Universitat Oberta de Catalunya

uoc.edu

Mejora para la imagen de Rayos X mediante el uso de *Deep Learning*

Patricia Moreno Berdón

Máster de Bioinformática y Bioestadística

Área de Machine Learning

Nombre Director: Edwin Santiago Alférez Baquero

Nombre Directora: Mónica García Abella

Nombre Profesor/a responsable de la assignatura: Laura Calvet Liñan

Fecha Entrega: 16/06/2021

FICHA DEL TRABAJO FINAL

Título del trabajo:	*Mejora para la imagen de Rayos X mediante el uso de Deep Learning*
Nombre del autor:	*Patricia Moreno Berdón*
Nombre del consultor/a:	*Edwin Santiago Alférez Baquero*
Nombre del PRA:	*Laura Calvet Liñan*
Fecha de entrega (mm/aaaa):	06/2021
Titulación:	*Máster de Bioinformática y Bioestadística*
Área del Trabajo Final:	*Machine Learning*
Idioma del trabajo:	Castellano
Número de créditos:	15
Palabras clave	*Rayos X, mejora de imagen, aprendizaje profundo*

Resumen del Trabajo:

La imagen de rayos X es una herramienta clave para el diagnóstico. Sin embargo, a pesar de su valor clínico, ofrece un contraste en tejido blando pobre así como una alta dosis de radiación. Es por ello que mejorar el contraste y tratar de llegar a una dosis mínima con una calidad de imagen compatible con el diagnóstico es necesario.

Los métodos de realce convencionales para imagen radiológica se basan en la concatenación de soluciones para cada uno de estos dos problemas (contraste y ruido), proporcionando buenos resultados pero con un alto coste computacional que dificulta el procesamiento en tiempo real, necesario en radiografía clínica.

En este trabajo se propone un método que integre la mejora de contraste y la reducción de ruido para imágenes radiológicas mediante aprendizaje profundo que permita eliminar

la necesidad de encontrar los parámetros adecuados para los distintos estudios y tener la imagen en un tiempo de ejecución por debajo del segundo.

Para ello se realiza un estudio bibliográfico de los métodos nuevos basados en aprendizaje profundo para mejora del ruido y contraste seleccionando la arquitectura UNet con distintas funciones de coste y distintas arquitecturas de codificador de redes neuronales convolucionales.

Tras evaluación visual de los resultados se establece que las mejores funciones de coste son el índice de similitud estructural multiescala y el índice de similitud estructural multiescala combinado con el error absoluto medio. Entre las arquitecturas probadas para el codificador, las que proporcionan mejores resultados son ResNet34 y EfficientNetB3.

Por lo tanto, se ha propuesto un nuevo método basado en aprendizaje profundo que permite mejorar el contraste y reducir el ruido de imágenes de radiología animal en menos de 1s.

Abstract:

X-ray imaging is an important diagnostic tool. However, despite its clinical value, it offers poor soft tissue contrast as well as a high radiation dose. That is why improving the contrast and trying to reach a minimum dose with an image quality compatible with the diagnosis is necessary.

Conventional enhancement methods for radiological imaging are based on the concatenation of solutions for each of these two problems (contrast and noise), providing good results but with a high computational cost that makes real-time processing, necessary in clinical radiography, difficult.

In this work, we proposed a method that integrates contrast enhancement and noise reduction for radiological images through deep learning, eliminating the need to find the appropriate parameters for the different studies and having the processed image in less than a second.

A bibliographic study of new methods based on deep learning for noise and contrast improvement was made, selecting the UNet architecture with different cost functions and different convolutional neural network encoder architectures.

After visual evaluation of the results, it is established that the best cost functions are the multiscale structural similarity index and the multiscale structural similarity index combined with the mean absolute error. Among the tested architectures for the encoder, those that provide the best results are ResNet34 and EfficientNetB3.

Therefore, a new method based on deep learning has been proposed that allows to improve the contrast and reduce the noise of animal radiology images in less than 1s.

Índice

Lista de figuras

Lista de tablas

1 Introducción

1.1 Contexto y justificación del Trabajo

Desde su descubrimiento, la imagen de rayos X ha sido una herramienta clave para el diagnóstico. Sin embargo, a pesar de su valor clínico, ofrece un contraste en tejido blando pobre. En la literatura se pueden encontrar diferentes métodos de realce de contraste para imagen radiológica centradas, normalmente, en mejorar el contraste en un rango determinado del histograma, lo que es insuficiente en estudios con un rango dinámico muy grande, como son los de tórax, donde tenemos tejido muy poco denso como los pulmones y tejido muy denso como la columna. Por otro lado, según el Consejo de Seguridad Nuclear (CSN), la dosis media para la población española se ha estimado en un total de 3,7 mSv (miliSievert) cada año, de los cuales la mayor parte corresponden a radiación por usos médicos (35,1%). La dosis media por usos médicos, para cada miembro de la población de un país del nivel sanitario I (como es el caso de España), se estima por la UNSCEAR en 1,28 mSv por año, de los cuales 1,2 mSv se deben a técnicas de diagnosis con rayos X [1]. Es por ello que se debe minimizar la dosis recibida en estos exámenes médicos alcanzando un compromiso entre una dosis mínima con una calidad de imagen compatible con el diagnóstico, ya que al disminuir la dosis el ruido en la imagen aumenta.

Los métodos de realce para imagen radiológica se basan en la concatenación de soluciones para cada uno de estos dos problemas (contraste y ruido). Para hacer frente al primero de los problemas, se han propuesto algoritmos basados en la ecualización de histograma [2-4]. Estos métodos producen una mejora de contraste global de la imagen, pero con un aspecto poco “realista”, pudiendo dificultar el diagnóstico. Otros métodos usados para la mejora del contraste son las transformaciones gamma [5] o sigma, que proporcionan buenos resultados, pero requieren de la selección de parámetros dependientes de la región anatómica. Para hacer frente al problema del ruido debido a la reducción de dosis, sin perder detalle en la imagen, se usan comúnmente métodos basados en la pirámide laplaciana [6] o la transformada wavelet [7]. Aunque estos métodos proporcionan buenos resultados, implican un alto coste computacional dificultando el procesamiento en tiempo real, necesario en radiografía clínica.

Más recientemente, se han usado los métodos de aprendizaje profundo, especialmente las redes neuronales de convolución profunda (CNN), para la reducción de ruido en radiografía clínica [8] o para la mejora de la calidad de la imagen en radiografía mecánica [9], así como en otras disciplinas de imagen médica como la Tomografía axial computarizada [8, 10].

Un método que integre la mejora de contraste y la reducción de ruido para imágenes radiológicas mediante aprendizaje profundo permitiría eliminar la necesidad de encontrar los parámetros adecuados para los distintos estudios y tener la imagen en un tiempo de ejecución por debajo del segundo.

1.2 Objetivos del Trabajo

El presente Trabajo Fin de Máster tiene como objetivo principal el diseño e implementación de una red neuronal convolucional que permita la eliminación del ruido y mejora del contraste en imágenes radiológicas.

De esta forma, se pueden distinguir los siguientes objetivos secundarios:

- Revisión de la literatura, estudiando las limitaciones de la imagen radiológica y explorando los métodos convencionales de procesado de imagen radiológica.
- Implementación de métodos de procesado convencionales.
- Selección de un modelo y creación de una base de datos para su entrenamiento y validación.
- Entrenamiento del modelo elegido.
- Evaluación del modelo en imágenes reales comparando con un método de procesado convencional y con otras arquitecturas de red.
- Implementación de un software funcional para la eliminación del ruido y mejora del contraste en imágenes radiológicas.

1.3 Enfoque y método seguido

La estrategia escogida consiste en la adaptación de la arquitectura de red UNet, ya que gracias a la estructura de codificador-decodificador y las *skip connections* (conexiones de salto), permite mantener los detalles de la imagen [11]. La red está previamente entrenada para el conjunto de imágenes de ImageNet [12]. ImageNet es un proyecto cuyo objetivo

es ser un banco de datos para la investigación y desarrollo de sofware especializado en reconocimiento de imágenes. Está compuesto por ILSVRC (*ImageNet Large Scale Visual Recognition Challenge*), un subconjunto de las imágenes de ImageNet que contiene 1,2 millones de imágenes de 1000 clases diferentes. Gracias a que la red está previamente entrenada, podemos usarla como punto de partida y adaptarla posteriormente a nuestro problema, de esta forma la red ya habrá aprendido a extraer las características complicadas.

En cuanto a la metodología, se distinguen cuatro fases principales:

Fase 1. Análisis

Esta primera fase se enfoca en la investigación del estado del arte del problema planteado, tanto en métodos convencionales como en los de aprendizaje profundo orientados a la reducción de ruido y mejora de contraste de imagen en imagen de Rayos X. También se hará una investigación de los métodos de aprendizaje profundo en mejora de imagen en general para tener una visión global de los métodos usados.

Fase 2. Diseño e implementación

Los modelos seleccionados deberán ser reentrenados para el conjunto de imágenes seleccionado y, si procede, modificarlos en busca del mejor rendimiento. Para ello, se incluirán las siguientes subtareas.

- Elección del modelo.
- Selección de imágenes con distinto voltaje y dosis (mAs: miliamperios por segundo) de distintos tamaños de animal y diferentes partes del cuerpo del Hospital Clínico Veterinario de la Universidad Complutense de Madrid (HCV).
- Preparación de la base de datos: creación de parches.

Fase 3. Evaluación

Evaluación de la imagen resultante con respecto a un “gold standard”. Esta imagen se obtiene del procesamiento de las adquisiciones por un especialista radiólogo mediante un software de procesamiento con métodos convencionales.

Fase 4. Documentación

La última fase consiste en la generación de la presente memoria y de contribución a un congreso nacional.

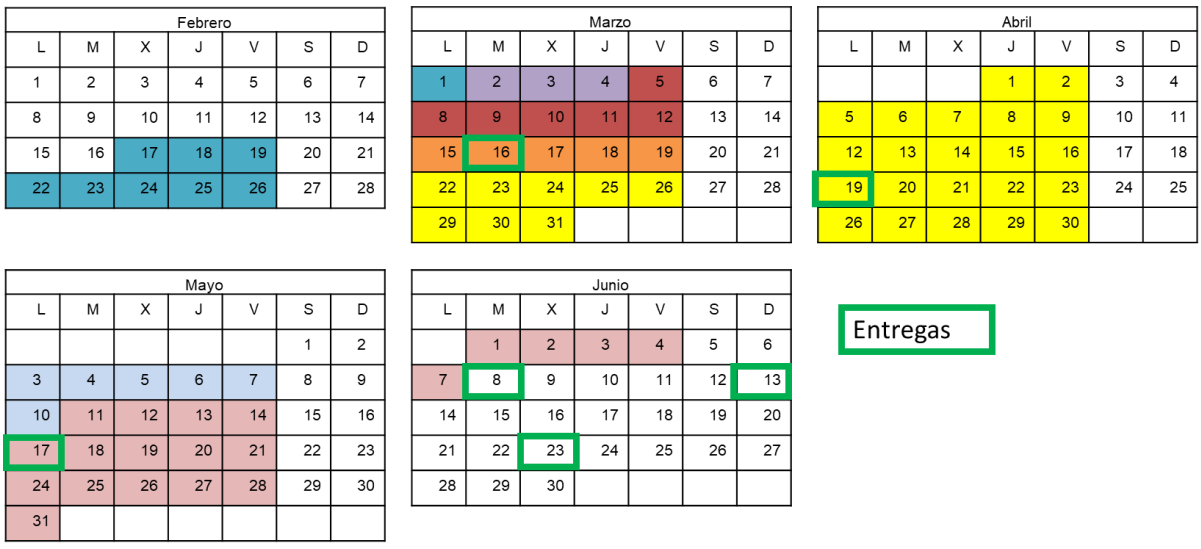
1.4 Planificación del Trabajo

1.4.1 Tareas, calendario e hitos.

La siguiente tabla muestra la duración de las tareas asignadas a las diferentes fases expuestas previamente.

Tarea	Días
1. Análisis	10
2.Diseño e implementación	50
Elección de la red	3
Selección de imágenes	6
Preparación de la base de datos	5
Adaptación del código de la red y entrenamiento	30
3.Evaluación	6
4.Documentación	20

Tabla 1. Duración de las tareas



Hitos

Análisis del estado del arte	1-Marzo
Elección de la red	4-Marzo
Selección de imágenes	12-Marzo
Preparación de la base de datos	19-Marzo
Adaptación del código de la red y entrenamiento	30-Abril
Evaluación	10-Mayo
Documentación	7-Junio

Tabla 2. Hitos del trabajo

1.4.2 Análisis de riesgos

El grupo donde se desarrolla el proyecto cuenta con experiencia en el desarrollo de herramientas de aprendizaje profundo como la que se propone en este trabajo y con la colaboración del HCV que proveerá de los datos radiológicos necesarios, reduciendo los riesgos potenciales. No obstante, se señalan a continuación los siguientes potenciales riesgos con su plan de contingencia:

- Imposibilidad de adquirir las imágenes debido a limitaciones en el acceso a la máquina de Rayos X del Hospital Clínico Veterinario por problemas en la certificación de la máquina. En tal caso se buscarán imágenes similares que hayan sido previamente adquiridas.
- Base de datos con un número insuficiente de imágenes debido a no poder adquirir más animales. Ello se solucionará mediante técnicas de data augmentation.
- No conseguir eliminar el ruido de las imágenes o mejorar el contraste de las imágenes. Se planteará la realización de dos redes separadas, una que mejore el ruido y otra el contraste.
- Problemas en la optimización de las redes debido al tamaño de las imágenes. Para hacer frente a este problema, se plantea el entrenamiento de los modelos en distintas máquinas para acelerar el proceso y la posibilidad de entrenar implementar un script para ejecutar el entrenamiento fuera de horas de trabajo.

- Si por cuestiones sanitarias no se puede acceder al lugar de trabajo se programará un acceso virtual al ordenador. Esta opción ya está implementada.

1.4.3 Costes

Los costes se analizan en las siguientes tablas:

Cargo	Tiempo invertido (horas)	Salario/hora (€)	Coste (€)
Ingeniero investigador	800	30	13000
Ingeniero investigador de apoyo	100	30	3000
Coordinador del proyecto	150	40	6000
Investigador informático	50	40	2000
		TOTAL:	**24000**

Tabla 3. Desglose de costes de personal

Para los cotes de material, se consideran cinco años de depreciación para el cálculo del costo de los equipos:

Material	Coste unitario	Coste/año (€)	Tiempo de uso	Coste (€)
Ordenador	1000	200	6 meses	100
GPU	439	87.8	6 meses	43.9
Licencia Matlab		800	6 meses	400
Licencia Python	0	0	0	0
			TOTAL:	**543.9**

Tabla 4. Desglose de costes materiales

Los costes de servicio incluyen los costes relacionados con la adquisición de las imágenes con un equipo de Rayos X.

