ejemplos, los cuales interfieren con una correcta fotometría como es el cálculo de magnitudes, y por ende del cálculo de colores. De eso también se podría explicar los malos resultados en los modelos entrenados, ya que si hay una cantidad considerable de artefactos en las imágenes aunque sea en una banda, esto influye directamente en las relaciones que existen con el resto. Ahora bien, esto podría generarse principalmente porque las imágenes *stack* son una combinación de imágenes *warp*, las cuales presentan cantidades considerable de artefactos. En la Figura B.5 se puede ver el problema descrito para las galaxias M99 y M101 usando un FoV = 60 ["], en donde los artefactos son considerablemente más significativos cuando se usa un campo de visión menor.

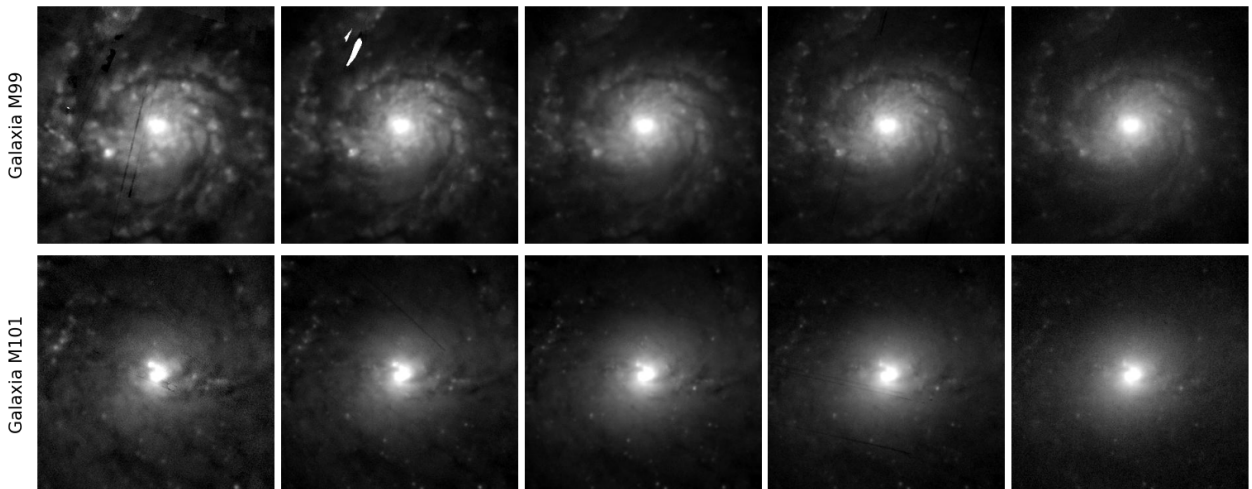


Figura B.5: Imágenes tipo *stack* extraídas de Pan-STARSS de galaxias. En la parte superior se encuentran las imágenes de la galaxia M99 en los filtros *grizy* (imágenes de izquierda a derecha), mientras que en la parte inferior se encuentran las de la galaxia M101. Ambos casos presentan artefactos en las mismas y/o distintas bandas.

Anexo C

Experimentos adicionales

C.1. Impacto de la banda u del SDSS en las predicciones

Para ver si realmente la banda u del SDSS influía en que las predicciones al usar las imágenes del SDSS fueran significativamente mejores que al usar imágenes de Pan-STARRS, se realizó un experimento en el que se ocuparon las imágenes del SDSS en todas las bandas $ugriz$ y otro en el que se usaron únicamente las bandas $griz$.

Este experimento se realizó antes de realizar todos los experimentos mencionados en esta memoria. La forma en el que se hizo fue utilizar los conjuntos de entrenamiento, validación y test usados por [6]. Además se utilizaron los mismos hiperparámetros que en la metodología, pero, como se tiene un conjunto de test, se utilizó el modelo en la mejor época de validación para obtener las métricas finales con dicho conjunto.

En la Figura C.1 se aprecian los resultados de las regresiones sobre el conjunto de test. Adicionalmente se presenta la Figura C.2 en donde se muestra el σ_{MAD} en atributos de interés, tales como; el *redshift* (z), el radio Petrosiano del 90 % en la banda r (PetroR90) y la magnitud en la banda r (MAG), tanto al utilizar todas las bandas como al eliminar la banda u .

