



ugr | Universidad
de Granada

TRABAJO FIN DE MÁSTER
MÁSTER EN INGENIERÍA INFORMÁTICA

Deconvolución Ciega de Imágenes Histológicas Usando Aprendizaje Profundo

Autor

José Alberto Gómez García

Director

Rafael Molina Soriano

Codirector

Fernando Pérez Bueno



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍAS INFORMÁTICA Y DE
TELECOMUNICACIÓN

Granada, enero de 2024

Agradecimientos

A mis inspiradores tutores, Rafa y Fernando, quienes no solo me introdujeron en el fascinante mundo de la investigación, sino que también demostraron una paciencia infinita a lo largo de este camino.

A los mejores amigos que Granada me ha podido dar, aquellos con los que me reencontraba una vez más y aquellos que conocí durante este curso. Fran, Alejandra, Jesús, Margalida, Bárbara, y todos los demás, gracias por haber hecho de este quinto año de vida universitaria el mejor de mi vida, los momentos y vivencias que hemos compartido no se olvidarán fácilmente.

A mis padres, gracias a la cultura de esfuerzo y superación que me inculcaron desde pequeño, he logrado enfrentar cada desafío que me propuse, incluyendo este exigente máster. A mi abuela, su apoyo inquebrantable y su vitalidad han sido un faro cuya luz hizo más llevaderos los momentos más difíciles. Gracias a los tres por soportarme todo este tiempo, lo cual no es fácil.

Deconvolución Ciega de Imágenes Histológicas Usando Aprendizaje Profundo

José Alberto Gómez García

Palabras clave: Deconvolución Ciega de Color, Imágenes Histológicas, Redes Neuronales, Aprendizaje Profundo, Deep Image Prior, BCD-Net, H&E

Resumen

Una imagen histológica es una fotografía microscópica de tejido biológico que ha sido preparado y teñido con dos o más tinciones para revelar sus estructuras celulares y tisulares. Las técnicas basadas en la deconvolución ciega de color (BCD) permiten separar el color de las tinciones y la información estructural (concentraciones) de los tejidos, lo cual es de utilidad a la hora de realizar tareas de procesamiento, aumento de datos o clasificación.

Las técnicas clásicas de BCD se basan en procesos analíticos de alta complejidad computacional, que deben aplicarse a cada una de las imágenes histológicas de forma independiente. Los métodos basados en aprendizaje profundo (DL) permiten, una vez han sido entrenados, procesar imágenes no observadas en un tiempo mucho menor. Sin embargo, los estudios que contemplan la aplicación del aprendizaje profundo a la deconvolución ciega de color son escasos, en tanto que encuentran un gran factor limitante, la falta de grandes conjuntos de datos que contengan “ground truth”.

El enfoque Deep Image Prior defiende que la estructura de una red neuronal convolucional profunda, por sí misma, sin entrenamiento de ningún tipo, es capaz de capturar un cantidad considerable de las características de bajo nivel de las imágenes, generando imágenes de una calidad comparable a las de redes neuronales entrenadas.

En este trabajo se propone la aplicación del enfoque Deep Image Prior a modelos de aprendizaje profundo basados en redes neuronales convolucionales con el objetivo de realizar BCD sobre imágenes histológicas. De esta manera, podemos obtener los beneficios de la aplicación de redes neuronales sin requerir de grandes conjuntos de datos que contengan “ground truth”. Para realizar la deconvolución ciega de color de una imagen histológica requeriremos única y exclusivamente de dicha imagen.

Se propondrán tres modelos diferentes, que emplearán arquitecturas, funciones de pérdida y tipos de entrada diferentes. Estos serán puestos a prueba sobre el conjunto de datos “Warwick Stain Separation Benchmark”. Los experimentos realizados indican que se pueden obtener resultados comparables a los proporcionados por modelos amortizados entrenados en grandes conjuntos de datos. Se propone también una variante de Deep Image Prior en la que las redes neuronales se inicializan con los pesos de un entrenamiento previo; lo que permite mejorar los resultados significativamente. Esto supone la apertura de una nueva línea de investigación, mediante la cual podríamos seguir mejorando los resultados obtenidos al realizar BCD sobre imágenes histológicas.

Blind Deconvolution of Histological Images Using Deep Learning

José Alberto Gómez García

Keywords: Blind Color Deconvolution, Histological Imaging, Neural Networks, Deep Learning, Deep Image Prior, BCD-Net, H&E

Abstract

A histological image is a microscopic photograph of biological tissue that has been prepared and stained with two or more stains to reveal its cellular and tissue structures. Techniques based on blind color deconvolution (BCD) allow separation of stains' colors and the structural information (concentrations) of tissues, which is useful for processing, data enhancement or classification.

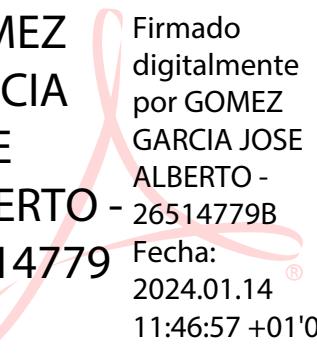
Classical BCD techniques are based on analytical processes of high computational complexity, which must be applied to each of the histological images independently. Deep learning (DL)-based methods allow, once trained, to process unobserved images in a much shorter time. However, studies contemplating the application of deep learning to color-blind deconvolution are scarce, as they encounter a major limiting factor, the lack of large data sets containing "ground truth".

The Deep Image Prior approach argues that the structure of a deep convolutional neural network, by itself, without training of any kind, is capable of capturing a considerable amount of the low-level features of the images, generating images of a quality comparable to those of trained neural networks.

In this work we propose the application of the Deep Image Prior approach to deep learning models based on convolutional neural networks in order to perform BCD on histological images. In this way, we can obtain the benefits of the application of neural networks without requiring large data sets containing "ground truth". To perform color blind deconvolution of a histological image we will require only and exclusively that image.

Three different models will be proposed, employing different architectures, loss functions and input types. These will be tested on the "Warwick Stain Separation Benchmark" data set. The experiments performed indicate that results comparable to those provided by amortized models trained on large data sets can be obtained. A variant of Deep Image Prior is also proposed in which the neural networks are initialized with weights from a previous training in a large dataset. This approach allows to significantly improve the results. This represents the opening of a new line of research, by which we could further improve the results obtained when performing BCD on histological images.

Yo, **José Alberto Gómez García**, alumno de la titulación Máster en Ingeniería Informática de la **Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación de la Universidad de Granada**, con DNI 26514779B, autorizo la ubicación de la siguiente copia de mi Trabajo Fin de Máster en la biblioteca del centro para que pueda ser consultada por las personas que lo deseen.

GOMEZ 
GARCIA
JOSE
ALBERTO -
26514779 Fecha:
B 11:46:57 +01'00'

Fdo: José Alberto Gómez García

Granada a 14 de enero de 2024.

D. **Rafael Molina Soriano**, Profesor del Departamento de Ciencias de la Computación e Inteligencia Artificial de la Universidad de Granada y D. **Fernando Pérez Bueno**, Investigador postdoctoral Juan de la Cierva en el Basque Center on Brain, Cognition, and Language.

Informan:

Que el presente trabajo, titulado *Deconvolución Ciega de Imágenes Histológicas Usando Aprendizaje Profundo*, ha sido realizado bajo su supervisión por **José Alberto Gómez García**, y autorizan la defensa de dicho trabajo ante el tribunal que corresponda.

Y para que conste, expiden y firman el presente informe en Granada a 14 de enero de 2024.

Los tutores:

MOLINA
SORIANO
RAFAEL -
24130240C

Firmado
digitalmente por
MOLINA
SORIANO RAFAEL
- 24130240C
Fecha: 2024.01.13
19:53:08 +01'00'

Rafael Molina Soriano

PEREZ
BUENO
FERNANDO
-
76440946D

Firmado
digitalmente por
PEREZ BUENO
FERNANDO -
76440946D
Fecha: 2024.01.13
18:41:59 +01'00'

Fernando Pérez Bueno

Índice general

1. Introducción	15
1.1. Objetivos	15
1.2. Imágenes histológicas y deconvolución de color	16
1.3. Estado del arte	18
2. Fundamentos teóricos	21
2.1. BCD-Net	21
2.1.1. Modelo bayesiano	21
2.1.2. Mecanismo de inferencia	23
2.1.3. Arquitectura de las redes neuronales	25
2.1.4. Entrenamiento	26
2.2. Deep Image Prior	27
3. Modelos propuestos	31
3.1. Arquitectura de los modelos y mecanismos de inferencia	31
3.1.1. Modelo A. C-Net _{MSE}	32
3.1.2. Modelo B. BCD-Net _{MSE}	33
3.1.3. Modelo C. BCD-Net _{MSE+KL}	34
3.2. Tipos de entrada fija	35
3.3. Entrenamiento de los modelos	36
4. Experimentación realizada	39
4.1. Métricas de rendimiento	39
4.2. Conjunto de datos de entrenamiento	41
4.3. Evolución de las métricas durante el entrenamiento	42
4.4. Comparativa de rendimiento entre modelos	44
4.5. Optimización del entrenamiento	54
5. Conclusiones y trabajo futuro	59
5.1. Conclusiones	59
5.2. Trabajo futuro	61
A. Anexo I	63

B. Anexo II	65
C. Anexo III	71
D. Anexo IV	73

Índice de figuras

1.1. Variación de color entre parches de imágenes histológicas tintadas con H&E en diferentes laboratorios.	17
2.1. Imagen histopatológica original (izquierda) frente a su representación en el espacio de densidad óptica (derecha)	22
2.2. Muestras de $p(m)$ en función de la varianza empleada en la ecuación 2.4	23
2.3. Arquitectura del modelo BCD-Net [38] [40].	25
2.4. Visualización del espacio de la imagen para un proceso de reconstrucción. Comparativa entre el uso convencional de la distribución a priori y Deep Image Prior. Extraído de [34].	29
2.5. Restauración de una imagen con artefactos debidos a la compresión JPEG haciendo uso de Deep Image Prior. Extraído de [34].	30
3.1. Arquitectura del modelo propuesto basado en el uso de C-Net, muestreo aleatorio sobre la matriz de Ruifrok [26], error cuadrático medio y ruido aleatorio como entrada fijada.	33
3.2. Arquitectura del modelo B, basado en el uso de BCD-Net [38] [40], error cuadrático medio y ruido aleatorio como entrada fijada.	34
3.3. Arquitectura del modelo C, basado en el uso de BCD-Net [38] [40], función de pérdida compuesta por error cuadrático medio y divergencia de Kullback-Leibler y ruido aleatorio como entrada fijada.	35
4.1. Subconjunto de imágenes de la base de datos WSSB [3], compuesto por una imagen de cada órgano (pulmón, mama y colon)	42
4.2. Evolución del PSNR para los entrenamientos de las imágenes seleccionadas al partir de ruido aleatorio para los diferentes modelos propuestos.	45
4.3. Evolución del SSIM para los entrenamientos de las imágenes seleccionadas al partir de ruido aleatorio para los diferentes modelos propuestos.	46
4.4. Evolución de los resultados obtenidos para la imagen “Breast_0” al entrenar utilizando el modelo A y ruido aleatorio como entrada.	47
4.5. Evolución de los resultados obtenidos para la imagen “Breast_48” al entrenar utilizando el modelo B y ruido aleatorio como entrada.	48

4.6. Evolución de los resultados obtenidos para la imagen “Breast_0” al entrenar utilizando el modelo C sin pre-entrenamiento y ruido aleatorio como entrada	49
B.1. Evolución del PSNR para los entrenamientos de las imágenes seleccionadas al partir de la imagen observada para los diferentes modelos propuestos.	66
B.2. Evolución del SSIM para los entrenamientos de las imágenes seleccionadas al partir de la imagen observada para los diferentes modelos propuestos. .	67
B.3. Evolución de los resultados obtenidos para la imagen “Breast_0” al entrenar utilizando el modelo A y la imagen observada como entrada.	68
B.4. Evolución de los resultados obtenidos para la imagen “Lung_0” al entrenar utilizando el modelo B y la imagen observada como entrada.	69
B.5. Evolución de los resultados obtenidos para la imagen “Colon_6” al entrenar utilizando el modelo C con pre-entrenamiento y la imagen observada como entrada.	70

Índice de tablas

4.1. Valores medios y desviaciones típicas del PSNR para los modelos propuestos en el capítulo 3 al hacer uso de ruido aleatorio como entrada. En negrita se marcan los mejores resultados de entre los proporcionados por los diferentes modelos.	50
4.2. Valores medios y desviaciones típicas del SSIM para los modelos propuestos en el capítulo 3 al hacer uso de ruido aleatorio como entrada. En negrita se marcan los mejores resultados de entre los proporcionados por los diferentes modelos.	50
4.3. Valores medios y desviaciones típicas del PSNR para los modelos propuestos en el capítulo 3 al hacer uso de la imagen observada como entrada. En negrita se marcan los mejores resultados de entre los proporcionados por los diferentes modelos.	51
4.4. Valores medios y desviaciones típicas del SSIM para los modelos propuestos en el capítulo 3 al hacer uso de la imagen observada como entrada. En negrita se marcan los mejores resultados de entre los proporcionados por los diferentes modelos.	51
4.5. Valores medios y desviaciones típicas del PSNR para el modelo A al emplear la imagen observada como entrada y los modelos B y C inicializados con los pesos de BCD-Net. En negrita se marcan los mejores resultados. .	53
4.6. Valores medios y desviaciones típicas del SSIM para el modelo A al emplear la imagen observada como entrada y los modelos B y C inicializados con los pesos de BCD-Net. En negrita se marcan los mejores resultados. .	53
4.7. Diferencias entre el PSNR máximo obtenido en la época de estabilización (iteraciones 1250 a 2000) y el PSNR máximo para todo el entrenamiento al entrenar los diferentes modelos empleando ruido aleatorio como entrada. .	56
4.8. Diferencias entre el PSNR máximo obtenido en la época de estabilización (iteraciones 250 a 1250) y el PSNR máximo para todo el entrenamiento al entrenar los diferentes modelos empleando la imagen observada como entrada.	57
C.1. Valores medios y desviaciones típicas del PSNR para cada órgano al emplear la variante de Deep Image Prior que inicializa el modelo B con los pesos disponibles de BCD-Net y usa como entrada ruido aleatorio.	71

C.2. Valores medios y desviaciones típicas del PSNR para cada órgano al emplear la variante de Deep Image Prior que inicializa el modelo C con los pesos disponibles de BCD-Net y usa como entrada ruido aleatorio.	72
D.1. Valores medios y desviaciones típicas del PSNR y SSIM para el modelo A al emplear ruido aleatorio y la imagen observada como entrada. En negrita se marcan los mejores resultados de entre los proporcionados por los diferentes modelos.	74
D.2. Valores medios y desviaciones típicas del PSNR y SSIM para el modelo B al emplear ruido aleatorio y la imagen observada como entrada. En negrita se marcan los mejores resultados de entre los proporcionados por los diferentes modelos.	74
D.3. Valores medios y desviación típica del PSNR y SSIM para el modelo C, con etapa de pre-entrenamiento enfocada al color, al emplear ruido aleatorio y la imagen observada como entrada. En negrita se marcan los mejores resultados	75

Capítulo 1

Introducción

1.1. Objetivos

En este Trabajo Fin de Máster se abordará el estudio y mejora de modelos de aprendizaje profundo propuestos en la literatura para llevar a cabo la deconvolución ciega de color (BCD, por sus siglas en inglés) en imágenes histológicas. El objetivo principal es optimizar y perfeccionar los resultados actuales en BCD mediante la combinación del innovador enfoque Deep Image Prior con la arquitectura de redes neuronales BCD-Net, destacada por su prometedor desempeño en trabajos previos [38] [40].

Los objetivos específicos de esta investigación son los siguientes:

- **Estudio de modelos propuestos en la literatura:** Realizar un análisis detallado de la arquitectura BCD-Net para comprender en profundidad su funcionamiento. Además, se abordará una explicación pormenorizada del enfoque de Deep Image Prior, explorando su aplicación para la deconvolución ciega de color.
- **Implementación de Deep Image Prior:** Integrar el enfoque Deep Image Prior dentro de la estructura de BCD-Net con el propósito de explorar cómo esta integración puede potenciar y perfeccionar la deconvolución de imágenes histológicas.
- **Evaluación comparativa:** Realizar una comparativa entre la versión convencional de BCD-Net y las distintas variantes propuestas en este trabajo, las cuales implementan Deep Image Prior. El objetivo es identificar mejoras significativas en la calidad y precisión de la deconvolución.
- **Búsqueda de futuras líneas de investigación:** En base a los resultados experimentales obtenidos se intentarán dilucidar posibles líneas de investigación con la que seguir mejorando la deconvolución ciega de color en imágenes histológicas.

En resumen, el enfoque central estará en comprender en detalle tanto la arquitectura de BCD-Net como el enfoque Deep Image Prior, proponer diversas variantes que combinen ambos conceptos y, finalmente, llevar a cabo una comparativa exhaustiva entre las implementaciones para identificar posibles mejoras y avances en el campo de la deconvolución ciega de color aplicada a imágenes histológicas.

1.2. Imágenes histológicas y deconvolución de color

En los últimos años los casos de cáncer diagnosticados se han incrementado sustancialmente [28]. Para realizar su diagnóstico se suelen utilizar muestras del tejido del paciente, obtenidas mediante una biopsia [7]. Tradicionalmente, cuando se toma una muestra de tejido para su examen se monta en una lámina de vidrio y se visualiza bajo un microscopio por un patólogo [18].

Gracias a los avances recientes, estas láminas de vidrio pueden digitalizarse en imágenes digitales de alta resolución, llamadas “Whole Slide Images” o “imágenes de porta-objetos completo” (WSI, por sus siglas en inglés). Estas imágenes proporcionan una representación integral de la muestra, lo que permite a los patólogos ver y analizar la sección de tejido completa a diversos aumentos [1][18]. Los patólogos pueden hacer zoom, desplazarse por la lámina y navegar por diferentes áreas de interés.

La disponibilidad de este tipo de imágenes ha generado un creciente interés en el desarrollo de herramientas de apoyo al diagnóstico (CAD, por sus siglas en inglés), e incluso en sistemas de diagnóstico automático [1]. Estos avances tecnológicos desempeñan un papel de suma importancia en la lucha contra el cáncer. La implementación de estos sistemas no solo aligera la carga de trabajo de los patólogos, sino que también mejora significativamente la precisión de los diagnósticos, (especialmente en situaciones complejas) y reduce los tiempos de espera, vitales en los tratamientos de este tipo de enfermedades [18].

Estos sistemas de patología computacional (CPATH, por sus siglas en inglés) dependen de grandes cantidades de imágenes, provenientes de WSI, empleadas para entrenarlos [12]. Utilizar imágenes inadecuadas o sin el debido procesamiento puede afectar muy negativamente al desempeño de estos modelos de aprendizaje automático. En la práctica, se ha observado que estos suelen verse severamente afectados cuando se utilizan imágenes de laboratorios u hospitalares que no se encontraban entre los datos del conjunto de entrenamiento [18].

Este hecho se atribuye principalmente a la variación de los colores de las imágenes generadas dentro de un mismo laboratorio, o en diferentes laboratorios. Resulta prácticamente imposible estandarizar las WSI debido a su proceso de adquisición, el cual está sujeto a multitud de variables difícilmente controlables [12][18] que pueden provocar variaciones en la tinción final de los tejidos.

En la figura 1.1 se puede apreciar la diferencia cromática entre diferentes secciones de imágenes histológicas obtenidas en diferentes laboratorios haciendo uso del mismo protocolo de tinción.

Este trabajo se centra en las imágenes resultantes de utilizar hematoxilina y eosina (H&E) para tintar los tejidos, al ser la combinación de compuestos químicos más utilizada en histología actualmente. La hematoxilina resalta los núcleos celulares, ADN y ARN en tonos morados y/o azulados, mientras que la eosina destaca los citoplasmas, proteínas y tejidos conectivos en tonos rosados [4].

Dadas las diferencias de colores que pueden existir, resulta necesario realizar un preprocesamiento de las imágenes para eliminar dichas variaciones de color [12]. Esto es

crucial para obtener modelos que funcionen adecuadamente en tantos centros sanitarios y/o laboratorios como sea posible, incluso si estos no aportaron imágenes al conjunto de entrenamiento.

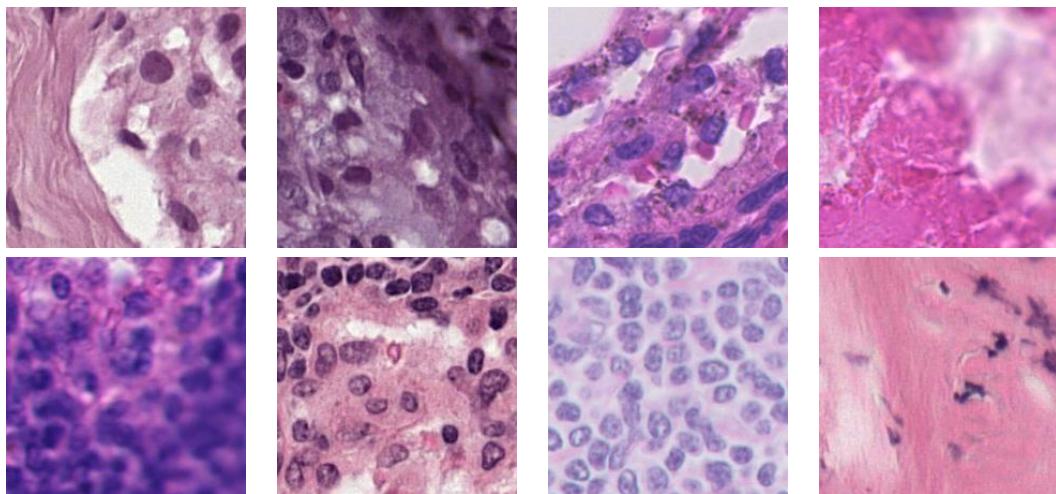


Figura 1.1: Variación de color entre parches de imágenes histológicas tintadas con H&E en diferentes laboratorios.

Para abordar la variación cromática en las WSI se pueden utilizar diferentes enfoques. Estos se pueden clasificar en normalización de color (CN), aumento de colores (CA) y deconvolución ciega de color.

Aquellos basados en la normalización de color buscan ajustar el color de las WSI, de manera que se simula que todas fueron tomadas siguiendo el mismo proceso de tintado y escaneado que una imagen de muestra dada [31]. Las imágenes normalizadas pueden usarse para reducir el error de generalización de sistemas que fueron entrenados sin tener en cuenta las variaciones de color [12]. La principal preocupación relativa a este enfoque es que el proceso de normalización de color debe preservar la estructura histológica contenida en la imagen [27]. Para conseguir esto, muchos trabajos que hacen uso de la normalización de color incluyen un paso previo de deconvolución ciega de color.

Por otra parte, puede usarse un enfoque basado en el aumento de colores [30], como particularización de las técnicas de aumento de datos. Se busca generar diferentes variaciones sintéticas de colores en los datos de entrenamiento, mientras se procura mantener la estructura histológica de la imagen. Todo ello con el objetivo de reducir el error de generalización de los modelos ante combinaciones de colores no observadas con anterioridad [30]. El aumento de colores básicos incluye cambios aleatorio de brillo, contraste, saturación y/o tono de la imagen; lo cual podrá ayudar al modelo a ser invariante a algunas de las variaciones de color introducidas durante el proceso de adquisición de la imagen. El aumento de colores puede combinarse con otras técnicas, por ejemplo, podría realizarse sobre el resultado de una normalización de color [43].

Por último, existe un enfoque basado en la separación de tinciones. Este planteamiento intenta replicar cómo los patólogos analizan las imágenes para identificar diferentes

estructuras celulares. Dado que los tejidos se tintan para facilitar la diferenciación de los diferentes tipos de estructuras, resulta razonable separar una imagen con varias tinciones en varias imágenes independientes [26]. Cada una de ellas sólo tendrá una tinción, y por tanto, tipo de estructura, diferente. Esta separación se realiza estimando el color y la cantidad (concentración) de cada tinción para cada píxel. Serán las concentraciones las que proporcionemos como entrada al CAD [20]. Para realizar esta separación de colores en función de la tinción se suelen utilizar técnicas basadas en la deconvolución ciega de color (BCD).

Será el enfoque de la separación de tinciones haciendo uso de BCD el que seguiremos este trabajo. Este enfoque se suele utilizar como primer paso a la hora de realizar normalización de color y aumento de colores, probando una mejora en los resultados obtenidos [31].

1.3. Estado del arte

En esta sección revisaremos las técnicas que se han propuesto en la literatura para realizar la separación de tinciones, y particularmente, aquellas relacionadas con la deconvolución ciega de color.