Servicio	Coste/número de estudios (€)	Número de estudios	Coste (€)
Equipo de Rayos X en Hospital Veterinario UCM	50	400	20000
Servicio	**Coste/hora (€)**	**Tiempo empleado**	**Coste (€)**
Técnico de Rayos X en Hospital Veterinario UCM	15	20	300
		TOTAL:	**23000**

Tabla 5. Desglose de costes del servicio

El coste indirecto se considera como el 15% de los costes previamente mencionados de acuerdo a la legislación del Instituto de Investigación Sanitaria Gregorio Marañón. Siendo la suma de los costes 47543.9 €, el coste indirecto es 0.15×47543.9 = 7131.585€.

Los costes totales se calculan como la suma de los costes previos.

Concepto	Coste (€)
Costes personales	24000
Costes de material	543.9
Costes de servicio	23000
Costes indirectos	7131.585
TOTAL	**54675.485**

Tabla 6. Desglose de costes totales

1.5 Breve sumario de contribuciones y productos obtenidos

Los productos obtenidos a la finalización del trabajo serán:

- Prototipo de software para la mejora de imagen de Rayos X.
- Memoria completa con la metodología e información de los resultados.
- Presentación PPT y vídeo explicativo.
- Presentación al CASEIB de 2021.

1.6 Breve descripción de los capítulos de la memoria

La memoria se ha estructurado en las siguientes secciones:

- Capítulo 1. Introducción: se describe la motivación del trabajo, indicando los objetivos y tareas asociadas, con un calendario que detalla la programación temporal.
- Capítulo 2. Estado del arte: breve introducción a los métodos de aprendizaje profundo y estudio bibliográfico de los métodos nuevos basados en aprendizaje profundo para mejora del ruido y contraste.
- Capítulo 3. Materiales y métodos: incluye el método propuesto y la generación de la base de datos.
- Capítulo 4. Resultados: se resumen los resultados obtenidos del análisis realizado por los diferentes modelos usados con la misma base de datos.
- Capítulo 5. Discusión y conclusiones: valoración global del trabajo y posibles líneas de investigación futuras.
- Glosario: definición de términos y acrónimos usados en el TFM.
- Bibliografía: bibliografía usada en el TFM.

2 Estado del arte

2.1 Redes Neuronales Artificiales

Las Redes Neuronales Artificiales (RNA) son sistemas de procesamiento de la información cuya estructura y funcionamiento están inspirados en las redes neuronales biológicas. Están compuestas por un conjunto nodos, llamados neuronas artificiales, conectados con otros nodos de diferente capa neuronal, de forma que, a partir de un vector de entrada, generan una respuesta única. Generalmente la RNA está organizada en capas.

- Capa de entrada: reciben la información desde el exterior de la red.
- Capa de salida: envían la información hacia el exterior de la red.
- Capas ocultas: se encuentran en el medio de las capas de entrada y de salida y no tienen contacto con el exterior.

Existen distintos tipos de RNA según la tipología de la red entre los que destacamos el Perceptrón Multicapa, las Redes Neuronales Recurrentes y las Redes Neuronales Convolucionales [13].

El perceptrón multicapa está compuesto de una o más capas de neuronas en las que la información pasa en una dirección y está orientado a problemas clasificación donde a las entradas se les asigna una clase o etiqueta. En la Figura 1 se muestra una arquitectura de perceptrón multicapa que consta de una capa de entrada, una capa oculta y una capa de salida. Un problema común de este tipo de redes es el problema de desvanecimiento y explotación del gradiente [14, 15]. Este problema está asociado a la retropropagación, método por el que aprenden las redes neuronales y que está basado en el descenso de gradiente y la regla de la cadena. En el caso de redes neuronales muy profundas, el gradiente se desvanece o explota conforme se retropropaga. Para solucionar este problema se proponen las Redes Neuronales Recurrentes y las Redes Neuronales Convolucionales.

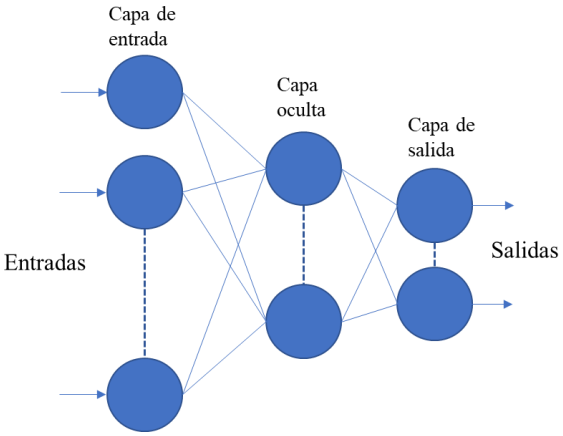


Figura 1. Arquitectura perceptrón multicapa

Las Redes Neuronales Recurrentes (RNN) pueden tener conexiones recurrentes en las neuronas ocultas que permiten recordar el estado de la neurona previa y son comúnmente usadas en reconocimiento de la voz o reconocimiento de la escritura a mano. En el caso de RNN profundas, también surge el problema de desvanecimiento y explotación de gradiente. La Figura 2 muestra una RNN en la que se pueden observar las conexiones recurrentes en las neuronas de la capa oculta.

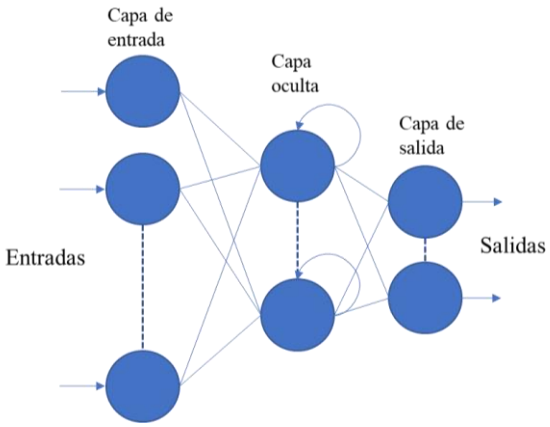


Figura 2 . Arquitectura Red Neuronal Recurrente

Las Redes Neuronales Convolucionales (CNN) [16], que contienen múltiples capas para procesar distintos aspectos de los datos de entrada y son las más usadas para reconocimiento y clasificación de imágenes. Los componentes básicos de las CNN son filtros en lugar de perceptrones, cuyos parámetros son los propios del filtro. Cada capa de una CNN se compone de varios filtros que extraen algunos mapas de características de la imagen de entrada. Estos mapas de características se pasan a través de una siguiente capa de filtros hasta que se alcanza la capa de salida.

Las capas convolucionales de CNN tienen parámetros específicos, generalmente llamados hiperparámetros, que determinan la salida de esta capa: tamaño del filtro, número de filtros, *padding* y *stride*.

- El tamaño del filtro. Controla el área de la imagen de entrada a la que el filtro es sensible. Un tamaño de filtro más grande tiene un campo receptivo más grande, mientras que los tamaños de filtro pequeños se enfocan en características en un campo más pequeño.
- El número de filtros especifica la dimensión del mapa de características de salida. Un número mayor de filtros implica un mapa de características más grande. Se usa para compensar modelos en los que el tamaño de la matriz disminuye con respecto a la CNN y se usa un mayor número de filtros para retener la mayor cantidad de información posible de la entrada. Sin embargo, una gran cantidad de filtros puede limitar el proceso de entrenamiento ya que la cantidad total de parámetros del modelo también aumentaría.
- El *padding* se utiliza para controlar la salida de una convolución en los bordes de la imagen de entrada, ya que algunos valores del filtro pueden no coincidir con valores de la imagen. El *zero padding* es la técnica más utilizada para asignar un valor cero a aquellos valores del filtro que no se pudieron asignar a ningún valor en la imagen.
- El *Stride* controla el paso de píxeles en cada convolución a lo largo de la entrada. Si *stride* se establece en uno, el filtro se mueve un píxel para aplicar otra convolución. Este parámetro se utiliza para evitar la superposición del campo receptivo de convoluciones secuenciales.

Las redes CNN se forman usando tres tipos de capas: capas convolucionales, capas de *pooling*,y capas totalmente conectadas (*fully connected*), aplicadas al final [17].

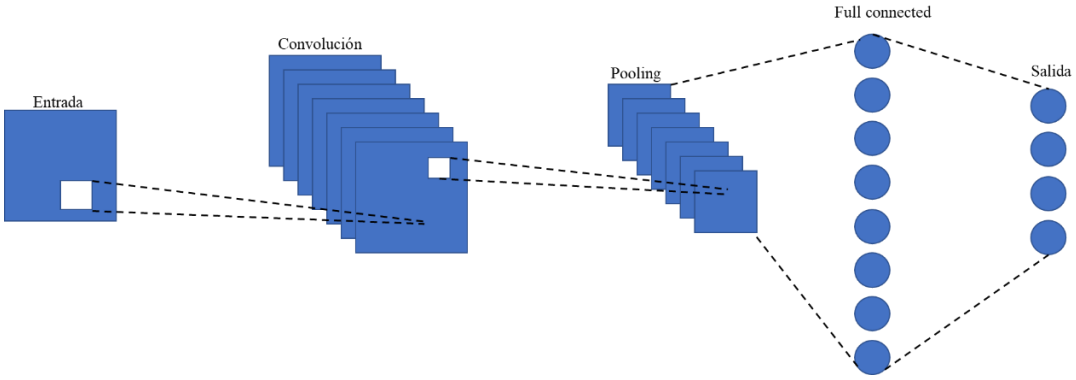


Figura 3. Arquitectura Red Neuronal Convolucional

La capa convolucional extrae distintas características de la imagen de entrada mediante la aplicación de filtros. La salida de esta capa convolucional son los mapas de características, que nos proporcionan información sobre los bordes de la imagen. Este mapa de características se pasa por la capa de *pooling*.

El *pooling* es un operador no lineal que se utiliza para reducir el tamaño del mapa de características y así reducir el corte computacional. Existen distintos tipos de *pooling* dependiendo de la operación utilizada. El *max pooling* es el más utilizado, ya que mantiene los valores de píxeles más brillantes en el mapa de características, generalmente relacionados con los bordes o la información relevante. Dada una ventana, coge el máximo valor dentro de esta. El *average pooling* calcula la media de los elementos en una ventana.

La capa totalmente conectada consiste en la conexión de los resultados de la capa anterior con todos los nodos de la capa siguiente.

2.2 Modelos de CNN

Existen distintos tipos de CNN diseñadas para procesado de imagen y reconocimiento de objetos entre las que podemos encontrar:

- LetNet [18], fue el primero en términos de clasificación de objetos. Inicialmente fue entrenada para clasificar dígitos del 0 al 9.
- AlexNet [19], ganador del desafío ImageNet ILSVCR 2012. Es uno de los modelos de referencia en las tareas de reconocimiento de objetos y visión por computadora. Una desventaja de esta red es que tiene muchos hiperparámetros.
- VGGNet [20], mostró el mejor desempeño en el desafío ImageNet 2014. La principal ventaja que introdujo VGGNet fue la reducción del número de hyperparámetros mediante la introducción del concepto de capas convolucionales dobles, cada una con un tamaño de kernel de 3x3. Se utiliza principalmente en tareas de clasificación de imágenes. Esta red tiene como desventajas un tiempo largo de entrenamiento computacionalmente caro.
- ResNet [21] fue introducido en ImageNet 2015 por Microsoft mostrando mejores resultados de rendimiento que los humanos en tareas de clasificación de objetos. El enfoque novedoso introducido por ResNet es que utilizan *skip connections [22]* y *batch normalization [23]*, permitiendo entrenar redes más profundas.

Sin embargo, en otras tareas de procesamiento de imágenes como la segmentación de imágenes o la transformación de imágenes, la salida deseada no es un vector de probabilidades sino la imagen original donde cada valor de píxel se ha transformado o indica la clase a la que pertenece. Esta idea condujo a la arquitectura UNet [11]. La idea básica es extraer características de alto nivel de la imagen y disminuir secuencialmente el tamaño de la matriz y finalmente usar estas características para restaurar el tamaño de la matriz original con la información deseada.

Su arquitectura se puede considerar en términos generales como una red de codificadores seguida de una red de decodificadores. El codificador es la primera mitad del diagrama de arquitectura (Figura 4). Por lo general, es una red de clasificación previamente entrenada como VGG / ResNet donde aplica bloques de convolución seguidos de una disminución de resolución que sirve para la extracción de características.

El decodificador es la segunda mitad de la arquitectura. El objetivo restaurar el mapa de características a la resolución original. El decodificador consta de muestreo y concatenación seguidos de operaciones de convolución.

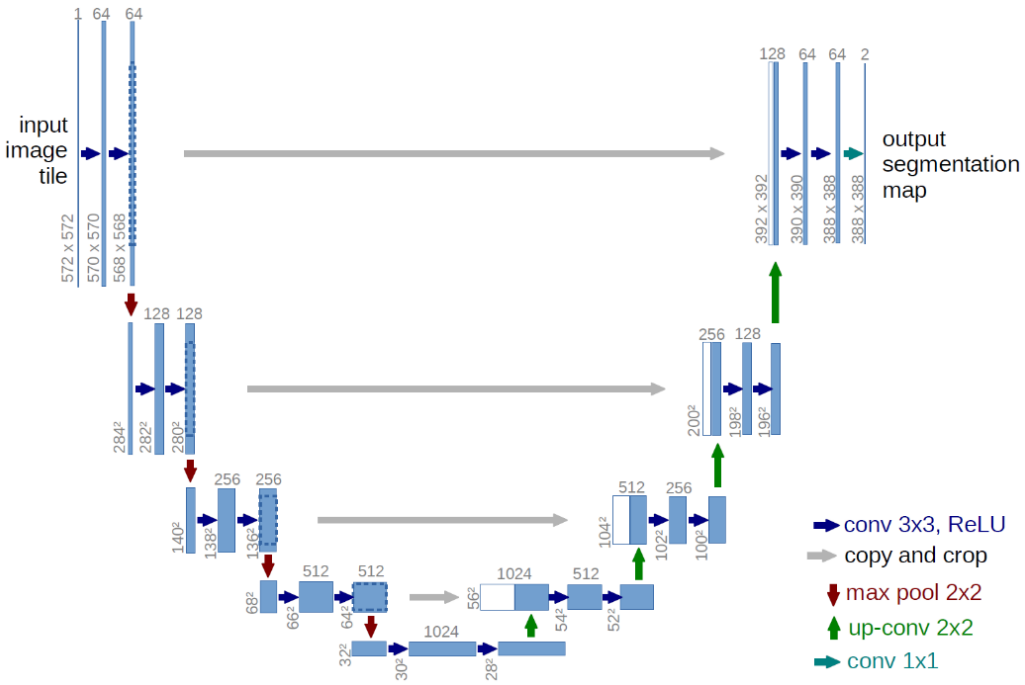


Figura 4. Arquitectura UNet [11]

2.3 Funciones de coste

Las funciones de coste se usan en las redes neuronales para medir la diferencia entre la imagen referencia y la imagen predicha.

