PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

Escuela de Posgrado

APRENDIZAJE PROFUNDO PARA TRANSCRIPCIÓN DE TEXTOS HISTÓRICOS MANUSCRITOS EN ESPAÑOL

Trabajo de investigación para obtener el grado académico de Maestro en Informática con mención en Ciencias de la Computación que presenta:

Gustavo Jorge Choque Dextre

asesor:

Dr. Cesar Armando Beltrán Castañón

Lima, 2024

Informe de Similitud

Yo, **César Armando BELTRÁN CASTAÑÓN**, docente de la Escuela de Posgrado de la Pontificia Universidad Católica del Perú, asesor de el trabajo de investigación titulado **"Aprendizaje profundo para transcripción de textos históricos manuscritos en español"** de el autor **Gustavo Jorge CHOQUE DEXTRE**, dejo constancia de lo siguiente:

- El mencionado documento tiene un índice de puntuación de similitud de 07%. Así lo consigna el reporte de similitud emitido por el software *Turnitin* el 28/05/2024.
- He revisado con detalle dicho reporte y el trabajo de investigación, y no se advierte indicios de plagio.
- Las citas a otros autores y sus respectivas referencias cumplen con las pautas académicas.

Lugar y fecha:

San Miguel, 28 de Mayo de 2024.

Apellidos y nombres del asesor / de la asesora: BELTRÁN CASTANÓN, César Armando	
DNI: 29561260	Firma 
ORCID: 0000-0002-0173-4140	

Dedicatoria

Dedico este trabajo a mis padres, quienes, enfrentando las vicisitudes de la vida, siempre lograron darme lo mejor. Asimismo, agradezco profundamente a mi esposa por su apoyo incondicional y su constante aliento.

Agradecimientos

Expreso mi agradecimiento al Grupo de Inteligencia Artificial IA-PUCP por proporcionar los servidores donde ejecutaron los modelos utilizados en esta investigación. También deseo agradecer a mi asesor en el curso de Seminario de Tesis II, el PhD Cesar Armando Beltran Castañon, así como al MSc Ferdinand Edgardo Pineda Ancco, por su valiosa orientación y apoyo.

Resumen

El reconocimiento de textos históricos es considerado un problema desafiante debido a los muchos factores que alteran el estado de los manuscritos y la complejidad de los diferentes estilos de escritura involucrados en este tipo de documentos; en los años recientes se han creado muchos modelos de Reconocimiento de textos manuscritos enfocados en diversos idiomas como el inglés, chino, árabe y japonés entre otros, sin embargo no se han encontrado muchas iniciativas de reconocimiento de texto orientadas al idioma español debido fundamentalmente a un escasez de datasets públicos disponibles para ayudar a solucionar la problemática en dicho idioma.

En esta publicación se presenta la aplicación de técnicas de Deep Learning basadas en una arquitectura de red neuronal encoder-decoder y convoluciones compuerta Gated-CNN las cuales en los últimos ha demostrado resultados sobresalientes para resolver dicha problemática, así mismo se propone la aplicación de mecanismos de Transferencia de Aprendizaje para el reconocimiento de textos históricos en español. Los experimentos demuestran que la aplicación de estos métodos puede brindar resultados sobresalientes, además la aplicación de otras técnicas tales como Aumentación de Datos y Modelos de Lenguaje conllevan a mejoras significativas en los resultados finales. Se propone además el uso de un nuevo dataset de textos históricos en español conformado por 1000 elementos tomados de textos históricos peruanos referentes al siglo XVIII.

ÍNDICE

Lista de Tablas

Lista de Figuras

Aprendizaje Profundo para transcripción de Textos Históricos manuscritos en español

1st Gustavo Jorge Choque Dextre
Escuela de Posgrado
Pontificia Universidad Católica del Perú
Lima, Perú
a20144012@pucp.pe

2st Cesar Beltran Castañon
Escuela de Posgrado
Pontificia Universidad Católica del Perú
Lima, Perú
cbeltran@pucp.edu.pe

***Abstract*—El reconocimiento de textos históricos es considerado un problema desafiante debido a los muchos factores que alteran el estado de los manuscritos y la complejidad de los diferentes estilos de escritura involucrados en este tipo de documentos; en los años recientes se han creado muchos modelos de Reconocimiento de textos manuscritos enfocados en diversos idiomas como el inglés, chino, árabe y japonés entre otros, sin embargo no se han encontrado muchas iniciativas de reconocimiento de texto orientadas al idioma español debido fundamentalmente a un escasez de datasets públicos disponibles para ayudar a solucionar la problemática en dicho idioma.**

En esta publicación se presenta la aplicación de técnicas de Deep Learning basadas en una arquitectura de red neuronal encoder-decoder y convoluciones compuerta Gated-CNN las cuales en los últimos ha demostrado resultados sobresalientes para resolver dicha problemática, así mismo se propone la aplicación de mecanismos de Transferencia de Aprendizaje para el reconocimiento de textos históricos en español. Los experimentos demuestran que la aplicación de estos métodos puede brindar resultados sobresalientes, además la aplicación de otras técnicas tales como Aumentación de Datos y Modelos de Lenguaje conllevan a mejoras significativas en los resultados finales. Se propone además el uso de un nuevo dataset de textos históricos en español conformado por 1000 elementos tomados de textos históricos peruanos referentes al siglo XVIII.

