

# Remoción de líneas en imágenes de textos manuscritos utilizando una red neuronal convolucional tipo U-Net

Diego A. Peralta Rodríguez, José E. Valdez Rodríguez,  
Nahum Carlos Alexis Rangel, Francisco Hiram Calvo Castro

Instituto Politécnico Nacional,  
Centro de Investigación en Computación,  
México

{dperaltar2022, jvaldezr2018, hcalvo}@cic.ipn.mx,  
rangelcarloss41@gmail.com

**Resumen.** Cuando se digitalizan documentos escritos a mano surgen diversos problemas, sobre todo en la etapa de preprocesamiento de imágenes. Uno de los problemas tiene que ver con la remoción de líneas horizontales que existen en las hojas, ya que muchas veces al traslaparse las líneas con las palabras se dificulta la extracción de información escrita. Anteriormente ya se han propuesto e implementado distintos enfoques para tratar de resolver este problema con técnicas y algoritmos clásicos de preprocesamiento. Sin embargo, pensamos que es importante el poder aprovechar las ventajas que ofrecen los métodos que se basan en Redes Neuronales Convolucionales (RNC), ya que estas tienen el potencial de mejorar significativamente la precisión y eficiencia en la eliminación de líneas. En este trabajo presentamos un método para remover líneas en textos manuscritos, sin afectar la información textual o la comprensión del mensaje escrito, implementando una red convolucional tipo U-Net. Para esto se llevaron a cabo experimentos utilizando conjuntos de imágenes especializados para tareas de remoción de pentagramas y para identificación de rasgos de personalidad, ya que no existen conjuntos de imágenes específicos para el problema de eliminación de líneas.

**Palabras clave:** Remoción de líneas, texto manuscrito, RNC.

## Ruled Line Removal in Handwritten Text Images Using a U-Net Type Convolutional Neural Network

**Abstract.** When handwritten documents are digitized, several problems arise, especially in the image preprocessing stage. One of the problems has to do with ruled line removal that exist on the sheets, since many times the overlapping of lines with words makes the extraction of written information difficult. Previously, different approaches have been proposed and implemented to try to solve this problem with classical preprocessing techniques and algorithms. However, we believe that it is important to take advantage of the benefits offered by methods based on Convolutional Neural Networks (CNN), since these have the

potential to significantly improve the accuracy and efficiency of line removal. In this paper we present a method to remove lines in handwritten texts, without affecting the textual information or the comprehension of the written message, by implementing a U-Net type convolutional network. For this purpose, experiments were carried out using specialized datasets for staff line removal in musical sheets and personality trait identification tasks, as there are no specific datasets for the ruled line removal problem.

**Keywords:** Ruled line removal, handwritten text, CNN.

## 1. Introducción

El reconocimiento automático de escritura a mano es una tarea importante en muchas áreas, como la educación, la medicina, la banca y la seguridad, entre otros. El éxito de los diversos métodos de reconocimiento que se puedan aplicar dependerá en gran medida de la precisión con la que se logre segmentar el texto. Esto implica una clara separación del texto, como primer plano, y el fondo para garantizar una alta tasa de precisión en el reconocimiento. Para mejorar el rendimiento en esta tarea, usualmente se aplican técnicas de procesamiento tales como la mejora de contraste, algoritmos de umbralado, eliminación de ruido, detección de manchas, etc.

La remoción de líneas en documentos con hojas rayadas es un problema común en el preprocesamiento de imágenes. El propósito de las líneas es ayudar a escribir en línea recta y mantener una cierta uniformidad en el tamaño y espaciado de las letras. Remover las líneas es importante para una variedad de aplicaciones, sobre todo en el Reconocimiento Óptico de Caracteres (OCR, por sus siglas en inglés) [1], por ejemplo, en tareas para la extracción de información escrita de documentos clínicos o la digitalización de documentos. Un trabajo en donde adquiere relevancia la remoción de líneas es en el reconocimiento de rasgos de personalidad a partir del procesamiento de imágenes de textos escritos a mano [2].