La mayoría de las técnicas de evaluación de la calidad de la imagen se basan en la cuantificación de errores entre una imagen referencia y una imagen de muestra. Una métrica común es cuantificar la diferencia en los valores de cada uno de los píxeles correspondientes entre la muestra y las imágenes de referencia, utilizando por ejemplo el error cuadrático medio o el error absoluto medio.

2.3.1 Error cuadrático medio

El error cuadrático medio (MSE) es la suma de las diferencias cuadráticas entre la imagen referencia (y) y la salida de la red esta ($\hat{y}$).

$$MSE = \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N} \quad (1)$$

Siendo N el tamaño de la imagen.

2.3.2 Error absoluto medio

El error absoluto medio (MAE) es la suma de las diferencias absolutas entre la imagen referencia y la imagen predicha.

$$MAE = \frac{\sum_{i=1}^{N}|y_i - \hat{y}_i|}{N} \quad (2)$$

Estas funciones de coste tienen el problema de que puede haber dos imágenes con el mismo MSE pero una más distorsionada que la otra [24].

2.3.3 Índice de similitud estructural, índice de similitud estructural medio e índice de similitud estructural multiescalaa

El sistema de percepción visual humana es altamente capaz de identificar información estructural de una imagen y, por lo tanto, identificar las diferencias entre la información extraída de una referencia y una escena de muestra. Por lo tanto, una función de coste que

replica este comportamiento funcionará mejor en tareas que implican diferenciar entre una muestra y una imagen de referencia. Es por ello que surge el índice de similitud estructural (SSIM) [24].

Este método extrae 3 características clave de una imagen: la luminancia, el contraste y la estructura. La comparación entre las dos imágenes se realiza teniendo como base estas 3 características.

- La luminancia en una imagen X se mide promediando todos los valores de píxeles. Se denota por μ_x y la fórmula se da a continuación:

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{3}$$

- El contraste se mide tomando la desviación de todos los valores de píxeles. Se denota por σ y se representa mediante la fórmula siguiente:

$$\sigma_x = \left(\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)^2\right)^{\frac{1}{2}} \tag{4}$$

- La comparación estructural se realiza mediante la división de la señal de entrada por su desviación estándar para que el resultado tenga una desviación estándar unitaria.

$$\frac{(X - \mu_x)}{\sigma_x} \tag{5}$$

Tomando estos 3 parámetros y siendo X la imagen predicha, Y la imagen de referencia, L el rango dinámico de la imagen y $K_1 = 0.01, K_2 = 0.03$, se define la función de comparación de luminancia como:

$$l(X,Y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \tag{6}$$

Siendo c_1 una constante para asegurar que el denominador no es 0. c_1 de define como:

$$c_1 = (K_1 L)^2 \tag{7}$$

La función de comparación de contraste viene dada por:

$$c(X,Y) = \frac{2\sigma_x\sigma_y \ + \ c_2}{\sigma_x^2 \ + \ \sigma_y^2 \ + \ c_2} \tag{8}$$

Siendo c_2 una constante para asegurar que el denominador no es 0. c_2 de define como:

$$c_2 = (K_2 L)^2 \tag{9}$$

La función de comparación estructural se define como:

$$s(X,Y) = \frac{\sigma_{xy} \ + \ c_3}{\sigma_x\sigma_y \ + c_3} \tag{10}$$

Siendo σ_{xy},

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y) \tag{11}$$

Con todo ello, la función del SSIM final viene dada por:

$$SSIM(X,Y) = [l(X,Y)]^{\alpha}[c(X,Y)]^{\beta}[s(X,Y)]^{\gamma} \tag{12}$$

Donde $\alpha > 0, \beta > 0, \gamma > 0$ denotan la importancia de cada métrica. Para simplificar la ecuación se asume que $\alpha = \beta = \gamma = 1$ y $c_3 = \frac{c_2}{2}$, resultando el SSIM por:

$$SSIM(X,Y) = \frac{(2\mu_x\mu_y + c_1)}{(\mu_x^2 + \mu_y^2 + c_1)} \frac{(2\sigma_x\sigma_y + c_2)}{(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (13)$$

Quedando definida la función de coste como:

$$L_{SSIM} = 1 - SSIM(X,Y) \quad (14)$$

En [24] los autores señalan que para evaluar la calidad de la imagen es mejor aplicar el SSIM localmente, es decir, en pequeñas regiones y tomando luego la media de todas, en vez de globalmente. Este método se llama normalmente índice de similitud estructural medio (MSSIM). Los autores utilizan una función de ponderación gaussiana simétrica circular de tamaño de matriz 11x11 que se mueve por la imagen. Usan una ventana suavizada para evitar artefactos de bloque en la imagen [25].

En cada paso de movimiento de la ventana se calcula el SSIM. Cuando se han recorrido toda la imagen se toma la media de los valores del SSIM.

$$MSSIM(X,Y) = \frac{1}{M}\sum_{j=1}^{M} SSIM(x_j - y_j) \quad (15)$$

Quedando definida la función de coste como:

$$L_{MSSIM} = 1 - MSSIM(X,Y) \quad (16)$$

Este método tiene el inconveniente de que no incorpora detalles de la imagen a distintas resoluciones. Con este fin se define en [26] un SSIM multiescala (multi-scale SSIM), que se basa en aplicar de forma iterativa un filtro de paso bajo y un submuestreo de la imagen por un factor de 2. Se calcula la comparación de contraste y de la comparación estructural después de cada submuestreo. La comparación de luminancia se calcula solo al final, después de haber realizado el bucle iterativo M veces. El SSIM final se obtiene combinando las medidas a diferentes escalas usando:

$$MS\ SSIM(X,Y) = [\mathrm{l_M}(x,y)]^{\alpha_M} \prod_{j=1}^{M} [c_j(X,Y)]^{\beta_j} [s_j(X,Y)]^{\gamma_j} \tag{17}$$

Para simplificar la definición de parámetros definen $\alpha_j = \beta_j = \gamma_j$ para todas las j.

Con ello la función de coste queda definida como:

$$L_{MS\ SSIM} = 1 - MS\ SSIM(X,Y) \tag{18}$$

Tanto con el MS-SSIM como SSIM puede suceder que haya cambios en el brillo de la imagen. Dado que el MS-SSIM conserva el contraste y el MAE conserva la luminancia, en [27] proponen combinarlos ambos.

En [28] sugirió usar una red CNN previamente entrenada (VGG16) para medir la similitud entre dos imágenes. Esta nueva métrica se llama *perceptual loss* y se basa en pasar la imagen predicha X y la imagen de referencia Y por esta nueva red (VGG16) y medir la diferencia entre las resultantes. Esta nueva función de coste se define como:

$$L_{VGG} = \frac{1}{N} \sum_{i=1}^{N} (VGG16(x_i) - VGG16(y_i))^2 \tag{19}$$

2.4 Redes Neuronales Artificiales en mejora de imagen médica

Recientemente se han comenzado a usar los métodos de aprendizaje profundo para la mejora de imagen en Rayos X. En [8] se ha usado una red neuronal convolucional con arquitectura UNet y la función de coste MSE para eliminar el ruido en imágenes de radiografía clínica simulada.

Para la mejora de la calidad de la imagen en radiografía mecánica se usa en [9] una Residual to Residual Network que se basa en una cascada de CNN, formada por un CNN profundo de la imagen de baja calidad, un CNN de imagen residual y una operación de interpolación bicúbica.

Respecto a otras disciplinas de imagen, como la Tomografía axial computarizada, usan en redes CNN con la función de coste AverageMSE para eliminar el ruido [29]. Para imágenes de intervención percutánea coronal a baja dosis usan una CNN para reducir el ruido en combinación con el SSIM [30].

3 Materiales y métodos

El método propuesto se basa en obtener a partir de una imagen de radiografía sin procesar, la correspondiente imagen mejorada. En la Figura 5 se observa el workflow del método. Primero, a la imagen sin procesar se le realiza un *padding* para que su tamaño sea el del múltiplo de 2 más cercano, debido a las características de la red usada. Posteriormente se pasa la imagen por la red y por último se elimina el *padding* realizado.

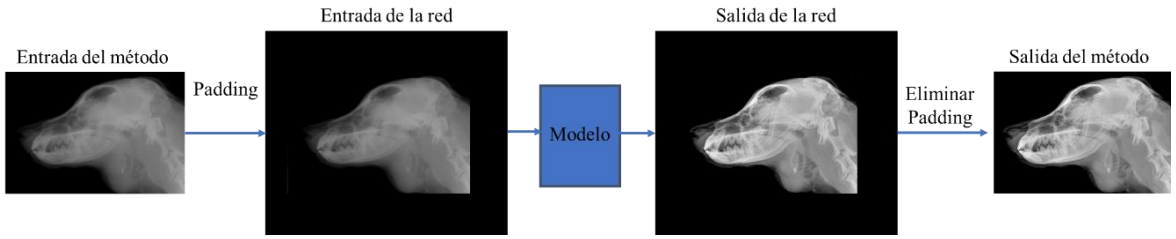


Figura 5. Workflow seguido

3.1 Arquitectura de la red

El método propuesto hace uso de la arquitectura U-Net [11], basada en la arquitectura codificador-decodificador diseñada para la segmentación de imágenes biomédicas. La razón de elegir esta red es que permite mantener la resolución espacial entre la imagen de entrada y salida. En este trabajo se sustituyó el codificador de la red por el modelo ResNet34, que gracias a las conexiones residuales, demostró mejorar el proceso de aprendizaje [21]. La arquitectura de la red se muestra en la Figura 6.

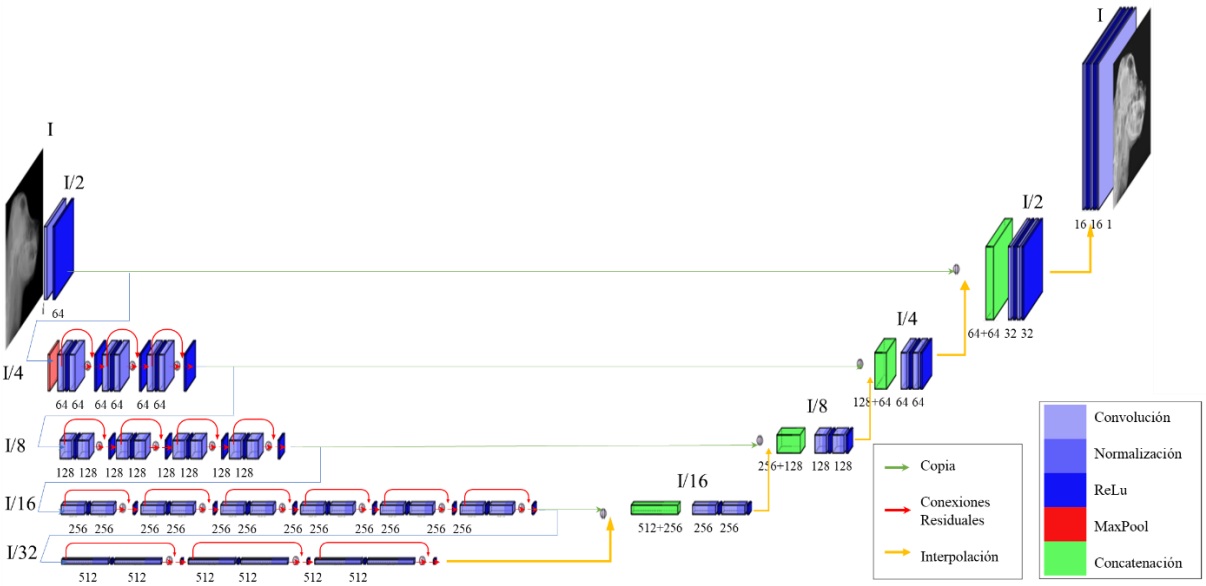


Figura 6. Arquitectura de la red

El codificador del método propuesto se basa el modelo ResNet34, entrenado para imágenes del conjunto ImageNet, aunque posteriormente para evaluación en el Capítulo 4 se cambiará por otras arquitecturas. La arquitectura de la red ResNet34 está formada por 34 capas, distribuidas en 5 bloques. El bloque inicial está formado por una capa de convolución 7×7, una capa de normalización y la capa de la función de activación (ReLU). La capa de normalización realiza una estandarización de las entradas a la siguiente capa por cada conjunto de imágenes que se le pasa a la red.

El resto de los bloques están formados por combinaciones de capa de convolución 3×3, capa de normalización, la capa de la función de activación (ReLU), capa de convolución 3×3 y una capa final de normalización. Después de cada bloque el número de filtros se dobla para aprender mapas de características más complejos.

El componente básico de esta arquitectura de codificador son los bloques residuales, que corresponden a los bloques formados por una capa de convolución, normalización, capa de función de activación (ReLU), otra capa de convolución y normalización. Estos bloques vienen representados en la Figura 7 por la siguiente estructura:

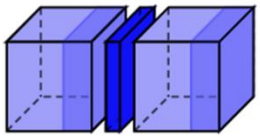


Figura 7. Representación de bloque residual de la Figura 2

En el bloque residual, la entrada X se suma directamente a la salida de la red, como se puede observar en la Figura 8.