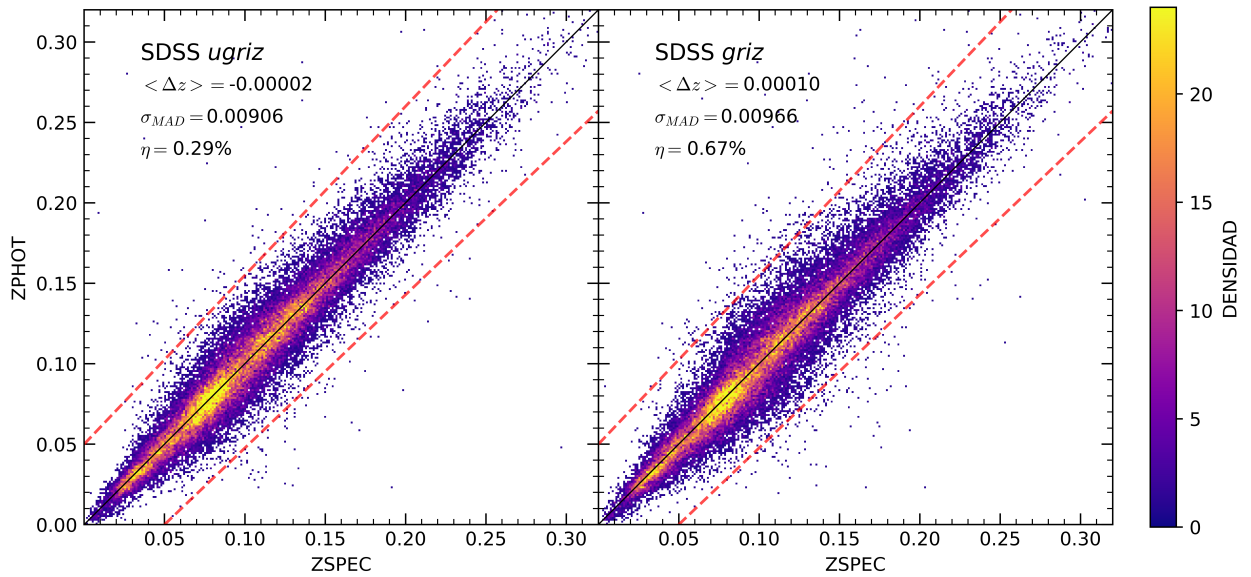


Figura C.1: Resultados de las regresiones al usar imágenes en resolución simple del SDSS provenientes de HiPS2FITS sobre el conjunto de test. En la figura de la izquierda se utilizaron las bandas *ugriz* mientras que en la derecha se usaron las bandas *griz*. Las líneas rojas punteadas corresponden a los límites que definen los errores catastróficos.