***Index Terms*—Aprendizaje Profundo, Reconocimiento de Texto Manuscrito, Red Neuronal Convolucional, Red Neuronal Recurrente, Transferencia de Aprendizaje.**

I. Introducción

En los últimos años el Reconocimiento de Textos Manuscritos (RTM) ha experimentado un notable aumento en su relevancia, impulsado tanto por su diversidad de aplicaciones a nivel de industria como por el creciente interés de la comunidad de investigación en visión computacional.

El Reconocimiento de textos aplicado a documentos históricos es considerado un problema desafiante debido a las variaciones en el estilo de escritura de los escritores y las degradaciones que sufren los documentos como resultado del paso del tiempo, manchas, trazos de lápiz e iluminación desigual.

Durante muchos años los Modelos Ocultos de Markov (HMM) [10] fueron muy populares para resolver el problema de los sistemas de RTM sin embargo en los últimos años, los métodos basados en Aprendizaje Profundo (AP) precisamente las Redes Neuronales Recurrentes Convolucionales (CRNN) y las Redes Convolucionales Compuerta (Gated-CNN) han tenido resultados sobresalientes para reconocimiento de este tipo de textos [1] [2].

Es importante destacar que la realización de experimentaciones con nuevos modelos de Reconocimiento de Texto Manuscrito está estrechamente ligada a la disponibilidad de datasets. Como resultado, la mayoría de las investigaciones y experimentos se han centrado en idiomas como el inglés [14], chino [15] , árabe [16], japonés [17] y varias lenguas indias [4]. Esto se debe en gran medida a la existencia de datasets públicos para estos idiomas, lo que ha facilitado la realización de estas investigaciones. Adicionalmente, al tratarse de RTM para documentos históricos, la disponibilidad de conjunto de datos se reduce aún más significativamente.

En este trabajo se propone utilizar una arquitectura basada en Convoluciones Compuertas (Gated-CNN) [2] debido principalmente a que este tipo de convoluciones han demostrado mejores resultados, como se detalla en este estudio. Ademas, se aplicaron algunos ajustes especificos, como la modificación en la cantidad de capas recurrentes apiladas LSTM. Este modelo fue entrenado con el dataset IAM [8], para poder realizar el reconocimiento de textos regulares de años recientes, posteriormente se realizó un procedimiento de Transferencia de Aprendizaje (TA) utilizando el dataset de textos históricos en español propuesto.

Las contribuciones están basadas en los siguientes aspectos: a) aprovechar el Aprendizaje por Transferencia para permitir RTM de manuscritos históricos, donde basado en el entrenamiento con un dataset público IAM en idioma inglés, se realizó un fine tuning de la red para su entrenamiento con un dataset más pequeño en idioma español , b) la creación de un dataset de textos históricos basados en documentos del siglo XVIII tomados de

documentos históricos peruanos conformado por 1000 imágenes de palabras debido a la escasez de dataset publico históricos en el idiomas español.

El resto del paper está organizado de la siguiente manera: en la parte 2 se referencia los modelos RTM basados en la literatura, en la parte 3 se explican las características del dataset propuesto, en la parte 4 se detalla la metodología, en la sección 5 los resultados experimentales obtenidos por el modelo. Finalmente, en la sección 6 las conclusiones que resumen este trabajo.

II. Estado del Arte

El objetivo de esta sección es el de investigar sobre los avances más recientes en aplicaciones relacionadas a RTM para textos históricos.

Durante algunos años las redes recurrentes tipo LSTM mostraron buenos resultados, sin embargo el entrenamiento de estos modelos conllevaba un alto costo computacional lo que motivó la propuesta presentada en [1] donde se propone el uso de redes convolucionales en reemplazo de redes multidimensionales para la decodificación, alcanzando resultados sobresalientes y optimizando la cantidad de tiempo de procesamiento. Por otro lado, debido a la profundidad de la red ocurre que se pierden los features mientras se va llega capas posteriores lo que motivó que en [2] se proponga el uso de Convoluciones Compuerta Gated-CNN, las cuales controlan la propagación de features a capas posteriores; la compuerta es implementada como una red convolucional más. El modelo Atención basado en Transformers ha mostrado excelentes resultados para muchos ámbitos de la visión computacional, el uso de este modelo se plantea en [3] donde se explora la aplicación de una arquitectura encoder-decoder utilizando un Enfoque de Atención para la decodificación.