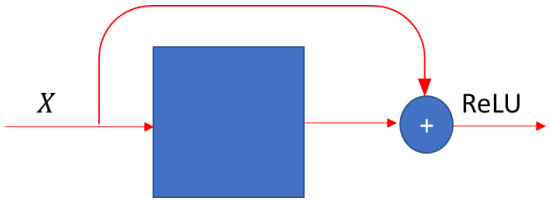


Figura 8. Bloque residual

El decodificador está formado por el decodificador original de UNet, que se basa en la convolución transpuesta, aumentando así el tamaño de la imagen hasta tener el de la original. Está compuesto de 4 bloques compuestos por una capa de interpolación, concatenación con el correspondiente mapa de características del codificador, una capa de convolución 3×3, capa de normalización, la capa de la función de activación (ReLU), otra capa de convolución 3×3, capa de normalización, la capa de la función de activación (ReLU). La idea principal del decodificador es restaurar la resolución espacial original usando información contextual extraída de la parte del codificador.

El tamaño de la imagen de entrada tiene que ser múltiplo de 2 ya que en cada bloque se realiza una reducción del tamaño de la imagen a la mitad. El número de canales de características inicial que se seleccionó es de 64, que es el que venía predeterminado en el modelo de ResNet34 escogido. Para el entrenamiento de la red se usó el optimizador de Adam, que converge más rápido y es más estable que otros optimizadores [31]. El tamaño de lote (*batch-size*) que se empleó es de 2 imágenes debido al gran tamaño de las imágenes y a las limitaciones técnicas y un número de épocas de 100. Como función de coste se aplicó el MS-SSIM combinado con el MAE entre la imagen salida de la red y el "gold standard" de forma que la función de coste resultante quedaba de la forma:

$$L_{MS\ SSIM+MAE} = 0.8 * L_{MS\ SSIM} + 0.2 * MAE \tag{20}$$

Para seleccionar la tasa de aprendizaje más adecuada se utilizó la prueba diseñada por Leslie N. Smith en [30], cuyo resultado se puede ver en la Figura 9 para el entrenamiento del predictor. Observando el resultado de la Figura 9, se decidió un valor para la tasa de aprendizaje de $3 \cdot 10^{-5}$ (zona de máxima pendiente).

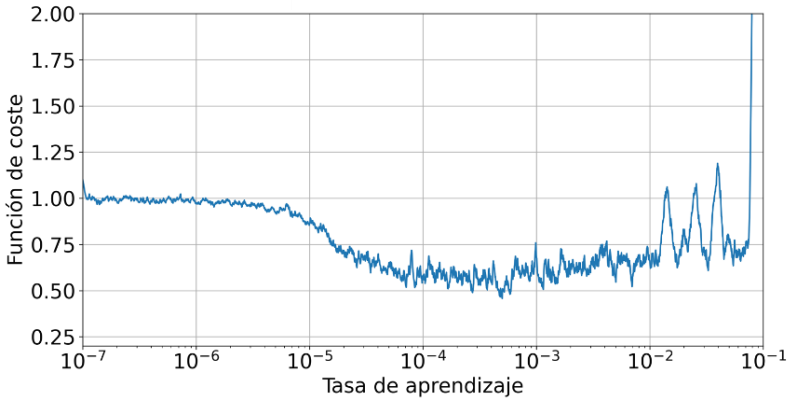


Figura 9. Resultado de la prueba de Leslie N. Smith para determinar el valor óptimo de la tasa de aprendizaje para el predictor.

3.2 Generación de la base de datos

Los datos que se usaron para la generación de la base de datos corresponden a distintas imágenes de radiografía animal de diversas partes anatómicas, adquiridas con distintos voltajes y dosis (mAs) en el Hospital Clínico Veterinario de la Universidad Complutense de Madrid.

Cada una de estas imágenes posee un "gold standard. Para la obtención del "gold standard", el especialista radiólogo del HCV usó una herramienta de procesado de imagen manual creada en Matlab, en la que procesa las imágenes de forma que la parte anatómica de interés se vea bien, aunque sature parte de la imagen.

Esta herramienta aplica primero una mejora de contraste, luego una eliminación del ruido y por último un realce de bordes. En cada paso solo se puede aplicar uno método de los disponibles. Permite, además de procesar la imagen, guardar el protocolo seguido en el procesado de la imagen, es decir, guardar los métodos que se han usado en cada paso por si se quiere aplicar el mismo procesamiento a otra imagen. La herramienta se puede observar en la Figura 10.

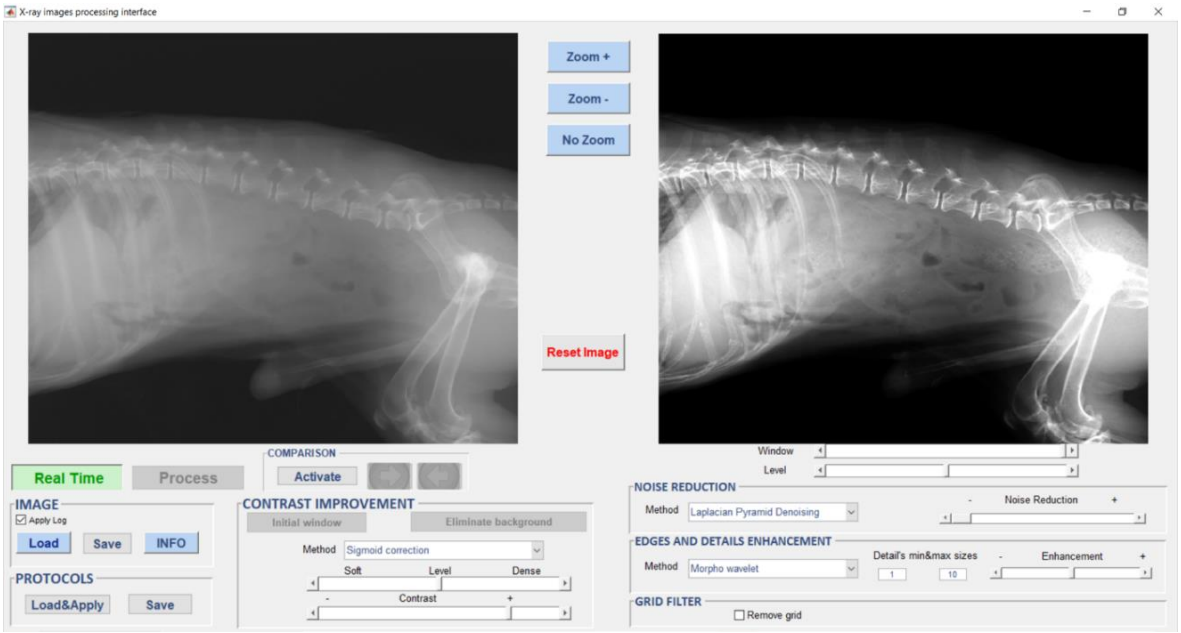


Figura 10. Herramienta para la corrección de imágenes

A partir de un estudio de métodos convencionales de procesado de imagen, se seleccionaron los que mejor funcionan para este tipo de imágenes y son los que se implementaron en la herramienta. Los métodos seleccionados son los siguientes:

- Para la mejora del contraste se dispone de un método de ecualización del histograma [32], que consiste en uniformar el histograma de la imagen de forma que todos los niveles de gris tengan la misma probabilidad de ocurrencia; un ajuste mediante una transformación sigma [33] o una gamma [5], que consiste en una transformación de los niveles de gris de la imagen de acuerdo a una función exponencial cuya curvatura se puede modular y una corrección de subhistogramas [34], que consiste en aplicar las técnicas anteriores a sub-histogramas, de forma que permite mejorar la visualización simultánea de tejido blando y denso, combinando linealmente las dos imágenes resultantes.
- Con objetivo de reducir el ruido se puede aplicar una pirámide laplaciana, que permite representar una imagen en múltiples escalas, permitiendo extraer los bordes de las diferentes escalas y eliminar el ruido de ellos.
- Para el realce de bordes se puede aplicar: una máscara de enfoque (*Unsharp mask*) [35] a la imagen, que se basa en detectar los bordes y luego sumárselos a la imagen; una máscara de enfoque dentro de la transformada wavelet [36] que se aplica a la imagen de baja frecuencia; un filtro morfológico en la transformada wavelet [37] basado en combinaciones de operaciones morfológicas a la imagen de baja frecuencia

o la pirámide laplaciana, que consigue el realce de bordes mediante la ponderación de estos al realizar la pirámide laplaciana inversa.

En la Figura 11 se pueden observar distintas partes anatómicas y su correspondiente “gold standard” para imágenes de dosis estándar.

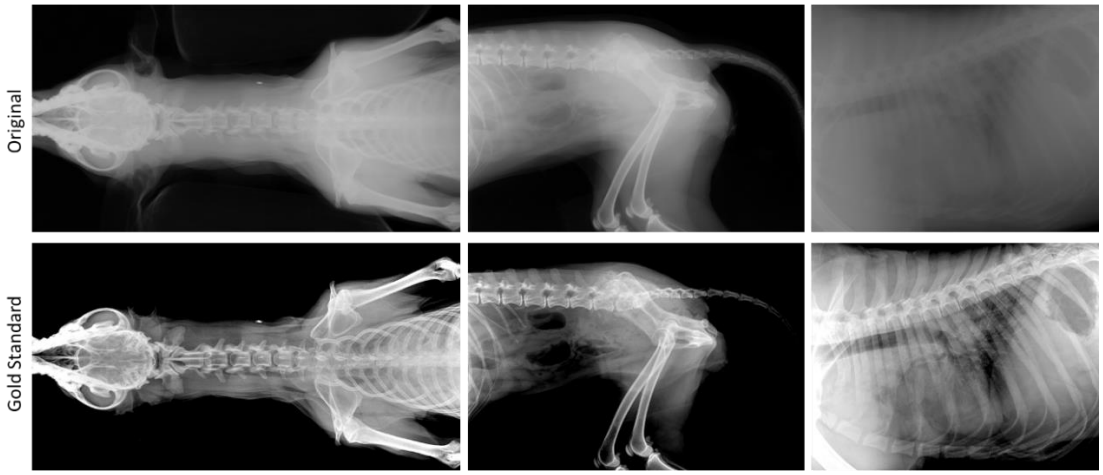


Figura 11. Radiografía animal de distintas partes anatómicas con dosis estándar

Para seleccionar las imágenes con distintos niveles de ruido se calculó el coeficiente de variación (CV) de distintas imágenes con distintos mAs. En la Figura 12 se muestra una ROI de una imagen de Cabeza Lateral en la que se puede observar el aumento de ruido a medida que se disminuye la dosis.

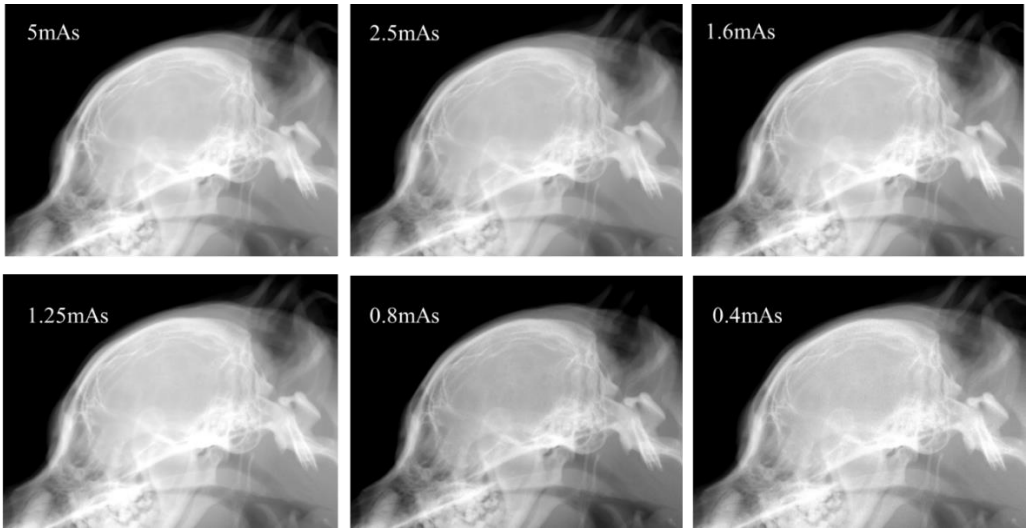


Figura 12. Radiografía animal de una ROI con variación de dosis

En la Figura 13 se observa el coeficiente de variación de las imágenes, que muestra un aumento del CV para mAs más bajos.

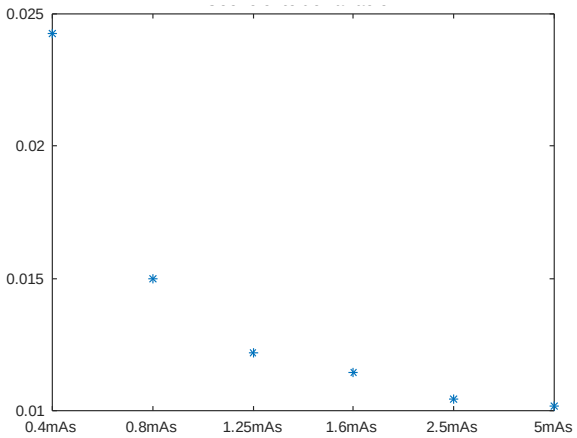


Figura 13. Coeficiente de variación

Las imágenes de radiografía tienen un tamaño aproximado de 4000x3000 píxeles, dependiendo de lo que se haya colimado luego para cada parte anatómica, y son de tipo 16-bit sin signo, por lo que cada imagen ocuparía 22.8 MB, y dado que tenemos 164 imágenes serían en total 4 GB aproximadamente. Si sumamos esto a que solo los mapas de características del modelo se estima que ocupan un 80% de la memoria [38], el modelo no cabría en la GPU RTX 2060 Super de 8 GB de capacidad, por lo que la red se entrenó con parches.

De cada imagen se realizaron 7 parches con un tamaño de matriz de 1024×1024 píxeles para que contengan suficiente información relevante de la imagen, inicializando los centros de manera aleatoria sin repetirse mediante un generador de números por permutación. Un ejemplo de los parches resultantes se muestra en la Figura 14.