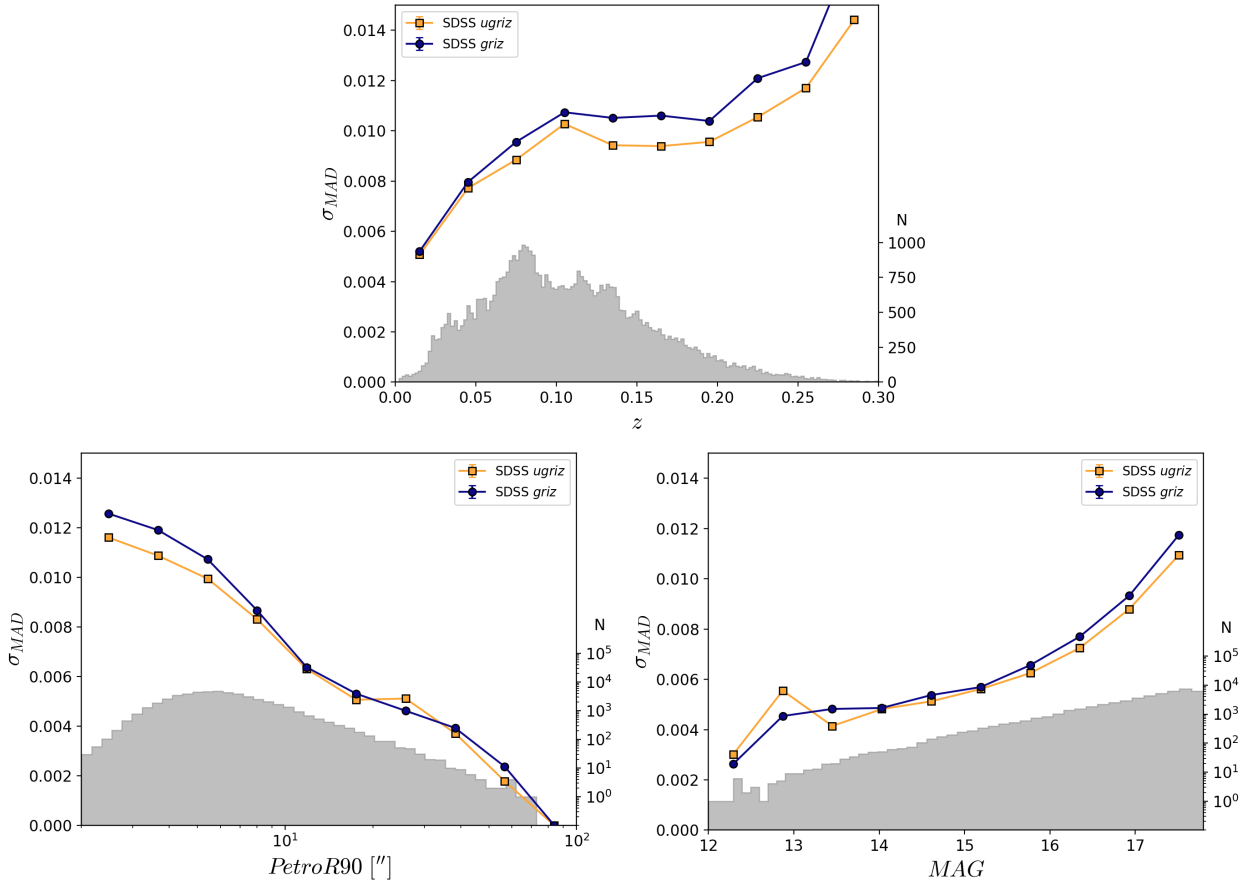


Figura C.2: Dispersión de los errores (σ_{MAD}) sobre el conjunto de test en distintas características usando las imágenes de resolución simple del SDSS provenientes de HiPS2FITS al usar las bandas *ugriz* y *griz*. La ilustración superior es a lo largo del *redshift*, la inferior izquierda a lo largo del radio de la galaxia, mientras que la inferior derecha es a lo largo de su magnitud en la banda *r*. Los histogramas corresponden a la distribución de cada característica en el conjunto de test.

De la Figura C.1 se puede ver que al eliminar la banda *u* las predicciones empeoran de forma general, repercutiendo en las métricas obtenidas. El σ_{MAD} aumentó aproximadamente un 6.6%, mientras que el porcentaje de errores catastróficos aumentó un 131% con respecto a utilizar todas las bandas.

Un fenómeno similar se aprecia con más detalle en la Figura C.2, en donde al eliminar esta banda las predicciones son peores de forma general, ya sea para *redshifts*, radios Petrosianos y magnitudes de las galaxias tanto altos como bajos.

C.2. Interpretando predicciones con SHAP

Para verificar si al compartir pesos las características extraídas en los cinco niveles son utilizadas en similar medida por el modelo, se utilizó SHAP (SHapley Additive exPlanations) [48]. Esta es una herramienta que sirve para dar interpretabilidad de las predicciones del modelo con respecto al enfoque que hace en la entrada.

Como no se quiere ver el enfoque que hace el modelo en las imágenes para realizar las predicciones sino en las características extraídas por los bloques convolucionales, se utilizaron éstas y se examinó el enfoque que les da en las capas *fully connected*, de esta forma se puede ver los niveles en lo que se enfoca para realizar la predicción.

Dicho esto se utilizó el explicador **DeepExplainer** y el modelo pre-entrenado del primer *fold*, tanto para el caso en el que se compartieron pesos como en el experimento en el que no se compartieron. Se usaron dos subconjuntos de datos, el primero fue de 9 galaxias con un $\text{PetroRad90} > 60$ ["] y el otro de 10 galaxias con un $\text{PetroRad90} < 15$ ["]. No se escogieron más datos por los tiempos que le toma a la librería realizar los cálculos.

En cada caso se utilizó el promedio de los valores absolutos de los valores SHAP para cada ejemplo, y finalmente se sumaron estos valores en cada clase y para cada característica. De esta forma se obtiene el $\text{mean}(|\text{shap value}|)$ para cada característica.

En la Figura C.3 se presentan los $\text{mean}(|\text{shap value}|)$ para cada característica de cada nivel usando 9 galaxias con un $\text{PetroRad90} > 60$ ["], mientras que en la Figura C.4 se presenta lo mismo pero usando 10 galaxias con un $\text{PetroRad90} < 15$ ["].