En 2001 se publica el trabajo pionero en este ámbito [26], el cual propuso el uso de la ley de Beer-Lambert, el espacio de densidad óptica y una determinada matriz de colores para las tinciones H&E, que obtuvieron experimentalmente a partir de muestras de laboratorio. Esta matriz se ha convertido en un estándar de facto que se usa en muchos trabajos posteriores hasta el día de hoy.

Métodos más recientes abordan las diferencias cromáticas entre muestras haciendo uso de otras técnicas. Por ejemplo, en 2003 se propuso utilizar factorización en matrices no negativas (NMF) [23], que serían complementadas en trabajos posteriores con regularización y términos de dispersión, que representan que cada tinción sólo se adhiere a un determinado tipo de tejido. En 2015 se propuso el uso del análisis de componentes independientes (ICA) [32], trabajo que sería extendido en [3], donde se propuso que las operaciones se realizaran en el dominio wavelet.

Otros trabajos, como [15] sugieren emplear la descomposición en valores singulares (SVD) para separar los canales H&E. También se ha propuesto el uso de técnicas de clustering; en [10] se hace uso del plano cromático de Maxwell para obtener los vectores de color de las tinciones.

Khan et al. proponen en [13] estimar la matriz de colores segmentando la imagen en píxeles de fondo y píxeles pertenecientes a una tinción usando máquinas de vector soporte supervisadas. El color medio de los píxeles de cada clase se utiliza como vector de color para dicha tinción.

En los últimos años, Zheng et al. [44] proponen utilizar la deconvolución usada por Ruifrok [26] como punto de partida y optimizar los vectores de colores y las matrices de concentraciones haciendo uso de una función objetivo basada en conocimiento previo. Salvi et al. presentaron en [27] un método en 3 pasos que utiliza kernels de Gabor, segmentación estructural y una deconvolución final.

También se han propuesto métodos que siguen un enfoque bayesiano. En particular, Hidalgo-Gavira et al [11] proponen una a priori de similitud entre los vectores de color con una referencia dada, así como una a priori basada en un modelo de suavizado auto-regresivo simultáneo para el cálculo de cada concentración. Este trabajo fue extendido en [20], donde se propuso el uso de una a priori basada en modelos de variación total. En [22], Pérez-Bueno et al. propusieron una a priori que hace uso de distribuciones súper gaussianas en un filtro paso alto durante el cálculo de las concentraciones. En [21], Pérez-Bueno et al. presentan un modelo bayesiano basado en el algoritmo de descomposición en K valores singulares (K-SVD), que plantea el problema como uno de aprendizaje de diccionarios.

Los métodos comentados hasta el momento no son amortizados. Es decir, para cada imagen se deben recalcular todos los parámetros del modelo para poder generar las estimaciones de la matriz de color y la matriz de concentraciones. Este proceso es sumamente costoso computacionalmente, lo que deriva en largos tiempos de procesamiento.

En un intento de reducir estos largos tiempos de cómputo, en los últimos años se ha propuesto el uso de técnicas de aprendizaje profundo (Deep Learning, DL). Estos modelos son amortizados, y por tanto proporcionan tiempos de procesamiento rápidos para cada imagen, pero tienen la desventaja de que requieren grandes conjuntos de datos para su entrenamiento.

En este ámbito, el aprendizaje profundo se ha utilizado mayoritariamente para abordar la normalización de color sin aplicar una deconvolución de color previa, en tanto que es difícil tener “ground truth” para la separación de tinciones. Sin embargo, existen trabajos que utilizan técnicas de Deep Learning para abordar la deconvolución ciega de color.

Duggal et al. [8] propusieron una capa de deconvolución de tinciones que opera en el espacio de densidad óptica a la entrada de redes neuronales convolucionales (CNN). Según los autores, esta capa proporciona una mejor representación de la interacción entre tejido y tinción, mejorando los resultados obtenidos. Los parámetros de la capa de deconvolución emulan la matriz de vectores de color y se optimizan durante el entrenamiento (inicializándose mediante lo propuesto en [15]). Sin embargo, este enfoque no aborda la variabilidad cromática entre las imágenes y requiere la normalización previa del conjunto de datos antes del entrenamiento.

De manera similar, en [16] Marini et al. propusieron una arquitectura de CNN para aprender características invariantes de las tinciones mediante la estimación de la matriz de color (utilizando lo propuesto en [15] como “ground truth”) y una etiqueta de clasificación de forma conjunta.

Zheng et al. [43] propusieron una “red cápsula” que produce varias separaciones de tinciones usando operaciones de convolución 1x1 y forma una salida basándose en una restricción de dispersión. La deconvolución se realiza a partir de los parámetros de la red, lo que hace que esta no se pueda adaptar a distribuciones de color no observadas con anterioridad.

Ninguno de los enfoques basados en aprendizaje profundo que acaban de ser mencionados evalúan la calidad de la separación de las tinciones después de haber realizado

la deconvolución ciega de color.

El trabajo realizado por Abousamra et al. [2] utiliza un autoencoder para la separación de tinciones en imágenes de inmunohistoquímica tintadas con seis tinciones diferentes. El autoencoder es entrenado con etiquetas de puntos colocadas manualmente como mecanismo de supervisión débil. Desafortunadamente, este trabajo no se puede extender a otros sistemas de tinción, como H&E, sin un conjunto de datos etiquetados, del cual no se suele disponer.

Uno de los aportes más recientes a este campo ha sido el modelo BCD-Net, propuesto en [38] y ampliado en [40]. Este modelo combina técnicas de deconvolución de imágenes ruidosas mediante aprendizaje profundo [41][42], y modelos analíticos empleados en BCD [11][20][22], para proponer un modelo amortizado que no requiere de los “ground truth” de los vectores de color ni de las matrices de concentración para su entrenamiento, los cuales no suelen estar disponibles.

BCD-Net ha demostrado un rendimiento prometedor en la separación de tinciones en diferentes tipos de tejidos, así como en la clasificación del cáncer de mama. BCD-Net es competitivo con los métodos más utilizados y estudiados en la literatura, a la vez que mejora significativamente el tiempo de cálculo requerido por los modelos no amortizados [38] [40].

Es por estos motivos, y por ser uno de los pocos trabajos que aboga por el procesamiento interpretable de imágenes histológicas mediante Deep Learning, que decidimos utilizar BCD-Net como arquitectura base para este trabajo.

En el siguiente capítulo abordaremos con mayor detalle los fundamentos teóricos en los que se basa este trabajo, la arquitectura BCD-Net y el enfoque Deep Image Prior.

Capítulo 2

Fundamentos teóricos

En este capítulo se abordaran los detalles teóricos en los que se basa este trabajo. En primer lugar se detallará el modelo BCD-Net, el cual combina técnicas de deconvolución de imágenes ruidosas mediante Deep Learning y modelos analíticos empleados en BCD. Este modelo demuestra un rendimiento prometedor en la separación de tinciones en diferentes tipos de tejidos, así como en la clasificación del cáncer de mama. Posteriormente, se explicará el enfoque Deep Image Prior, el cual defiende que la estructura de una red neuronal convolucional profunda puede capturar una cantidad considerable de las características de bajo nivel de las imágenes, de manera que las redes neuronales no necesitan estar entrenadas en grandes conjuntos de datos para ofrecer buenos resultados.

2.1. BCD-Net

En este apartado trataremos en detalle el modelo de aprendizaje profundo propuesto en [38] y ampliado en [40], BCD-Net, ya que será la arquitectura base cuyos resultados intentaremos mejorar a lo largo de este trabajo. El objetivo final será el de obtener un modelo que realice una separación de tinciones de mayor calidad, respetando la estructura del tejido.

2.1.1. Modelo bayesiano

Este modelo primer realiza una conversión del espacio de color RGB al espacio de densidad óptica (logarítmicamente inverso al RGB). En la figura 2.1 pueden verse un parche de una imagen histológica en el espacio de color RGB y el de densidad óptica.

El uso del espacio de densidad óptica (OD) posibilita la aplicación de la ley de Beer-Lambert [26], que establece una relación lineal entre la intensidad observada y la concentración de cada tinción; es decir, permite representar la absorción de un elemento óptico por unidad de distancia para una longitud de onda dada [35].

La ley de Beer-Lambert establece que cada píxel y^k de una imagen Y en el espacio de densidad óptica sigue la ecuación 2.1. Para facilitar el procesamiento, la imagen se

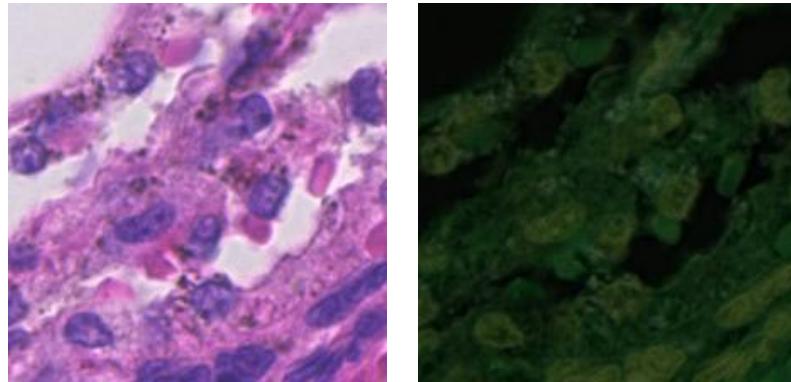


Figura 2.1: Imagen histopatológica original (izquierda) frente a su representación en el espacio de densidad óptica (derecha)

distribuye como una matriz de 3 filas, de manera que en cada fila tendremos los valores de cada canal para cada píxel.

$$y^k = M c^k + N \quad (2.1)$$

donde $\mathbf{M} \in \mathbb{R}^{3 \times S}$ representa la matriz de colores, $\mathbf{c}^k \in \mathbb{R}^S$ equivale a la concentración de la tinción en el k -ésimo píxel y $N \in \mathbb{R}^{3 \times 1}$ es un vector de ruido. S es el número de tinciones empleadas. El valor de la densidad óptica correspondiente a cada canal está relacionado linealmente con la concentraciones y con las tinciones; por lo tanto, las tinciones de la muestra se puede separar en el espacio OD siguiendo la ecuación

$$\mathbf{y}^k = \sum_{s=1}^S c_s^k \mathbf{m}_s + \epsilon \quad (2.2)$$

en la que $\mathbf{m}_s \in \mathbb{R}^{3 \times S}$ es el vector de color asociado a la s -ésima tinción y c_s^k es la concentración de la s -ésima tinción para el píxel k . De ahora en adelante denotaremos como $\mathbf{C} \in \mathbb{R}^{S \times HW}$ a la matriz con la concentración para todos los píxeles, siendo H y W el número de píxeles vertical y horizontalmente en la imagen original.

En [38] [40] se considera el siguiente modelo generativo a partir de distribuciones de probabilidad:

$$p(C, M, Y) = p(C)p(M)p(Y|M, C). \quad (2.3)$$

A la hora de definir las distribuciones a priori, $p(C)$ y $p(M)$, sería interesante seguir un enfoque basado en los datos, de manera que la información que nos proporcionen dichas distribuciones sea lo más real y relevante posible. Sin embargo, no existen grandes conjuntos de datos con “ground truth” para las concentraciones de cada tinción ni el color de las tinciones, lo que complica definir las distribuciones a priori.

Para $p(C)$ podrían usarse distribuciones a priori que consideren información general de las concentraciones, como se propone en [11] [20] [22]. Sin embargo, esto complicaría el modelo, así que se decide utilizar $p(C)$ constante por simplicidad.

Para $p(M)$ se elige la distribución a priori de la ecuación 2.4. Esta decisión se toma en base a que el protocolo de tinción de H&E es conocido y se acepta de forma general que los colores suelen ser cercanos a los proporcionados por la matriz de referencia de Ruifrok [26], por lo que podemos considerar como combinaciones válidas aquellas con una cierta variación respecto de dicha matriz.

$$p(\mathbf{M}) = \prod_{s=1}^{N_s} p(\mathbf{m}_s) = \prod_{s=1}^{N_s} \mathcal{N} \left(\mathbf{m}_s \mid \mathbf{m}_s^{\text{Rui}}, (\sigma_s^{\text{Rui}})^2 \mathbf{I} \right) \quad (2.4)$$

De la ecuación 2.4 es importante destacar la varianza, dado que este valor controla la cantidad de variación permitida. En función del valor empleado (véase la figura 2.2) restringiremos las combinaciones de colores que se pueden generar. Valores bajos de varianza fuerzan a la combinación generada a parecerse a la matriz de referencia, mientras que valores demasiado altos podrían generar combinaciones alejadas de la realidad, y por tanto inútiles. En [38] [40] se emplea como valor óptimo $\sigma = 0.05$.

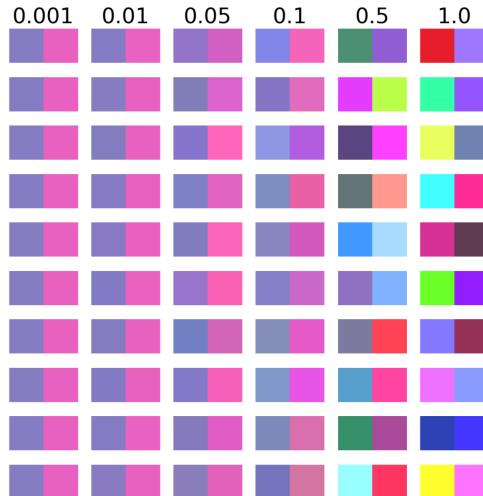


Figura 2.2: Muestras de $p(m)$ en función de la varianza empleada en la ecuación 2.4

Finalmente, el modelo de observación dado por la Ley de Beer-Lambert puede ser reescrito como sigue:

$$p(\mathbf{Y} \mid \mathbf{M}, \mathbf{C}) \propto \exp \left(-\frac{1}{2\lambda_n^2} \left\| \mathbf{Y}^\top - \mathbf{MC} \right\|_F^2 \right) \quad (2.5)$$

donde $\|\cdot\|_F$ denota la norma de Frobenius y λ_n^2 la varianza del ruido.

2.1.2. Mecanismo de inferencia

Recordemos que el objetivo final es realizar una estimación de las matrices C y M para cada observación Y . Para ello se necesita calcular la distribución a posteriori $p(C, M|Y)$.

Sin embargo, esta expresión no se puede calcular de forma analítica, por lo que debemos realizar aproximaciones basadas en inferencia bayesiana. En particular, utilizaremos

$$q(C, M|Y) = q_\alpha(C|Y)q_\beta(M|Y) \quad (2.6)$$

En [38] [40] se decide utilizar dos redes neuronales profundas, llamadas C-Net y M-Net; cuyos parámetros se denotan por α y β respectivamente. La primera de las redes se encargará de estimar $q_\alpha(C|Y)$, correspondiente a la matriz de concentraciones. La segunda red calculará $q(M|Y)$, correspondiente a la matriz de colores.

Las salidas de ambas redes se emplearán para generar una imagen con dos canales, estando en uno de ellos la información correspondiente a la concentración y tinción de la parte del tejido teñida de hematoxilina, mientras que en el otro canal se encontrará la separación relativa a la eosina.

Ambas redes neuronales se entrena para maximizar el límite inferior de la evidencia de las observaciones (ELBO). La fórmula matemática de dicha función de pérdida dada una imagen Y perteneciente a la base de datos \mathcal{Y} queda descrita en [2.7]

$$\begin{aligned} \text{ELBO}(\mathbf{Y}) = & -\underbrace{\mathbb{E}_{q_\alpha(\mathbf{C}|\mathbf{Y})} \left[\log \frac{q_\alpha(\mathbf{C}|\mathbf{Y})}{p(\mathbf{C})} \right]}_{A_1} - \underbrace{\mathbb{E}_{q_\beta(\mathbf{M}|\mathbf{Y})} \left[\log \frac{q_\beta(\mathbf{M}|\mathbf{Y})}{p(\mathbf{M})} \right]}_{A_2} \\ & - \underbrace{\frac{1}{2\lambda_n^2} \mathbb{E}_{q_\beta(\mathbf{M}|\mathbf{Y})} \left[\left\| \mathbf{Y}^\top - \mathbf{MC}^\alpha(\mathbf{Y}) \right\|_F^2 \right]}_{A_3} + \text{const.} \end{aligned} \quad (2.7)$$

El término A_1 se ignora al ser infinito, dado que $p(C)$ es impropia y q_α es degenerada.

El término A_2 corresponde al valor negativo de la divergencia de Kullback-Leibler entre las distribuciones gaussianas $p(M)$ y $q_\beta(M|Y)$. Este puede reescribirse como:

$$A_2 = \mathcal{L}_{\text{KL}}^{\beta}(\mathbf{Y}) = \frac{1}{2} \sum_{s=1}^S \frac{\left\| \boldsymbol{\mu}_s^{\beta}(\mathbf{Y}) - \mathbf{m}_s^{\text{Rui}} \right\|^2}{\gamma_s^{\text{Rui}}} + \frac{3}{2} \sum_{s=1}^S \left(\frac{\sigma_s^{\beta}(\mathbf{Y})^2}{\gamma_s^{\text{Rui}}} - \log \frac{\sigma_s^{\beta}(\mathbf{Y})^2}{\gamma_s^{\text{Rui}}} - 1 \right) \quad (2.8)$$

En [38] [40], este término se emplea para medir la diferencia entre la distribución de color a posteriori y la distribución de color a priori. Durante el entrenamiento de la red, buscaremos que $q(M)$ se parezca a la matriz de referencia de Ruifrok [26] admitiendo una determinada varianza.

Por último, el término A_3 de la ecuación [2.7] es aproximado de la siguiente manera: $A_3 \approx -0.5\lambda_n^{-2}\mathcal{L}_{\text{MSE}}^{\alpha,\beta}(\mathbf{Y})$, siendo:

$$\mathcal{L}_{\text{MSE}}^{\alpha,\beta}(\mathbf{Y}) = \frac{1}{N_M} \sum_{i=1}^{N_M} \left[\left\| \mathbf{Y}^\top - \mathbf{MC}_i^{\beta}(\mathbf{Y}) \right\|_F^2 \right] \quad (2.9)$$

Esta aproximación se utiliza dado que durante el entrenamiento llevado a cabo en [40] se observó que utilizar el truco de la reparametrización proporcionaba mejores resultados que su expresión en forma cerrada.

La componente $\mathcal{L}_{\text{MSE}}^{\alpha,\beta}(\mathbf{Y})$ representa el error cuadrático medio (MSE, por sus siglas en inglés) entre la imagen original y la reconstrucción generada. El MSE se utiliza para cuantificar la diferencia píxel por píxel entre dos imágenes, y por tanto, nos permitirá medir la calidad de la reconstrucción de la imagen observada de acuerdo con su modelo de observación.

Con ello llegamos a la ecuación 2.10, empleada como función de pérdida en la práctica a partir de la ecuación 2.7 para el conjunto de imágenes \mathcal{Y} . En dicha expresión se emplea un valor θ para ponderar la importancia que se da a cada uno de los términos. En [38] [40] se realiza un estudio del valor óptimo para este parámetro, fijándose el mismo en 0.3 de acuerdo a los resultados experimentales obtenidos.

$$\mathcal{L}(\mathcal{Y}) = \sum_{\mathbf{Y} \in \mathcal{Y}} \left[\theta \mathcal{L}_{\text{KL}}^{\beta}(\mathbf{Y}) + (1 - \theta) \mathcal{L}_{\text{MSE}}^{\alpha,\beta}(\mathbf{Y}) \right] \quad (2.10)$$

2.1.3. Arquitectura de las redes neuronales

Después de discutir la función de pérdida empleada en el entrenamiento de la red, hablaremos ahora sobre la arquitectura de las redes neuronales que conforman BCD-Net. En la figura 2.3 se ilustra el modelo.

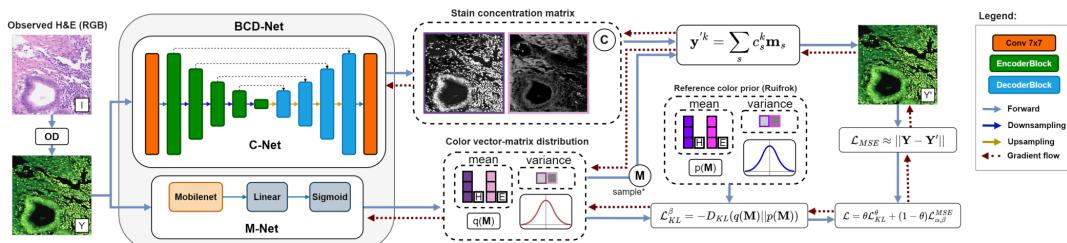


Figura 2.3: Arquitectura del modelo BCD-Net [38] [40].

La red neuronal C-Net tiene forma de U-Net. Tanto el codificador como el decodificador tienen cuatro bloques, cada uno de ellos compuestos por ResBlocks que emplean LeakyRelu como función de activación. Los bloques de submuestreo utilizan capas convolucionales 3x3, mientras que los sobremuestreadores utilizan convoluciones transpuestas 5x5. Se utiliza “skip-connection” para permitir que el sobremuestreador pueda ver la entrada de fases anteriores. Esta red proporciona como salida una imagen de dos canales, uno para la concentración de hematoxilina y otro para la concentración de eosina. El uso de redes con arquitectura U-Net para segmentación de imágenes biológicas se propuso y usó con anterioridad en otros trabajos, como [25], [29] y [42].

La red neuronal M-Net utiliza un backbone MobileNet V3 pequeño y una capa totalmente conectada para estimar la media y la varianza de distribución a posteriori

$q_\beta(M|Y)$. La elección de esta red está motivada por el reducido tamaño de la matriz de color.

2.1.4. Entrenamiento

Para llevar acabo el entrenamiento de BCD-Net se hizo uso del conjunto de datos CAMELYON17, empleado por primera vez como parte del reto CAMELYON17 [5], consistente en la detección de metástasis de cáncer en los ganglios linfáticos de las mamas. Este conjunto de datos contiene 1000 WSI, procedentes de cinco centros médicos diferentes de los Países Bajos.

Al igual que en [21, 44], se utilizaron las 500 WSI (100 por centro) del conjunto de entrenamiento para extraer 500 parches de cada una de ellas. Estos parches no se superponen entre sí, son de tamaño 224x224 píxeles y contienen al menos un 70 % de tejido. No se dispone de “ground truth” para ninguna de las imágenes del conjunto de datos. BCD-Net se entrenó haciendo uso de 60.000 parches provenientes de los centros 0, 2 y 4, empleándose para validación las imágenes de los dos centros restantes.

Para validar los resultados se empleó la base de datos Warwick Stain Separation Benchmark (WSSB) [3]. Este es un conjunto de datos compuesto por 24 imágenes con tinciones H&E de tejidos de diferentes órganos, concretamente, de mama, colon y pulmón. Estas imágenes proceden de diferentes laboratorios y han sido capturadas empleando diferentes microscopios, por lo que se presentan varianza tanto intra como inter-laboratorio. Las imágenes RGB “ground truth” fueron generadas a partir de vectores de color determinados por expertos patólogos; siendo las concentraciones calculadas haciendo uso de la ley de Beer-Lambert.

Los resultados obtenidos al entrenar BCD-Net son destacables por su capacidad para alcanzar un equilibrio óptimo entre precisión y tiempo de ejecución. Esto lo posiciona como una alternativa atractiva cuando se busca un procesamiento eficiente con resultados de calidad aceptable en el ámbito de imágenes histológicas.

Una de sus mejoras significativas es su naturaleza amortizada: se entrena una vez y luego se puede aplicar a multitud de imágenes histológicas no incluidas en los conjuntos de datos originales. Otra ventaja de BCD-Net es que su entrenamiento no requiere del “ground truth” de la separación de tinciones de las imágenes, el cual no suele estar disponible. Además, no es necesaria de una tarjeta gráfica para ejecutar el modelo, en tanto que su ejecución en un procesador de propósito general se realiza en un tiempo asumible y similar al que tomaría ejecutar otros modelos de rendimiento similar en una tarjeta gráfica.

Aunque BCD-Net ha demostrado ser un modelo prometedor, creemos que aún puede mejorar. Con el objetivo de intentar obtener mejores resultados, probaremos a aplicar el enfoque Deep Image Prior, cuyos fundamentos teóricos se detallan en la siguiente sección.

2.2. Deep Image Prior

En esta sección, exploraremos una perspectiva divergente en el uso de redes neuronales para el procesamiento de imágenes: el enfoque Deep Image Prior (DIP) propuesto por Ulyanov et al [34]. Su principal singularidad radica en la capacidad de funcionar sin depender de grandes conjuntos de datos durante el proceso de entrenamiento.