La identificación de personalidad a partir de la escritura es una tarea que va cobrando cada vez más importancia en el área de visión por computadora, ya que analizar los trazos y estilo, si es que se parte desde un enfoque grafológico, requiere poder separar las letras de las líneas de fondo; o si se aborda la identificación desde un enfoque de análisis léxico, entonces se necesitará tomar en cuenta únicamente las palabras para poder clasificarlas de acuerdo a un tipo de personalidad. En resumen, actualmente se sigue enfrentando con el desafío de detectar y remover líneas con un alto grado de eficacia en diferentes tipos de documentos.

Dicha eficacia dependerá en gran medida de factores como la calidad de la imagen, la complejidad del contenido y la resolución de la misma. A su vez, la implementación de técnicas de preprocesamiento y postprocesamiento puede mejorar significativamente los resultados finales. En este trabajo se propone un método para la remoción de líneas conservando el texto sin pérdida significativa de información, utilizando una combinación de técnicas de procesamiento de imágenes y una red neuronal convolucional [3] de tipo U-Net.



Fig. 1. Ejemplo de imágenes incluidas en el dataset CVC-MUSCIMA [13].

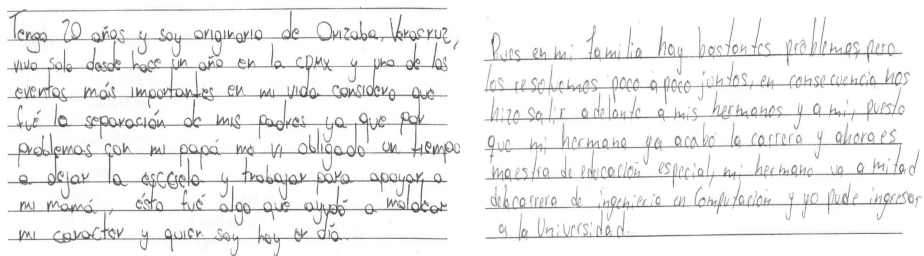


Fig. 2. Ejemplo de imágenes incluidas en el dataset HWxPI [12].

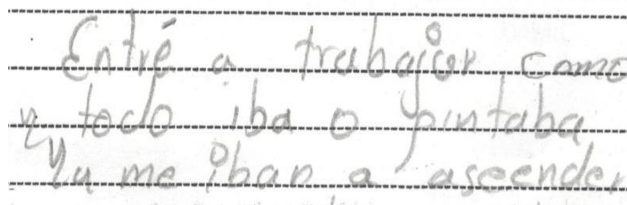
## 2. Trabajo previo

Diversos trabajos han propuesto métodos para lidiar con el ruido que añaden las líneas en imágenes de documentos. La mayoría de los trabajos, sobre todo los realizados antes del auge de las redes neuronales convolucionales, se centran en metodologías basadas en un preprocesamiento utilizando técnicas clásicas o estándar. Los métodos se dividían en tres importantes grupos [4]:

1. Métodos basados en morfología matemática [5].
2. Métodos que emplean la transformada de Hough para extraer características del texto y encontrar líneas en cualquier dirección [6].
3. Métodos que utilizan perfiles de proyección [7] para estimar líneas y reducir las dimensiones del problema.

No obstante, aun con la aparición de las redes neuronales convolucionales, hoy en día hay una exploración limitada en la remoción de líneas utilizando modelos neuronales. Un trabajo reciente que explora la remoción de líneas usando redes neuronales convolucionales es el presentado por [8]. Los autores crean un conjunto de datos sintéticos donde se generan líneas y se colocan en secuencias de imágenes concatenadas de palabras manuscritas.

La arquitectura propuesta consta de 3 capas convolucionales. Otros autores como Zhixin y sus colegas [9], proponen un método que se basa en un enfoque de perfilado local direccional para la detección de las ubicaciones de líneas en documentos árabes escritos a mano. Para eliminar los píxeles de las líneas de reglas, sin dañar demasiado el texto, realizan una búsqueda vertical adaptativa.



**Fig. 3.** Muestra parcial de imagen previa al filtro.

En el trabajo de Konstantinos [10], los autores presentan un sistema que implementa una técnica para el entrenamiento automático de la eliminación de líneas. Los parámetros de un algoritmo existente se ajustan en función de las características de una colección de documentos y se utiliza el algoritmo de recocido simulado para estimar los valores más adecuados de los parámetros.