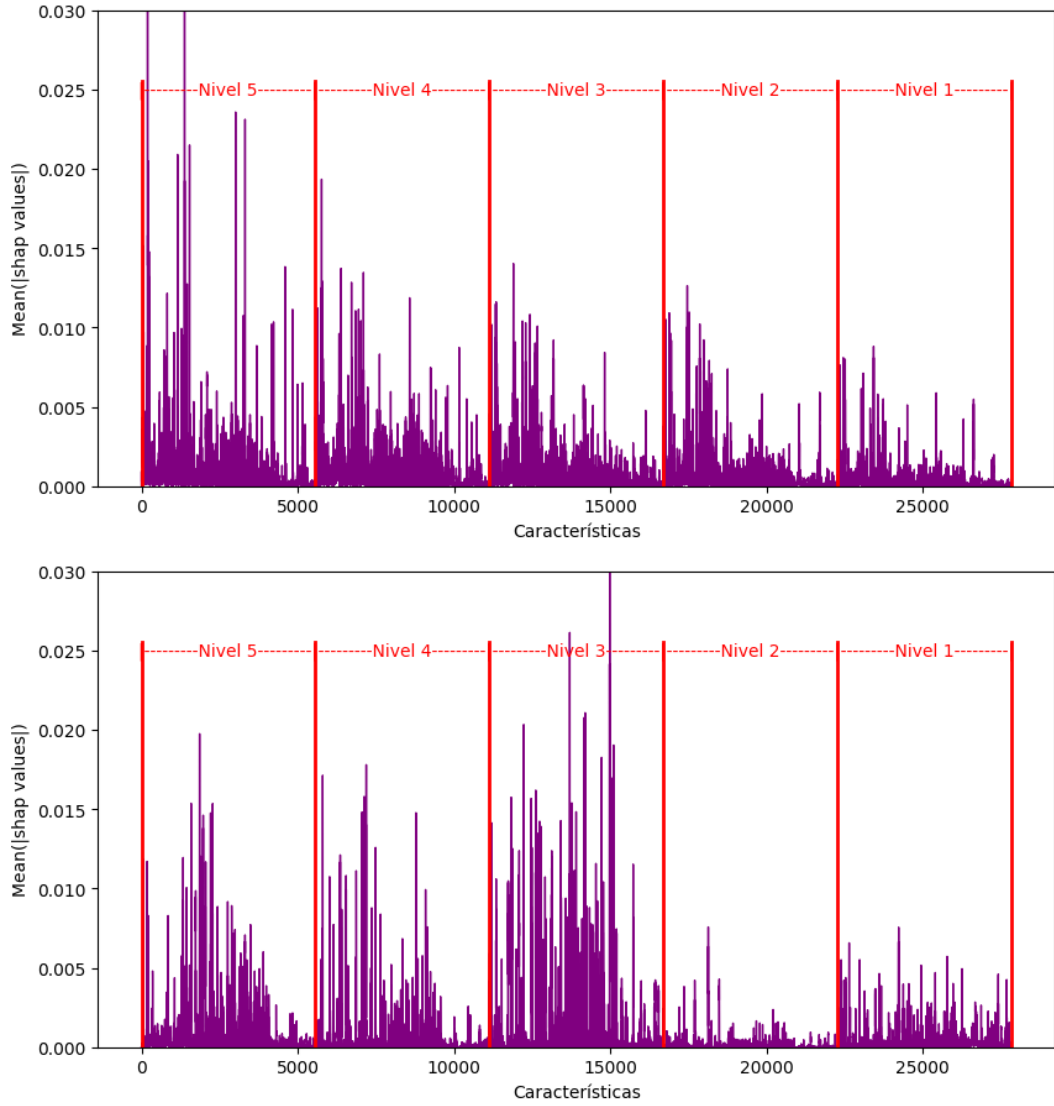


Figura C.3: Mean(|shap values|) obtenidos al usar las características extraídas de los bloques convolucionales en cada nivel con el clasificador del modelo, usando 9 galaxias con un $\text{PetroRad90} > 60$ ["]. La figura superior corresponde al caso de compartir pesos mientras que la inferior al caso de no compartir pesos.

Se puede apreciar que al no compartir pesos el modelo se enfoca en las características de niveles específicos, por ejemplo para predecir galaxias grandes el modelo se enfoca en las características de los últimos 3 niveles (Nivel 3, 4 y 5), en donde los niveles 1 y 2 casi no influyen en la predicción final. Por otro lado, al compartir pesos las características que extraen los bloques convolucionales influyen en la predicción de forma similar, por lo que las características de los primeros niveles si influyen en la predicción del *redshift* y al parecer sirven para realizar una mejor predicción, lo cual era una de las hipótesis que se tenía.