Las redes neuronales convolucionales llevan años utilizándose para tareas de procesamiento de imágenes; tales como su restauración o generación sintética, ya sea desde cero o en tareas de super-resolución (generar imágenes de mayor resolución dada una imagen de partida). Normalmente, se requiere del uso de un gran número de imágenes durante el entrenamiento del modelo para que este sea capaz de adquirir conocimiento a priori.

Sin embargo, Ulyanov et al [34] defienden que la estructura de una red neuronal convolucional profunda, por sí misma, sin entrenamiento de ningún tipo, es capaz de capturar una cantidad considerable de las características de bajo nivel de las imágenes, generando imágenes de una calidad comparable a las de redes neuronales entrenadas [33].

En lo que sigue, emplearemos x para referirnos a la imagen original sin alteraciones, que es el objetivo del proceso de restauración; \hat{x} hace referencia a la misma imagen después de haber sufrido algún tipo de deterioro, siendo esta la versión comúnmente disponible para el procesamiento. Por otro lado, x^* denota la imagen restaurada que generamos como resultado de nuestro tratamiento. La imagen deteriorada \hat{x} se obtuvo al aplicar un operador de degradación específico, representado por la función $d(\cdot)$, sobre la imagen original x , matemáticamente expresado como $\hat{x} = d(x)$. Estos operadores de degradación pueden abarcar la introducción de ruido, pérdida de valores en ciertos píxeles, entre otros efectos.

En el contexto típico del uso de redes neuronales, se suministra a la red un conjunto de imágenes degradadas junto con sus versiones originales correspondientes, estas últimas libres de degradación. Esto permite que la red neuronal extraiga y aprenda el conocimiento a priori necesario para llevar a cabo el proceso de restauración de las imágenes.

Alternativamente, podemos utilizar conocimiento a priori explícito, diseñado a mano y expresado como un problema de optimización donde se busca conseguir una imagen x que minimice la función $\|d(x) - \hat{x}\|$, sujeto a que la imagen x sea una cara, un paisaje, etc. Este segundo enfoque es complicado de representar matemáticamente, por lo que recurrimos a otras aproximaciones, como que la imagen sea suave, las diferencias entre píxeles tengan unas determinadas características, etc.

Podemos definir el proceso de restauración de la imagen degradada como la búsqueda de una imagen x^* tal que maximice la distribución a posteriori $p(x | \hat{x})$. Este concepto queda matemáticamente definido en la ecuación 2.11.

$$x^* = \arg \max_x p(x | \hat{x}) \quad (2.11)$$

Usando la regla de Bayes tenemos la siguiente aproximación de la distribución a

posteriori $p(x | \hat{x})$:

$$p(x | \hat{x}) = \frac{p(\hat{x} | x)p(x)}{p(\hat{x})} \propto p(\hat{x} | x)p(x) \quad (2.12)$$

siendo $p(\hat{x} | x)$ la verosimilitud y $p(x)$ la distribución a priori explícita.

Combinando las ecuaciones 2.11 y 2.12 se obtiene la ecuación:

$$x^* = \arg \max_x p(\hat{x} | x)p(x) \quad (2.13)$$

Sin embargo, es común emplear una notación alternativa usando un mínimo en lugar del máximo. Además, en lugar de trabajar con distribuciones de probabilidad explícitamente, empleamos una notación que nos permite tratar tareas como la eliminación de ruido o “in-painting” como problemas de minimización de energía. Así pues, $E(x; \hat{x})$ es un término de datos dependiente de la tarea y $R(x)$ es un regularizador. En la ecuación 2.14 se desarrollan estas equivalencias.

$$\begin{aligned} x^* &= \arg \max_x p(\hat{x} | x)p(x) = \arg \min_x -\log[p(\hat{x} | x)p(x)] \\ &= \arg \min_x -\log p(\hat{x} | x) - \log p(x) \\ &= \arg \min_x E(x; \hat{x}) + R(x) \end{aligned} \quad (2.14)$$

En [34] se propone que, en lugar de realizar una búsqueda o trabajo de optimización en el espacio de la imagen, realicemos dicha tarea en un espacio de parámetros diferente. Dicho espacio de parámetros queda denotado como θ . Por lo tanto, necesitaremos una función g que establezca correspondencia entre el espacio de parámetros θ y el espacio de la imagen.

Si g es una función sobreyectiva, es decir, $\forall x \exists \theta : g(\theta) = x$, entonces ambos problemas son equivalentes en la teoría. Sin embargo, al realizar este cambio pasamos a tratar con un problema de optimización local, por lo que la función utilizada importa, en tanto que en la práctica las soluciones serán diferentes dependiendo de la función g escogida.

La idea es que si la función g está diseñada para favorecer el tipo de imágenes que deseamos obtener, esta función actúa como conocimiento a priori por sí misma. Por tanto, podríamos optar por minimizar únicamente el término E dependiente de los datos y tarea a realizar, obviando el regularizador R , cuyo óptimo es difícil de formular. Así pues, trataremos a la función g como un hiper-parámetro más que deberemos ajustar.

Si se realiza la reparametrización que se acaba de mencionar, se obtiene la ecuación 2.15.

$$x^* = \arg \min_{\theta} E(g(\theta); \hat{x}) + R(g(\theta)) \quad (2.15)$$

La reparametrización de la ecuación 2.15 implica elegir una función g . El enfoque Deep Image Prior [34] propone establecer la equivalencia $g(\theta) \equiv f_{\theta}(z)$, siendo f una red neuronal convolucional con una serie de parámetros θ . Así pues el espacio de θ será el espacio de los parámetros de la red. Dada una red neuronal cualquiera, tenemos infinitas funciones, una para cada combinación de parámetros θ de la red.

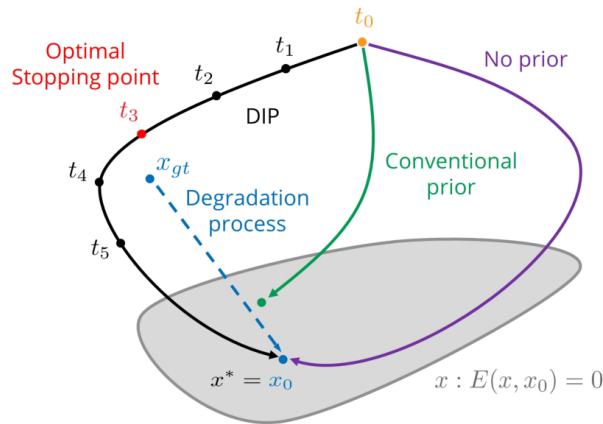


Figura 2.4: Visualización del espacio de la imagen para un proceso de reconstrucción. Comparativa entre el uso convencional de la distribución a priori y Deep Image Prior. Extraído de [34].

En [34] se decidió dejar la entrada de la red fija con una inicialización cualquiera compatible con la salida, por ejemplo, ruido gaussiano de mismas dimensiones que la imagen objetivo, la cual se sigue deseando que sea la reconstrucción de una imagen “corrupta”. A diferencia del enfoque convencional, en el que se fijan los pesos y se proporcionan distintas entradas para obtener diferentes salidas; el enfoque Deep Image Prior [34] opta por fijar la entrada e ir variando los pesos para obtener diferentes salidas.

El tener un conocimiento a priori, es decir, una red neuronal afín a la tarea que queremos realizar, hará que se obtenga con mayor sencillez la salida deseada, evitando atascarse en mínimos locales no deseados durante el entrenamiento.

La clave de este enfoque reside en que, con el tiempo y número de iteraciones suficientes, puede reconstruir cualquier imagen deseada dada una entrada cualquiera. Si se marca como objetivo una imagen degradada, acabará llegando a reconstruir dicha imagen degradada a la perfección; pero por el camino habrá pasado por una combinación de parámetros que generaría la salida sin degradación que deseamos. Solamente después de haber aprendido los componentes estructurales básicos de la imagen se comenzará a introducir en ella la degradación que se pretende eliminar [33], por lo que realizar una parada temprana del proceso de entrenamiento es fundamental en este caso.

En la imagen 2.4 se ilustra este comportamiento. La solución generada por Deep Image Prior (en negro) en el instante t_3 sería la mejor reconstrucción que este método nos puede ofrecer. De no aplicar una parada temprana y seguir entrenando, la calidad de la reconstrucción irá empeorando, hasta eventualmente igualar a la imagen degradada x_0 . En la imagen 2.5 se ilustra el proceso de restauración de una imagen con artefactos debidos a la compresión JPEG. Como podemos observar, a medida que el proceso avanza se puede recuperar la mayor parte de la señal a la vez que se eliminan los halos y

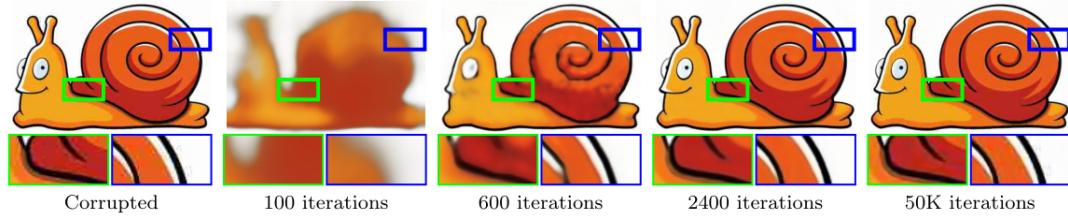


Figura 2.5: Restauración de una imagen con artefactos debidos a la compresión JPEG haciendo uso de Deep Image Prior. Extraído de [34].

artefactos (alrededor de las 2.400 iteraciones) antes de llegar a sobreajustar una vez más a la entrada deteriorada (en 50.000 iteraciones).

En el capítulo 3 propondremos nuestro modelo para la separación de tinciones en imágenes histológicas, el cual estará basado en la arquitectura de redes neuronales BCD-Net y el enfoque Deep Image Prior. Posteriormente, en el capítulo 4 se abordará la experimentación realizada y los resultados obtenidos.

Capítulo 3

Modelos propuestos

En este capítulo se presentarán y discutirán los modelos propuestos para mejorar los resultados obtenidos por BCD-Net al realizar la deconvolución ciega de color en imágenes histológicas.

Como se mencionó en el capítulo 2, el enfoque Deep Image Prior se suele aplicar a la restauración de imágenes, por lo que parece posible aplicarlo a la separación de tinciones en imágenes histológicas.

Partiremos de la premisa de que podemos entender las imágenes histológicas de las que se dispone como imágenes que han sufrido un cierto proceso de degradación, en tanto que provendrían de unas imágenes en las que las tinciones (y sus concentraciones) se encontraban separadas en canales distintos. Son estas imágenes las que precisamente se desea reconstruir, ya que no disponemos de ellas.

Para poder emplear el enfoque Deep Image Prior debe fijarse la arquitectura de red neuronal a emplear, cuyos pesos tendrán una inicialización aleatoria cualquiera, y una única imagen de entrada a la red que se mantendrá fija durante todo el entrenamiento, y que denotaremos de ahora en adelante como Z . Se suele utilizar una entrada compuesta por ruido aleatorio uniforme, aunque son posibles otras opciones, que se discuten más adelante, en la sección 3.2. Comenzaremos discutiendo la arquitectura de los modelos propuestos para realizar la deconvolución ciega de color.

3.1. Arquitectura de los modelos y mecanismos de inferencia

Respecto de la arquitectura de red neuronal base a utilizar, en este trabajo se decide utilizar BCD-Net, la cual está compuesta por dos redes neuronales, C-Net y M-Net para estimar las matrices de concentraciones y los vectores de color, respectivamente. En 34 se mencionan diversos métodos de reconstrucción de imágenes, en los que se deben estimar los valores de los píxeles de la imagen para poder generar la reconstrucción. Sin embargo, nuestra problemática es algo más compleja en tanto que deseamos reconstruir los canales de la imagen (en este caso dos, H y E) por separado, y los valores de cada

canal dependen a su vez de dos componentes, una relacionada con la concentración de la tinción (matriz C) y otra con el color de la misma (matriz M).

Se expondrán tres modelos, detallados a continuación: (i) el Modelo A, también conocido como C-Net_{MSE}; (ii) el Modelo B, denominado BCD-Net_{MSE}; y (iii) el Modelo C, identificado como BCD-Net_{MSE+KL}.

Así pues, en un primer lugar buscaremos confirmar que la conjunción de Deep Image Prior y BCD-Net es capaz de generar matrices de concentraciones adecuadas a partir de una entrada fija conformada por ruido aleatorio muestreado de una distribución uniforme. Para ello, emplearemos el modelo A, descrito en el siguiente apartado.

3.1.1. Modelo A. C-Net_{MSE}

Este primer modelo estará compuesto únicamente por la red neuronal C-Net, cuya función es generar predicciones de las matrices de concentraciones. Sin embargo, para poder generar la reconstrucción de la imagen, que será empleada por la función de pérdida a lo largo del proceso de optimización del modelo, necesitamos también de una matriz de color M . Asumimos que M puede muestrearse directamente de la distribución a priori introducida en la ecuación 2.4 donde $\mathbf{m}_s^{\text{Rui}}$ proviene de la matriz de referencia de Ruifrok [26] y σ_s^{Rui} representa la desviación típica máxima que admitiremos en el proceso de muestreo aleatorio. Siguiendo las conclusiones expuestas en [40] se decide emplear $\sigma_s^{\text{Rui}} = 0.05$, de manera que se obtengan variaciones razonables y verosímiles respecto de la referencia de Ruifrok [26].

Definida esta primera arquitectura, debemos establecer también la función de pérdida que empleará, en tanto que es la principal encargada de dictar el grado en que se deberán cambiar los pesos de la red.

Dado que la red neuronal empleada se limita a generar una predicción de las concentraciones, mientras que los colores son muestreados aleatoriamente a partir de una referencia, optaremos por no emplear un término de regularización basado en el color (\mathcal{L}_{KL}). Esto carecería de sentido en tanto que dicho muestreo aleatorio no se ve influenciado de ninguna manera por los parámetros de la C-Net ni la función de pérdida que dicta su modificación.

Recordemos también que al seguir el enfoque Deep Image Prior los pesos de la red se optimizan en función de una única imagen de entrada fija durante todo el entrenamiento, no se dispone de un dataset con varias imágenes.

Así pues, la función de pérdida empleada por este modelo durante su entrenamiento queda definida por la ecuación 3.1. Esta representa el error cuadrático medio entre la imagen observada y la reconstrucción generada. Nótese que esta ecuación es una variación de 2.10 que tiene en cuenta las consideraciones expuestas.

En 3.1 emplearemos el término Z para hacer referencia a la imagen de entrada de la red y fija durante todo el proceso de entrenamiento. Y' representará la imagen reconstruida a partir de la salida de la red neuronal y la matriz M muestreada aleatoriamente. El término Y denotará la imagen observada y cuya reconstrucción queremos generar, representada en el espacio de densidad óptica. N_M denota el número de muestras de los vectores de color.

$$\mathcal{L}(Z) = \mathcal{L}_{\text{MSE}}^{\alpha}(\mathbf{Z}) = \frac{1}{N_M} \sum_{i=1}^{N_M} \left[\left\| \mathbf{Y}^T - \mathbf{M}_i \mathbf{C}^{\alpha}(\mathbf{Z}) \right\|_F^2 \right] \quad (3.1)$$

Este primer modelo propuesto, C-Net_{MSE}, queda descrito gráficamente en la figura 3.1 en el caso de emplear ruido aleatorio como entrada. Podría también emplearse como entrada la imagen observada, representada en el espacio de densidad óptica.

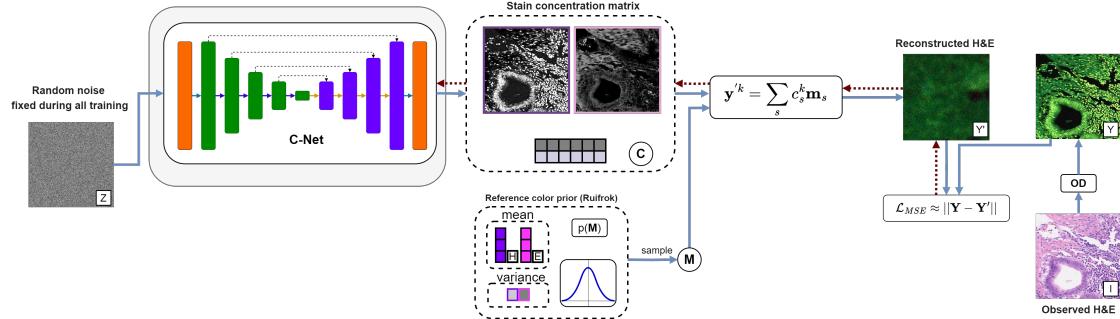


Figura 3.1: Arquitectura del modelo propuesto basado en el uso de C-Net, muestreo aleatorio sobre la matriz de Ruifrok [26], error cuadrático medio y ruido aleatorio como entrada fija.

Para construir los siguientes modelos, B y C, ampliaremos la arquitectura de la red neuronal descrita en este apartado.

3.1.2. Modelo B. BCD-Net_{MSE}

En este apartado ampliaremos la arquitectura de la red propuesta, incluyendo el uso de la red neuronal M-Net, lo cual nos permitirá estimar la matriz de color, pudiendo prescindir del muestreo aleatorio sobre la referencia de Ruifrok [26] empleada anteriormente. De esta manera, la imagen reconstruida será generada a partir de información predicha en su totalidad por BCD-Net.

En esta ocasión se empleará la ecuación 3.2 como función de pérdida. Esta es una ligera variación de la ecuación 3.1, en la que el error de reconstrucción influye en el entrenamiento y optimización de los pesos tanto de la C-Net como de la M-Net. Nótese que en este modelo no se emplea la matriz de referencia de Ruifrok [26].

$$\mathcal{L}(Z) = \mathcal{L}_{\text{MSE}}^{\alpha, \beta}(\mathbf{Z}) = \frac{1}{N_M} \sum_{i=1}^{N_M} \left[\left\| \mathbf{Y}^T - \mathbf{M}_i(Z) \mathbf{C}^{\alpha}(\mathbf{Z}) \right\|_F^2 \right] \quad (3.2)$$

El modelo B, BCD-Net_{MSE}, queda ilustrado en la figura 3.2 en el caso de emplear ruido aleatorio como entrada. Alternativamente, se podría modificar la entrada para utilizar la imagen observada, la cual debería estar representada en el espacio de densidad óptica.

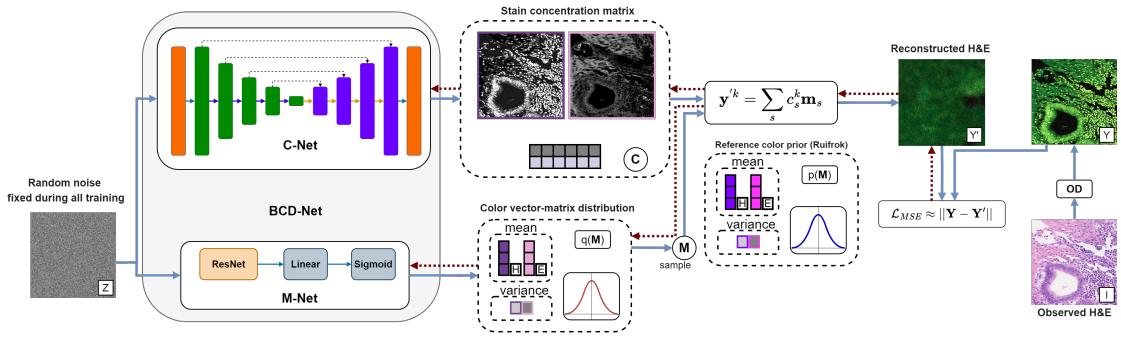


Figura 3.2: Arquitectura del modelo B, basado en el uso de BCD-Net [38] [40], error cuadrático medio y ruido aleatorio como entrada fija.

3.1.3. Modelo C. BCD-Net_{MSE+KL}

Dado que en el modelo B sí se conforma una matriz de color gracias a la predicción generada por la M-Net, podemos incluir un término que actue como regularizador del color. De esta manera, favoreceremos que las predicciones de la M-Net tiendan a parecerse en cierta medida a la matriz de Ruifrok usada como conocimiento a priori, pero sin una restricción de varianza fija, como en el modelo A. Estas consideraciones dan lugar al llamado modelo C, o BCD-Net_{MSE+KL}.

Así pues, y partiendo de la ecuación 2.7, realizamos las modificaciones necesarias para tener en cuenta el hecho de que la entrada de la red es ruido aleatorio (Z) y el objetivo es reconstruir una única imagen observada Y . La nueva función de pérdida queda descrita en la ecuación 3.3.

$$\text{ELBO}(\mathbf{Z}) = \underbrace{-\mathbb{E}_{q_\alpha(\mathbf{C}|\mathbf{Z})} \left[\log \frac{q_\alpha(\mathbf{C}|\mathbf{Z})}{p(\mathbf{C})} \right]}_{A_1} - \underbrace{\mathbb{E}_{q_\beta(\mathbf{M}|\mathbf{Z})} \left[\log \frac{q_\beta(\mathbf{M}|\mathbf{Z})}{p(\mathbf{M})} \right]}_{A_2} - \underbrace{\frac{1}{2\lambda_n^2} \mathbb{E}_{q_\beta(\mathbf{M}|\mathbf{Z})} \left[\left\| \mathbf{Y}^\top - \mathbf{MC}^\alpha(\mathbf{Z}) \right\|_F^2 \right]}_{A_3} + \text{const.} \quad (3.3)$$

Al igual que en el caso de la ecuación 2.7 original, el término A_1 se ignora al ser infinito.

El término A_2 , correspondiente a la divergencia de Kullback-Leibler entre la priori $p(M)$ y la posteriori $q(M|Z)$, puede ser calculado mediante la ecuación 3.4.

$$A_2 = \mathcal{L}_{\text{KL}}^\beta(\mathbf{Z}) = \frac{1}{2} \sum_{s=1}^S \frac{\left\| \boldsymbol{\mu}_s^\beta(\mathbf{Z}) - \mathbf{m}_s^{\text{Rui}} \right\|^2}{\gamma_s^{\text{Rui}}} + \frac{3}{2} \sum_{s=1}^S \left(\frac{\sigma_s^\beta(\mathbf{Z})^2}{\gamma_s^{\text{Rui}}} - \log \frac{\sigma_s^\beta(\mathbf{Z})^2}{\gamma_s^{\text{Rui}}} - 1 \right) \quad (3.4)$$

Por su parte, el término A_3 correspondiente al error cuadrático medio entre la imagen

observada Y y la reconstrucción Y' a partir de ruido aleatorio Z puede ser aproximado de la siguiente forma: $A_3 \approx -0.5\lambda_n^{-2}\mathcal{L}_{\text{MSE}}^{\alpha,\beta}(\mathbf{Z})$, siendo:

$$\mathcal{L}_{\text{MSE}}^{\alpha,\beta}(\mathbf{Z}) = \frac{1}{N_{\mathbf{M}}} \sum_{i=1}^{N_{\mathbf{M}}} \left[\left\| \mathbf{Y}^\top - \mathbf{M}_i(Z) \mathbf{C}^\alpha(\mathbf{Z}) \right\|_{\text{F}}^2 \right] \quad (3.5)$$

A partir de los términos anterior se obtiene la ecuación 3.6, que actuará como función de pérdida para el entrenamiento de la red neuronal en su intento de reconstruir la imagen observada Y a partir del ruido aleatorio denotado como Z . El valor θ servirá para ponderar la importancia que se da a cada uno de los términos.

$$\mathcal{L}(Z) = \theta \mathcal{L}_{\text{KL}}^{\beta}(\mathbf{Z}) + (1 - \theta) \mathcal{L}_{\text{MSE}}^{\alpha,\beta}(\mathbf{Z}) \quad (3.6)$$

El modelo C (BCD-Net_{MSE+KL}), descrito anteriormente, queda ilustrado en la figura 3.3 en el caso de emplear ruido aleatorio como entrada. Nótese que podría modificarse la entrada para emplear la imagen observada, representada en el espacio de densidad óptica.

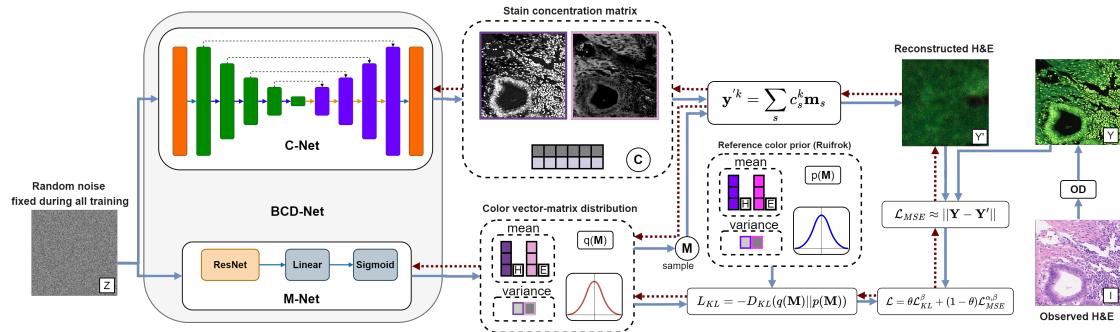


Figura 3.3: Arquitectura del modelo C, basado en el uso de BCD-Net [38] [40], función de pérdida compuesta por error cuadrático medio y divergencia de Kullback-Leibler y ruido aleatorio como entrada fija.