El trabajo de Refay [11], utiliza la transformada de Hough en cuatro subventanas cuadradas. La etapa de eliminación emplea el histograma de intensidad y su entropía para aislar el texto. A la etapa de remoción le sigue una mejora mediante técnicas morfológicas.

### 3. Metodología

#### 3.1. Datasets

Actualmente no hay un conjunto de datos diseñado específicamente para la tarea de remoción de líneas en documentos. Esto significa que no existe un conjunto de imágenes de documentos con y sin líneas disponibles para este propósito. La falta de un conjunto de datos de esta naturaleza limita la capacidad para mejorar y evaluar algoritmos y modelos de remoción de líneas en documentos.

Lo que más se le asemeja a datasets con este propósito son los utilizados para la remoción de pentagramas en partituras. Para superar esta carencia y limitación, se experimentaron con dos datasets diferentes en este trabajo. El primero es el conjunto de datos HWxPI [12], creado por investigadores de la UAM con el fin de identificar rasgos de personalidad a partir de ensayos escritos a mano.

Este conjunto de datos está compuesto por 418 imágenes de ensayos manuscritos (Fig. 1) y 418 transcripciones de los mismos. Aunque este dataset no fue creado específicamente para la remoción de líneas en documentos, se utilizó para entrenar y evaluar modelos debido a la presencia de líneas que contenían todas las imágenes de los ensayos. El segundo dataset utilizado en este trabajo es el CVC-MUSCIMA, diseñado para la eliminación de pentagramas en partituras [13].

Este conjunto de datos consta de 1000 imágenes de partituras con pentagramas y 1000 imágenes de partituras sin pentagramas, únicamente conformadas por las notas (Fig. 2). Aunque este conjunto de datos tampoco fue creado con el propósito específico de la remoción de líneas en documentos, se decidió utilizarlo porque las líneas de pentagrama pueden ser tratadas de manera similar a las líneas en documentos escritos a mano.

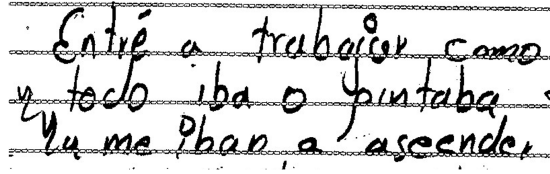


Fig. 4. Resultado después de pasar por el filtro.

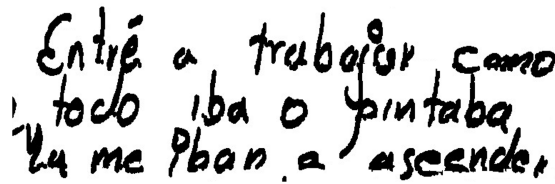


Fig. 5. Resultado después de aplicar apertura morfológica.

Por lo tanto, se determinó que este conjunto de imágenes podría resultar de gran utilidad en la realización de experimentos con el objetivo de desarrollar un modelo capaz de generalizar la tarea de remoción de líneas en diferentes tipos de documentos y no solo en partituras.

### 3.2. Preprocesamiento

El objetivo de trabajar con el dataset HWxPI no fue obtener la remoción de líneas del conjunto total de imágenes, sino solo de un porcentaje.

Lo que se buscaba era poder aparear las imágenes originales con sus respectivas etiquetas (también conocidas como 'ground truth' en inglés) o imágenes limpias de líneas para entrenar un modelo de red neuronal. Para conseguir esto se aplicaron técnicas clásicas de preprocesamiento y morfología matemática. Se tomaron 100 imágenes como punto de partida, las cuales se recortaron para conservar únicamente el área donde había texto y se convirtieron a valores en escala de grises.

A continuación, se revisó el histograma y el valor de los píxeles que conformaban las líneas de los ensayos para definir un umbral adecuado. Es bien sabido que es complejo o casi imposible definir un valor único de umbralado para imágenes que difieran una de otra incluso en detalles que puedan parecer triviales. Por ejemplo, el hecho de que la persona al escribir haya recargado más o menos el lápiz, ocasiona un cambio grande en los valores de píxeles; si las líneas son continuas o son punteadas, si se escribe con bolígrafo o lápiz, si hay manchas en la hoja, etc.