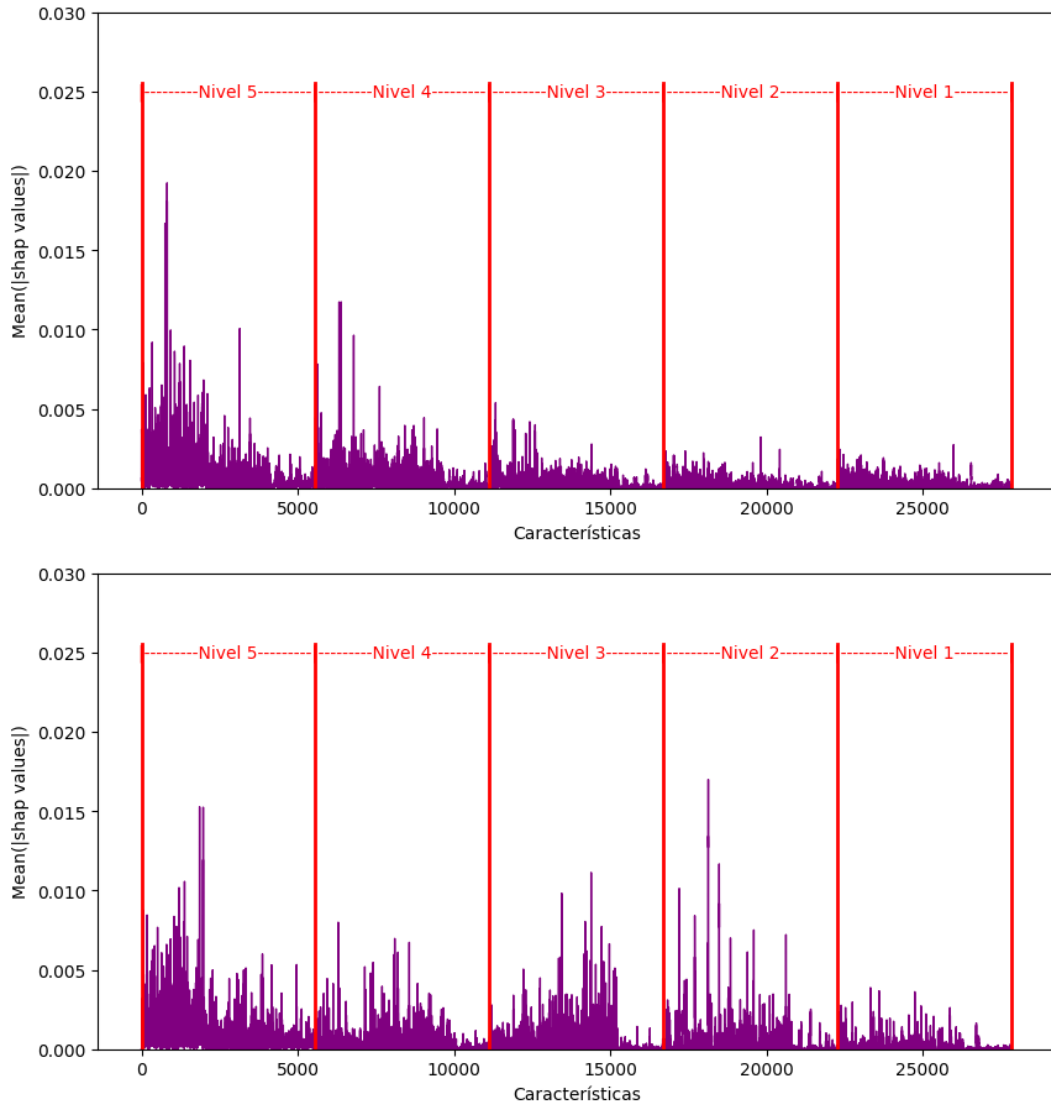


Figura C.4: Mean(|shap values|) obtenidos al usar las características extraídas de los bloques convolucionales en cada nivel con el clasificador del modelo, usando 10 galaxias con un $\text{PetroRad90} < 15''$. La figura superior corresponde al caso de compartir pesos mientras que la inferior al caso de no compartir pesos.

Otra hipótesis que se tenía era que si bien al compartir pesos se regulaba que la red fuera capaz de extraer características con el mismo grado de importancia en cada nivel, igualmente las características de los niveles superiores iban a ser más influyentes en las predicciones de galaxias pequeñas debido a que en los primeros niveles estas galaxias estaban representadas casi en un único pixel, por lo que era más difícil que pueda extraer características igual de complejas e influyentes que en los niveles superiores. Y esto fue lo que precisamente ocurrió. Se observa en la Figura C.4 que la influencia de las características de los niveles en las predicciones va aumentando progresivamente a medida que aumenta el nivel.