El uso de Modelos de Lenguaje estadísticos también ha tenido grandes aportes, por ejemplo en [3] se utiliza un modelo de encoder decoder con un módulo semántico para tratar textos en lenguas de la India, este módulo está basado en modelos de lenguaje pre-entrenados.

Si bien es cierto la mayoría de propuestas abarca el reconocimiento a nivel de líneas, en [12] se presenta un sistema end-to-end para reconocimiento de párrafos utilizando una red de Atención Vertical alcanzando resultados competitivos. La variabilidad de estilos de escritura es todo un reto en el RTM de documentos históricos es por eso que en [13] se propone el uso de redes Convolucionales Deformables para sobrellevar este problema dado que las convoluciones convencionales se mueven sobre cuadrillas fijas ignorando que los caracteres escritos a mano pueden variar de forma, escala y orientación.

III. Dataset Propuesto

En este trabajo de investigación se propone la creación de un dataset para textos históricos basado en documentos generados en el siglo XVIII en el Perú tomados de la Biblioteca Nacional del Perú [9], el cual se ha nombrado Dataset Phi.

Las características de este dataset se muestran en la Tabla I.

TABLE I
CARACTERÍSTICAS DEL DATASET PHI.

Característica	valor
Cantidad de imágenes	1000
Siglo de los documentos	XVIII
Cantidad de escritores	12
Nro. de Páginas promedio por documento	60
Nro. de tipos de letras y signos	80

En la tabla II se brinda la vista de los diferentes títulos, autores de los escritos y año de los diferentes documentos que fueron considerados para la creación del dataset.

TABLE II
MUESTRA DE LA LISTA DE DOCUMENTOS QUE FUERON USADOS PARA LA CONSTRUCCIÓN DEL DATASET

Código	Título	Escritor	Año
91	'Expediente sobre la merced de Título ... '	Josef Antonio Becerra	1796
92	'Expediente formado sobre una ...'	Ruiz de Castilla Gobernador Provincia de Puno	1800
97	'Superior oficio de 13 de abril del...'	Virrey Barón de Ballenary	1796

Además, en la Tabla III se brinda la vista de algunas páginas extraídas de los diferentes documentos históricos que corresponden a una variedad de estilos de escritura.

El dataset será publicado en el siguiente repositorio github : https://github.com/gustvjor2005/dataset-phi .

IV. Metodología

En esta sección se detalla el método propuesto para el reconocimiento de textos manuscritos en idiomas español. En primer lugar, es detallada la arquitectura, luego los procedimientos para realizar Aumento de Datos (AD) y finalmente los métodos para el entrenamiento de la red.

A. Arquitectura

Esta arquitectura es resultado de realizar un benchmark de arquitecturas de RTM que tienen los mejores resultados, en la tabla IV se muestran la información detallada de este análisis, los valores mostrados fueron tomados de otros estudios.

Como se puede apreciar en la tabla IV, existen muchos modelos que tratan de abarcar la problemática del RTM, sin embargo para el presente trabajo se tomó como referencia el trabajo “Gated Convolutional Recurrent Neural Networks“ [2] debido a que mostró resultados

TABLE III
MUESTRA DE EJEMPLO DE LOS DIFERENTES TIPOS DE IMÁGENES USADA EN EL DATASET.

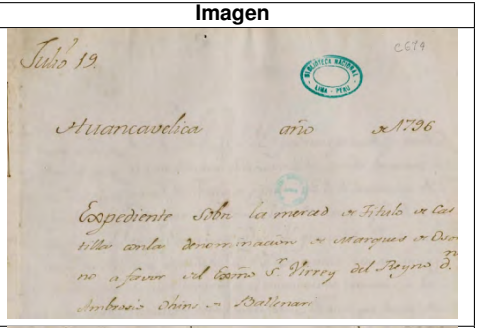
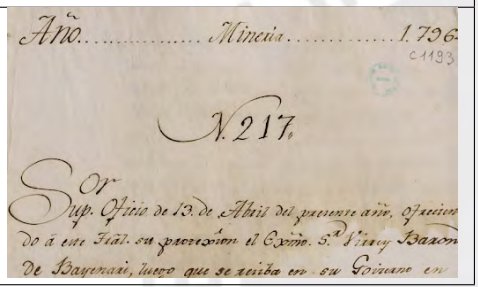
Código	Imagen
91	
92	
97	

TABLE IV
BENCHMARK DE ARQUITECTURAS Y MODELOS CON LOS MEJORES RESULTADOS.