Después de exponer los modelos propuestos, las próximas secciones se centrarán en consideraciones adicionales sobre el proceso de entrenamiento. Esto incluirá aspectos como el tipo de entrada, optimizador utilizado y los hiper-parámetros empleados en el desarrollo de dichos modelos.

3.2. Tipos de entrada fija

Ulyanov et al. en su trabajo [34] argumentan que la entrada de una red neuronal que haga uso de Deep Image Prior puede ser cualquier tensor compatible en tamaño con la imagen deseada como salida. En nuestro caso, dicho tamaño se representa mediante: $S \times HW$. H y W hacen referencia al número de píxeles verticales y horizontales en la imagen observada; siendo S el número de tinciones.

No obstante, señalan que la estructura de la entrada seleccionada puede afectar los resultados del entrenamiento, especialmente en las iteraciones iniciales. Como ejemplo, se menciona que en la eliminación de ruido el uso de imágenes “naturales” como entrada parece mejorar la optimización de los parámetros de la red.

En este trabajo se contempla el uso de dos tipos de entradas, denotadas como Z : (i) imágenes desprovistas de estructura, generadas a partir de ruido aleatorio muestrado de una distribución uniforme en el intervalo $[0,1]$; y (ii) imágenes estructuradas o “naturales”. Se entiende como imagen “natural” aquella que puede ser obtenida de forma normal a partir de algo observable en la realidad mediante una cámara, como una fotografía de un gato.

La utilización de imágenes generadas a partir de ruido aleatorio garantiza la ausencia de estructura y, por ende, de información sobre la salida deseada. Esto demanda que la red aprenda progresivamente la información estructural de la imagen observada y sus particularidades desde cero. La ejecución del entrenamiento durante el tiempo suficiente, potencialmente considerable, permitirá una reconstrucción de calidad. En el contexto de la deconvolución ciega de color, una reconstrucción de calidad será aquella que en la que se separa adecuadamente las concentraciones y colores en canales diferentes para cada tinción.

La propia imagen histológica a reconstruir puede ser considerada una imagen “natural”, en tanto que corresponde a una realidad captada mediante un sensor. De emplearla como entrada, podría acelerar la obtención de resultados, en tanto que la red dispondría de un punto inicial del que aprender rápidamente alguna información estructural. En nuestro caso concreto, esta podrían ser la forma general de grandes grupos de tejidos o las principales tonalidades cromáticas presentes en la imagen. . Sin embargo, esta información inicial podría condicionar negativamente el resto del entrenamiento al inducir suposiciones potencialmente incorrectas o inexactas.

Tras haber comentado la importancia del tipo de entrada a emplear, y haber decidido cuáles emplearemos en este trabajo, en la siguiente sección, se presentarán en detalle los consideraciones adicionales del proceso de entrenamiento, como el optimizador utilizado, los hiper-parámetros empleados, o el grado de paralelización.

3.3. Entrenamiento de los modelos

La eficacia de los modelos propuestos no depende únicamente de sus arquitecturas, funciones de pérdida empleadas y/o entradas fijadas, sino también del optimizador usado durante el entrenamiento y otros parámetros y estrategias relevantes.

Tanto Ulyanov et al [34] como Pérez-Bueno et al [38] [40] deciden utilizar Adam (Adaptive Moment Estimation) como optimizador. Sin embargo, en este trabajo se empleará el optimizador AdamW [14], en tanto que ha demostrado mejorar a Adam en términos de la generalización realizada y el rango de hiper-parámetros óptimos capaz de calcular [36] al desacoplar la caída de los pesos de la actualización del gradiente.

Es importante destacar que al emplear Deep Image Prior, no estamos realizando un proceso de generalización directa como tal. En este caso, el modelo se convierte en uno no

amortizado, lo que significa que se entrena para reconstruir una única imagen observada cada vez. Para ello, se hace uso de una inicialización aleatoria de los parámetros de la red seleccionada y una entrada fija, que bien puede ser una imagen conformada por ruido aleatorio o bien la propia imagen a reconstruir. Los pesos de la red serán modificados para hacer que la entrada fijada sea capaz de convertirse en la reconstrucción deseada dada la imagen observada.

Respecto de la tasa de aprendizaje, el valor empleado durante los experimentos que se detallarán en el capítulo 4 es el mismo que se utilizó en [33] [40], 0.0001. Se decide utilizar este valor en tanto que empleamos la misma arquitectura de red neuronal, o al menos una parte de ella, y dicho valor fue el que mejores resultados proporcionó en la práctica. Este valor de la tasa de aprendizaje se mantendrá constante durante todo el entrenamiento.

Debido a la naturaleza del enfoque Deep Image Prior, que deconvoluciona en una única imagen a la vez, el tamaño del lote para el procesamiento es de 1. En consecuencia, no hay ventaja computacional al emplear múltiples tarjetas gráficas (GPU) para entrenar el modelo en una imagen individual, en tanto que el proceso de entrenamiento es secuencial. El uso de varias GPUs sería beneficioso para entrenar distintas instancias del modelo para diferentes imágenes o para procesar imágenes de mayor tamaño. Sin embargo, en este caso, sólo una GPU estaría realizando trabajo computacional útil.

Dada la limitación de utilizar una sola GPU por imagen para evitar la infrautilización de recursos hardware, se decide procesar el conjunto de datos, detallado en la sección 2.1.4, para uniformar todas las imágenes a un tamaño de 500x500 píxeles. Esta adaptación se justifica por la incapacidad de almacenar el modelo para imágenes de resolución 2000x2000 píxeles en una única tarjeta gráfica NVIDIA GeForce RTX 3090. La elección de dividir las imágenes en otras de 500x500 píxeles se basa en el tamaño estándar de las imágenes de colon utilizadas, aunque la GPU mencionada podría manejar imágenes de hasta 1000x1000 píxeles sin problema.

Es relevante destacar que al ajustar las imágenes al tamaño de 500x500 píxeles, el uso de memoria se reduce aproximadamente a 5.2GB cuando se utiliza la arquitectura BCD-Net completa, y a 3.5GB al emplear solo la C-Net. Este cambio permite ejecutar el modelo en GPUs de gama más accesible o con menor capacidad, ampliando la viabilidad de implementación en sistemas con recursos de hardware más limitados.

Cada proceso de entrenamiento deberá ejecutarse durante un determinado número de iteraciones definido previamente. En la deconvolución ciega de color no existe un defecto visual inherente a eliminar, que pueda reintroducirse al ejecutar el bucle de entrenamiento durante demasiadas iteraciones, no haremos uso de mecanismos de parada temprana.

Así pues, se decide que cada entrenamiento conste de 4000 iteraciones. Este valor se justifica a partir de una fase inicial de pruebas realizada con una ejecución prolongada para una pareja de imágenes de cada órgano; así como el análisis de los resultados expuestos en [34] al intentar eliminar ruido de varias imágenes.

Los resultados obtenidos, tanto en términos cualitativos como cuantitativos, así como consideraciones adicionales a las aquí mencionadas, se expondrán en el siguiente

capítulo. En este también se realizará una comparativa de rendimiento entre los modelos propuestos y la implementación original de BCD-Net [38] [40], así como una variante de Deep Image Prior en la que se utilizan los pesos disponibles de BCD-Net para inicializar las redes de los modelos B y C.

Capítulo 4

Experimentación realizada

Este capítulo detalla los experimentos realizados y los resultados obtenidos al emplear los distintos modelos propuestos en el capítulo 3.

Esta sección experimental se concibe como un estudio de ablación, destinado a investigar la influencia de la arquitectura y la función de pérdida de cada uno de los modelos propuestos en las reconstrucciones generadas. Como parte del estudio también se explorará la influencia del tipo de entrada utilizada, pudiendo ser esta desestructurada (ruido) o “natural” (imagen observada).

Comenzaremos la sección introduciendo y definiendo las métricas que se emplearán para evaluar cuantitativamente el desempeño de cada uno de los modelos evaluados. A continuación, se comentarán los detalles del conjunto de datos que se va a utilizar. Seguidamente, se incluyen los siguientes experimentos: (i) entrenamiento de los modelos propuestos haciendo uso de ruido aleatorio; (ii) entrenamiento haciendo uso de la imagen observada como entrada; (iii) entrenamiento de los modelos B y C haciendo uso de los pesos de BCD-Net como inicialización de la red, (iv) optimización del tiempo de entrenamiento de los modelos propuestos. Finalmente, se analizan y comparan los resultados obtenidos para dichos experimentos.

4.1. Métricas de rendimiento

Para evaluar cuantitativamente el desempeño de los modelos propuestos en el capítulo 3 se emplearán dos métricas: el Pico de Relación Señal-Ruido (PSNR, por sus siglas en inglés) y el Índice de Similitud Estructural (SSIM, por sus siglas en inglés).

El PSNR determina la relación entre la máxima energía teóricamente posible de una señal y el ruido presente en su representación [17]. En el contexto de restauración de imágenes, la imagen original se considera la señal y la discrepancia entre esta y la imagen estimada se interpreta como ruido. El PSNR es un valor, medido en decibelios (dB), a maximizar, ya que un valor más alto indica un menor nivel de ruido en la imagen estimada, reflejando así una mayor calidad en la misma.

La definición matemática del PSNR [17] es la siguiente:

$$PSNR = 20 \cdot \log_{10} \left(\frac{MAX_I}{\sqrt{MSE}} \right) \quad (4.1)$$

siendo el error cuadrático medio (MSE) entre las imágenes I y K :

$$MSE = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [I(i, j) - K(i, j)]^2 \quad (4.2)$$

y MAX_I el máximo valor que puede tomar un píxel en la imagen. De usar B bits por muestra, su valor será $MAX_I = 2^B - 1$.

En nuestro caso, reconstruiremos dos imágenes RGB con la información de los canales de la hematoxilina y eosina respectivamente. Así pues, el error cuadrático medio se calcula como la media aritmética de los MSE para cada una de las imágenes RGB generadas [17].

Aunque el PSNR es una métrica comúnmente utilizada, no siempre se correlaciona bien con la percepción humana de la calidad de la imagen. Por tanto, nos apoyaremos también en el índice de similitud estructural (SSIM) al realizar la evaluación.

El Índice de Similitud Estructural (SSIM) representa una métrica perceptual que evalúa la degradación de la imagen considerando cambios percibidos en la información estructural, incluyendo términos que abordan el enmascaramiento de luminancia y contraste [19][6]. El SSIM produce un valor dentro del rango $[-1, +1]$, donde $+1$ indica que las dos imágenes son idénticas y -1 señala una gran diferencia entre las mismas. Generalmente, estos valores se ajustan al rango $[0, 1]$ para uso práctico. En el contexto específico de nuestro estudio, se buscará maximizar este valor, el cual oscilará en el rango $[0, 1]$.

El SSIM se representa matemáticamente mediante la ecuación:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (4.3)$$

donde:

- x e y son las dos imágenes que se están comparando.
- μ_x y μ_y son las medias de x e y respectivamente.
- σ_x^2 y σ_y^2 son las varianzas de x e y respectivamente.
- σ_{xy} es la covarianza entre x e y .
- $C_1 = (k_1 \times L)^2$ y $C_2 = (k_2 \times L)^2$ son constantes para estabilizar la división en caso de denominadores muy pequeños. L hace referencia al valor máximo que puede tomar cada píxel, mientras que k_1 y k_2 son dos constantes que toman los valores 0.01 y 0.03 por defecto.

Para los experimentos de este capítulo se usará una variación del conjunto de imágenes Warwick Stain Separation Benchmark (WSSB) [3], detallada en la sección 4.2. A

diferencia de la práctica común, para este conjunto de imágenes se cuenta con datos de referencia (“ground truth”) para cada imagen. Así pues, podemos comparar los valores del PSNR y SSIM con respecto al “ground truth”. Se presentarán tres valores para cada métrica:

- PSNR_GT_H y SSIM_GT_H: para comparar la imagen generada para la información de la hematoxilina con el correspondiente canal en el “ground truth”.
- PSNR_GT_E y SSIM_GT_E: para comparar la imagen generada para la información de la eosina con el correspondiente canal en el “ground truth”.
- PSNR_GT y SSIM_GT: entre las imágenes generadas y el “ground truth”. Para calcular las métricas se compara la media de las imágenes RGB generadas para cada tinción frente a la media de la imagen “ground truth”.

Tras comentar las métricas que se utilizarán para medir el rendimiento de los modelos y cómo se calculan, en la siguiente sección se detallará el conjunto de datos de entrenamiento que se empleará para los experimentos.

4.2. Conjunto de datos de entrenamiento

Para evaluar el rendimiento de los modelos propuestos se empleará el conjunto de imágenes Warwick Stain Separation Benchmark (WSSB) [\[3\]](#). Este es un conjunto de datos compuesto por 24 imágenes con tinciones H&E de tejidos de diferentes órganos, concretamente, de mama, colon y pulmón. Estas imágenes proceden de diferentes laboratorios y han sido capturadas empleando diferentes microscopios, por lo que se presentan varianza tanto intra como inter-laboratorio. Las imágenes RGB “ground truth” fueron generadas a partir de vectores de color determinados por expertos patólogos; siendo las concentraciones calculadas haciendo uso de la ley de Beer-Lambert. Las imágenes de colon tienen una resolución de 500x500 píxeles, mientras que las de mama y pulmón son de 2000x2000 píxeles.

Como se comentó en el capítulo [\[3\]](#), ante la incapacidad de almacenar el modelo para imágenes de resolución 2000x2000 píxeles en una única tarjeta gráfica NVIDIA GeForce RTX 3090, se decide procesar el conjunto de datos para que todas las imágenes sean del mismo tamaño. Así pues, las imágenes de 2000x2000 píxeles se subdividen en 16 imágenes de 500x500 píxeles. De esta manera, se conforma un conjunto de datos compuesto por un total de 174 imágenes: 14 de colon, 64 de pulmón (a partir de las 4 imágenes originales) y 96 de mama (a partir de las 6 imágenes originales).

Dado este conjunto de imágenes, deberemos realizar 174 procesos de entrenamiento individuales e independientes para cada uno de los modelos expuestos anteriormente. Con el fin de mantener la reproducibilidad de los experimentos y asegurar la comparabilidad de los resultados obtenidos, se fija una determinada semilla para el generador de números aleatorios, común a todos los experimentos.

Es crucial señalar que, aunque se emplea el “ground truth” para calcular la calidad de la reconstrucción durante cada iteración del entrenamiento con el objetivo de evaluar el

desempeño de los modelos, no se utiliza esta información para el entrenamiento de la red neuronal en sí. La función de pérdida únicamente considera la imagen generada por la red y la imagen observada, ambas en el espacio de densidad óptica, para ir aprendiendo y modificando los pesos de las redes.

En la siguiente sección realizaremos un análisis de la evolución de las métricas durante el proceso de entrenamiento. Para ello, emplearemos como ejemplo una imagen de cada órgano, las cuales se muestran en la figura 4.1.

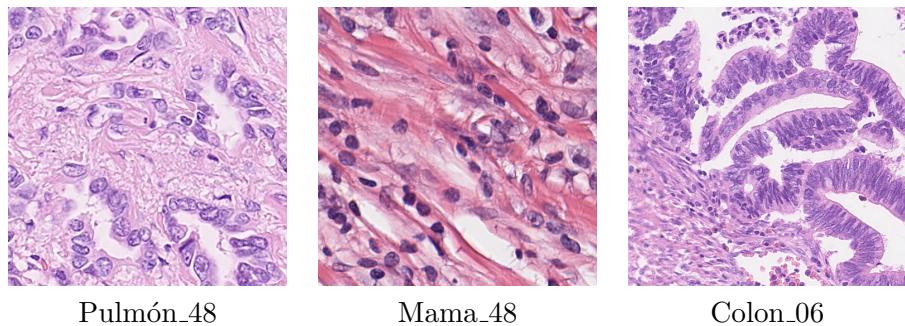


Figura 4.1: Subconjunto de imágenes de la base de datos WSSB [3], compuesto por una imagen de cada órgano (pulmón, mama y colon)

Este enfoque nos permitirá estudiar el comportamiento de los modelos para cada tipo de órgano de una forma más pormenorizada, en tanto que cada órgano tiene peculiaridades que lo diferencia de los demás. Por ejemplo, el colon suele contar con una mayor cantidad de grupos celulares de muy reducido tamaño, mientras que las imágenes de mama suelen tener colores más intensos en tanto que sus tejidos más densos absorben más tinción.

Además, esto nos permitirá establecer una correspondencia entre los valores numéricos de las métricas y algunas de las imágenes generadas durante los entrenamientos, lo cual resulta muy ilustrativo del desempeño de los modelos.

4.3. Evolución de las métricas durante el entrenamiento

En esta sección se discutirá la evolución de las métricas al entrenar los modelos propuestos para una imagen de ejemplo de cada órgano. De esta manera, podremos analizar el comportamiento que muestran los diferentes modelos en distintas etapas del entrenamiento.

Se realizarán cuatro procesos de entrenamiento para cada una de las imágenes. Se utilizan los modelos A, B y C, más una versión del modelo con etapa de pre-entrenamiento.

Durante la experimentación con la primera versión del modelo C se observó que la red neuronal tiene mayores dificultades al realizar predicciones correctas de los colores de las tinciones. Por tanto, se requiere una atención especial para garantizar la precisión de estas predicciones. Con este fin, se ha optado por ajustar la ponderación de la función de pérdida durante la fase inicial del proceso de entrenamiento. Durante el primer 20 % del

entrenamiento (800 iteraciones), se empleará un valor de $\theta = 0.99$ en la ecuación 3.6 con el propósito de favorecer la minimización de la discrepancia entre los colores predichos y la matriz de referencia propuesta por Ruifrok [26]; se tendrá una fase de pre-entrenamiento enfocada a la predicción del color. Se espera que este ajuste contribuya a mejorar la precisión de las predicciones cromáticas. Posteriormente, se restablecerá una ponderación equilibrada (50-50) para el resto del proceso de entrenamiento. Esta estrategia permite tanto minimizar el error asociado a la reconstrucción de las concentraciones como otorgar cierta flexibilidad en la predicción del color, posibilitando así variaciones leves en caso de ser consideradas necesarias.

En las figuras 4.2 y 4.3 se muestra la evolución del PSNR y SSIM para todos los modelos propuestos al emplear ruido aleatorio como entrada fija. El lector interesado puede consultar las figuras para la evolución de las métricas al emplear la imagen observada como entrada en el anexo B. El comportamiento en dicho caso es muy similar al expuesto en esta sección, con la salvedad de que las primeras etapas son de una duración menor.

En el caso del modelo A se observa un rápido crecimiento inicial de las métricas, en tanto que los colores son muestreados de una referencia y ya se dispone de una cantidad considerable de conocimiento. Después, se observa una segunda etapa de crecimiento algo más dilatada en el tiempo, correspondiente al aprendizaje de las concentraciones. Posteriormente, se entra en una etapa de estabilización en la que la red no es capaz de mejorar los resultados obtenidos, existiendo algunas oscilaciones debido al muestreo aleatorio del color.

El modelo B muestra un comportamiento más errático. Recordemos que ahora la red tiene que estimar también el color, pero carece de regularización de los colores a utilizar. Esto derivará en que, en la mayoría de los casos, uno de los canales tenga la mayoría de la información cromática, o que los colores de las tinciones sean poco fieles a la realidad. En [38] [40] se observó que la red neuronal suele considerar más relevante la información de la hematoxilina.

Este comportamiento es especialmente notable en la caída de rendimiento alrededor de la iteración 1300 para el caso de la imagen 'Colon_6'. Otro posible comportamiento es el mostrado para la imagen 'Lung_48', para la cual los colores y concentraciones predichas son bastante acertadas, pero se asignan al canal incorrecto.

El modelo C exhibe un comportamiento similar al del modelo A. En las primeras etapas se enfoca en aprender los colores para posteriormente reducir el error de reconstrucción dadas las concentraciones predichas. En la subfigura d), correspondiente al caso en que se usa una etapa de pre-entrenamiento enfocada al color, se puede apreciar una ligera disminución de rendimiento en el instante en que se cambia la ponderación de la función de pérdida. Esto se debe a que al empezar a contemplar el error de reconstrucción de las concentraciones pueden realizarse ligeros ajustes al color. Estos en un primer momento pueden disminuir la calidad de la reconstrucción generada, pero conforme se aprenda la información de las concentraciones el rendimiento irá mejorando rápidamente, hasta llegar a un punto de pseudo-estabilización en que las mejoras se producen de forma muy dilatada en el tiempo. Como se verá posteriormente en la comparativa de la

sección 4.4, se obtienen resultados ligeramente superiores al utilizar pre-entrenamiento.

Para los modelos B y C puede observarse cómo durante las primeras etapas del entrenamiento se tiene una reconstrucción de mayor calidad para el canal de la eosina (azul en las gráficas) de la que se obtendrá conforme avance el entrenamiento y se llegue al punto de convergencia. En contraposición, la reconstrucción de la hematoxilina suele mejorar de forma constante antes de llegar a la etapa de estabilización. Idealmente, buscariamos tener las mejores reconstrucciones posibles para ambos canales, por lo que podría ser factible quedarnos con la reconstrucción de la eosina de las primeras iteraciones y la de la hematoxilina para etapas más avanzadas. Esta idea se plantea como trabajo a realizar en el futuro.

En las figuras 4.4 y 4.5 se presentan ejemplos visuales de la correspondencia entre la evolución de las métricas y las reconstrucciones generadas. Dichas figuras permiten al lector relacionar de manera más clara y visual los valores numéricos de las métricas, especialmente el PSNR, con el resultado proporcionado por la red neuronal.

En la figura 4.4 se puede observar cómo, partiendo de ruido aleatorio, el modelo A es capaz de ir aprendiendo poco a poco la información de las concentraciones. En torno a la iteración 800 se obtiene una reconstrucción de bastante calidad.

En la figura 4.5 se ilustra el comportamiento del modelo B, no deseado en tanto que los colores predichos no son fieles a la realidad. Puede observarse como, conforme pasan las iteraciones, uno de los canales pasa a tener la mayoría de la información cromática, mientras que el otro actúa como un ajuste para que la media de ambos se parezca a la imagen observada. Además, puede observarse que las concentraciones predichas se encuentran intercambiadas de canal.

El comportamiento del modelo C para la imagen “Breast_0” es relativamente similar al expuesto para el caso del modelo A, con la diferencia de que en las primeras 100-200 iteraciones parte de colores algo distintos en tanto que no los muestrea de una referencia, como el modelo A, sino que los aprende poco a poco. Esto puede observarse en la figura 4.6.

El lector interesado puede encontrar ejemplos adicionales del comportamiento de los modelos al usar la imagen observada como entrada en el anexo B.

Tras haber analizado la evolución de las métricas y su impacto en las reconstrucciones generadas para cada modelo, en la siguiente sección se realizará una comparativa de rendimiento entre los mejores resultados alcanzados por cada uno de los modelos propuestos.