Por otro lado al no compartir pesos la influencia de las características no es tan definida, debido a que ahora el modelo puede extraer características más importantes en cada nivel

priorizando aquellos que mayor información útil tengan para realizar la predicción.

C.3. Uso de niveles de forma independiente

A continuación se muestran los resultados al utilizar cada uno de los cinco niveles de forma independiente, en donde se llevó a cabo la misma metodología de los experimentos principales. En la Tabla C.1 se pueden encontrar las métricas promedio de los cinco *folds* de test agregando el σ_{MAD} dados distintos radios Petrosianos de las galaxias (R_P).

Tabla C.1: Resumen de métricas promedio de los 5 *folds* de test obtenidos al usar los niveles de forma independiente.

| Experimento | $\langle \Delta z \rangle [10^{-4}]$ | $\eta [\%]$ | $\sigma_{\text{MAD}} [10^{-3}]$ | $\sigma_{\text{MAD}}(R_p > 30) [10^{-3}]$ | $\sigma_{\text{MAD}}(R_p > 50) [10^{-3}]$ | $\sigma_{\text{MAD}}(R_p > 70) [10^{-3}]$ |
|-------------|--------------------------------------|--------------|---------------------------------|---|---|---|
| Nivel 1 | 3.088 | 1.633 | 13.832 | 5.415 | 5.659 | 4.502 |
| Nivel 2 | 2.464 | 1.105 | 12.430 | 4.771 | 5.069 | 4.471 |
| Nivel 3 | 1.811 | 0.663 | 10.747 | 4.455 | 4.683 | 3.822 |
| Nivel 4 | 1.449 | 0.441 | 9.752 | 4.784 | 5.517 | 5.448 |
| Nivel 5 | 1.166 | 0.332 | 9.295 | 5.474 | 7.361 | 7.377 |

En la Figura C.5 se presenta el σ_{MAD} obtenido en cada uno de estos experimentos sobre atributos de interés, tales como; el *redshift* (z), el radio Petrosiano del 90% en la banda r (PetroR90) y la magnitud en la banda r (MAG).