Por lo tanto, se sabía que solo un porcentaje de las imágenes podrían tener éxito en la remoción total o parcial de las líneas, debido a la variabilidad en las características de las imágenes. Una vez realizadas las modificaciones necesarias, se determinó que el rango de píxeles más adecuado para umbralar se encontraba entre 120 y 230. En consecuencia, se diseñó un filtro para preservar únicamente los valores de gris dentro de ese rango. Tras aplicar el filtro, de las 100 imágenes se obtuvieron 40 con una remoción casi total de líneas.

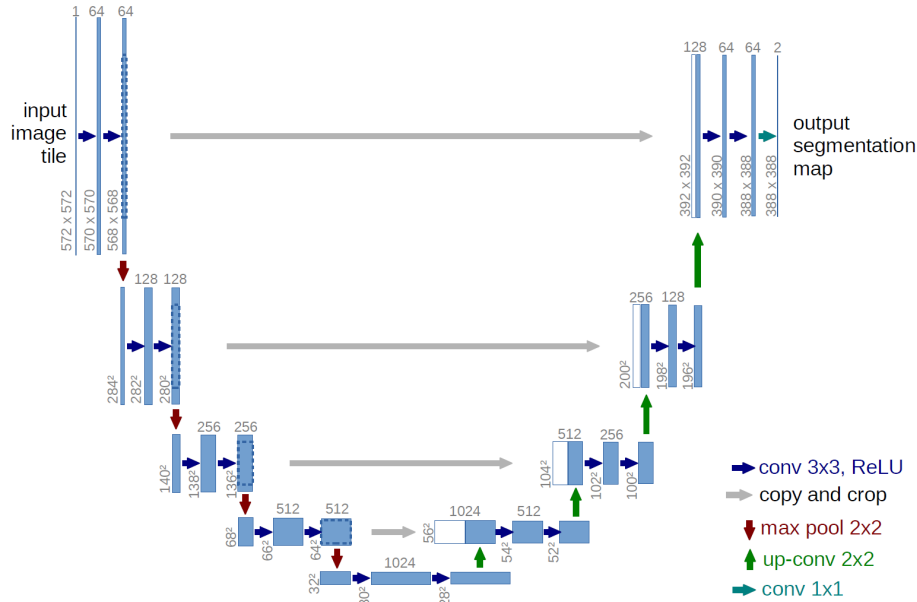


Fig. 6. Arquitectura U-Net [15].

En las figuras 3 y 4 se muestra el ejemplo de una sección de imagen antes y después de pasarse por el filtro. Gran parte del relleno en las líneas se pudo eliminar quedando solamente su contorno. Para eliminar el sobrante, se aplicó una operación morfológica de apertura, dada por la ecuación 1:

$$A \circ B = (A \ominus B) \oplus B. \quad (1)$$

Esta operación realiza una erosión seguida de una dilatación utilizando un mismo elemento de estructura para ambas operaciones. Se experimentó con diferentes elementos de estructura y el que mejor funcionó fue de forma circular y tamaño 3 x 3. La erosión eliminó por completo los contornos sobrantes y la dilatación ayudó a conservar la estructura original que tenían las letras (Fig. 5).

Por último, ya que en algunas imágenes quedaban puntos de cierta área que no se pudieron remover con una erosión, se aplicó una apertura por área. Esta es una operación útil de filtro que consiste en remover todos los componentes conectados cuya área en número de píxeles sea más pequeña que el valor de umbralado que se proponga [14].

Se agregaron 10 imágenes adicionales sin líneas al conjunto inicial de 40, pero estas fueron obtenidas mediante la eliminación manual de líneas en lugar de utilizar operaciones morfológicas. Como resultado, en total se generó un pequeño conjunto de datos compuesto por 100 imágenes: 50 imágenes originales y 50 imágenes sin líneas. Con respecto al conjunto de imágenes en el dataset CVC-MUSCIMA, primero se ensancharon las líneas del pentagrama utilizando una técnica de dilatación con un elemento de estructura rectangular de 3x5.