4.4. Comparativa de rendimiento entre modelos

En esta sección se realizará una comparativa del rendimiento proporcionado por todos los enfoques abordados en este trabajo. En primer lugar, se enfrentarán los modelos propuestos en este trabajo basados en Deep Image Prior (A, B y C en su variante con pre-entrenamiento enfocado al color) entre sí. No sé tendrá en cuenta el modelo C sin pre-entrenamiento para el color en tanto que la versión que lo utiliza ofrece mejor rendimiento. Posteriormente, se examinarán los resultados proporcionados por el en-

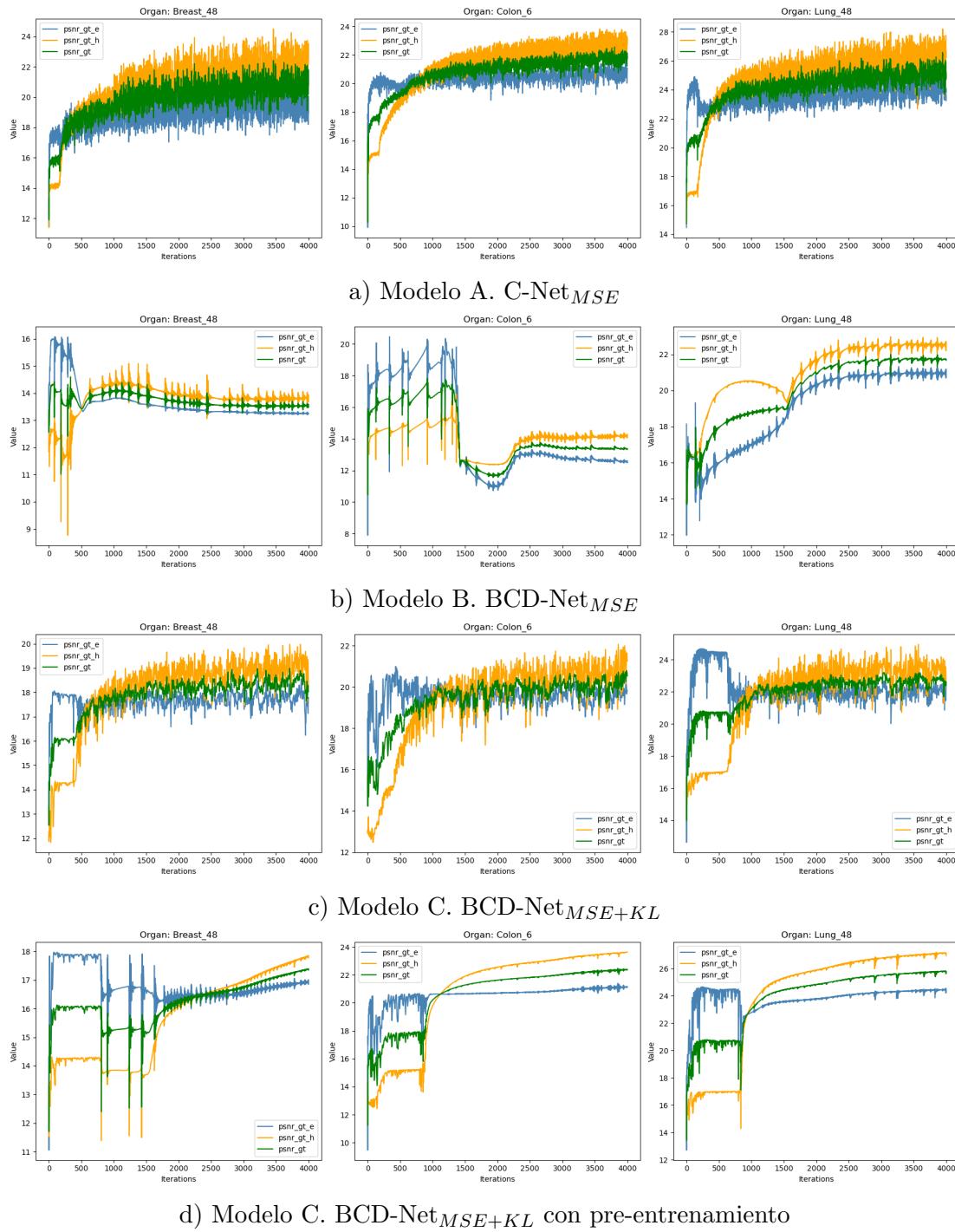


Figura 4.2: Evolución del PSNR para los entrenamientos de las imágenes seleccionadas al partir de ruido aleatorio para los diferentes modelos propuestos.

4.4. Comparativa de rendimiento entre modelos

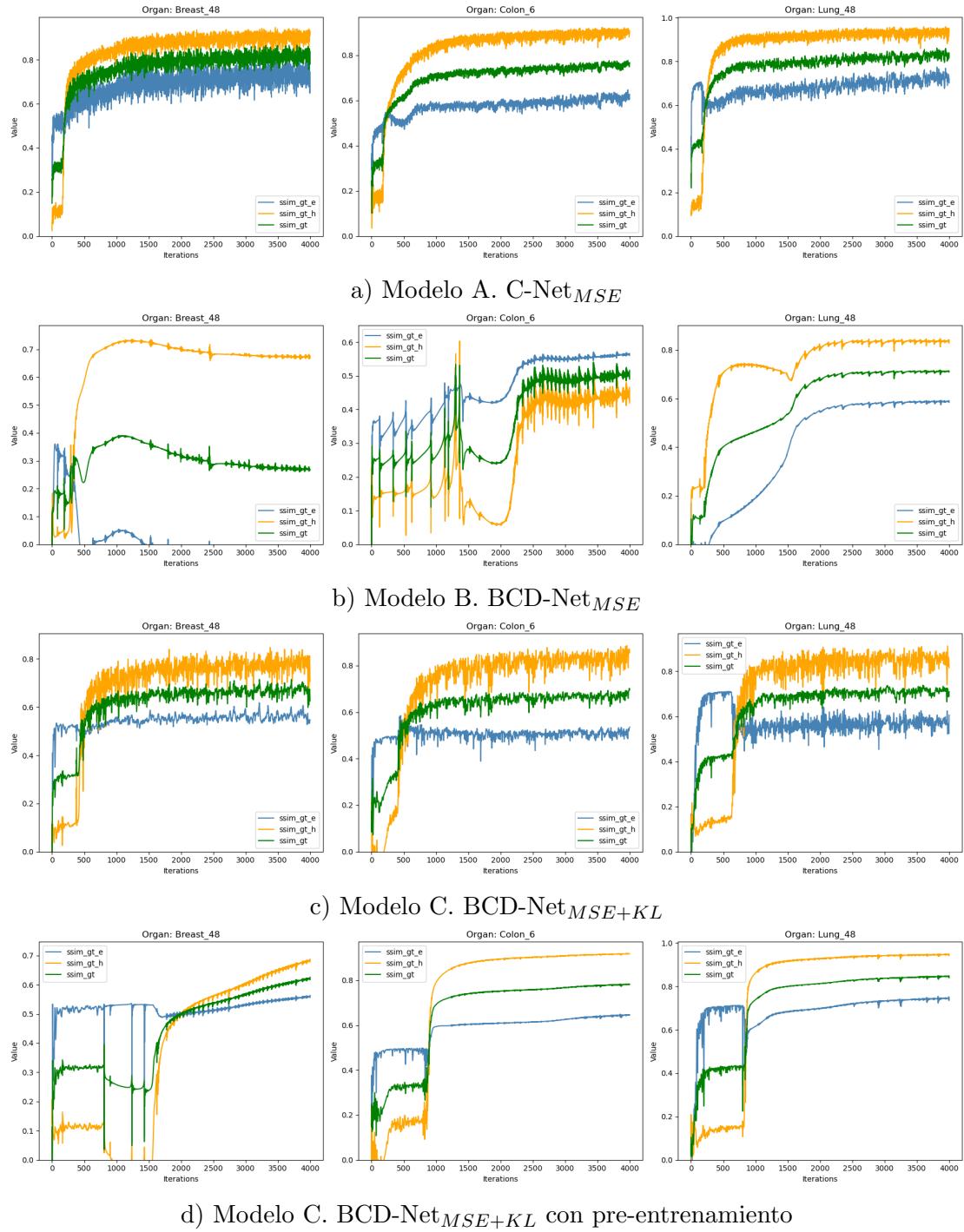


Figura 4.3: Evolución del SSIM para los entrenamientos de las imágenes seleccionadas al partir de ruido aleatorio para los diferentes modelos propuestos.

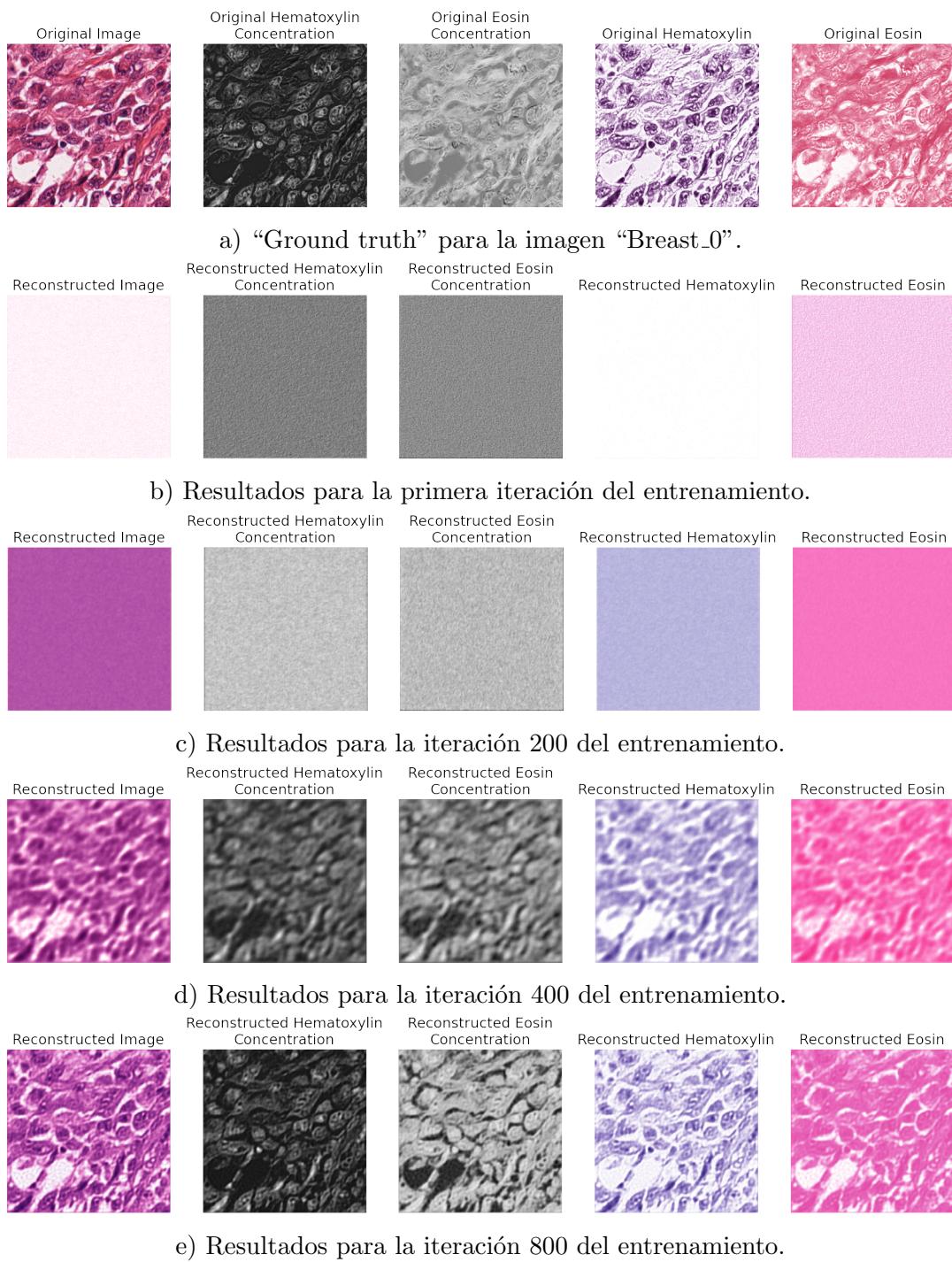


Figura 4.4: Evolución de los resultados obtenidos para la imagen "Breast_0" al entrenar utilizando el modelo A y ruido aleatorio como entrada.

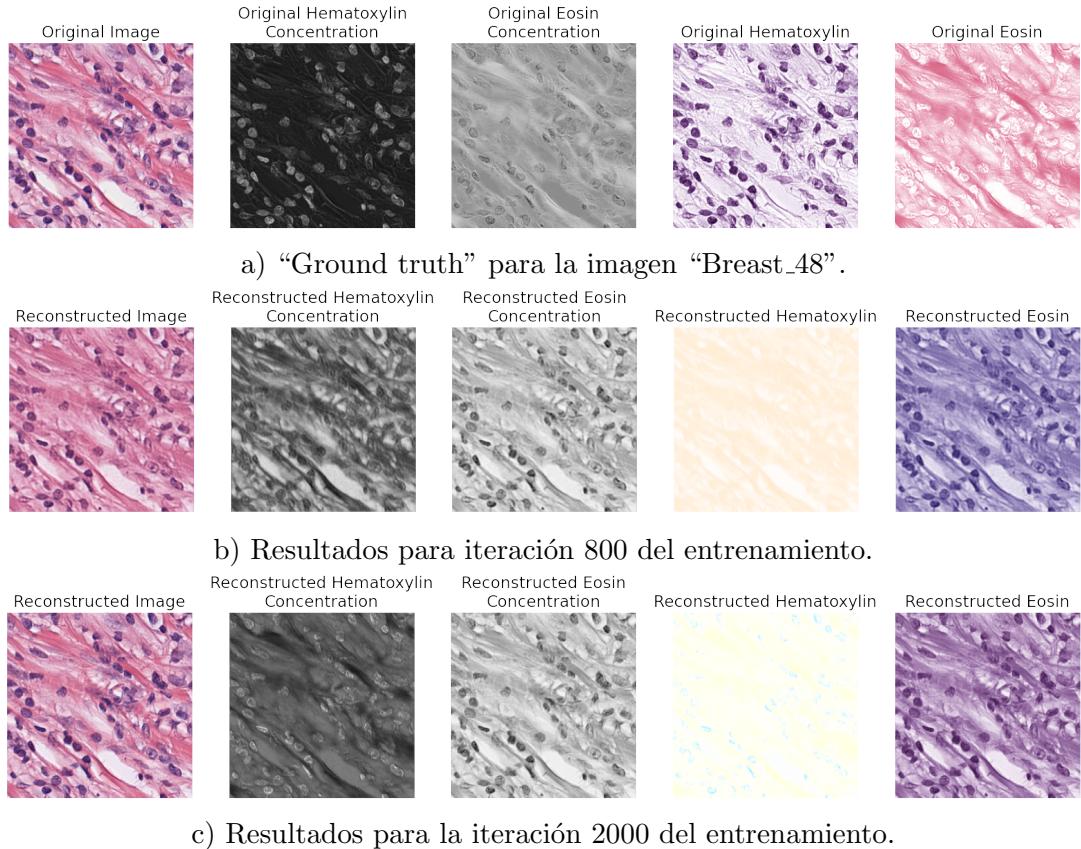


Figura 4.5: Evolución de los resultados obtenidos para la imagen “Breast_48” al entrenar utilizando el modelo B y ruido aleatorio como entrada.

trenamiento disponible de BCD-Net y los resultantes de realizar dos procesos de “fine tuning”, en los que se emplean ponderaciones diferentes para los términos de la función de pérdida. Finalmente, se compararán los mejores resultados obtenidos en las dos discusiones anteriormente mencionadas.

Comencemos exponiendo un resumen de los resultados para los modelos propuestos en el capítulo 3 al hacer uso de ruido aleatorio como entrada; estos se muestran en las tablas 4.1 (PSNR) y 4.2 (SSIM).

En las tablas mencionadas se puede apreciar como el modelo A, que hace uso de C-Net y el error cuadrático medio, obtiene los mejores resultados medios para todos los modelos propuestos en el capítulo 3. Este comportamiento puede sorprender al lector, en tanto que los modelos B y C hacen uso de BCD-Net, una arquitectura más compleja; y el modelo C utiliza mecanismos adicionales para intentar mejorar las predicciones de color y por tanto el resultado final.

En [24], Ren et al. estudian la aplicación del enfoque Deep Image Prior para el desembarronamiento de imágenes. En dicho trabajo se llega a la conclusión de que emplear

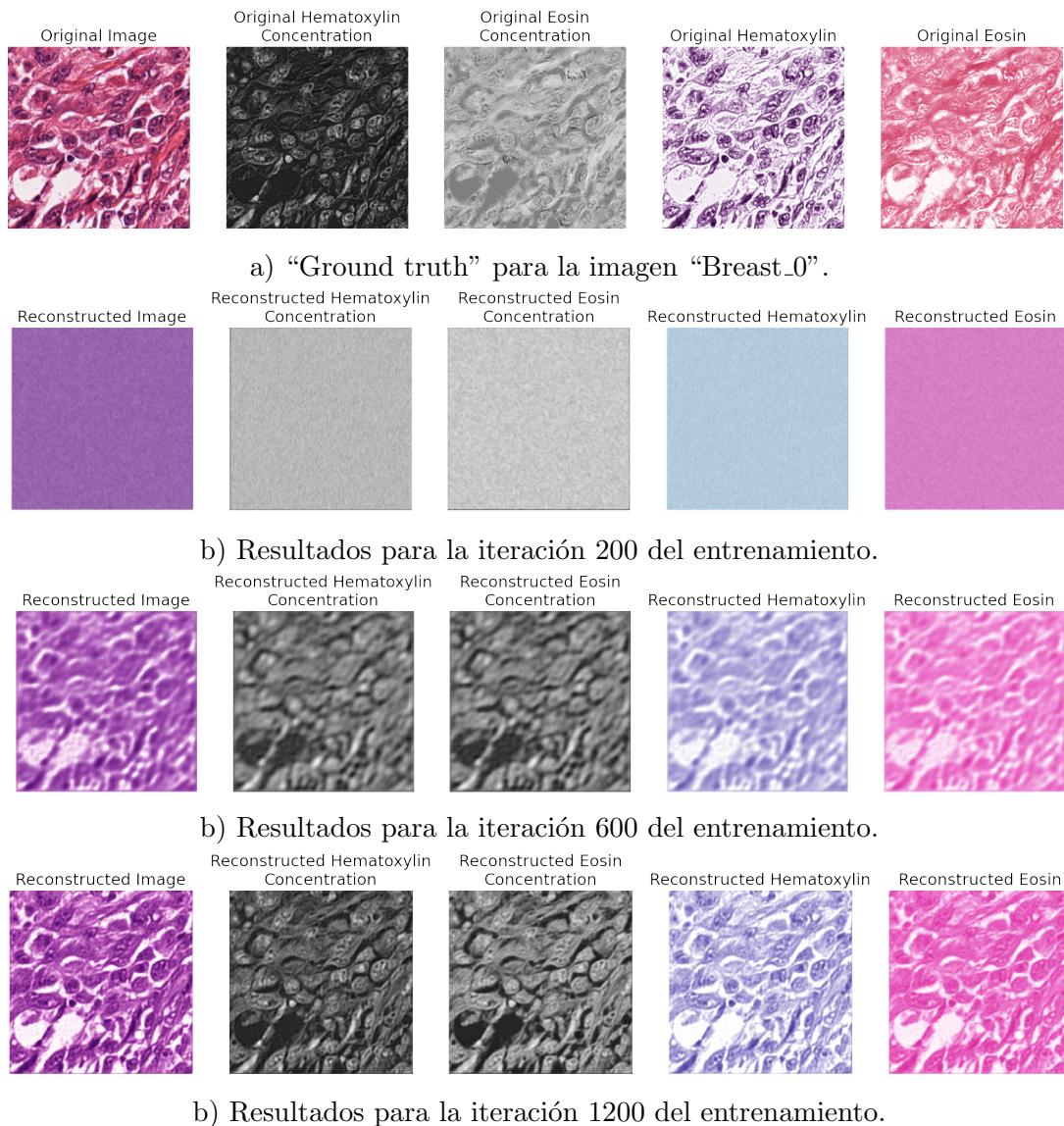


Figura 4.6: Evolución de los resultados obtenidos para la imagen “Breast_0” al entrenar utilizando el modelo C sin pre-entrenamiento y ruido aleatorio como entrada.

una red simple para el núcleo de deconvolución permite obtener mejores resultados al aplicar Deep Image Prior que arquitecturas más complejas con mayor cantidad de capas convolucionales.

Este comportamiento se encuentra en línea con los resultados experimentales obtenidos en nuestro trabajo. Así pues, es probable que la arquitectura de la red M-Net (MobileNet V3) sea demasiado compleja y no permita extraer el máximo potencial del enfoque Deep Image Prior a la hora de predecir los colores de las tinciones. Por tanto,

Órgano	Tinción	Modelos		
		A	B	C con pre entrenamiento
Colon	H	24.656 ± 0.905	17.375 ± 2.054	24.051 ± 0.510
	E	22.080 ± 0.713	19.711 ± 0.596	21.902 ± 0.507
Pulmón	H	27.686 ± 0.346	19.371 ± 3.607	26.710 ± 0.386
	E	24.505 ± 0.609	22.327 ± 1.318	24.362 ± 0.189
Mama	H	22.927 ± 1.564	16.639 ± 2.240	18.491 ± 0.715
	E	19.170 ± 1.156	16.302 ± 1.490	17.240 ± 0.228
Media	H	25.089 ± 0.938	17.795 ± 2.633	23.084 ± 0.537
	E	21.918 ± 0.826	19.447 ± 1.135	21.168 ± 0.308
	Media	23.504 ± 0.882	18.621 ± 1.884	22.126 ± 0.423

Tabla 4.1: Valores medios y desviaciones típicas del PSNR para los modelos propuestos en el capítulo 3 al hacer uso de ruido aleatorio como entrada. En negrita se marcan los mejores resultados de entre los proporcionados por los diferentes modelos.

Órgano	Tinción	Modelos		
		A	B	C con pre entrenamiento
Colon	H	0.916 ± 0.007	0.630 ± 0.121	0.921 ± 0.004
	E	0.759 ± 0.112	0.688 ± 0.117	0.762 ± 0.113
Pulmón	H	0.944 ± 0.006	0.586 ± 0.261	0.938 ± 0.007
	E	0.765 ± 0.008	0.573 ± 0.015	0.750 ± 0.004
Mama	H	0.934 ± 0.003	0.799 ± 0.072	0.780 ± 0.093
	E	0.817 ± 0.022	0.450 ± 0.396	0.683 ± 0.120
Media	H	0.931 ± 0.005	0.672 ± 0.151	0.880 ± 0.027
	E	0.708 ± 0.142	0.570 ± 0.176	0.732 ± 0.079
	Media	0.855 ± 0.025	0.595 ± 0.164	0.806 ± 0.053

Tabla 4.2: Valores medios y desviaciones típicas del SSIM para los modelos propuestos en el capítulo 3 al hacer uso de ruido aleatorio como entrada. En negrita se marcan los mejores resultados de entre los proporcionados por los diferentes modelos.

emplear Deep Image Prior únicamente sobre la C-Net y optar por una estrategia más simple para la estimación de color, como realizar un muestreo aleatorio sobre una referencia, puede permitir mejorar las calidad de las reconstrucciones generadas de forma significativa. Sin embargo, esta hipótesis requerirá de mayor estudio en trabajos futuros.

Puede apreciarse que el modelo B es aquel que peor rendimiento ofrece en la práctica. Esto se debe principalmente a las variaciones poco realistas que pueden sufrir los colores predichos conforme avanza el entrenamiento para algunas imágenes, de ahí las desviaciones típicas tan elevadas. Disponer de un mecanismo de parada temprana resulta fundamental cuando se usa este modelo, aunque en la práctica sería más sencillo optar por el uso del modelo C en su lugar.

A continuación, realizaremos una comparativa del rendimiento de cada modelo en caso de utilizar como entrada fija ruido aleatorio o la imagen observada.

Órgano	Tinción	Modelos		
		A	B	C con pre entrenamiento
Colon	H	25.025 ± 0.385	21.849 ± 4.186	22.408 ± 0.670
	E	23.654 ± 1.127	19.774 ± 3.182	20.422 ± 0.328
Pulmón	H	27.992 ± 0.638	18.664 ± 4.803	24.740 ± 0.979
	E	24.861 ± 0.668	20.837 ± 0.356	22.523 ± 0.440
Mama	H	23.078 ± 0.832	19.941 ± 2.140	19.562 ± 0.686
	E	20.784 ± 1.876	18.861 ± 1.430	17.980 ± 1.329
Media	H	25.365 ± 0.618	20.151 ± 3.710	22.237 ± 0.778
	E	23.010 ± 1.224	19.832 ± 1.656	20.315 ± 0.699
	Media	24.188 ± 0.921	19.992 ± 2.683	21.276 ± 0.739

Tabla 4.3: Valores medios y desviaciones típicas del PSNR para los modelos propuestos en el capítulo 3 al hacer uso de la imagen observada como entrada. En negrita se marcan los mejores resultados de entre los proporcionados por los diferentes modelos.

Órgano	Tinción	Modelos		
		A	B	C con pre entrenamiento
Colon	H	0.936 ± 0.002	0.877 ± 0.064	0.857 ± 0.012
	E	0.804 ± 0.096	0.480 ± 0.399	0.653 ± 0.135
Pulmón	H	0.953 ± 0.012	0.579 ± 0.281	0.882 ± 0.008
	E	0.759 ± 0.024	0.518 ± 0.084	0.641 ± 0.002
Mama	H	0.940 ± 0.003	0.846 ± 0.086	0.843 ± 0.008
	E	0.851 ± 0.006	0.791 ± 0.121	0.726 ± 0.034
Media	H	0.943 ± 0.006	0.767 ± 0.144	0.861 ± 0.009
	E	0.805 ± 0.042	0.596 ± 0.201	0.673 ± 0.057
	Media	0.874 ± 0.024	0.682 ± 0.173	0.767 ± 0.330

Tabla 4.4: Valores medios y desviaciones típicas del SSIM para los modelos propuestos en el capítulo 3 al hacer uso de la imagen observada como entrada. En negrita se marcan los mejores resultados de entre los proporcionados por los diferentes modelos.