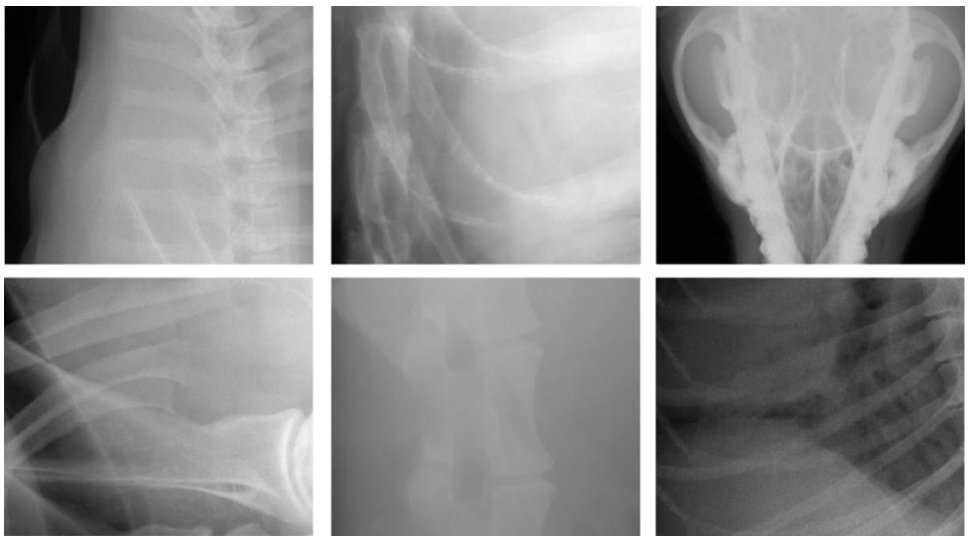


Figura 14. Ejemplo de parches de las imágenes

Tras ello se procedió a eliminar los parches repetidos y saturados del dataset mediante la visualización los parches correspondientes al "gold standard", eliminando mayoritariamente parches correspondientes a huesos de extremidades en imágenes que contenían mucho tejido blando, ya que el experto para ver bien el tejido blando saturó el hueso. En total se eliminaron 166 parches. La eliminación de parches saturados no se realizó tras la obtención del "gold standard" ya que no era necesario desechar la totalidad de la imagen, en algunos casos era posible sacar un parche que contuviera una zona no saturada.

Además, como el número de imágenes de las que se disponía no era muy grande, se realizaron técnicas de data augmentation con el fin de aumentar la base de datos como flip horizontal y vertical de cada imagen. En total se dispuso de 1252 parches de 1024×1024 píxeles de entrenamiento y 304 de validación.

4 Evaluación

La evaluación del modelo se hizo en radiografías animales que no se han habían usado para el entrenamiento de la red.

Para evaluar la mejora de contraste se hizo uso de tres estudios de perro de: Cervical Lateral de 3891×2992 píxeles, Cabeza Lateral de 3883×2564 píxeles y de Abdomen Lateral de 3269×2066.

Para evaluar la reducción de ruido se usan cinco imágenes que corresponden a la parte del cuerpo de Tórax Lateral, con un tamaño de matriz de 3203×2222 píxeles. Las imágenes están adquiridas con el mismo voltaje, pero con distintos mAs. Para cada mAs se calcula el coeficiente de variación, siendo el CV para 0.4mAs de 0.0612, para 0.8mAs de 0.0550, para 1.25mAs de 0.0539, para 2.5mAs de 0.0538 y para 5mAs de 0.0532

El modelo propuesto se compara con el mismo entrenado con las siguientes funciones de coste:

- MSE
- L_{SSIM}
- $L_{MS\ SSIM}$
- L_{VGG16}
- $L_{VGG16+SSIM} = 0.6 * VGG16 + 0.4 * SSIM$

La Tabla 7 muestra un resumen de las combinaciones de función de coste usadas, siendo la del modelo propuesto la primera de ellas.

Modelo	Función de coste
Arquitectura UNet y codificador sustituido por el modelo ResNet34 entrenado para imágenes del conjunto de ImageNet	$\boldsymbol{L_{MS\ SSIM+MAE}}$
	MSE
	L_{SSIM}
	$L_{MS\ SSIM}$
	L_{VGG16}
	$L_{VGG16+SSIM}$

Tabla 7. Resumen de las distintas funciones de coste usadas para el modelo de arquitectura UNet y codificador sustituido por el modelo ResNet34 entrenado para imágenes del conjunto de ImageNet

Tras la elección de la mejor función de coste, se compara el modelo de arquitectura UNet y codificador sustituido por el modelo ResNet34 entrenado para imágenes del conjunto de ImageNet con los siguientes modelos:

- Un modelo de red neuronal convolucional con arquitectura UNet preentrenada para segmentación de imágenes de Resonancia Magnética [39] con la función de coste L_{SSIM}. En adelante este modelo se denominará "UNet".
- Un modelo de red neuronal convolucional con arquitectura UNet y sustitución del codificador por la arquitectura EfficientNet B3 [40] con la función de coste L_{SSIM}. El codificador está preentrenado en el conjunto de imágenes de ImageNet. En adelante este modelo se denominará "EfficientNetB3".
- Un modelo de red neuronal convolucional con arquitectura UNet y sustitución del codificador por la arquitectura ResNet50 con la función de coste L_{SSIM}. El codificador está preentrenado en el conjunto de imágenes de ImageNet. En adelante este modelo se denominará "ResNet50".
- Un modelo de red neuronal convolucional con arquitectura UNet y sustitución del codificador por la arquitectura ResNet34 con la función de coste L_{SSIM}. El codificador no está preentrenado. En adelante este modelo se denominará "ResNet34 sin pret".

También se comparará con un método convencional que usa la pirámide laplaciana [6].

La Tabla 8 muestra un resumen de los distintos modelos usados.

Modelo	Función de coste
UNet	L_{SSIM}
EfficientNetB3	L_{SSIM}
ResNet50	L_{SSIM}
ResNet34 sin pret	L_{SSIM}

Tabla 8. Resumen los modelos usados en la segunda parte de la comparativa

En ambos casos se compara en base a la Raíz del Error Cuadrático Medio (RMSE), al Índice de Similitud Estructural (SSIM), al Índice de Similitud Estructural Multiescala (MS-SSIM) y al Índice de Similitud Estructural Multiescala combinado con el Error Absoluto Medio (MAE). También se realiza una evaluación visual por un radiólogo especializado del Hospital Clínico Veterinario mediante una herramienta de comparación

implementada en Matlab. En la Figura 15 podemos observar la herramienta para la elección del "gold standard".

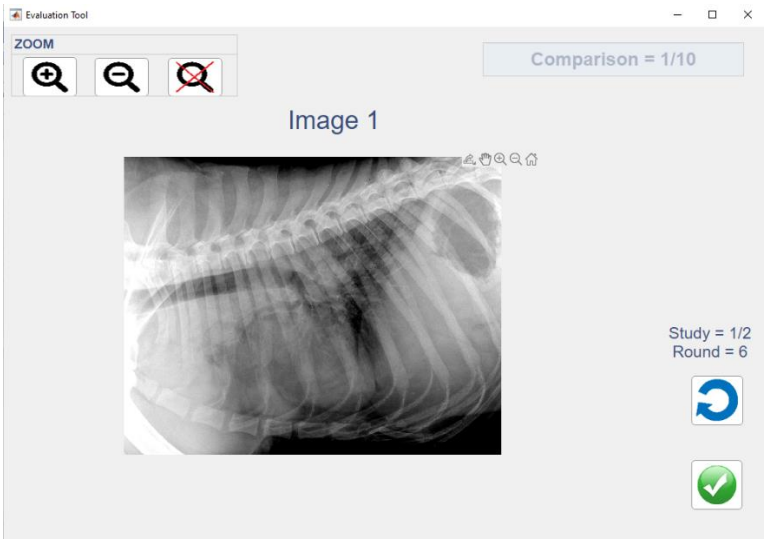


Figura 15. Herramienta para la evaluación

El funcionamiento de esta herramienta se basa en ir comparando las imágenes de dos en dos hasta obtener la imagen definitiva, es decir, si se tienen por ejemplo 4 imágenes, se compararía la imagen 1 con la imagen 2. La mejor de esas dos se seleccionaría, por ejemplo la imagen 1, y se compararía con la imagen 3. La mejor imagen entre la 1 y la 3 se seleccionaría, por ejemplo la 3 y se compararía con la imagen 4. Entre la imagen 3 y la imagen 4 se selecciona una, la imagen 4 por ejemplo (Figura 16A). Una vez se ha obtenido la imagen definitiva, la imagen 4 en el caso del ejemplo, se compara otra vez con todas las imágenes por si el experto hubiera cambiado de opinión, es decir, se compararía con la imagen 1 y con la 2 (Figura 16B). Con la imagen 3 no se compara porque ya ha sido comparada antes.

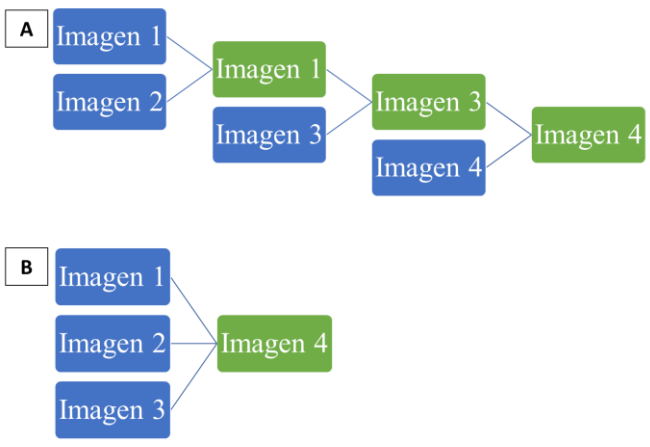


Figura 16. Diagrama herramienta de comparación

5 Resultados

La Figura 17 muestra los resultados obtenidos para el modelo de arquitectura UNet y codificador sustituido por el modelo ResNet34 entrenado para imágenes del conjunto de ImageNet con distintas funciones de coste para distintas partes del cuerpo con dosis estándar. Visualmente se observa como el contraste se ha recuperado mejor con las funciones de coste de $L_{MS\ SSIM}$ y $L_{MS\ SSIM+MAE}$. Además, las funciones de coste MSE y L_{VGG16} son las que peor contraste presentan.

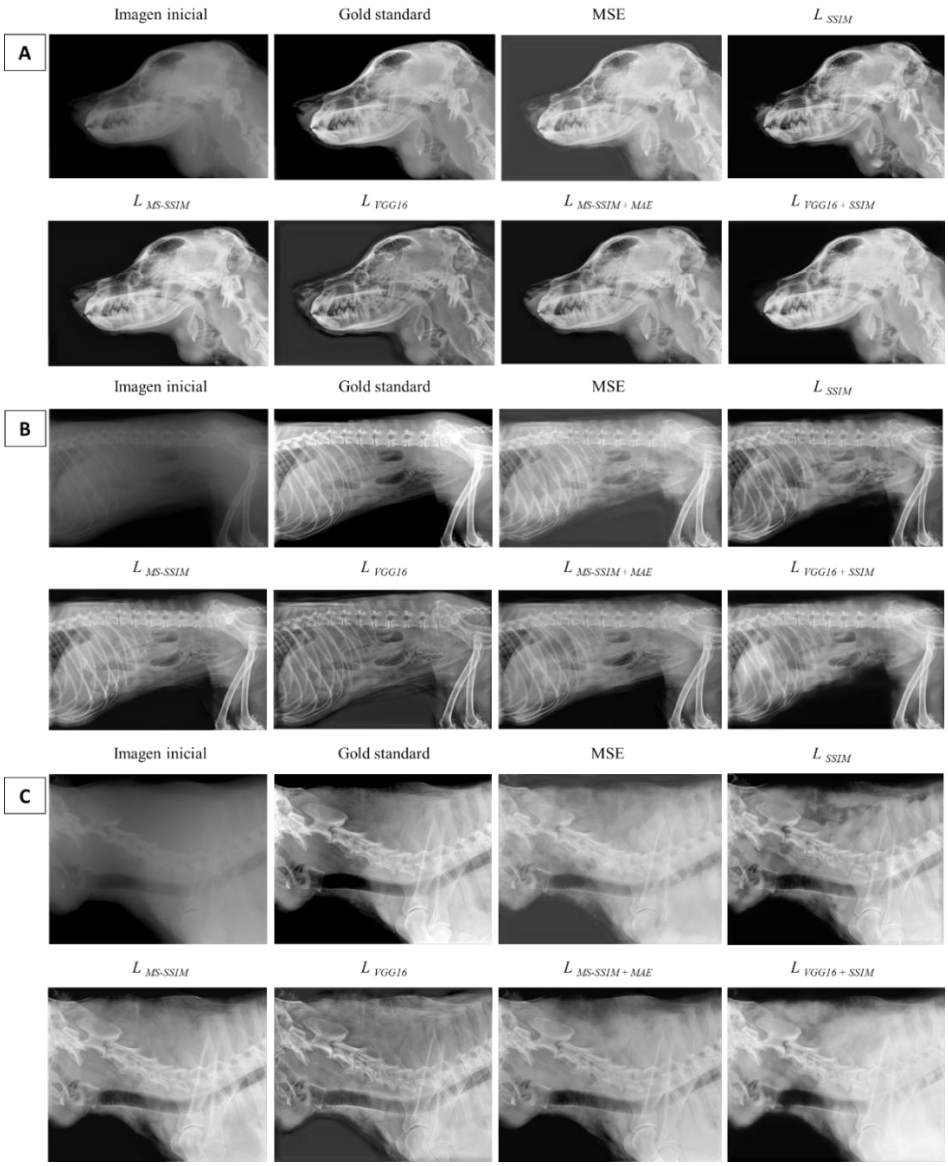


Figura 17. Resultados con distintas funciones de coste para imágenes con dosis estándar con el modelo de arquitectura UNet y codificador sustituido por el modelo ResNet34 entrenado para imágenes del conjunto de ImageNet

En la Figura 18 se pueden observar los resultados de realizar perfiles en distintas zonas de interés de las imágenes de la Figura 17 para ver la similitud con el "gold standard". Todos los perfiles han detectado bien los bordes y siguen la tendencia del "gold standard", aunque los de las funciones de coste MSE, L_{VGG16} y $L_{VGG16+SSIM}$ son los más similares, ya que el resto de los perfiles están más suavizados.