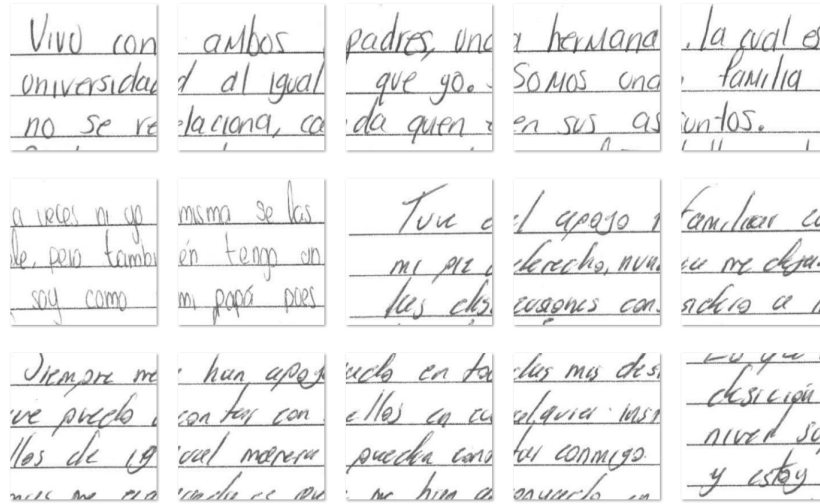


Fig. 7. Muestra de parches generados.

También se obtuvo el negativo (ya que originalmente el fondo era color negro) y se convirtieron a escala de grises. Estas modificaciones permitieron homologar, en cierto grado, las características con el conjunto de datos HWxPI, para poder después utilizarlos de manera conjunta en la tarea de entrenamiento y evaluación.

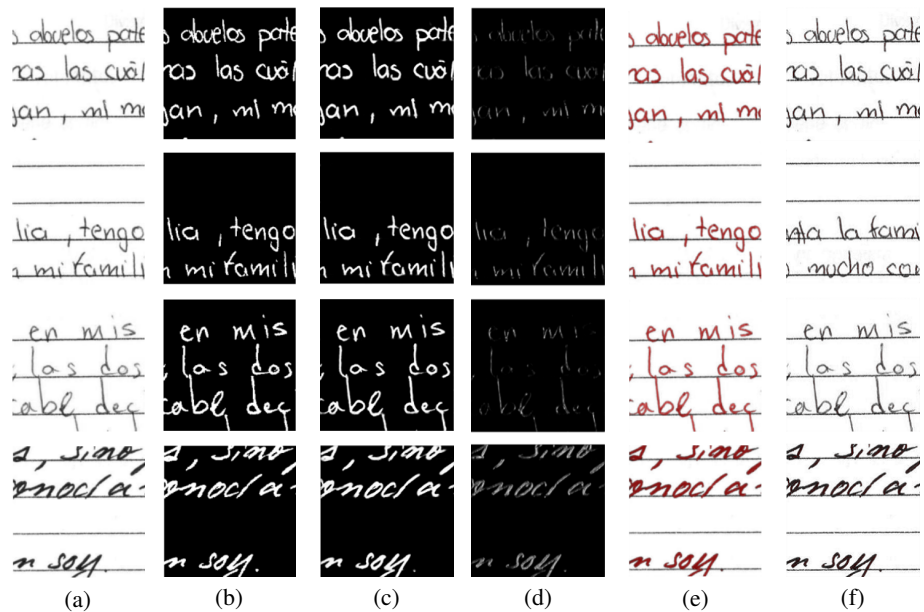
### 3.3. U-Net

La arquitectura U-Net (Fig. 6), ha demostrado ser altamente efectiva sobre todo en la segmentación de imágenes médicas debido a su capacidad para realizar predicciones precisas de la máscara de segmentación, incluso en situaciones en las que las regiones de interés son pequeñas o poco claras [15]. Esta es una Red Neuronal Convolutiva (RNC) profunda que consta de dos partes principales: la ruta de codificación (encoder) y la ruta de decodificación (decoder).

La ruta de codificación es similar a la arquitectura de una CNN típica, donde la imagen de entrada se reduce gradualmente en tamaño mediante la aplicación de capas de convolución y submuestreo (pooling). Por otro lado, la ruta de decodificación aumenta gradualmente el tamaño de la imagen mediante la aplicación de capas de convolución y sobremuestreo (upsampling).

Además, U-Net utiliza una técnica llamada “salto de conexiones” (skip connections), que conecta las capas de codificación y decodificación, permitiendo que la información de alta resolución se transmita directamente a las capas de decodificación, lo que ayuda a evitar la pérdida de detalles importantes en la imagen.