En las tablas 4.3 y 4.4 se muestran los valores medios, en términos de PSNR y SSIM, para las mejores reconstrucciones para todas las imágenes del conjunto de datos. Al igual que al emplear una entrada desprovista de estructura, puede observarse como el modelo A es aquel con mejor rendimiento. Es destacable como la diferencia entre el rendimiento de los modelos B y C se estrecha para el caso de colon y mama, aunque, en media, el modelo C con pre-entrenamiento enfocado al color sigue ofreciendo mejores resultados que el modelo B.

A la luz de los resultados expuestos en las tablas 4.1 y 4.2 (PSNR y SSIM para ruido aleatorio como entrada), 4.3 y 4.4 (PSNR y SSIM para imagen observada como entrada); podemos concluir lo siguiente:

- El modelo A proporciona mejores resultados para todos los órganos cuando emplea la imagen observada como entrada, tanto en términos de PSNR como de SSIM. Este

modelo se beneficia de la información adicional inicial sobre las concentraciones que proporciona el uso de la imagen observada.

- El modelo B proporciona mejores resultados para colon y mama cuando emplea la imagen observada como entrada; mientras que para el caso del pulmón emplear ruido aleatorio permite alcanzar un mayor rendimiento. En media, funciona mejor emplear la imagen observada. Recordemos que este modelo presenta un comportamiento errático en sus predicciones del color al carecer de un regularizador que las guíe hacia una referencia.
- El modelo C funciona mejor para pulmón y colon al emplear ruido aleatorio, mientras que para el caso de mama emplear la imagen observada permite generar mejores reconstrucciones. Este comportamiento se justifica en tanto que la mama tiene colores más alejados de la referencia, y la información que pueda extraerse de la imagen observada al inicio del entrenamiento permite una mejor predicción de los colores. En término medio para todos los órganos, funciona mejor al emplear ruido aleatorio como entrada.

El lector interesado puede encontrar una comparativa más exhaustiva de la influencia del tipo de entrada sobre cada modelo y órgano en el anexo D.

Después de analizar los resultados de los modelos propuestos en función del tipo de entrada y compararlos entre ellos, pasemos a detallar la segunda parte de los experimentos realizados.

Al utilizar el enfoque Deep Image Prior se parte de una entrada fija y una inicialización aleatoria de los pesos de la red. Sin embargo, en tanto que se dispone de pesos para BCD-Net, podemos aprovechar dicho conocimiento para inicializar las redes de los modelos B y C. Así pues, se propone una variación del enfoque Deep Image Prior “puro” en la que se aprovecha el conocimiento generado en otro conjunto de datos para facilitar el entrenamiento de una única imagen observada cada vez, que será usada como entrada de la red. De esta manera, podríamos obtener resultados de mayor calidad, en tanto que nos basamos de un conocimiento aprendido en grandes conjuntos de datos, que particularizaremos para cada imagen concreta que se deba procesar.

Utilizando este nuevo enfoque, se propone un nuevo conjunto de experimentos para los modelos B y C. Al igual que hasta ahora, se realizará un entrenamiento individual para cada imagen del conjunto de datos, en la que dicha imagen será la entrada fija de la red. La diferencia radicará en la inicialización de los pesos de la red neuronal, la cual ahora se realizará a partir de los pesos de BCD-Net publicados y disponibles para su descarga en [39].

En las tablas 4.5 (PSNR) y 4.6 (SSIM) se exponen los resultados obtenidos mediante la aplicación del enfoque que se acaba de proponer para los modelos B y C. Además, se incluyen los resultados derivados del entrenamiento original del BCD-Net y los resultados del Modelo A (haciendo uso de la imagen observada como entrada), que destacó por su rendimiento superior frente al resto de modelos propuestos al adoptar el enfoque Deep Image Prior de manera “pura”.

Órgano	Tinción	Modelos		
		A (imagen observada)	BCD-Net amortizado	Modelo B con pesos entrenados
Colon	H	25.03 ± 0.39	24.71 ± 0.28	24.23 ± 0.76
	E	23.65 ± 1.13	22.44 ± 0.10	21.60 ± 0.82
Pulmón	H	27.99 ± 0.64	27.22 ± 0.50	28.73 ± 0.53
	E	24.81 ± 0.67	24.76 ± 0.19	25.57 ± 0.62
Mama	H	23.08 ± 0.83	24.35 ± 1.20	28.62 ± 1.82
	E	20.78 ± 1.88	22.18 ± 1.12	25.44 ± 2.17
Media	H	25.37 ± 0.65	25.43 ± 0.59	27.22 ± 1.04
	E	23.10 ± 1.22	23.13 ± 0.46	24.20 ± 1.20
	Media	24.23 ± 0.94	24.28 ± 0.53	25.71 ± 1.13

Tabla 4.5: Valores medios y desviaciones típicas del PSNR para el modelo A al emplear la imagen observada como entrada y los modelos B y C inicializados con los pesos de BCD-Net. En negrita se marcan los mejores resultados.

Órgano	Tinción	Modelos		
		A (imagen observada)	BCD-Net amortizado	Modelo B con pesos entrenados
Colon	H	0.936 ± 0.002	0.925 ± 0.013	0.951 ± 0.002
	E	0.804 ± 0.096	0.792 ± 0.026	0.843 ± 0.008
Pulmón	H	0.953 ± 0.012	0.952 ± 0.003	0.974 ± 0.000
	E	0.759 ± 0.024	0.875 ± 0.016	0.893 ± 0.000
Mama	H	0.940 ± 0.003	0.936 ± 0.002	0.980 ± 0.004
	E	0.851 ± 0.006	0.708 ± 0.015	0.955 ± 0.004
Media	H	0.943 ± 0.006	0.938 ± 0.006	0.968 ± 0.002
	E	0.805 ± 0.042	0.792 ± 0.015	0.897 ± 0.003
	Media	0.874 ± 0.024	0.865 ± 0.001	0.933 ± 0.003

Tabla 4.6: Valores medios y desviaciones típicas del SSIM para el modelo A al emplear la imagen observada como entrada y los modelos B y C inicializados con los pesos de BCD-Net. En negrita se marcan los mejores resultados.

Se puede observar como el modelo A, siguiendo el enfoque Deep Image Prior “puro” y empleando la imagen observada como entrada, ofrece un mejor resultado que la versión amortizada de BCD-Net para las imágenes de colon y pulmón, de hasta un 5 % en PSNR. Por otra parte, BCD-Net obtiene mejores resultados en las imágenes de mama, entre un 5 y 7 % en términos de PSNR. En media, las diferencias entre ambos modelos son ínfimas (menores al 0.5 %) y pueden considerarse despreciables.

Los modelos B y C inicializados con los pesos disponibles de BCD-Net permiten obtener resultados experimentales superiores. Para el caso del colon y el pulmón las mejoras son modestas respecto del modelo A con inicialización aleatoria, oscilando estas entre el 1 % y 3.5 % en términos de PSNR; frente a la versión de BCD-Net la mejora es de un 2.5 % y 5 % (PSNR), aunque ofrecen peor desempeño en el canal de eosina del colon.

La mayor diferencia de rendimiento se encuentra en el caso de las imágenes de mama. Para ellas, el modelo B inicializado con los pesos de BCD-Net consigue una mejora

sustancial respecto a sus competidores, de entre 3.3 y 5.5 decibelios (PSNR). Frente al modelo A esta mejora es de un 23 % en media. En tanto que la inicialización de los pesos proporciona una gran información sobre los colores de tejidos teñidos y escaneados en diferentes centros, se puede compensar la mayor diferencia de color que existe para los tejidos de mama del dataset WSSB [3] frente a la referencia de Ruifrok [26]. Así pues, prescindir del regularizador de color para el entrenamiento de la M-Net y confiar en el conocimiento genérico resulta en una mejora de rendimiento considerable.

Esta gran mejora en la reconstrucción de las imágenes de mama hará que el modelo B inicializado con los pesos de BCD-Net sea el que mejor resultado global proporcione de todos los modelos propuestos en este trabajo, tanto en términos de PSNR como de SSIM. La mejora de PSNR y SSIM es de un 6 % y 7 % respecto del modelo A que sigue Deep Image Prior “puro” y usa la imagen observada como entrada. Además, existe una gran ventaja en tiempo de computación, en tanto que los mejores resultados para los modelos B y C inicializados con pesos se obtuvieron entre las primeras 50 y 100 iteraciones del bucle de entrenamiento; frente a las más de 1500 que suelen ser necesarias para los modelos que siguen Deep Image Prior “puro”.

Los resultados expuestos nos permiten concluir que el enfoque Deep Image Prior resulta aplicable a la deconvolución ciega de color y permite proporcionar buenos resultados. En el caso del modelo A, estos pueden ser comparables a los de modelos amortizados, con la gran ventaja de no requerir de grandes conjuntos de datos para el entrenamiento.

De disponer de pesos entrenados para un conjunto de datos compatible con el que se desea procesar, estos pesos pueden ser empleados para mejorar el rendimiento de los modelos, siguiendo la variante de Deep Image Prior propuesta. Así pues, la mejora de rendimiento puede ser muy significativa, requiriendo también de un tiempo de ejecución mucho menor.

La utilización de Deep Image Prior “puro” puede resultar adecuada en casos en que se deban analizar pocas imágenes o no se disponga de modelos amortizados cuyo conocimiento sea aplicable al tipo de imágenes del que se dispone, lo cual es habitual. Cuando se emplea Deep Image Prior, resulta fundamental detener el entrenamiento de forma temprana para ahorrar tiempo de cómputo, dado que como se mostró en la sección 4.3, los resultados tienden a estabilizarse pasado un determinado número de iteraciones del bucle de entrenamiento.

4.5. Optimización del entrenamiento

En esta sección se abordarán las diferencias en calidad y tiempo que podrían tener lugar si hubiéramos detenido el entrenamiento de los diferentes modelos de forma temprana, en lugar de ejecutar las 4000 iteraciones completas que se proponen como cantidad suficiente para la convergencia de las predicciones de los modelos.

Como se pudo observar en las figuras 4.2 y 4.3, los modelos A y C (en sus dos variantes) sufren una pseudo-estabilización del rendimiento entre la iteración 1250 y 2000 para las imágenes de ejemplo mostradas (y otras adicionales de las que se disponen); siendo la mejora de rendimiento pasada la iteración 2000 poco significativa en la mayoría

de casos. El comportamiento del modelo B es algo más errático, pero se puede observar este mismo comportamiento (incluso en iteraciones más tempranas) en algunos casos.

En la tabla 4.7 se muestran los máximos valores medios del PSNR, así como la iteración media en que se obtuvieron de haber detenido el entrenamiento entre las iteraciones 1250 y 2000; frente al PSNR medio máximo durante todo el entrenamiento, y en que iteración se obtuvo. Posteriormente, se muestra la fracción del rendimiento que se obtiene por el óptimo local frente al óptimo global, así como la fracción de tiempo empleado. En dicha tabla se muestran los resultados al emplear ruido aleatorio como entrada. No se adjunta una tabla para los resultados obtenidos en caso de considerar el SSIM, en tanto que los resultados son similares a los expuestos para el caso del PSNR.

Podremos observar como la disminución de rendimiento en media es de apenas un 2.5 % y 5 % en función del modelo. Sin embargo, emplearemos hasta un 50 % de tiempo de ejecución menos en media. La excepción a este caso es el modelo B, en que la reducción de rendimiento es cercana a un 20 % en media, el cual se justifica porque el comportamiento del modelo B es errático al carecer de regularizador de color. De haber contemplado un intervalo de iteraciones menor, por ejemplo, entre 250 y 1250, se podría haber obtenido una reducción del rendimiento medio de en torno al 7 % y el 11 % para los modelos A y C, a cambio de una disminución del tiempo de entrenamiento medio cercana al 65 %. En esta ocasión, la disminución de rendimiento del modelo B sería de entorno a un 4 % en media. Nótese que el modelo B funciona mejor si se detiene en etapas muy tempranas, en tanto que puede comenzar a introducir error dadas sus predicciones de colores cada vez más alejadas de la realidad conforme avanza el entrenamiento.

En el Anexo B se pueden observar la evolución de las métricas para los entrenamientos de los modelos al usar una entrada estructurada para las imágenes de ejemplo. En dichas figuras se observa que la etapa de convergencia del modelo se comienza a dar en un número menor de iteraciones que al emplear ruido aleatorio como entrada. Así pues, de utilizar la imagen estructurada la parada deberá ser aún más temprana. En la tabla 4.8 se exponen los mejores resultados proporcionados en términos de PSNR por los modelos en un intervalo de 250 a 1250 iteraciones al fijar como entrada la imagen observada, así como para el entrenamiento completo. No se adjunta una tabla para los resultados obtenidos en caso de considerar el SSIM, en tanto que los resultados son similares a los expuestos para el caso del PSNR.

Se puede observar que las diferencias de rendimiento para los modelos A y C oscilan entre un 1.5 % y 4 %, a cambio de una considerable reducción del tiempo de entrenamiento de entre un 60 % y 70 %. La excepción a este comportamiento es el modelo B. Para este, los mejores resultados globales suelen alcanzarse durante las primerísimas iteraciones del entrenamiento, pero con una variabilidad altísima. Se dan casos en los que las mejores reconstrucciones se generan antes de la iteración 250, extremo inferior del intervalo considerado para la parada temprana; por lo que no se obtiene mejoría alguna al dejar ejecutar el entrenamiento por más tiempo.

Como muestran los resultados expuestos, la aplicación de un mecanismo de parada temprana permite reducir considerablemente el tiempo de ejecución con repercusiones mínimas en el rendimiento obtenido para todos los modelos, sea cual fuere la entrada

empleada. En [34] se plantea cómo y cuándo realizar la parada temprana como una pregunta a resolver en el futuro. A la hora de realizar BCD en un caso real no se dispone del “ground truth” para calcular PSNR y SSIM respecto de este, por lo que el cálculo de un momento óptimo de parada se torna difícil. Alternativamente, podría usarse el cálculo del PSNR o SSIM respecto de la imagen observada; de manera que si las métricas no mejoran una determinada cantidad en un número dado de iteraciones se detenga el entrenamiento. El valor de estos parámetros podría diferir significativamente en función de las características de la imagen concreta empleada, lo que dificulta determinarlos para una aplicación genérica. La efectividad de esta hipótesis debe ser estudiada en trabajos futuros.

Modelo	Órgano	Métricas					
		PSNR _{fast}	Iter _{fast}	PSNR _{max}	Iter _{max}	$\frac{PSNR_{fast}}{PSNR_{max}}$	$\frac{t_{fast}}{t_{max}}$
A	Colon	22.529 ± 0.759	1969.5 ± 28.5	23.368 ± 0.809	3761.0 ± 42.0	0.964	0.5237
	Mama	20.300 ± 1.498	1951.5 ± 31.5	21.049 ± 1.360	3470.5 ± 11.5	0.964	0.5614
	Pulmón	25.201 ± 0.456	1822.5 ± 56.5	26.096 ± 0.477	3527.5 ± 416.5	0.966	0.5167
	Media	22.677 ± 0.904	1914.5 ± 38.8	23.504 ± 0.882	3586.0 ± 156.6	0.965	0.5339
B	Colon	17.820 ± 0.373	1563.0 ± 267.0	18.543 ± 0.729	2095.5 ± 1177.5	0.961	0.7477
	Mama	16.298 ± 1.947	1434.5 ± 57.5	16.470 ± 1.865	2112.5 ± 1777.5	0.990	0.6748
	Pulmón	19.406 ± 1.626	1624.0 ± 363.0	20.849 ± 1.145	1721.0 ± 1424.0	0.931	0.9457
	Media	17.841 ± 1.315	1540.5 ± 229.2	18.621 ± 1.246	1976.3 ± 1459.7	0.961	0.7894
C con $\theta = 0$	Colon	20.645 ± 0.270	1660.0 ± 110.0	21.117 ± 0.328	3885.5 ± 101.5	0.978	0.4276
	Mama	17.455 ± 1.182	1874.0 ± 57.0	17.935 ± 1.051	3832.0 ± 123.0	0.973	0.4894
	Pulmón	22.656 ± 0.087	1897.0 ± 47.0	23.391 ± 0.200	3344.5 ± 304.5	0.969	0.5672
	Media	20.252 ± 0.513	1810.3 ± 71.3	20.634 ± 0.526	3687.3 ± 176.3	0.973	0.4947
C con $\theta = 0.5$	Colon	22.221 ± 0.594	1959.0 ± 37.0	22.977 ± 0.509	3912.5 ± 59.5	0.967	0.4998
	Mama	16.657 ± 0.330	1972.5 ± 5.5	17.865 ± 0.472	3998.5 ± 0.5	0.932	0.4938
	Pulmón	24.369 ± 0.456	2000.0 ± 0.0	25.536 ± 0.288	3973.5 ± 19.5	0.954	0.5039
	Media	21.082 ± 0.460	1977.2 ± 14.2	22.126 ± 0.436	3961.5 ± 26.5	0.951	0.4992

Tabla 4.7: Diferencias entre el PSNR máximo obtenido en la época de estabilización (iteraciones 1250 a 2000) y el PSNR máximo para todo el entrenamiento al entrenar los diferentes modelos empleando ruido aleatorio como entrada.

Habiendo terminado la exposición de los resultados obtenidos en este trabajo, en el siguiente capítulo se establecerán las conclusiones a las que se ha llegado en este trabajo y posibles líneas de investigación en un futuro.

Modelo	Órgano	Métricas					
		PSNR _{fast}	Iter _{fast}	PSNR _{max}	Iter _{max}	PSNR _{fast} PSNR _{max}	t _{fast} t _{max}
A	Colon	23.551 ± 0.773	1203.0 ± 41.0	24.339 ± 0.756	3166.5 ± 463.5	0.968	0.3802
	Mama	21.367 ± 1.119	1081.0 ± 10.0	21.931 ± 1.354	3741.0 ± 247.0	0.974	0.2883
	Pulmón	25.285 ± 0.623	1169.0 ± 11.0	26.427 ± 0.653	3592.5 ± 389.5	0.957	0.3273
	Media	23.401 ± 0.838	1151.0 ± 20.7	24.232 ± 0.921	3499.8 ± 366.7	0.966	0.3319
B	Colon	20.811 ± 3.684	919.5 ± 212.5	20.811 ± 3.684	919.5 ± 212.5	1.000	1.0000
	Mama	19.228 ± 1.612	677.0 ± 427.0	19.401 ± 1.785	637.5 ± 466.5	0.991	1.0620
	Pulmón	18.970 ± 3.360	837.5 ± 99.5	19.751 ± 2.580	468.5 ± 468.5	0.960	1.7989
	Media	19.700 ± 2.885	608.0 ± 246.3	19.988 ± 2.683	675.2 ± 382.5	0.984	1.2870
C con $\theta = 0$	Colon	21.100 ± 0.480	1166.5 ± 51.5	21.260 ± 0.361	3130.5 ± 789.5	0.992	0.3735
	Mama	17.905 ± 0.967	1147.0 ± 29.0	18.505 ± 0.936	3203.5 ± 167.5	0.968	0.3603
	Pulmón	23.137 ± 0.181	1173.0 ± 34.0	23.528 ± 0.091	2652.0 ± 933.0	0.991	0.4439
	Media	20.774 ± 0.543	1162.2 ± 38.2	21.097 ± 0.472	2995.3 ± 630.0	0.984	0.3926
C con $\theta = 0.5$	Colon	20.720 ± 0.596	1041.0 ± 54.0	21.415 ± 0.499	3080.5 ± 424.5	0.968	0.3372
	Mama	17.977 ± 0.812	1017.5 ± 223.5	18.771 ± 1.007	3198.5 ± 383.5	0.958	0.3184
	Pulmón	23.260 ± 0.228	777.5 ± 430.5	23.631 ± 0.270	3446.0 ± 65.0	0.984	0.2262
	Media	20.772 ± 0.545	945.3 ± 236.0	21.272 ± 0.592	3241.7 ± 291.0	0.970	0.2939

Tabla 4.8: Diferencias entre el PSNR máximo obtenido en la época de estabilización (iteraciones 250 a 1250) y el PSNR máximo para todo el entrenamiento al entrenar los diferentes modelos empleando la imagen observada como entrada.

Capítulo 5

Conclusiones y trabajo futuro

5.1. Conclusiones

Este Trabajo Fin de Máster propone la aplicación del enfoque Deep Image Prior a arquitecturas de redes neuronales convolucionales con el objetivo de realizar la deconvolución ciega de color sobre imágenes histológicas mediante aprendizaje profundo sin requerir de grandes conjuntos de imágenes con “ground truth”.

Para poder emplear el enfoque Deep Image Prior debe fijarse la arquitectura de red neuronal a emplear, el tipo de imagen que se utilizará como entrada, que se mantendrá fija durante todo el entrenamiento; así como demás hiper-parámetros propios del entrenamiento, como la función de pérdida, optimizador, tasa de aprendizaje, etc.

En lo relativo a la arquitectura de red neuronal y función de pérdida se proponen tres modelos: (i) el modelo A, que hace uso de la red C-Net para la predicción de las concentraciones, un muestreo sobre la matriz de Ruifrok [26] para los colores y el error cuadrático medio como función de pérdida; (ii) el modelo B, que hace uso de la arquitectura BCD-Net completa para estimar concentraciones y colores, y una función de pérdida basada en el error cuadrático medio; y (iii) el modelo C, que amplía el modelo B al incorporar un regularizador con el objetivo de hacer que las predicciones de los colores se parezcan más a la referencia de Ruifrok [26]. Para el modelo C se proponen dos variantes, de manera que se estudia la influencia de utilizar o no una etapa de pre-entrenamiento enfocada a la correcta predicción de los colores en los resultados finales.

Se estudia también la influencia del tipo de entrada empleada en los resultados finales. Para ello, se entrena los tres modelos propuestos haciendo uso tanto de: (i) una imagen carente de estructura, conformada por ruido aleatorio muestreado de una distribución uniforme en el intervalo $[0,1]$; (ii) una imagen “natural”, concretamente, se usa la propia imagen a reconstruir.

Se propone también una variante del enfoque Deep Image Prior que utiliza los pesos resultantes de entrenar BCD-Net en otro conjunto de datos para inicializar las redes de los modelos B y C. Se estudian los resultados obtenidos frente a los proporcionados por los modelos propuestos al seguir Deep Image Prior “puro” y la versión amortizada de BCD-Net.

Según los resultados experimentales presentados, podemos concluir que los modelos de Aprendizaje Profundo entrenados mediante el enfoque Deep Image Prior, utilizando únicamente una imagen como referencia e inicializados aleatoriamente, logran generar resultados equiparables en la deconvolución ciega de color a los obtenidos por una red neuronal entrenada en extensos conjuntos de datos. Esto es un hito destacable, en tanto que acerca la aplicación de modelos de Aprendizaje Profundo a la deconvolución ciega de color de imágenes histológicas sin requerir de grandes conjuntos de datos, raramente disponibles.

Si se dispone de pesos resultantes de haber entrenado BCD-Net en grandes conjuntos de datos, la inicialización de los modelos B y C con dichos pesos permite obtener una mejora sustancial de rendimiento al realizar BCD, en un tiempo significativamente menor, para cada una de las imágenes procesadas individualmente.

Para lograr un rendimiento óptimo, es esencial seleccionar redes neuronales que se adapte a la tarea de la deconvolución ciega de color y una función de pérdida que considere las características particulares de las imágenes histológicas teñidas con hematoxilina y eosina.

Además, extraemos las siguientes conclusiones:

- Redes neuronales con núcleos de deconvolución demasiado complejos no permiten extraer el máximo potencial del enfoque Deep Image Prior. Quedó demostrado que el modelo A, con una arquitectura más sencilla y que muestran los colores sobre una referencia, permite obtener mejores resultados que modelos con arquitecturas más complejas que estiman tanto las concentraciones como el color de las tinciones.
- El tipo de entrada fija durante el entrenamiento del modelo influye en los resultados. En el caso del modelo A funciona mejor emplear la imagen observada, en tanto que esto le permite extraer un conocimiento general respecto de las concentraciones al comienzo del entrenamiento. Modelos más complejos que deben estimar concentraciones y tinciones, como el C, proporcionan un mejor rendimiento al emplear ruido aleatorio como entrada. No partir de ninguna información permite no asumir ciertas hipótesis iniciales que podrían degradar el rendimiento.
- Cuando se emplea la variante de Deep Image Prior en la que los modelos B y C se inicializan con los pesos resultantes del entrenamiento de BCD-Net sobre el dataset CAMELYON17 [5], se observan mejores resultados al prescindir del término regularizador de color y “confiar” en el conocimiento general extraído del gran conjunto de datos.
- Se observa que a partir de cierta etapa del entrenamiento, el rendimiento se estabiliza y no mejora significativamente más allá de un determinado umbral. Por lo tanto, sería conveniente determinar algún mecanismo de parada temprana para evitar emplear más tiempo de ejecución del necesario. Este es un aspecto complejo, en tanto que no se dispone de una referencia con la que calcular un punto de parada óptimo y diferentes imágenes pueden requerir de más o menos tiempo de

procesamiento para obtener BCDs de calidad. En la siguiente sección se propone una posible idea que es conveniente explorar en un futuro.