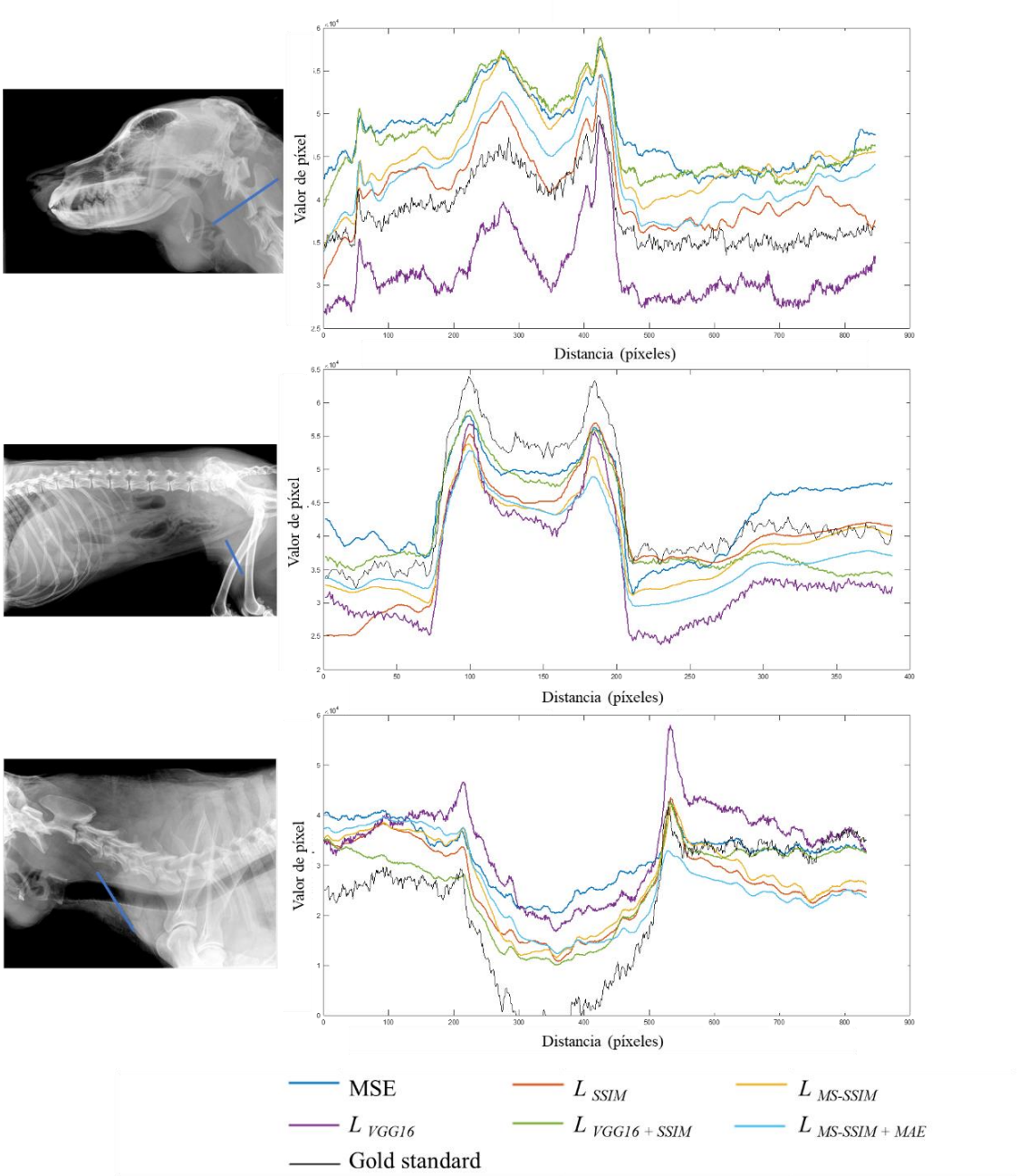


Figura 18. Perfiles en distintas zonas de interés para imágenes con distintas funciones de coste con el modelo de arquitectura UNet y codificador sustituido por el modelo ResNet34 entrenado para imágenes del conjunto de ImageNet

La Tabla 9 muestra el RMSE, SSIM, MS-SSIM y MS-SSIM+MAE entre el "gold standard" y las imágenes obtenidas con distintas funciones de coste. Para cada métrica se señala el menor valor de entre las funciones de coste probadas. Se observa como dependiendo de la imagen y la métrica difiere la función de coste seleccionada, habiendo sido seleccionado el L_{SSIM} en un 42% de las ocasiones, aunque las imágenes de la Figura 17 mostraban que el contraste era peor.

		MSE	L_{SSIM}	$L_{MS\ SSIM}$	L_{VGG16}	$L_{VGG16+SSIM}$	$L_{MS\ SSIM+MAE}$
Cabeza Lateral	*RMSE*	11575	4867.1	5675.3	9081.7	5896.6	**4698.9**
	SSIM	0.790	**0.757**	0.758	0.768	0.773	0.759
	MS-SSIM	0.604	0.559	0.570	0.574	0.560	**0.55**
	MS-SSIM+MAE	2057.5	**636.87**	966.52	1648.6	952.93	813.22
Abdomen Lateral	*RMSE*	11013	9176.3	7303.3	11621	**6645.2**	9189.1
	SSIM	0.774	**0.723**	0.725	0.725	0.743	0.744
	MS-SSIM	0.556	0.478	**0.459**	0.466	0.496	0.480
	MS-SSIM+MAE	1881.5	1445.5	1210.8	2000.9	**1068.4**	1452.6
Cervical Lateral	*RMSE*	10380	**5350**	6507.2	9945.1	7996.9	6735
	SSIM	0.873	0.832	0.853	**0.82**	0.858	0.860
	MS-SSIM	0.719	0.628	0.639	**0.621**	0.665	0.652
	MS-SSIM+MAE	1675.5	**787.65**	998.28	1651.9	1300.5	1075

Tabla 9. RMSE, SSIM, MS-SSIM y MS-SSIM+L1 entre el "gold standard" y las imágenes obtenidas con distintas funciones de coste con el modelo de arquitectura UNet y codificador sustituido por el modelo ResNet34 entrenado para imágenes del conjunto de ImageNet

La Figura 19 muestra el resultado de las distintas funciones de coste a una imagen con reducción de dosis para el modelo de arquitectura UNet y codificador sustituido por el modelo ResNet34 entrenado para imágenes del conjunto de ImageNet. Se muestra una ROI de la imagen inicial en la que se puede observar el ruido. Al igual que pasaba en la Figura 16, se observa que el contraste se ha recuperado mejor con las funciones de coste de $L_{MS\ SSIM}$ y $L_{MS\ SSIM+MAE}$. Visualmente no se nota el ruido en las imágenes de baja dosis en comparación con la imagen original.

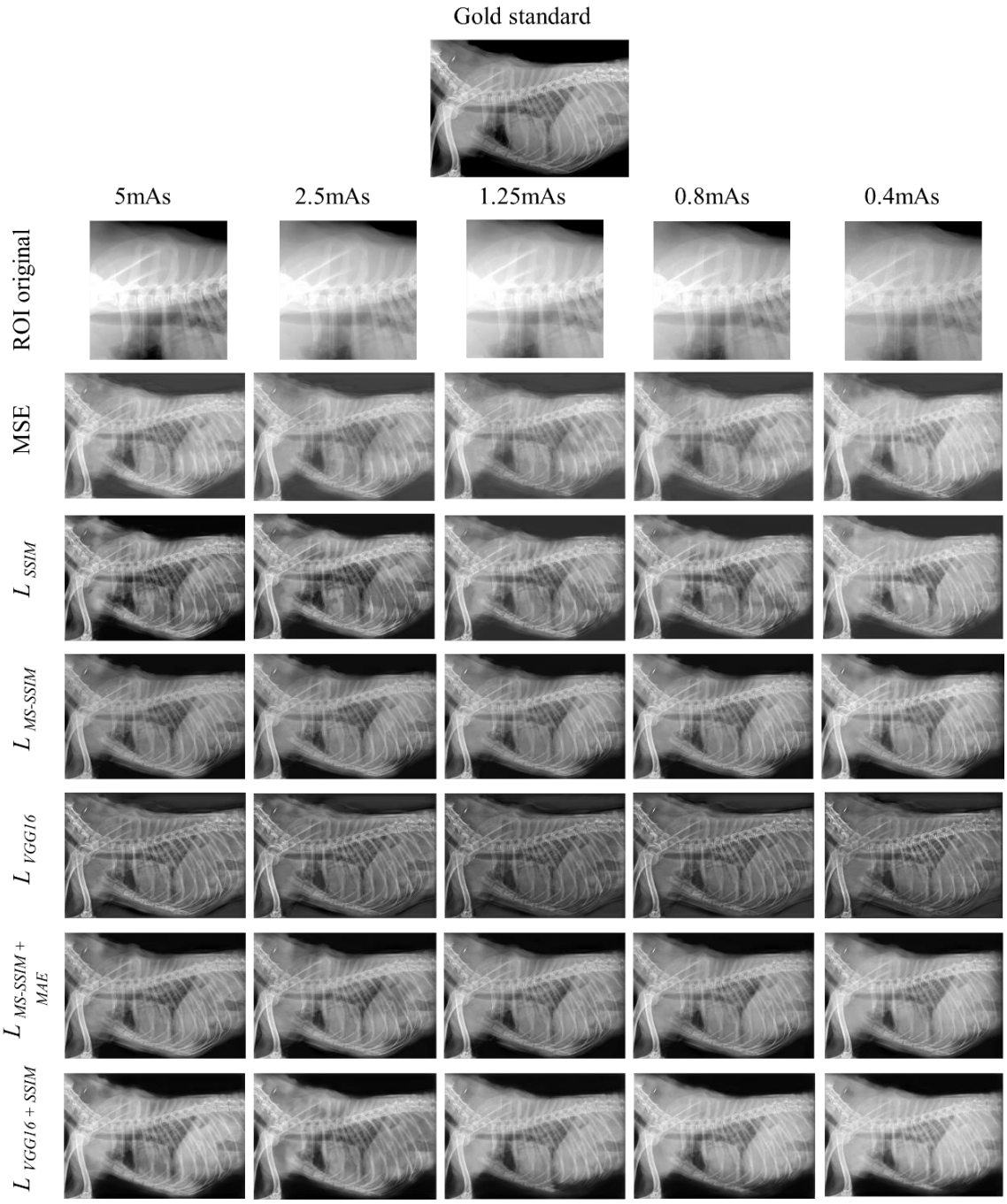


Figura 19. Resultados con distintas funciones de coste para imágenes disminuyendo la dosis

El cálculo del coeficiente de variación en cada imagen de la Figura 19 se observa en la Figura 20. Todos los coeficientes de variación se encuentran por encima del de la imagen original. Entre las distintas funciones de coste evaluadas, el MSE es el que menor coeficiente de variación tiene.

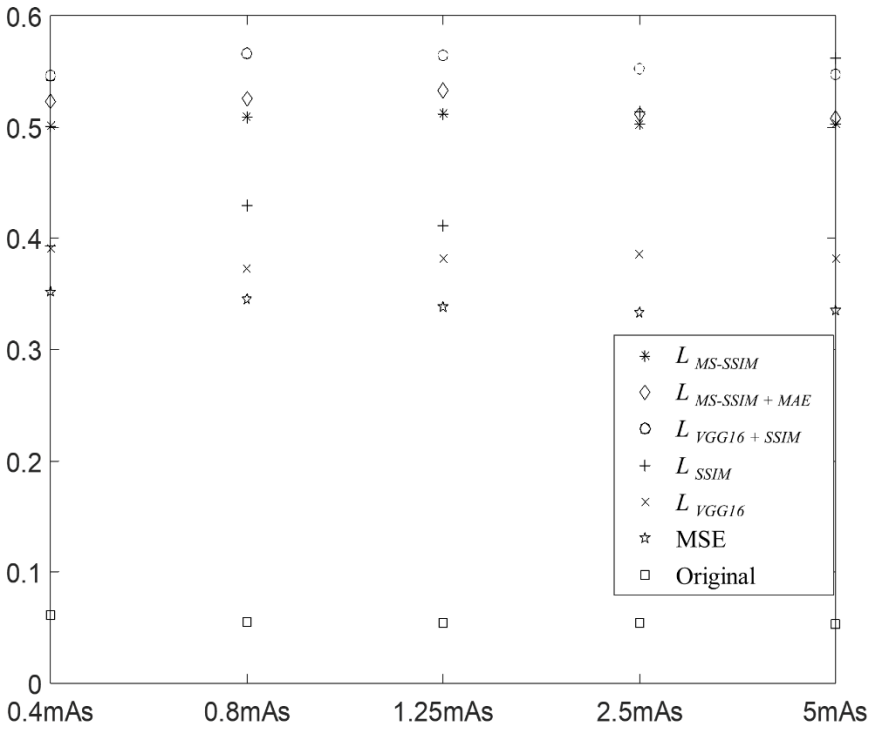


Figura 20. Coeficiente de variación imágenes con distintas dosis y distintas funciones de coste

Tras evaluación visual por el radiólogo especializado, determina que ninguna de las imágenes probadas de 0.4 mAs, 0.8 mAs y 1.25 mAs son buenas. Entre las distintas funciones de coste, las mejores para el modelo de arquitectura UNet y codificador sustituido por el modelo ResNet34 entrenado para imágenes del conjunto de ImageNet son $L_{MS\ SSIM}$ y $L_{MS\ SSIM+MAE}$, por lo que el resto se desechan para posteriores comparaciones.

La Figura 21 muestra los resultados obtenidos con los distintos métodos para distintas partes del cuerpo con dosis estándar. Visualmente se observa que, por lo general, con los modelos de ResNet34 sin pret y UNet no se aprecia buen contraste en las imágenes. Los modelos ResNet50 y EfficientNetB3, aunque parece que presentan buen contraste en general, se observa que hay zonas con poco contraste. El modelo previamente estudiado con las funciones de coste $L_{MS\ SSIM}$ y $L_{MS\ SSIM+MAE}$ y el método convencional parece que dan buen contraste en todas las imágenes.