El modelo U-Net es muy útil porque puede entrenarse con relativamente pocas imágenes, lo que la hace ideal para aplicaciones en las que se dispone de un conjunto de datos pequeño o se necesita una segmentación rápida en tiempo real. Debido a esta razón se optó por trabajar con esta arquitectura, ya que el conjunto de imágenes de HWxPI obtenido después de la eliminación de líneas fue limitado.



**Fig. 8.** Muestras de resultados en la segmentación usando los dataset HWxPI y CVC-MUSCIMA, a) Parches original con líneas, b) Máscara, c) Predicción con HWxPI, d) Predicción con CVC-MUSCIMA, e) Intersección de a y c, f) Intersección de a y d.

#### 4. Experimentos

Se realizaron experimentos con las imágenes obtenidas del dataset HWxPI y con el fin de ampliar el conjunto de entrenamiento, se propuso dividir cada imagen en parches de 256 x 256 píxeles, lo que permitió generar un conjunto de datos compuesto por 450 imágenes (Fig. 7). Además, se realizó un aumento de datos, que dio como resultado 64 imágenes adicionales. Este aumento incluyó técnicas de rotación, espejo, acercamiento, estiramiento en lo alto y ancho.

Una de las ventajas del aumento de datos es que ayuda a reducir el sobreajuste al introducir variaciones en el conjunto de entrenamiento, lo que hace que el modelo sea menos propenso a memorizar las muestras y tenga una mayor capacidad de generalización. Para la definición de hiperparámetros, se utilizaron 64 filtros, un dropout de 0.3, batch size de 32 y una función de activación sigmoide. Por otro parte, la métrica utilizada para evaluar el modelo fue la intersección sobre la unión (Ec. 2):

$$IoU = \frac{\text{Área de intersección}}{\text{Área de unión}}. \quad (2)$$

Esta métrica consiste en dividir el área de intersección entre las predicciones del modelo y las etiquetas reales por el área de unión de ambas. Es ampliamente utilizada en tareas de segmentación y permite medir la similitud entre la segmentación realizada por el modelo y la segmentación real, siendo un valor de 1 indicativo de una segmentación perfecta y un valor de 0 indicativo de una segmentación completamente errónea.



**Tabla 1.** Resultados cuantitativos (mayor es mejor).

	IoU del conjunto HWxPI de prueba
Entrenado con CVC-MUSCIMA	0.60
Entrenado con HWxPI	<b>0.73</b>

Soy una persona intrínseca, le ocasiona estar solo en su totalidad, está a la par y las personas de su alrededor. Soy muy poco involucrada tanto sus decisiones y que necesite ir al lado de continuamente, es multicolor con el y sus actividades cotidianas, ya que su relación con sus amigos no fue en totalidad. Llena, ya que nunca me permitieron expresarse sobre mis padres afecto ellos mismos. Siempre estoy fuerte y sobrio no tener sentimientos nada real. Siempre estoy adelante y nunca soy confundida. Prefiero con ellos no haberlo nunca y al momento de cosas que soy homosexual. Mi padre me va a ayudar, pero mi madre no.

(a) Imagen original.

Soy una persona intrínseca, le ocasiona estar solo en su totalidad, está a la par y las personas de su alrededor. Soy muy poco involucrada tanto sus decisiones y que necesite ir al lado de continuamente, es multicolor con el y sus actividades cotidianas, ya que su relación con sus amigos no fue en totalidad. Llena, ya que nunca me permitieron expresarse sobre mis padres afecto ellos mismos. Siempre estoy fuerte y sobrio no tener sentimientos nada real. Siempre estoy adelante y nunca soy confundida. Prefiero con ellos no haberlo nunca y al momento de cosas que soy homosexual. Mi padre me va a ayudar, pero mi madre no.

(b) Imagen después de ser procesada.

**Fig. 9.** Resultado de modelo U-Net para una imagen.

Para la optimización del modelo se empleó el algoritmo de descenso de gradiente estocástico, junto con una función de pérdida de entropía cruzada binaria. Durante el entrenamiento, se obtuvo un valor de pérdida de 0.0667, mientras que el coeficiente de intersección sobre la unión (IoU) arrojó un resultado de 0.6847.