- Tener en cuenta las características propias de los diversos tejidos resulta vital para obtener el mejor rendimiento posible. De esta manera, diferentes modelos, funciones de pérdida, y ponderaciones de los términos de estas pueden emplearse para imágenes de distintos órganos.

5.2. Trabajo futuro

El presente trabajo propone una primera aplicación del enfoque Deep Image Prior a arquitecturas de redes neuronales convolucionales para la deconvolución ciega de color sobre imágenes histológicas. Este ha demostrado la capacidad de igualar e incluso mejorar ligeramente los resultados proporcionados por modelos amortizados entrenados en grandes conjuntos de imágenes. También se propone una variante de Deep Image Prior capaz de aprovechar el conocimiento de los modelos amortizados que permite mejorar el rendimiento. Así pues, creemos que los resultados obtenidos pueden ser mejorados en un futuro. A continuación, se plantean varias líneas que consideramos de interés para futuros trabajos.

- El modelo A debe muestrear aleatoriamente a partir de la referencia de Ruifrok [26]. En este trabajo se ha usado una varianza constante, pero creemos que los resultados podrían mejorar de variar dinámicamente la varianza en función de la etapa de entrenamiento.
- Se comprobó que la M-Net de BCD-Net podría ser demasiado compleja para aprovechar completamente el potencial del enfoque Deep Image Prior. Así pues, la utilización de una red neuronal convolucional más sencilla podría ayudar a mejorar los resultados obtenidos. En futuros trabajos podrían proponerse arquitecturas alternativas para la red encargada de la predicción de los colores y evaluar los resultados que proporcionen.
- Para los modelos B y C se observó que las mejores reconstrucciones del canal de la eosina se obtienen en las primeras etapas del entrenamiento, disminuyendo la calidad de estas en favor de las reconstrucciones de hematoxilina posteriormente. Sería interesante encontrar un mecanismo para obtener la reconstrucción más favorable para cada canal en diferentes etapas del entrenamiento, teniendo en cuenta que no se dispone de un “ground truth” que usar como referencia.
- El diseño de un mecanismo de parada para el entrenamiento de modelos que hacen uso de Deep Image Prior es una pregunta ya planteadas en [34]; en tanto que no podemos conocer el punto óptimo en que parar. En el caso de BCD, podría usarse el cálculo del PSNR o SSIM respecto de la imagen observada; de manera que si las métricas no mejoran una determinada cantidad en un número dado de iteraciones se detenga el entrenamiento. El valor de estos parámetros podría

diferir significativamente en función de las características de la imagen concreta empleada, lo que dificulta determinarlos para una aplicación genérica. Se requiere de más estudio a este respecto.

- En este trabajo se ha empleado el conjunto de datos “Warwick Stain Separation Benchmark”, compuesto por imágenes de colon, mama y pulmón. En futuros trabajos sería interesante estudiar el rendimiento de los modelos propuestos para imágenes de tejidos de otros órganos.

Apéndice A

Anexo I

Detallamos en este anexo las herramientas software y hardware que se han utilizado. También se expondrá una pequeña guía de uso del software empleado para ejecutar el entrenamiento de los modelos propuestos.

Para poder desarrollar este trabajo se ha empleado como lenguaje de programación *Python*, en tanto que permite trabajar con redes neuronales, manejar y analizar datos de forma sencilla y cómoda. La implementación base del modelo BCD-Net puede ser encontrada en [37], (correspondiente a las publicaciones [38] [40]) mientras que el código fuente desarrollado en el marco de este Trabajo Fin de Máster se encuentra publicado como software libre en el repositorio [9].

Para realizar dicha implementación, así como las variantes propuestas en este trabajo, se utilizó el paquete software *PyTorch*, que proporciona las herramientas necesarias para trabajar con redes neuronales. Para la realización del análisis de los datos obtenidos se empleó la liberaría *Scipy*, la cual proporciona gran cantidad de funciones matemáticas. Para la generación de la gran mayoría de los gráficos expuestos se utilizó el paquete software *Matplotlib*; mientras que para algunas figuras más complejas, como las arquitecturas de los modelos propuestos, se empleó *Draw.io*.

Los requisitos necesarios más importantes para ejecutar el software desarrollado se listan a continuación. Nótese que en el fichero “enviroment.yaml” del repositorio [9] se encuentran listados todos los paquetes que se han empleado junto a sus versiones específicas.

- Python 3.7 o superior.
- PyTorch 1.12 o 1.13.
- Scipy 1.10.1 o superior.
- Matplotlib 3.7.1 o superior.
- Jupyter 1.0.0 o superior.
- Tarjeta gráfica con un mínimo de 6 GB de memoria VRAM. Se recomienda emplear una tarjeta gráfica con 8 GB de VRAM o más.

Téngase en cuenta que de utilizar la versión 2.0 o superior de *PyTorch* puede que se deban realizar ajustes menores al código para su correcto funcionamiento. Se recomienda utilizar los paquetes que proporcionan soporte para CUDA o ROCm, en tanto que la ejecución de los modelos en un procesador de propósito general puede demorar una cantidad de tiempo muy elevada y difícilmente asumible.

Para ejecutar los experimentos necesarios se utilizó un servidor del “Visual Information Processing Group” de la Universidad de Granada, equipado con procesadores Intel Xeon y tarjetas gráficas NVIDIA GeForce RTX 3090. Durante la etapa de desarrollo, se utilizó un equipo doméstico equipado con un procesador Intel Core i7-8700K y una tarjeta gráfica NVIDIA GeForce GTX 1080.

Para ejecutar el bucle de entrenamiento de los modelos propuestos, el lector deberá moverse a la carpeta “*code/dip*”, donde se encuentran diversos scripts *Python* y cuadernos de *Jupyter Notebook*. De querer ejecutar el entrenamiento para un número de imágenes considerable se recomienda usar el script *Python* provisto con tal fin.

El archivo llamado “*options.py*” contiene una lista de todos los parámetros que pueden ser utilizados como argumentos para los scripts de *Python* a través de la terminal. En el caso de los cuadernos de *Jupyter*, estos valores están definidos como constantes en las primeras celdas. Algunos de estos parámetros incluyen: el dispositivo para ejecutar el entrenamiento, el modelo a entrenar, el tipo de entrada a utilizar, el conjunto de imágenes a emplear (individuales, por órgano o todo el conjunto de datos), su ruta, el número de iteraciones del entrenamiento y el valor θ usado para ponderar términos en la ecuación de pérdida [3.6](#), entre otros.

Es importante destacar que en el código fuente se emplea una nomenclatura distinta para hacer referencia a cada uno de los modelos propuestos. A continuación, se establecen las equivalencias:

- “cnet_e2”. En este trabajo recibe el nombre de modelo A, o C-Net_{MSE}.
- “bcdnet_e1”. Este modelo es identificado en este trabajo como modelo B, o BCD-Net_{MSE}.
- “bcdnet_e2”. Denotado como modelo C, o BCD-Net_{MSE+KL}, si no se hace uso de pre-entrenamiento enfocado al color.
- “bcdnet_e3”. Caso en que el modelo C tenga una fase de pre-entrenamiento enfocado al color.

En el caso de querer generar diversos tipos de gráficas y reportes de métricas, el lugar al que acudir será la carpeta “*code/doc_generation*”. El fichero más relevante de esta carpeta es “*generate_performance_metrics.py*”. Este cuenta con funciones para generar los informes con los valores de las métricas expuestas en este trabajo, tanto para ficheros CSV individuales como para todos los generados en el marco de un entrenamiento para todo el conjunto de imágenes; así como funciones para generar diversos típicos de gráficas y realizar análisis estadísticos. Todas las funciones cuentan con la documentación correspondiente que explica como utilizarlas. Además, se proporciona un ejemplo de utilización para los métodos más relevantes.

Apéndice B

Anexo II

En este segundo anexo se detalla la evolución de las métricas y las imágenes generadas por los diferentes modelos al emplear la imagen observada como entrada.

En las figuras B.1 y B.2 se muestra la evolución de PSNR y SSIM durante el entrenamiento de las imágenes seleccionadas para los modelos propuestos.

Si el lector compara dichas figuras con las expuestas para el caso del ruido aleatorio en la sección 4.3, podrá observar un comportamiento sumamente similar. La diferencia radica principalmente en que las etapas iniciales, en las que las reconstrucciones mejoran su calidad rápidamente y posteriormente algo más lentamente, son de una duración significativamente menor. En tanto que la red no debe aprender desde una imagen desprovista de estructura los modelos tienden a reconstruir las concentraciones con una mayor facilidad y rapidez.

En el caso del modelo C, que cuenta con un regularizador para el color, las tonalidades propias de cada tinción también se aprenden con mayor velocidad. Para el modelo B, aún sin regularizador, la información inicial que aporta la imagen observada permite disminuir ligeramente los casos en que se predicen colores poco fieles a la realidad o un canal tiende a contener toda la información cromática.

En la figura B.3 se puede observar la evolución del proceso de entrenamiento para la imagen “Breast_0” al usar la propia imagen como entrada del modelo A. Es destacable como en la primera iteración ya se tiene una cantidad considerable de las concentraciones, mientras que los resultados en apenas 200 iteraciones son bastante satisfactorios. En el caso de haber empleado ruido aleatorio se hubiera requerido una cantidad mayor de iteraciones para alcanzar reconstrucciones similares. Para las reconstrucciones del modelo C se observa un comportamiento sumamente parecido al expuesto en la figura B.3.

En el caso de la figura B.4 podemos observar como el modelo B sigue generando reconstrucciones cuyos colores no son factibles, en tanto que se carece de un regularizador de color que guíe el grado de corrección de los colores predichos. Los colores verdes de la iteración 250 se tornan azules en torno a la iteración 500, no variando para el resto del proceso de entrenamiento.

En la figura B.5 se puede observar el proceso de entrenamiento para una imagen de colon al hacer uso del modelo C con pre-entrenamiento. Se puede apreciar como en

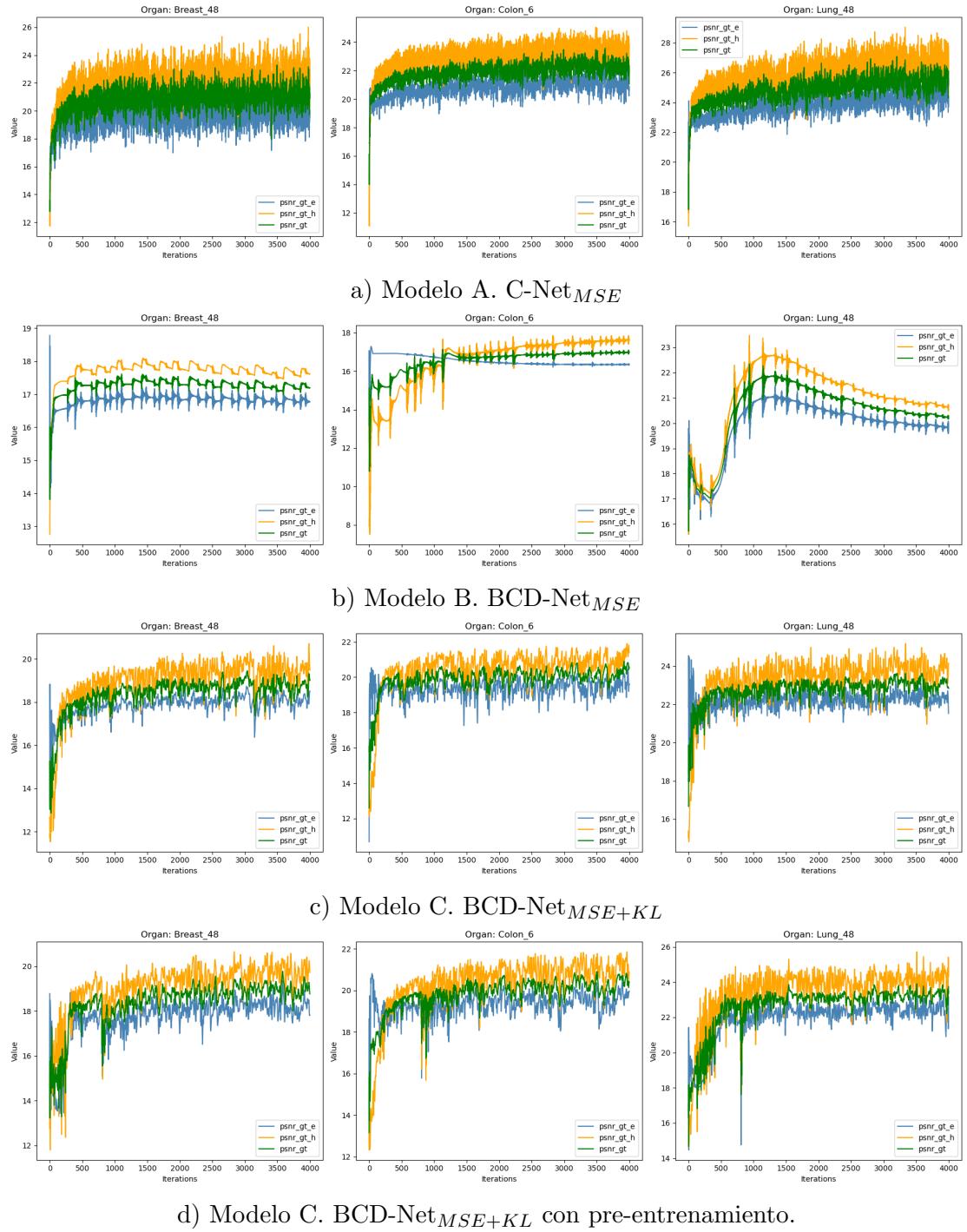


Figura B.1: Evolución del PSNR para los entrenamientos de las imágenes seleccionadas al partir de la imagen observada para los diferentes modelos propuestos.

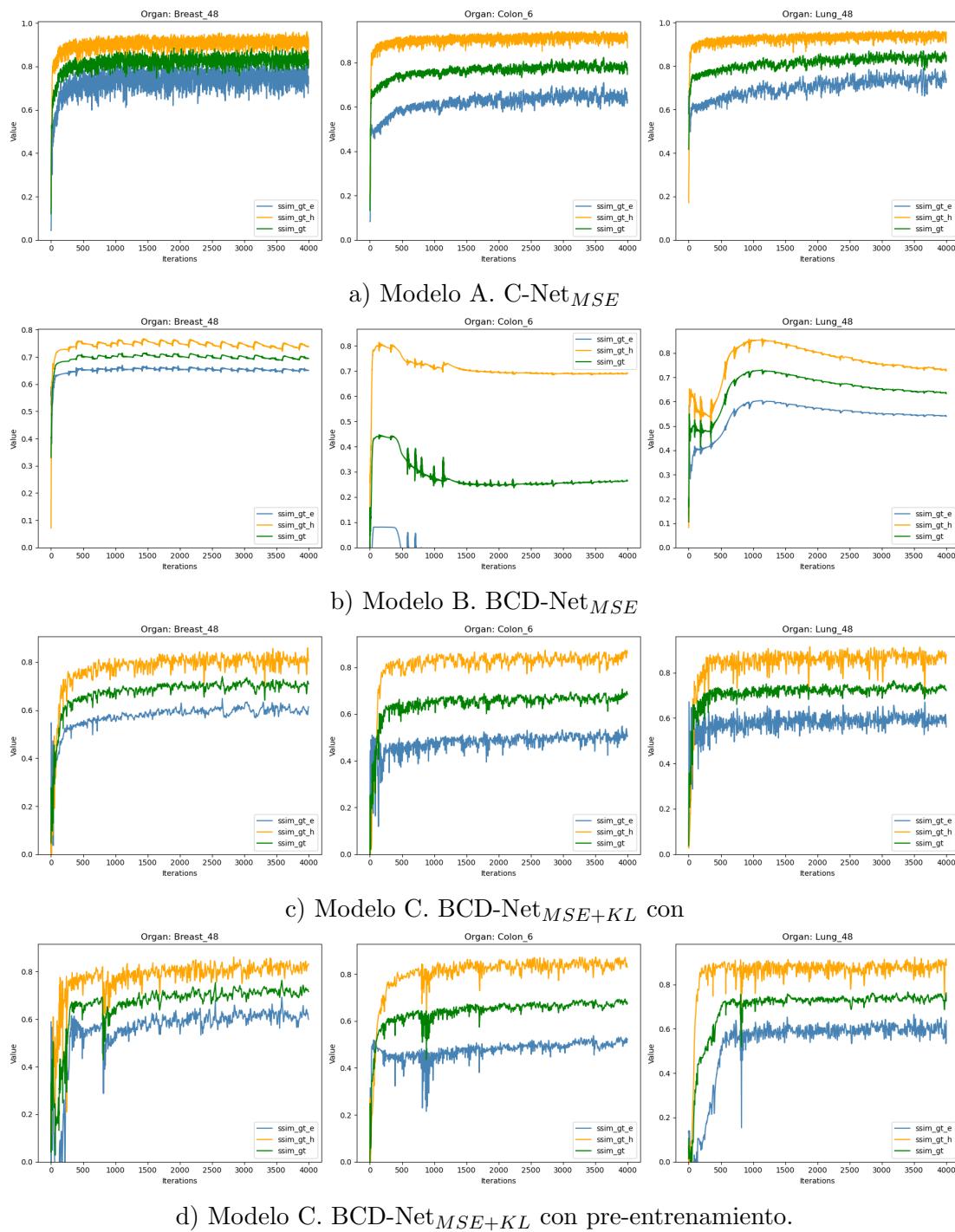
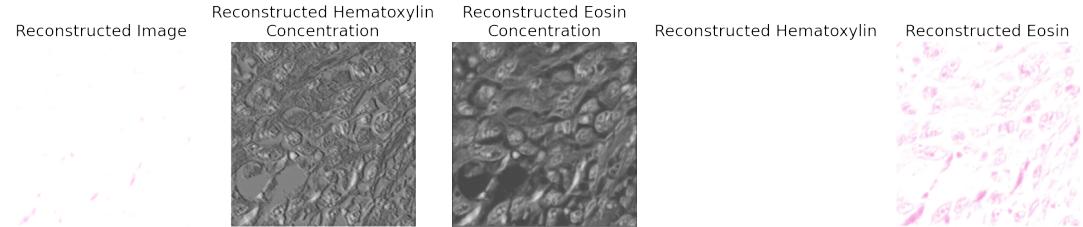


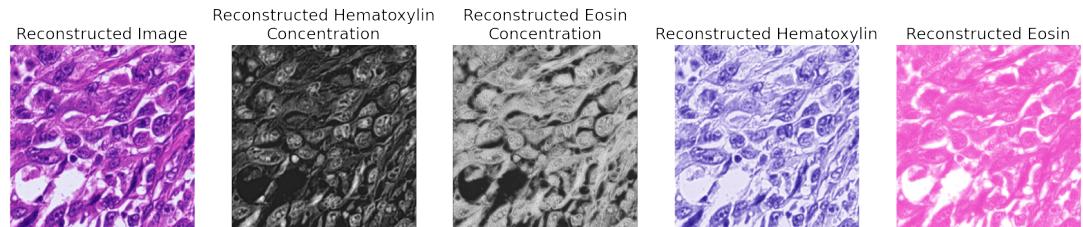
Figura B.2: Evolución del SSIM para los entrenamientos de las imágenes seleccionadas al partir de la imagen observada para los diferentes modelos propuestos.



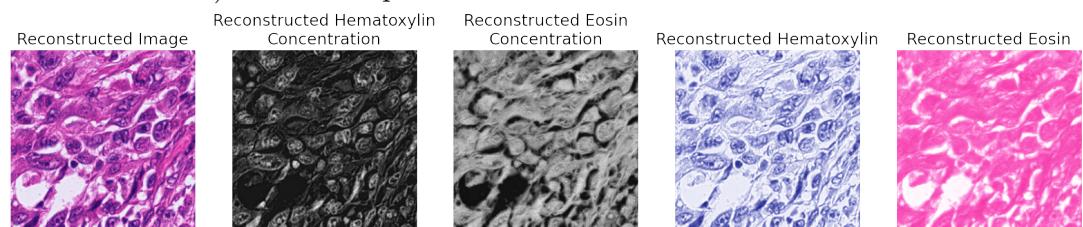
a) "Ground truth" para la imagen "Breast_0".



b) Resultados para la primera iteración del entrenamiento.



c) Resultados para la iteración 200 del entrenamiento.



d) Resultados para la iteración 600 del entrenamiento.

Figura B.3: Evolución de los resultados obtenidos para la imagen "Breast_0" al entrenar utilizando el modelo A y la imagen observada como entrada.

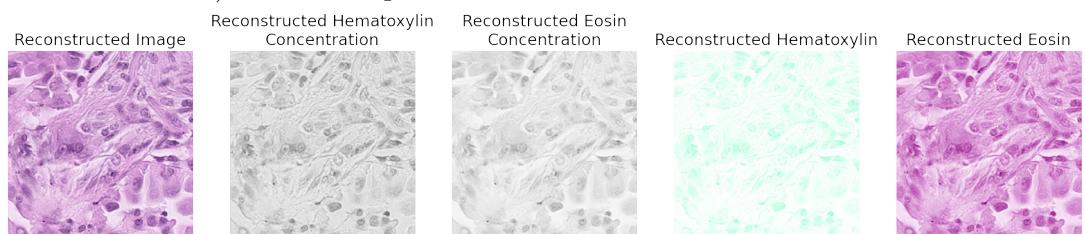
la primera iteración ya se dispone de información básica sobre las concentraciones. En 250 iteraciones se obtiene un resultado bastante correcto, el cual mejorará ligeramente conforme avance el entrenamiento. Si el lector compara detenidamente las imágenes de las iteraciones 250 y 750 se dará cuenta de que en la iteración 750 existe una mayor nitidez en las reconstrucciones de las concentraciones.



a) “Ground truth” para la imagen “Lung_0”.



b) Resultados para la iteración 250 del entrenamiento.



c) Resultados para la iteración 500 del entrenamiento.

Figura B.4: Evolución de los resultados obtenidos para la imagen “Lung_0” al entrenar utilizando el modelo B y la imagen observada como entrada.