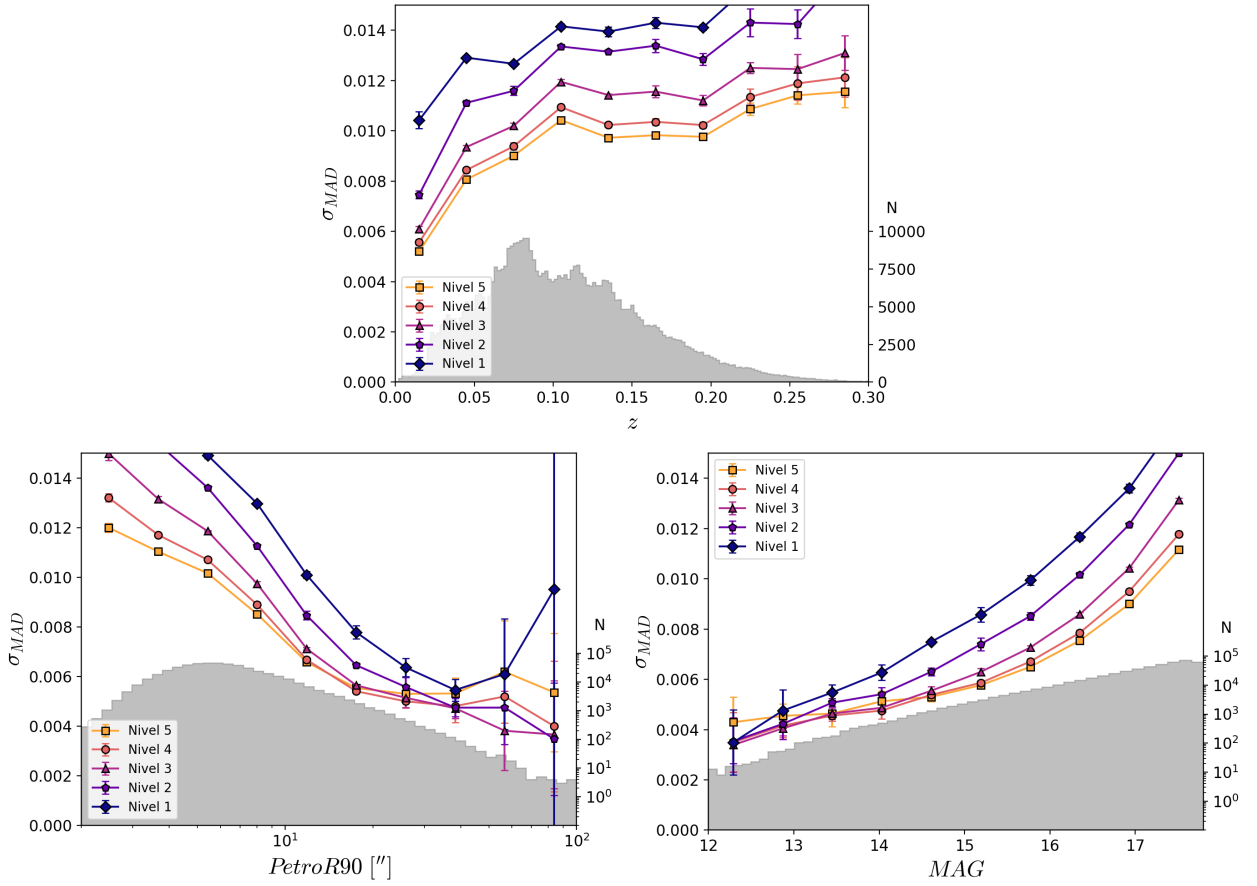


Figura C.5: Promedio de la dispersión de los errores (σ_{MAD}) en los 5 *folds* de test y su respectiva desviación estándar en distintas características usando los niveles de forma independiente. La ilustración superior es a lo largo del *redshift*, la inferior izquierda a lo largo del radio de la galaxia, mientras que la inferior derecha es a lo largo de su magnitud en la banda *r*. Los histogramas corresponden a la distribución de cada característica en todo del conjunto de datos.

Anexo D

Resumen de resultados e impacto del SNR

A continuación se muestran las métricas obtenidas en todos los experimentos realizados, en donde se pueden encontrar algunos que no se mostraron anteriormente. En la Tabla D.1 se encuentra un resumen de métricas promedio en cada uno, las cuales se dividieron respecto a los cuatro experimentos principales.

Tabla D.1: Resumen de métricas promedio de los 5 *folds* de test obtenidos en todos los experimentos realizados.

| Experimento | $\langle \Delta z \rangle [10^{-4}]$ | η [%] | $\sigma_{\text{MAD}} [10^{-3}]$ | $\sigma_{\text{MAD}}(R_p > 30) [10^{-3}]$ | $\sigma_{\text{MAD}}(R_p > 50) [10^{-3}]$ | $\sigma_{\text{MAD}}(R_p > 70) [10^{-3}]$ |
|------------------------|--------------------------------------|--------------|---------------------------------|---|---|---|
| PS1 Compresión 1 | 5.099 | 3.236 | 19.345 | 7.142 | 7.399 | 13.236 |
| PS1 Compresión 2 | 4.316 | 3.068 | 19.054 | 6.924 | 6.998 | 15.092 |
| Pasquet et al. | 0.880 | 0.309 | 8.968 | 4.5 | 4.374 | 7.726 |
| HiPS2FITS | 1.215 | 0.307 | 9.057 | 4.6 | 4.553 | 7.454 |
| Nivel 1-2-3-4-5 W/O SW | 0.982 | 0.291 | 9.003 | 4.384 | 5.015 | 6.120 |
| Nivel 1-2-3-4-5 | 1.282 | 0.343 | 9.141 | 4.206 | 4.262 | 3.694 |
| Nivel 2-3-4-5 | 1.258 | 0.347 | 9.174 | 4.325 | 4.241 | 3.043 |
| Nivel 3-4-5 | 1.338 | 0.342 | 9.154 | 4.468 | 3.884 | 3.866 |
| Nivel 4-5 | 1.404 | 0.354 | 9.240 | 4.887 | 5.595 | 3.486 |
| Nivel 5 | 1.166 | 0.332 | 9.295 | 5.474 | 7.361 | 7.377 |
| Nivel 4 | 1.449 | 0.441 | 9.752 | 4.784 | 5.517 | 5.448 |
| Nivel 3 | 1.811 | 0.663 | 10.747 | 4.455 | 4.683 | 3.822 |
| Nivel 2 | 2.464 | 1.105 | 12.430 | 4.771 | 5.069 | 4.471 |
| Nivel 1 | 3.088 | 1.633 | 13.832 | 5.415 | 5.659 | 4.502 |

Finalmente la Figura D.1 muestra el $\langle \Delta z \rangle$ y σ_{MAD} promedio de los cinco *folds* de test en función del SNR en cada banda utilizada del SDSS (*ugriz*), respecto a los experimentos de usar imágenes en resolución simple con HiPS2FITS y también con el uso de imágenes en multi-resolución con compartimiento de pesos.