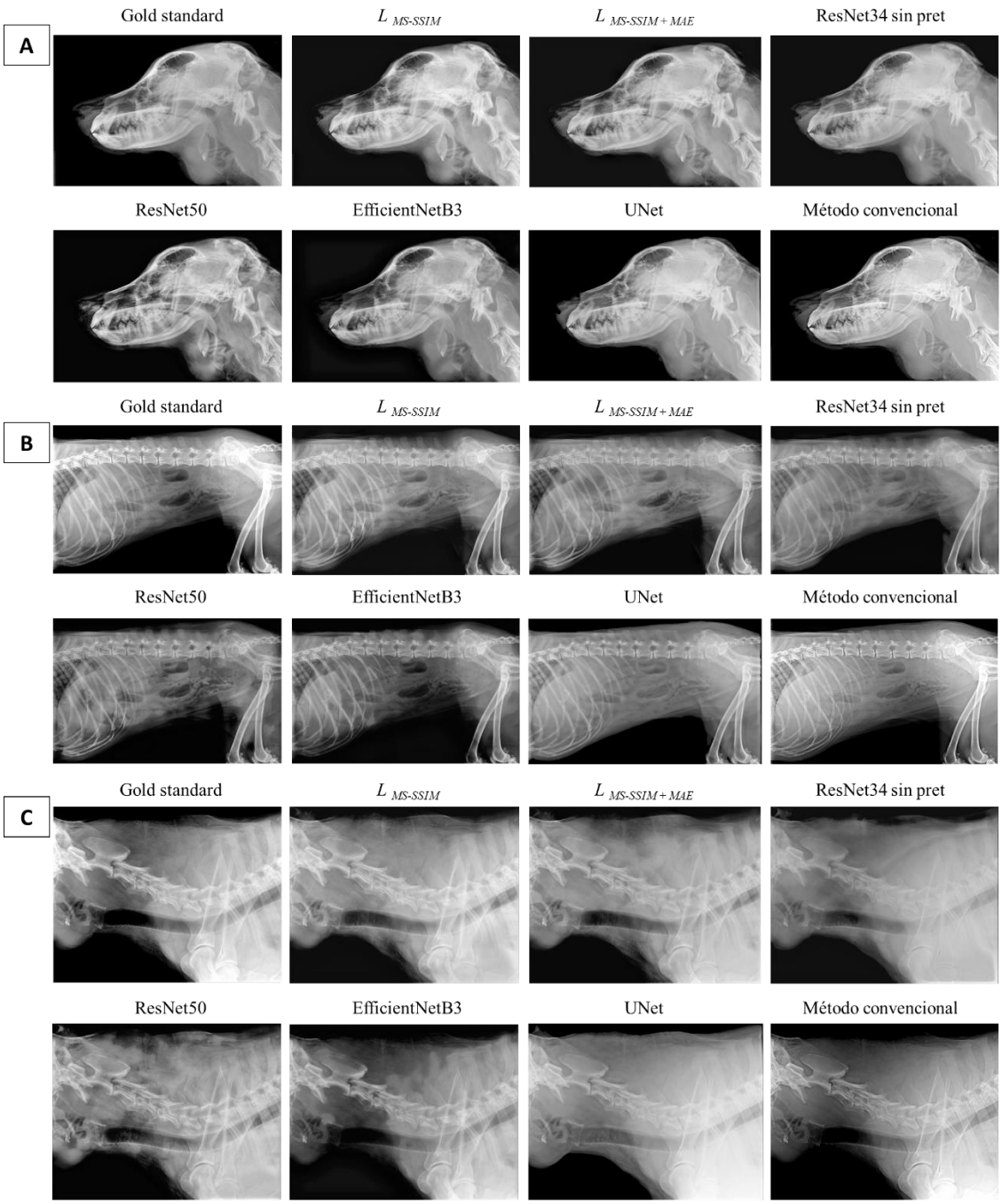


Figura 21. Resultados con distintos métodos para imágenes con dosis estándar

En la Figura 22 se pueden observar los resultados de realizar perfiles en distintas zonas de interés para ver la similitud con el "gold standard". El método convencional es el que tiene un perfil más distinto respecto del "gold standard". El resto de los modelos han detectado bien los bordes y siguen la tendencia del "gold standard", aunque la ResNet34 Sin Pret y la UNet son los que tienen unos perfiles más suavizados.

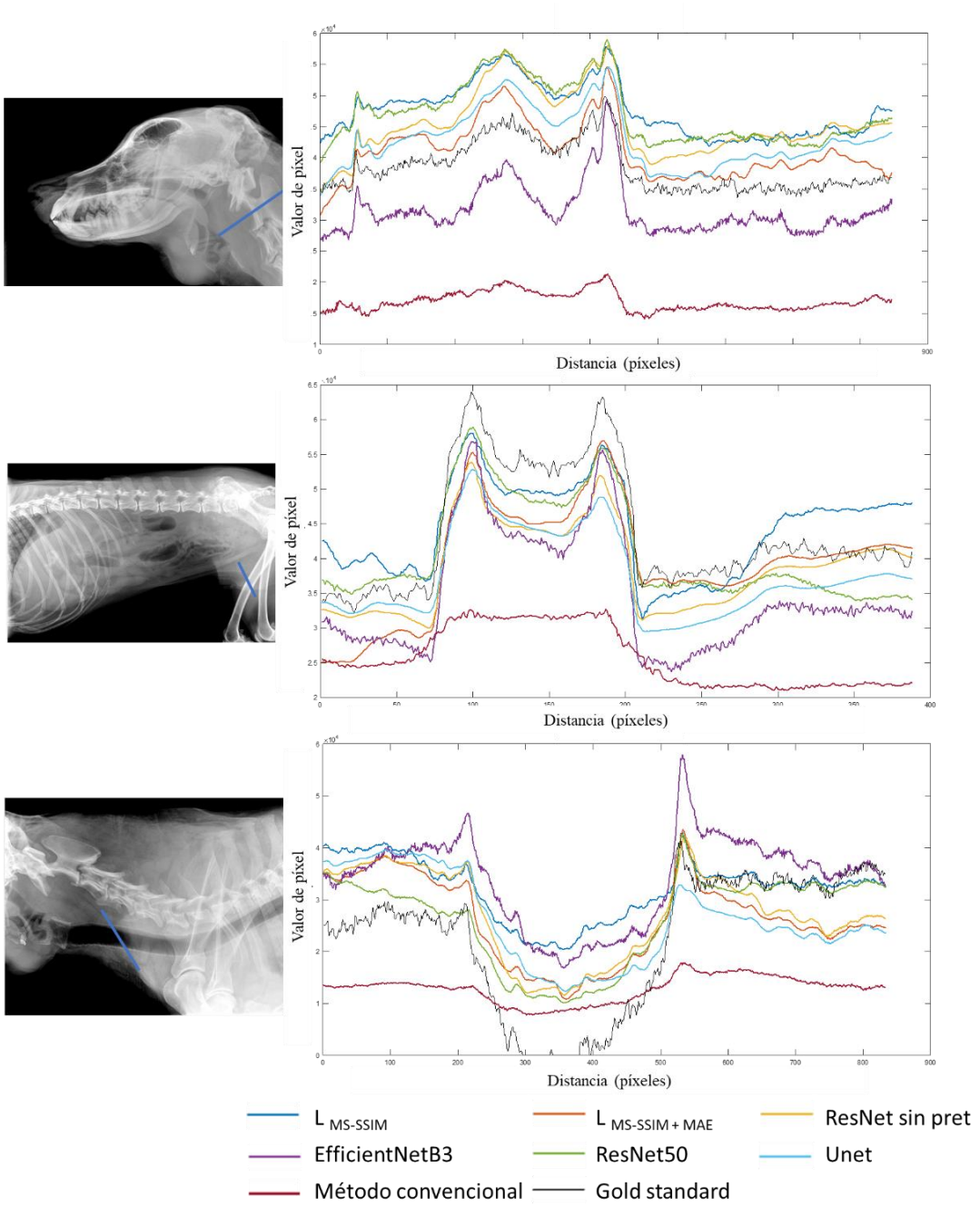


Figura 22. Perfiles en distintas zonas de interés para imágenes con distintos métodos

La Tabla 10 muestra el RMSE, SSIM, MS-SSIM y MS-SSIM+MAE entre el "gold standard" y las imágenes obtenidas con los distintos métodos. Al igual que sucedía en la Tabla 9, se observa como dependiendo de la imagen y la métrica difiere el método seleccionado, habiendo sido seleccionado el método convencional el 67% de las veces.

		$L_{MS\ SSIM}$	$L_{MS\ SSIM+MAE}$	ResNet sin pret	Efficient NetB3	ResNet50	UNet	Método convencional
Cabeza Lateral	*RMSE*	5675.3	**4698.9**	5125.4	4734.2	**4460.3**	5221.4	5350.4
	SSIM	0.758	0.759	0.79	0.749	0.745	0.415	**0.385**
	MS-SSIM	0.570	**0.55**	0.533	0.567	0.538	0.265	**0.227**
	MS-SSIM+MAE	966.52	813.22	886.25	684.32	**620.68**	670.45	727.59
Abdomen Lateral	*RMSE*	7303.3	9189.1	8941.8	10132	10399	7477.8	**4416.3**
	SSIM	0.725	0.744	0.786	0.725	0.7253	0.584	**0.522**
	MS-SSIM	**0.459**	0.480	0.548	0.461	0.478	0.362	**0.309**
	MS-SSIM+MAE	1210.8	1452.6	1449.4	1754.4	1565.4	1104.9	**628.36**
Cervical Lateral	*RMSE*	**6507.2**	6735	6750.9	7903.5	6725.6	6776.6	8045.2
	SSIM	0.853	0.860	0.890	0.850	0.831	0.742	**0.627**
	MS-SSIM	0.639	0.652	0.715	0.638	0.619	0.576	**0.488**
	MS-SSIM+MAE	998.28	1075	1083.6	1338.6	1046.5	**878.43**	1223.1

Tabla 10. RMSE, SSIM, MS-SSIM y MS-SSIM+L1 entre el "gold standard" y las imágenes obtenidas con los distintos métodos

La Figura 23 muestra el resultado de los distintos métodos a una imagen con reducción de dosis. Visualmente se observa que para el modelo de ResNet34 sin pret, para dosis estándar el contraste de la imagen no es el deseado. El modelo UNet tampoco presenta buen contraste en ninguna de las imágenes. Los modelos ResNet50 y EfficientNetB3, aunque parece que presentan buen contraste en general, se observa que hay zonas con poco contraste. El método convencional parece que da buen contraste en todas las imágenes, aunque se nota más el ruido conforme se disminuye la dosis.

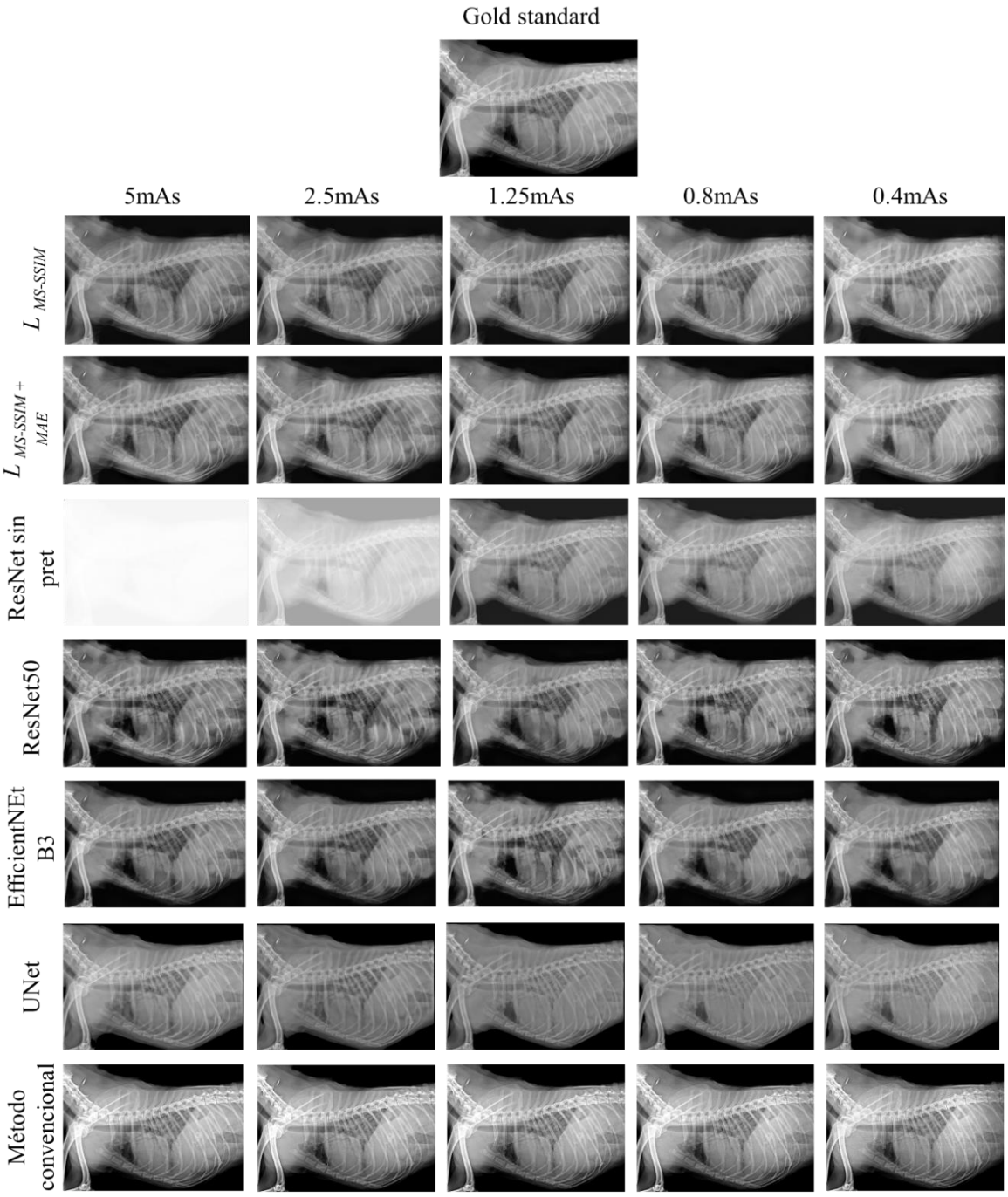


Figura 23. Resultados con distintos métodos para imágenes disminuyendo la dosis

El cálculo del coeficiente de variación en cada imagen de la Figura 23 se observa en la Figura 24. Todos los coeficientes de variación se encuentran por encima del de la imagen original. Entre los distintos modelos evaluados, el $L_{MS\ SSIM}$ es el que menor coeficiente de variación tiene.