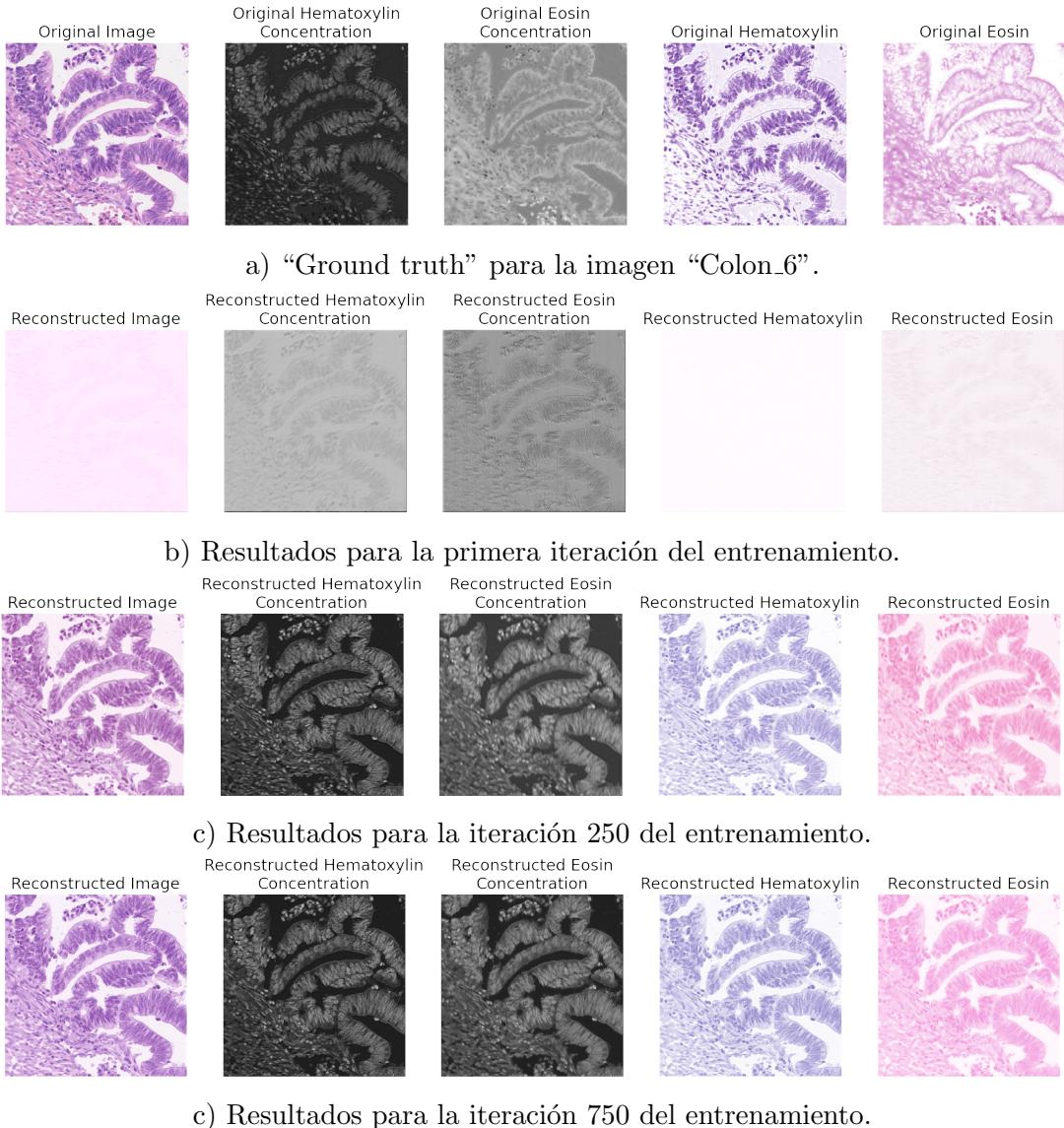


Figura B.5: Evolución de los resultados obtenidos para la imagen "Colon_6" al entrenar utilizando el modelo C con pre-entrenamiento y la imagen observada como entrada.

Apéndice C

Anexo III

En este apéndice se detalla un experimento adicional que se realizó.

Se probó a emplear la variante de Deep Image Prior que emplea como inicialización los pesos disponibles de BCD-Net, haciendo uso como entrada de la red de ruido aleatorio muestreado de una distribución uniforme en el rango [0,1]. Este experimento se realizó tanto para el modelo B (que sólo tiene en cuenta el error cuadrático medio en su función de pérdida), como para el modelo C (que considera tanto el MSE como la divergencia de Kullback Leibler, en una ponderación 50-50).

Los resultados en términos de PSNR y SSIM para sendos experimentos se muestran en las tablas [C.1](#) y [C.2](#).

Órgano	Métricas			
	PSNR_GT_H	PSNR_GT_E	SSIM_GT_H	SSIM_GT_E
Colon	12.830 ± 1.691	10.797 ± 1.453	0.071 ± 0.057	0.291 ± 0.004
Pulmón	11.860 ± 0.060	11.267 ± 0.003	0.046 ± 0.003	0.232 ± 0.007
Mama	11.353 ± 0.688	9.993 ± 0.575	0.454 ± 0.136	0.523 ± 0.161
Media	12.014 ± 0.813	10.686 ± 0.677	0.190 ± 0.065	0.349 ± 0.057

Tabla C.1: Valores medios y desviaciones típicas del PSNR para cada órgano al emplear la variante de Deep Image Prior que inicializa el modelo B con los pesos disponibles de BCD-Net y usa como entrada ruido aleatorio.

En el caso de emplear el modelo B los resultados son absolutamente desastrosos. La red es incapaz de predecir correctamente dado un ruido en blanco y negro empleando únicamente el MSE, y esto lastra la calidad de las reconstrucciones de las concentraciones, de las cuales únicamente se distingue una estructura muy general sin ningún detalle reseñable. Sin embargo, al utilizar el modelo C (que incorpora el regularizador de color) la red sí que es capaz de realizar una reconstrucción coherente de la imagen dada una entrada de ruido, aunque los resultados proporcionados son de una menor calidad que las reconstrucciones generadas al usar Deep Image Prior “puro” o realizar este mismo proceso empleando como entrada la propia imagen histológica a reconstruir.

Órgano	Métricas			
	PSNR_GT_H	PSNR_GT_E	SSIM_GT_H	SSIM_GT_E
Colon	21.385 ± 0.820	20.376 ± 0.155	0.840 ± 0.019	0.639 ± 0.133
Pulmón	23.633 ± 0.653	22.587 ± 0.091	0.874 ± 0.012	0.625 ± 0.005
Mama	18.765 ± 0.837	16.836 ± 1.184	0.820 ± 0.003	0.648 ± 0.065
Media	21.261 ± 0.770	19.933 ± 0.477	0.845 ± 0.011	0.637 ± 0.067

Tabla C.2: Valores medios y desviaciones típicas del PSNR para cada órgano al emplear la variante de Deep Image Prior que inicializa el modelo C con los pesos disponibles de BCD-Net y usa como entrada ruido aleatorio.

Apéndice D

Anexo IV

En este anexo se encuentra una comparativa más exhaustiva de la influencia del tipo de entrada sobre cada uno de los modelos propuestos para cada órgano.

En la tabla **D.1** se presentan los resultados obtenidos en términos de PSNR y SSIM al entrenar el modelo A con ambos tipos de entrada. En esta se puede observar como la utilización de la imagen observada permite al modelo proporcionar un mejor desempeño en la gran mayoría de las situaciones. Las mayores mejoras en términos de PNSR se observan para el caso de colon; siendo estas algo más modestas para pulmón y mama. En media, emplear la imagen observada como entrada permite obtener unas ganancias de PSNR y SSIM en torno al 3 % y 2 % respectivamente para todo el conjunto de datos.

En la tabla **D.2** se presentan los resultados obtenidos al entrenar el modelo B con ambos tipos de entrada, ruido e imagen observada. Los resultados obtenidos son variados.

Para el caso de colon se obtiene una gran mejora al reconstruir el canal de hematoxilina, tanto en términos de PNSR como de SSIM al emplear la imagen observada; mientras que para el canal de eosina se obtiene una gran mejora de SSIM al emplear ruido aleatorio. Para el caso de pulmón se obtienen mejores resultados al emplear ruido aleatorio como entrada, especialmente para el canal de la eosina. Por lo contrario, al reconstruir imágenes de mama se obtienen mejoras significativas al emplear la imagen observada; especialmente en el caso de la eosina. Al no poseer regularizador para el color, las predicciones del color de las tinciones varían significativamente en función del órgano, lo que propicia los resultados expuestos.

En media para todo el conjunto de datos, emplear la imagen como entrada proporciona un desempeño mayor del modelo; en torno de un 7.4 % para el PSNR y un 14.6 % para el SSIM.

En la tabla **D.3** se presentan los resultados correspondientes al entrenar el modelo C, con etapa de pre-entrenamiento enfocada a la predicción del color, para ambos tipos de entrada.

En este caso, podemos observar un comportamiento completamente distinto al presentado por el modelo B. Para el caso de las imágenes de colon y pulmón se obtiene una notable mejoría de entre un 7 y 8.2 % para el caso del PSNR; y de entre un 6.5 y 17 % para el SSIM en función del canal. En contraposición, las imágenes de mama parecen

Órgano	Tinción	Métricas			
		PSNR (ruido aleatorio)	PSNR (imagen observada)	SSIM (ruido aleatorio)	SSIM (imagen observada)
Colon	H	24.656 ± 0.905	25.025 ± 0.385	0.916 ± 0.007	0.936 ± 0.002
	E	22.080 ± 0.713	23.654 ± 1.127	0.759 ± 0.112	0.804 ± 0.096
Pulmón	H	27.686 ± 0.346	27.992 ± 0.638	0.944 ± 0.006	0.953 ± 0.012
	E	24.505 ± 0.609	24.861 ± 0.668	0.765 ± 0.008	0.759 ± 0.024
Mama	H	22.927 ± 1.564	23.078 ± 0.832	0.934 ± 0.003	0.940 ± 0.003
	E	19.170 ± 1.156	20.784 ± 1.876	0.817 ± 0.022	0.851 ± 0.006
Media	H	25.089 ± 0.938	25.365 ± 0.618	0.931 ± 0.005	0.943 ± 0.006
	E	21.918 ± 0.826	23.010 ± 1.224	0.708 ± 0.142	0.805 ± 0.042
	Media	23.504 ± 0.882	24.188 ± 0.921	0.855 ± 0.025	0.874 ± 0.024

Tabla D.1: Valores medios y desviaciones típicas del PSNR y SSIM para el modelo A al emplear ruido aleatorio y la imagen observada como entrada. En negrita se marcan los mejores resultados de entre los proporcionados por los diferentes modelos.

Órgano	Tinción	Métricas			
		PSNR (ruido aleatorio)	PSNR (imagen observada)	SSIM (ruido aleatorio)	SSIM (imagen observada)
Colon	H	17.375 ± 2.054	21.849 ± 4.186	0.630 ± 0.121	0.877 ± 0.064
	E	19.711 ± 0.596	19.774 ± 3.182	0.688 ± 0.117	0.480 ± 0.399
Pulmón	H	19.371 ± 3.607	18.664 ± 4.803	0.586 ± 0.261	0.579 ± 0.281
	E	22.327 ± 1.318	20.837 ± 0.356	0.573 ± 0.015	0.518 ± 0.084
Mama	H	16.639 ± 2.240	19.941 ± 2.140	0.799 ± 0.072	0.846 ± 0.086
	E	16.302 ± 1.490	18.861 ± 1.430	0.450 ± 0.396	0.791 ± 0.121
Media	H	17.795 ± 2.633	20.151 ± 3.710	0.672 ± 0.151	0.767 ± 0.144
	E	19.447 ± 1.135	19.832 ± 1.656	0.570 ± 0.176	0.596 ± 0.201
	Media	18.621 ± 1.884	19.992 ± 2.683	0.595 ± 0.164	0.682 ± 0.173

Tabla D.2: Valores medios y desviaciones típicas del PSNR y SSIM para el modelo B al emplear ruido aleatorio y la imagen observada como entrada. En negrita se marcan los mejores resultados de entre los proporcionados por los diferentes modelos.

reconstruirse mejor de emplear la imagen como entrada. La mejora de PSNR será de entre un 4 y 6.5 %; mientras que la de SSIM oscilará entre el 6 y 8 % en función del canal. Este comportamiento puede deberse a la capacidad de la red de aprender mejor los colores al contar con algo de información adicional al inicio, en tanto que usa la imagen observada y los colores de los tejidos de mama suelen alejarse más de la referencia.

Para todo el conjunto de imágenes, la utilización de ruido aleatorio en la entrada proporciona mejores resultados medios; cercanos al 4 % para el PSNR y 5 % para el SSIM.

Órgano	Tinción	Métricas			
		PSNR (ruido aleatorio)	PSNR (imagen observada)	SSIM (ruido aleatorio)	SSIM (imagen observada)
Colon	H	24.051 ± 0.510	22.408 ± 0.670	0.921 ± 0.004	0.857 ± 0.012
	E	21.902 ± 0.507	20.422 ± 0.328	0.762 ± 0.113	0.653 ± 0.135
Pulmón	H	26.710 ± 0.386	24.740 ± 0.979	0.938 ± 0.007	0.882 ± 0.008
	E	24.362 ± 0.189	22.523 ± 0.440	0.750 ± 0.004	0.641 ± 0.002
Mama	H	18.491 ± 0.715	19.562 ± 0.686	0.780 ± 0.093	0.843 ± 0.008
	E	17.240 ± 0.228	17.980 ± 1.329	0.683 ± 0.120	0.726 ± 0.034
Media	H	23.084 ± 0.537	22.237 ± 0.778	0.880 ± 0.027	0.861 ± 0.009
	E	21.168 ± 0.308	20.315 ± 0.699	0.732 ± 0.079	0.673 ± 0.057
	Media	22.126 ± 0.423	21.276 ± 0.739	0.806 ± 0.053	0.767 ± 0.330

Tabla D.3: Valores medios y desviación típica del PSNR y SSIM para el modelo C, con etapa de pre-entrenamiento enfocada al color, al emplear ruido aleatorio y la imagen observada como entrada. En negrita se marcan los mejores resultados

Bibliografía

- [1] Esther Abels, Liron Pantanowitz, Famke Aeffner, Mark D Zarella, Jeroen van der Laak, Marilyn M Bui, Venkata NP Vemuri, Anil V Parwani, Jeff Gibbs, Emmanuel Agosto-Arroyo, Andrew H Beck, and Cleopatra Kozlowski. Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the Digital Pathology Association. *The Journal of Pathology*, 249(3):286–294, 2019.
- [2] Shahira Abousamra, Danielle Fassler, Le Hou, Yuwei Zhang, Rajarsi Gupta, Tahsin Kurc, Luisa F. Escobar-Hoyos, Dimitris Samaras, Beatrice Knudson, Kenneth Shroyer, Joel Saltz, and Chao Chen. Weakly-Supervised Deep Stain Decomposition for Multiplex IHC Images. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 481–485, 2020.
- [3] Najah Alsubaie, Nicholas Trahearn, Shan E. Ahmed Raza, David Snead, and Nasir M. Rajpoot. Stain Deconvolution Using Statistical Analysis of Multi-Resolution Stain Colour Representation. *PLOS ONE*, 12(1):1–15, 01 2017.
- [4] John D. Bancroft and Christopher Layton. 10 - The hematoxylins and eosin. In S. Kim Suvarna, Christopher Layton, and John D. Bancroft, editors, *Bancroft's Theory and Practice of Histological Techniques (Eighth Edition)*, pages 126–138. Elsevier, eighth edition edition, 2019.
- [5] Péter Bárdi, Oscar G. F. Geessink, Quirine F. Manson, Marcory Crf van Dijk, Maschenka C. A. Balkenhol, Meyke Hermsen, Babak Ehteshami Bejnordi, Byungjae Lee, Kyunghyun Paeng, Aoxiao Zhong, Quanzheng Li, Farhad Ghazvinian Zanjani, Svitlana Zinger, Keisuke Fukuta, Daisuke Komura, Vlado Ovtcharov, Shenghua Cheng, Shaoqun Zeng, Jeppe Thagaard, Anders Bjorholm Dahl, Huangjing Lin, Hao Chen, Ludwig Jacobsson, Martin Hedlund, Melih Çetin, Eren Halici, Hunter Jackson, Richard Chen, Fabian Both, Jörg K.H. Franke, Heidi V.N. Küsters-Vandervelde, W. Vreuls, Peter Bult, Bram van Ginneken, Jeroen A. van der Laak, and Geert J. S. Litjens. From Detection of Individual Metastases to Classification of Lymph Node Status at the Patient Level: The CAMELYON17 Challenge. *IEEE Transactions on Medical Imaging*, 38:550–560, 2019.

- [6] Pranjal Datta. All about Structural Similarity Index (SSIM): Theory + Code in PyTorch. <https://medium.com/srm-mic/all-about-structural-similarity-index-ssim-theory-code-in-pytorch-6551b455541e>, 2020.
- [7] Izak B. Dimenstein. Grossing biopsies: an introduction to general principles and techniques. *Annals of Diagnostic Pathology*, 13(2):106–113, 2009.
- [8] Rahul Duggal, Anubha Gupta, Ritu Gupta, and Pramit Mallick. Sd-layer: Stain deconvolutional layer for cnns in medical microscopic imaging. In Maxime Descoteaux, Lena Maier-Hein, Alfred Franz, Pierre Jannin, D. Louis Collins, and Simon Duchesne, editors, *Medical Image Computing and Computer Assisted Intervention. MICCAI 2017*, pages 435–443, Cham, 2017. Springer International Publishing.
- [9] José Alberto Gómez García. Deconvolución Ciega de Imágenes Histológicas Usando Aprendizaje Profundo. <https://github.com/modejota/BCD-Net-Deep-Image-Prior>, 2024.
- [10] Milan Gavrilović, Jimmy C. Azar, Joakim Lindblad, Carolina Wählby, Ewert Bengtsson, Christer Busch, and Ingrid Carlbom. Blind Color Decomposition of Histological Images. *IEEE Transactions on Medical Imaging*, 32:983–994, 2013.
- [11] N. Hidalgo-Gavira, J. Mateos, M. Vega, R. Molina, and A.K. Katsaggelos. Blind Color Deconvolution of Histopathological Images using a variational Bayesian Approach. In *IEEE International Conference on Image Processing (ICIP 2018)*, pages 983–987. Athens (Greece), October 2018.
- [12] Neel Kanwal, Fernando Pérez-Bueno, Arne Schmidt, Kjersti Engan, and Rafael Molina. The Devil is in the Details: Whole Slide Image Acquisition and Processing for Artifacts Detection, Color Variation, and Data Augmentation: A Review. *IEEE Access*, 10:58821–58844, 2022.
- [13] Adnan Khan, Nasir Rajpoot, Darren Treanor, and Derek Magee. A Non-Linear Mapping Approach to Stain Normalisation in Digital Histopathology Images using Image-Specific Colour Deconvolution. *IEEE Transactions on Biomedical Engineering*, 61, 06 2014.
- [14] Ilya Loshchilov and Frank Hutter. Decoupled Weight Decay Regularization, 2019.
- [15] Marc Macenko, Marc Niethammer, J. Marron, David Borland, John Woosley, Xiaojun Guan, Charles Schmitt, and Nancy Thomas. A Method for Normalizing Histology Slides for Quantitative Analysis. In *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, volume 9, pages 1107–1110, 06 2009.
- [16] Niccolò Marini, Manfredo Atzori, Juan Otálora Montenegro, Stephane Marchand-Maillet, and Henning Müller. H&E-adversarial network: a convolutional neural network to learn stain-invariant features through Hematoxylin & Eosin regression, 10 2021.

- [17] MathWorks. Compute peak signal-to-noise ratio (PSNR) between images - MathWorks. <https://es.mathworks.com/help/vision/ref/psnr.html>, 2023.
- [18] Sandra Morales, Kjersti Engan, and Valery Naranjo. Artificial Intelligence in Computational Pathology – Challenges and Future Directions. *Digit. Signal Process.*, 119(C), dec 2021.
- [19] Gabriel Prieto Renieblas. Uso de SSIM como índice de calidad de imagen médica. Docta Complutense, 2009. Accessed: 28 dic 2023.
- [20] Fernando Pérez-Bueno, Miguel López Pérez, Miguel Vega, Javier Mateos, Valery Naranjo, Rafael Molina, and Aggelos Katsaggelos. A TV-based image processing framework for blind color deconvolution and classification of histological images. *Digital Signal Processing*, 101:102727, 03 2020.
- [21] Fernando Pérez-Bueno, Juan G. Serra, Miguel Vega, Javier Mateos, Rafael Molina, and Aggelos K. Katsaggelos. Bayesian K-SVD for H and E blind color deconvolution. Applications to stain normalization, data augmentation and cancer classification. *Computerized Medical Imaging and Graphics*, 97, 2022.
- [22] Fernando Pérez-Bueno, Miguel Vega, Valery Naranjo, Rafael Molina, and Aggelos Katsaggelos. Super Gaussian Priors for Blind Color Deconvolution of Histological Images. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 3010–3014, 10 2020.
- [23] Rabinovich, Andrew and Agarwal, Sameer and Laris, Casey and Price, Jeffrey and Belongie, Serge. Unsupervised color decomposition of histologically stained tissue samples. In S. Thrun, L. Saul, and B. Schölkopf, editors, *Advances in Neural Information Processing Systems*, volume 16. MIT Press, 2003.
- [24] Ren, Dongwei and Zhang, Kai and Wang, Qilong and Hu, Qinghua and Zuo, Wangmeng. Neural blind deconvolution using deep priors, 08 2019.
- [25] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. *CoRR*, abs/1505.04597, 2015.
- [26] Arnout Ruifrok and Dennis Johnston. Ruifrok ac, johnston da. quantification of histochemical staining by color deconvolution. anal quant cytol histol 23: 291-299. *Analytical and quantitative cytology and histology / the International Academy of Cytology [and] American Society of Cytology*, 23:291–9, 09 2001.
- [27] Massimo Salvi, Nicola Michielli, and Filippo Molinari. Stain Color Adaptive Normalization (SCAN) algorithm: Separation and standardization of histological stains in digital pathology. *Computer Methods and Programs in Biomedicine*, 193:105506, 04 2020.
- [28] Rebecca L. Siegel, Kimberly D. Miller, Nikita Sandeep Wagle, and Ahmedin Jemal. Cancer statistics, 2023. *CA: A Cancer Journal for Clinicians*, 73(1):17–48, 2023.

- [29] David Tellez, Maschenka Balkenhol, Irene Otte-Holler, Rob van de Loo, Rob Vogels, Peter Bult, Carla Wauters, Willem Vreuls, Suzanne Mol, Nico Karssemeijer, Geert Litjens, Jeroen van der Laak, and Francesco Ciompi. Whole-Slide Mitosis Detection in H&E Breast Histology Using PHH3 as a Reference to Train Distilled Stain-Invariant Convolutional Networks. *IEEE transactions on medical imaging*, March 2018.
- [30] David Tellez, Geert Litjens, Péter Bárdi, Wouter Bulten, John-Melle Bokhorst, Francesco Ciompi, and Jeroen van der Laak. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. *Medical Image Analysis*, 58:101544, 2019.
- [31] Thaína Tosta, Paulo de Faria, Leandro Neves, and Marcelo Zanchetta do Nascimento. Computational normalization of H&E-stained histological images: Progress, challenges and future potential. *Artificial Intelligence in Medicine*, 95, 11 2018.
- [32] Nicholas Trahearn, David Snead, Ian Cree, and Nasir Rajpoot. Multi-class stain separation using independent component analysis. In Metin N. Gurcan and Anant Madabhushi, editors, *Medical Imaging 2015: Digital Pathology*, volume 9420 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 94200J, March 2015.
- [33] Dmitry Ulyanov. Dmitry Ulyanov - Deep Image Prior conference at MAN AHL. https://www.youtube.com/watch?v=-g1NsTuP1_I, 2018.
- [34] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep Image Prior. *arXiv:1711.10925*, 2017.
- [35] Jared Vicory, Heather D. Couture, Nancy E. Thomas, David Borland, J.S. Marron, John Woosley, and Marc Niethammer. Appearance normalization of histology slides. *Computerized Medical Imaging and Graphics*, 43:89–98, 2015.
- [36] Philipp Wirth. Which Optimizer should I use for my ML Project? *Lightly.ai*, 2020.
- [37] Shuowen Yang, Fernando Pérez-Bueno, Francisco M Castro-Macías, Rafael Molina, and Aggelos K Katsaggelos. Deep Bayesian Blind Color Deconvolution of Histological Images. <https://github.com/vipgugr/BCD-Net>, 2023.
- [38] Shuowen Yang, Fernando Pérez-Bueno, Francisco M. Castro-Macías, Rafael Molina, and Aggelos K. Katsaggelos. Deep Bayesian Blind Color Deconvolution of Histological Images. In *2023 IEEE International Conference on Image Processing (ICIP)*, pages 710–714, 2023.
- [39] Shuowen Yang, Fernando Pérez-Bueno, Francisco M. Castro-Macías, Rafael Molina, and Aggelos K. Katsaggelos. Huggingface - bcd-net weights. <https://huggingface.co/Franblueee/BCD-Net/tree/main>, 2023.

- [40] Shuowen Yang, Fernando Pérez-Bueno, Francisco M. Castro-Macías, Rafael Molina, and Aggelos K. Katsaggelos. BCD-net: Stain separation of histological images using deep variational Bayesian blind color deconvolution. *Digital Signal Processing*, 145:104318, 2024.
- [41] Zongsheng Yue, Hongwei Yong, Qian Zhao, Lei Zhang, and Deyu Meng. Variational Denoising Network: Toward Blind Noise Modeling and Removal. *CoRR*, abs/1908.11314, 2019.
- [42] Qian Zhao, Hui Wang, Zongsheng Yue, and Deyu Meng. A deep variational Bayesian framework for blind image deblurring. *Knowledge-Based Systems*, 249:109008, 2022.
- [43] Yushan Zheng, Zhiguo Jiang, Haopeng Zhang, Fengying Xie, Hu Dingyi, Shujiao Sun, Jun Shi, and Chenghai Xue. Stain Standardization Capsule for Application-Driven Histopathological Image Normalization. *IEEE Journal of Biomedical and Health Informatics*, PP:1–1, 03 2020.
- [44] Yushan Zheng, Zhiguo Jiang, Haopeng Zhang, Fengying Xie, Jun Shi, and Chenghai Xue. Adaptive Color Deconvolution for Histological WSI Normalization. *Computer Methods and Programs in Biomedicine*, 170, 03 2019.