Estos resultados comprueban lo mencionado en Pasquet et al.[1], en donde mencionan que mejorar los resultados es independiente del modelo que se utilice, ya que el principal factor que determina la calidad de las predicciones es el SNR de las imágenes. Lo cual es lo que se obtuvo, independiente del modelo las mejores predicciones son en imágenes con SNR alto, además de que no hay diferencias significativas entre los resultados de los modelos.

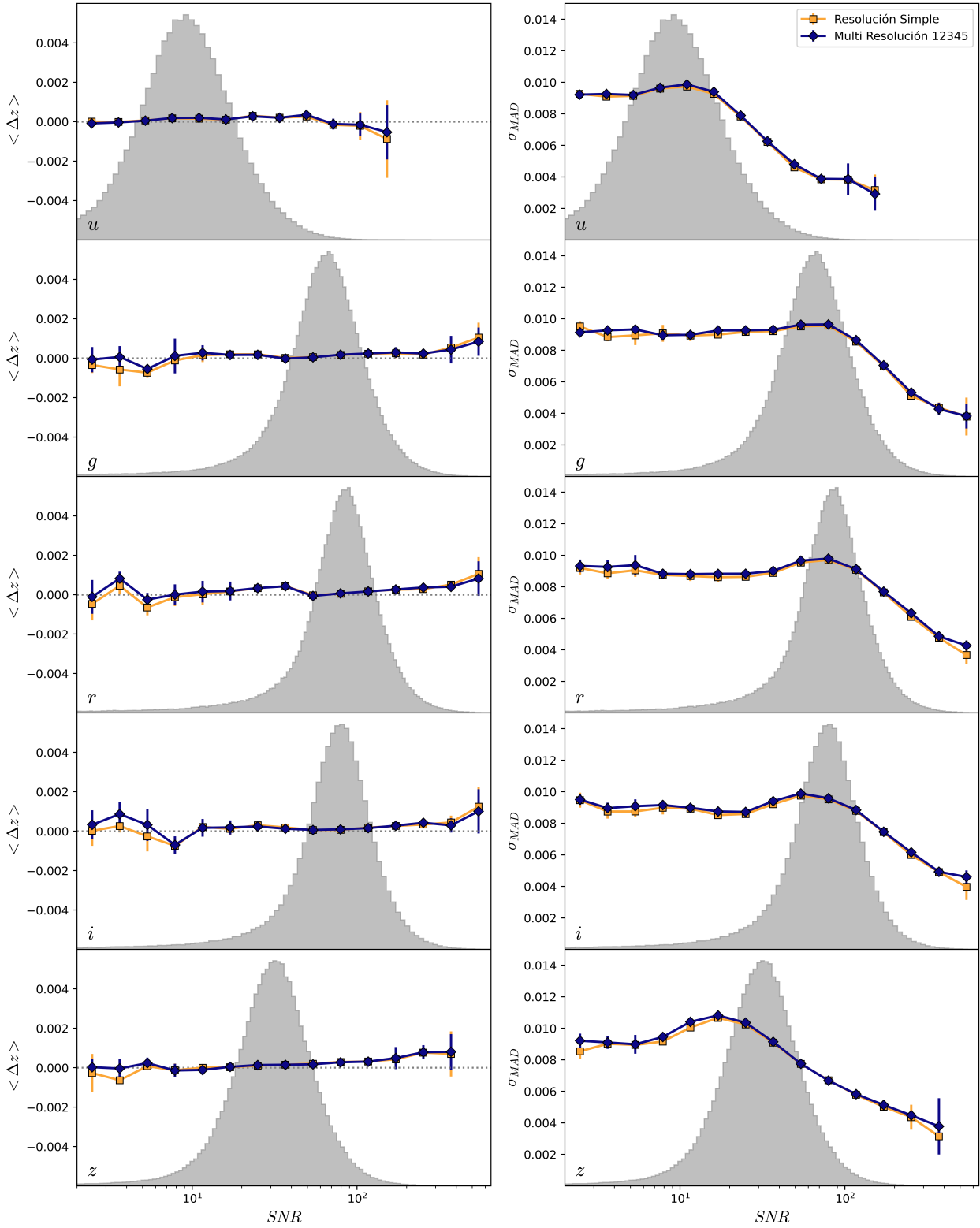


Figura D.1: $\langle \Delta z \rangle$ y σ_{MAD} promedio en función del SNR en cada banda del SDSS. Los resultados provienen de los experimentos de utilizar imágenes en resolución simple y de usar imágenes en multi-resolución con compartimiento de pesos. En ambos casos las imágenes provienen de HiPS2FITS.