Al evaluar el modelo con el conjunto de pruebas, se logró obtener un valor de IoU de 0.733. Las imágenes resultantes se muestran en la figura 8(c). En cuanto a los experimentos realizados con CVC-MUSCIMA, de igual manera se generaron parches de cada imagen,  $256 \times 256$  píxeles, y se realizó aumentación de datos. Con esto se pasó de tener 1000 imágenes a 5,955. Se cambiaron los valores de dropout a 0.4 y el tamaño de batch size a 64.

Todos los demás parámetros se conservaron igual que con el entrenamiento previo. En el entrenamiento se obtuvo un valor de pérdida de 0.02, mientras que el resultado de IoU fue 0.98. En cambio, con el conjunto de evaluación (test), que en este caso fueron los 450 parches obtenidos de las 50 imágenes de nuestro dataset que resultó del preprocesamiento, se consiguió un coeficiente de IoU de 0.60. Las imágenes resultantes se muestran en la figura 8(d).

## 5. Resultados

Con el dataset CVC-MUSCIMA en la fase de entrenamiento se lograron resultados por encima de 0.90 en la métrica IoU, pero en las pruebas y utilizando un conjunto de imágenes externo, no se pudo generalizar la tarea de eliminar únicamente las líneas dando como resultado una gran pérdida de texto.

Al contrario, con el dataset  $HW \times PI$  durante la fase de entrenamiento se obtuvieron resultados ligeramente superiores a 0.60 en la métrica IoU. Sin embargo, en las pruebas, se obtuvieron valores significativamente elevados y se tuvo una mayor precisión en la segmentación de objetos, lo cual se tradujo en una remoción de líneas correctamente y texto no distorsionado.

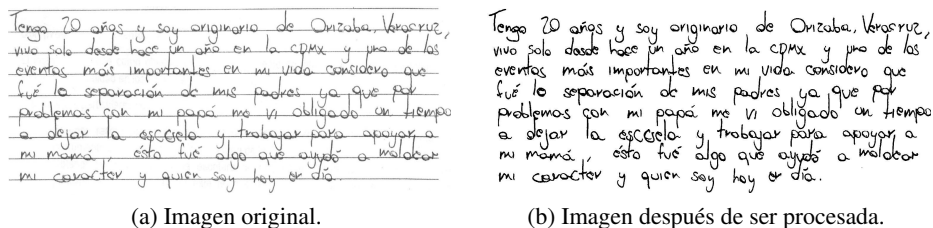


Fig. 10. Resultado de modelo U-Net para una imagen.

En la Tabla 1 se puede observar el resultado cuantitativo de los experimentos antes mencionados. En la figura 8(e) y 8(f), el color rojo en las letras resalta una intersección exitosa entre la imagen original y la predicción. Como puede verse, el color rojo en las predicciones resultantes del dataset de partituras es casi imperceptible.

Una vez establecido el modelo, se utilizaron todas las imágenes del dataset HW×PI. Cada imagen se dividió en parches de  $256 \times 256$  píxeles, generando un total de 6200 imágenes parche. Cada uno de estos parches fue procesado para remover las líneas presentes. Posteriormente, los parches de cada imagen se unieron nuevamente para restaurar la imagen original. Finalmente, todas las imágenes fueron binarizadas, ya que los resultados se obtuvieron en escala de grises. En la figura 9 y figura 10 se presentan ejemplos del resultado final.

## 6. Conclusiones y trabajo a futuro

Debido a que no existen datasets disponibles que contengan ejemplos de imágenes de textos con y sin líneas, se optó por experimentar con dos tipos de datasets, CVS-MUSIMA y HW×PI. El primero a pesar de ser un conjunto de imágenes para la remoción de líneas de pentagrama y de su similitud con el objetivo del presente trabajo, no tuvo un resultado favorable. Se encontró que el segundo pudo realizar una mejor generalización y sin una pérdida significativa de información.