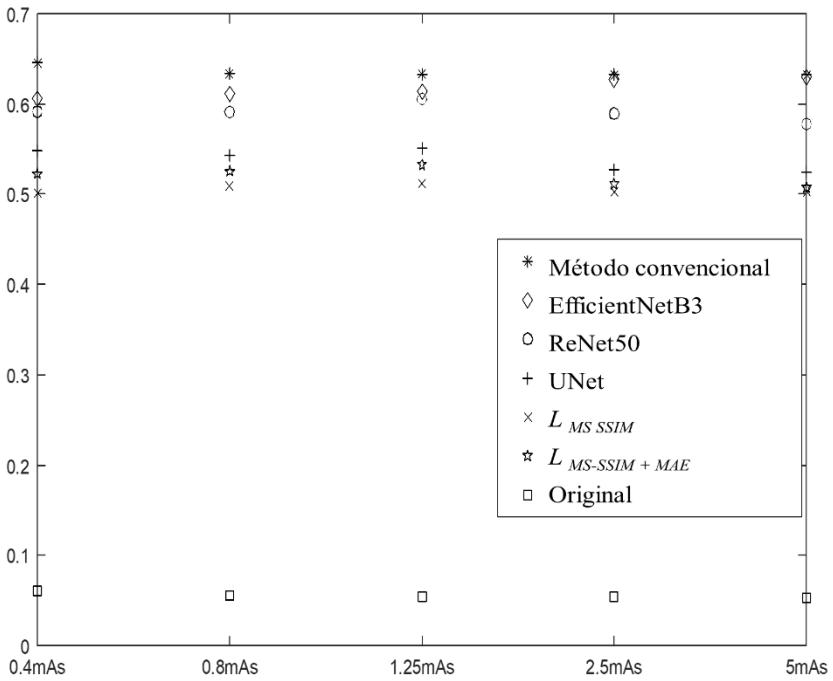


Figura 24. Coeficiente de variación imágenes con distintas dosis y distintos modelos

Tras evaluación visual por el radiólogo especializado, determina que ninguna de las imágenes de 0.4 mAs, 0.8 mAs y 1.25mAs son buenas. Entre los distintos modelos, los elegidos por el radiólogo son $L_{MS\ SSIM}$ y EfficientNetB3.

Analizando por último el tiempo de ejecución de los mejores métodos, podemos observar, en la Tabla 9 como para los métodos de aprendizaje profundo los tiempos son por debajo del segundo por lo general.

	Cervical Lateral de 3891×2992 píxels	**Cabeza Lateral de 3883×2564 píxels**	**Abdomen Lateral de 3269×2066 píxels**
$L_{MS\ SSIM}$	**0.979**	**0.991**	**0.776**
EfficientNetB3	1.02	0.998	0.792
Método convencional	1.3621	1.1384	0.8240

Tabla 11. Tiempos de ejecución de los distintos métodos

6 Discusión y conclusiones

En este trabajo se presenta un método que integra la mejora de contraste y la reducción de ruido para imágenes radiológicas mediante aprendizaje profundo.

La evaluación visual en datos de radiografía animal reales muestra cómo, de las distintas funciones de coste probadas, con el $L_{MS\ SSIM}$ y $L_{MS\ SSIM+MAE}$ se ofrece un mejor contraste. Los resultados de las métricas muestran que no hay una función de coste que sea la mejor por encima del resto. Respecto a los perfiles, los modelos con las funciones de coste MSE, L_{VGG} y $L_{VGG+SSIM}$ son los que tienen un perfil más similar al "gold standard", sin embargo, en las imágenes se ve mucho brillo en el fondo y los perfiles están desplazados. Esto puede deberse a un desplazamiento del histograma hacia valores altos causado por los valores en el fondo. Como trabajo futuro se probará a hacer una eliminación del fondo de las imágenes del conjunto de datos.

El coeficiente de variación de las imágenes con distinta dosis para las distintas funciones de coste muestra que éste es siempre superior al de la imagen original. Esto puede deberse a una diferencia de valores grande entre la imagen original y el resto de las imágenes.

Comparando visualmente los distintos modelos con el "gold standard", las que obtienen una imagen con mejor contraste son el modelo $L_{MS\ SSIM}$, $L_{MS\ SSIM+MAE}$ y el método convencional. Observando las métricas, el modelo que más se asemeja al "gold standard" es el método convencional, aunque los perfiles muestran que el método convencional es el que tiene un perfil más distinto respecto del "gold standard". Esto puede deberse a que los bordes obtenidos de la pirámide laplaciana estén suavizados.

Respecto al coeficiente de variación de las imágenes con distinta dosis y distintos modelos, muestra que éste es siempre superior al de la imagen original al igual que sucedía para la previa comparación.

Tras comparación de los resultados con la opinión del experto, se establece que las mejores funciones de coste son $L_{MS\ SSIM}$ y $L_{MS\ SSIM+MAE}$. Entre los modelos probados, los elegidos por el radiólogo son el $L_{MS\ SSIM}$ y EfficientNetB3.

El radiólogo además determina que ninguna de las imágenes de 0.4 mAs, 0.8 mAs y 1.25mAs son buenas. Como trabajo futuro se incluirán más imágenes de baja dosis en el dataset para tratar de solucionarlo.

Las imágenes de radiografía tienen un tamaño superior a 4000×3000 píxeles. Sin embargo, en este trabajo se han utilizado imágenes de 1024×1024 pixeles, debido a las limitaciones actuales en la memoria de la GPU. En futuros trabajos se hará uso de imágenes con el tamaño original, lo que podría lograrse mediante la paralelización de la red.

Por otra parte, la red neuronal se ha entrenado un número reducido de imágenes. El trabajo futuro también incluye el entrenamiento con un conjunto de datos más grande y diverso, lo que podría conducir a la optimización del número de capas y filtros, reduciendo los tiempos de entrenamiento.

Por último, actualmente la red se entrena con un ritmo de aprendizaje constante, siendo posible la utilización de una tasa de aprendizaje variable a lo largo de las épocas [41] en un futuro.

Dado que las métricas basadas en píxeles como funciones de pérdida tienen varios problemas como que si la imagen está desplazada por un píxel ya el error sería muy grande o que realizan un pesado igualitario para todos los píxeles de la imagen sin tener en cuenta que el ruido por ejemplo se puede enmascarar en los detalles de alta frecuencia, se considerará el uso de funciones de coste basadas en el funcionamiento del sistema visual humano [42]. Para reducir el error también se considerará la inclusión de arquitecturas basadas en Generative Adversarial Networks (GAN) [43].

En conclusión, se ha propuesto un nuevo método basado en aprendizaje profundo que permite mejorar el contraste y reducir el ruido de imágenes de radiología animal en menos de 1s.

7 Glosario

CNN	Redes neuronales convolucionales
mAs	Miliamperios por segundo
HCV	Hospital Clínico Veterinario de la Universidad Complutense de Madrid
RNA	Redes Neuronales Artificiales
RNN	Redes Neuronales Recurrentes
MSE	Error cuadrático medio
MAE	Error absoluto medio
SSIM	Índice de similitud estructural
MSSIM	Índice de similitud estructural medio
MS-SSIM	Índice de similitud estructural multiescala
ReLU	Rectified Linear Unit
CV	Coeficiente de variación

8 Bibliografía

[1] Consejo de Seguridad Nuclear. Dosis de radiación. 2010, SDB-04.07

[2] Abdullah-Al-Wadud M, Kabir MH, Dewan MAA, Chae O. A Dynamic Histogram Equalization for Image Contrast Enhancement. *IEEE Transactions on Consumer Electronics*. Vol. 53, Sup. 3, 2007, pp. 593-600.10.1109/TCE.2007.381734

[3] Pizer S, Amburn E, Austin J, Cromartie R, Geselowitz A, Greer T. Adaptive Histogram Equalization and Its Variations. *Computer Vision, Graphics, and Image Processing*. Vol. 39, 1987, pp. 355-68.10.1016/S0734-189X(87)80186-X

[4] Raman P. Fundamental Enhancement Techniques. *Handbook of Medical Image Processing and Analysis*. 2000, 10.1016/B978-012373904-9.50008-8

[5] Karuppanagounder S, Palanisamy K. Medical Image Contrast Enhancement based on Gamma Correction. *International Journal of Knowledge Management and e-Learning*. Vol. 3, 2011, pp. 15-8.

[6] Choudhary BK, Kumar N, And S, Shanker P. Pyramid method in image processing. 2012,

[7] Bouden T, Nibouche M. The Wavelet Transform for Image Processing Applications. 2012, 10.5772/35982

[8] Gnudi P, Schweizer B, Kachelrieß M, Berker Y. Denoising of X-ray projections and computed tomography images using convolutional neural networks without clean data. *The 6th International Conference on Image Formation in X-Ray Computed Tomography*. 2020, pp. 590-3.

[9] Sun Y, Li L, Cong P, Wang Z, Guo X. Enhancement of digital radiography image quality using a convolutional neural network. *Journal of X-ray science and technology*. Vol. 25, Sup. 6, 2017, pp. 857-68.10.3233/XST-17310

[10] Guo Z, Yu2 H. Low-dose CT Denoising with Convolutional Neural Network for Unknown Noise Levels. *The 6th International Conference on Image Formation in X-Ray Computed Tomography*. 2020, pp. 288-91.

[11] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, LNCS*. Vol. 9351, 2015, pp. 234-41.10.1007/978-3-319-24574-4_28

[12] Deng J, Dong W, Socher R, Li L-J, Li K, L. F-F. Imagenet: A large-scale hierarchical image database. *2009 IEEE conference on computer vision and pattern recognition*. 2009 pp. 248–55

[13] Mehar A. Multilayer Perceptrons vs CNN iq.opengenus.org [cited 2021 19-05]. Available from: https://iq.opengenus.org/multilayer-perceptrons-vs-cnn/.

[14] Pai A. CNN vs. RNN vs. ANN – Analyzing 3 Types of Neural Networks in Deep Learning analyticsvidhya.com2020 [cited 2021]. Available from: https://www.analyticsvidhya.com/blog/2020/02/cnn-vs-rnn-vs-mlp-analyzing-3-types-of-neural-networks-in-deep-learning/.

[15] Pykes K. The Vanishing/Exploding Gradient Problem in Deep Neural Networks 2020 [cited 2021]. Available from: https://towardsdatascience.com/the-vanishing-exploding-gradient-problem-in-deep-neural-networks-191358470c11.

[16] O'Shea K, Nash R. An Introduction to Convolutional Neural Networks. *ArXiv e-prints*. 2015,

[17] Gurucharan M. Basic CNN Architecture upGrad2020 [cited 2021]. Available from: https://www.upgrad.com/blog/basic-cnn-architecture/#:~:text=CNNs%20are%20a%20class%20of,vision%20and%20natural%20language%20processing][https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-convolutional-neural-networks.

[18] LeCun Y, Bottou L. Gradient-Based Learning Applied to Document Recognition. *IEEE*. Vol. 86, Sup. 11, 1998, pp. 2278 - 324.10.1109/5.726791

[19] Krizhevsky A, Sutskever I, Hinton G. ImageNet Classification with Deep Convolutional Neural Networks. *Neural Information Processing Systems*. Vol. 25, Sup. 2, 2012, 10.1145/3065386

[20] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv 14091556*. 2014,

[21] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 770-8.10.1109/CVPR.2016.90

[22] He K, Zhang X, Ren S, Sun J. Identity Mappings in Deep Residual Networks. *European Conference on Computer Vision*. Vol. 9908, 2016, pp. 630-45.10.1007/978-3-319-46493-0_38

[23] Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv*. 2015,

[24] Wang Z, Bovik A, Sheikh HR, Simoncelli E. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans Image Process*. Vol. 13, 2004, pp. 600-12.

[25] Wang Z, Bovik A, Sheikh H. Image Quality Assessment: From Error Measurement to Structural Similarity. *IEEE Trans Imgage Process*. Vol. 13, 2003,

[26] Wang Z, Simoncelli E, Bovik A. Multiscale structural similarity for image quality assessment. *Conference Record of the Asilomar Conference on Signals, Systems and Computers*. Vol. 2, 2003, pp. 1398 - 402.10.1109/ACSSC.2003.1292216

[27] Zhao H, Gallo O, Frosio I, Kautz J. Loss Functions for Image Restoration With Neural Networks. *IEEE Transactions on Computational Imaging*. 2016, 10.1109/TCI.2016.2644865

[28] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution. *Conference: European Conference on Computer Vision*. Vol. 9906, 2016, pp. 694-711.10.1007/978-3-319-46475-6_43

[29] Guo Z, Yu H. Low-dose CT Denoising with Convolutional Neural Network for Unknown Noise Levels. *The 6th International Conference on Image Formation in X-Ray Computed Tomography*. 2020,

[30] Pavoni M, Chang Y, Park S-h, Smedby O. Convolutional neural network-based image enhancement for X-ray percutaneous coronary intervention. *Journal of Medical Imaging*. Vol. 5, 2018, 10.1117/1.JMI.5.2.024006

[31] Diederik P. Kingma, Ba J. Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations*. 2014,

[32] Gonzalez R, Woods R. Digital Image Processing.

[33] Agarwal T, Bhupendra G. An Approach based on Parametric Sigmoid Function for Contrast Enhancement with Mean Brightness Preservation. *IEEE TechSym 2014 - 2014 IEEE Students' Technology Symposium*. 2014, 10.1109/TechSym.2014.6808055

[34] Huang SC. Efficient Contrast Enhancement Using Adaptive Gamma Correction With Weighting Distribution. *Ieee Transactions on Image Processing*. Vol. 22, 2013, pp. 1032-41.

[35] Kikinis C-FWR, Knutsson H. Handbook of Medical Imaging2000.

[36] Wang L. Noise removal for medical x-ray images in Multiwavelet domain. *Int J Image Graphics*. Vol. 8, 2008, 10.1142/S0219467808002952

[37] Wei Z, Hua Y, Hui-sheng S. X-Ray Image Enhancement Based on Multiscale Morphology. *1st International Conference on Bioinformatics and Biomedical Engineering, ICBBE*. 2007, 10.1109/ICBBE.2007.183

[38] Rhu M, Gimelshein N, Clemons J, Zulfiqar A, Keckler SW. vDNN: Virtualized deep neural networks for scalable, memory-efficient neural network design. *2016 49th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO), Taipei*. 2016, pp. 1-13.10.1109/MICRO.2016.7783721

[39] Buda M, Saha A, A.Mazurowski M. Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm. *Computers in Biology and Medicine*. Vol. 109, 2019, pp. 218-25.

[40] Tan M, Le QV. EfficientNet Rethinking Model Scaling for Convolutional Neural Networks. *International Conference on Machine Learning*. 2019,

[41] Smith LN. Cyclical learning rates for training neural networks. *IEEE Winter Conference on Applications of Computer Vision (WACV)*. 2017,

[42] Mikhailiuk A. Deep Image Quality Assessment [cited 2021]. Available from: https://towardsdatascience.com/deep-image-quality-assessment-30ad71641fac.

[43] Wolterink JM. Generative adversarial networks for noise reduction in low-dose CT. *IEEE transactions on medical imaging*. 2017,