Utilizar una arquitectura tipo U-Net demostró ser una herramienta valiosa permitiendo obtener buenos resultados en la eliminación de líneas en textos manuscritos, sin la necesidad de un realizar un trabajo adicional de reconstrucción. Aunque en algunos casos hubo que hacer una dilatación seguida de una erosión para acabar de cerrar algunos espacios, en general no hubo alteraciones del texto. A diferencia de trabajos previos en los que, utilizando técnicas estándar de procesamiento, después de eliminar las líneas tuvieron que implementar algoritmos de reconstrucción.

Si bien este trabajo ha demostrado el potencial de las redes neuronales convolucionales en la eliminación de líneas, es importante destacar que aún existen áreas que pueden ser exploradas en futuras investigaciones. Por ejemplo, se pueden proponer diferentes arquitecturas de red para mejorar aún más la precisión del modelo. Para el trabajo a futuro, se considera combinar conjuntos de imágenes de ensayos y partituras para el entrenamiento, así como evaluar con diversas imágenes de documentos. Por último, se planea desarrollar un dataset específico para este tipo de tarea y ponerlo a disposición de aquellos que trabajen en áreas afines.

## Referencias

1. Schantz, H. F.: History of OCR, optical character recognition. Recognition Technologies Users Association (1982)
2. Valdez-Rodríguez, J. E., Calvo, H., Felipe-Riverón, E. M.: Handwritten texts for personality identification using convolutional neural networks. *Pattern Recognition and Information Forensics*, pp. 140–145 (2018) doi: 10.1007/978-3-030-05792-3\_13
3. LeCun, Y., Kavukcuoglu, K., Farabet, C.: Convolutional networks and applications in vision. In: *Proceedings of IEEE International Symposium on Circuits and Systems* (2010) doi: 10.1109/iscas.2010.5537907
4. Farahmand, A., Sarrafzadeh, A., Shanbezadeh, J.: Document image noises and removal methods. *Lecture Notes in Engineering and Computer Science*, vol. 2202, pp. 436–440 (2013)
5. Soille, P.: *Morphological image analysis: Principles and applications*. 2nd edn., Springer, Berlin (2004) doi: 10.1007/978-3-662-05088-0
6. Duda, R. O., Hart, P. E.: Use of the hough transformation to detect lines and curves in pictures. *Communications of the Association for Computing Machinery*, vol. 15, no. 1, pp. 11–15 (1972) doi: 10.1145/361237.361242
7. Namboodiri, A. M., Jain, A. K.: Document structure and layout analysis. *Digital Document Processing*, pp. 29–48 (2007) doi: 10.1007/978-1-84628-726-8\_2
8. Gold, C., Zesch, T.: CNN-based ruled line removal in handwritten documents. *Frontiers in Handwriting Recognition*, pp. 530–544 (2022) doi: 10.1007/978-3-031-21648-0\_36
9. Shi, Z., Setlur, S., Govindaraju, V.: Removing rule-lines from binary handwritten arabic document images using directional local profile. In: *20th International Conference on Pattern Recognition* (2010) doi: 10.1109/icpr.2010.472
10. Prokopiou, K., Kavallieratou, E., Stamatatos, E.: An image processing self-training system for ruling line removal algorithms. In: *18th International Conference on Digital Signal Processing* (2013) doi: 10.1109/icdsp.2013.6622767
11. Refaey, M. A.: Ruled lines detection and removal in grey level handwritten image documents. In: *6th International Conference on Information and Communication Systems* (2015) doi: 10.1109/iacs.2015.7103230
12. Ramírez, G., Villatoro, E., Ionescu, B., Escalante, H. J., Escalera, S., Larson, M., Müller, H., Guyon, I.: Overview of the multimedia information processing for personality and social networks analysis contest. *Pattern Recognition and Information Forensics*, pp. 127–139 (2018) doi: 10.1007/978-3-030-05792-3\_12
13. Fornés, A., Dutta, A., Gordo, A., Lladós, J.: CVC-MUSCIMA: A ground truth of handwritten music score images for writer identification and staff removal. *International Journal on Document Analysis and Recognition*, vol. 15, no. 3, pp. 243–251 (2011) doi: 10.1007/s10032-011-0168-2
14. Vincent, L.: Morphological area openings and closings for grey-scale images. *Shape in Picture*, pp. 197–208 (1994) doi: 10.1007/978-3-662-03039-4\_13
15. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science*, pp. 234–241 (2015) doi: 10.1007/978-3-319-24574-4\_28