Paper Title	CER	WER	DATASET
Are Multidimensional Recurrent Layers Really Necessary for Handwritten Text Recognition? [1]	5.8	18.14	IAM
Gated Convolutional Recurrent Neural Networks for Multilingual Handwriting Recognition [2]	3.2	10.5	IAM
HTR-Flor++ [5]	5.1	16.2	IAM
Boosting Modern and Historical Handwritten Text Recognition with Deformable Convolutions [13]	6.8	24.7	IAM
AttentionHTR: Handwritten Text Recognition Based on Attention Encoder-Decoder Networks [3]	6.5	15.4	IAM

CER and WER metrics indicate the error level of the models.

sobresalientes y se acopla sin problemas al nuevo modelo propuesto el cual está basado en una arquitectura Encoder-Decoder, una vista general de la arquitectura es presentada en la Figura 1.

El detalle de la arquitectura propuesta es el siguiente:

- Bloque Convolucional de Codificación: El primer bloque de convolución está formado por una capa de convolución con un kernel de 5x5 con 32 características y se aplica una Normalización de batches (BN). Como función de activación se utilizan Unidades Lineares Rectificadoras (ReLU) y finalmente se aplica un Max Pooling con un kernel de 1x2.
  El segundo bloque de convolución es similar al anterior, pero cuenta con 64 características y además se aplica una Convolución Compuerta (Gated-CNN) de 5x5 con igual cantidad de características.
  En el tercer bloque de convolución a diferencia del anterior se aplica un kernel de 3x3 con 128 características y una Convolución Compuerta de 3x3.
  El cuarto bloque es similar al tercero y finalmente para el quinto bloque de convolución se aplica un kernel de 3x3 con 256 características y una Convolución Compuerta de 3x3 con igual cantidad de características.
- Bloque Convolucional de Decodificación: Este bloque está formado por una capa LSTM bidireccional conformada por tres capas apiladas con 256 unidades LSTM y se establece un dropout de 0.5 para evitar sobreajuste.
  Finalmente, para mapear la salida se aplica una capa de Convolución con un kernel de 1x1 con una cantidad de características de salida igual a la cantidad de clases correspondiente al tamaño de lista de caracteres considerado para el presente trabajo.
- CTC Beam Search: Después de entrenar el modelo, en la etapa de inferencia se utilizó el algoritmo de CTC Beam Search para encontrar la salida más probable y de esta forma no obviar el hecho que una salida puede tener múltiples alineaciones [18].
- Modelo de Lenguaje N-gram: Para mejorar el resultado del modelo se aplicó un Modelo de Lenguaje basado en N-gram utilizando la biblioteca KemLM1, esto para estimar la ocurrencia de secuencia de

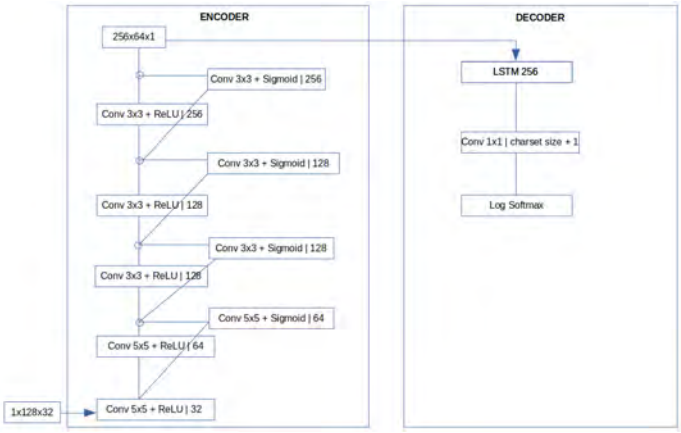


Fig. 1. Arquitectura de tipo Encoder-Decoder Propuesta.

palabras. Este modelo de lenguaje fue entrenado con un corpus conformado por 842 líneas de texto tomadas de documentos históricos de la Biblioteca Nacional del Perú (BNP) que incluye el dataset y tiene como objetivo minimizar la tasa de error a nivel de palabra.

B. Preprocesamiento y Aumento de datos

En el pre-procesamiento de las imágenes se utilizó la función convertScaleAbs para aumentar el contraste en las imágenes que están en escala de grises, dado que el fondo de las imágenes del dataset Phi no es totalmente claro.

Para el aumento de datos en la etapa de entrenamiento se utilizaron transformaciones fotométricas de las imágenes, así como también se aplicaron filtros de suavizamiento con el fin de reducir el nivel de ruido de las imágenes tales como filtros mediana y gaussiano. Finalmente se aplicaron operaciones morfológicas de erosión y dilatación [21] con el fin de aumentar la data.

En la Figura 2. se puede apreciar la aplicación de erosión sobre un elemento de entrenamiento.

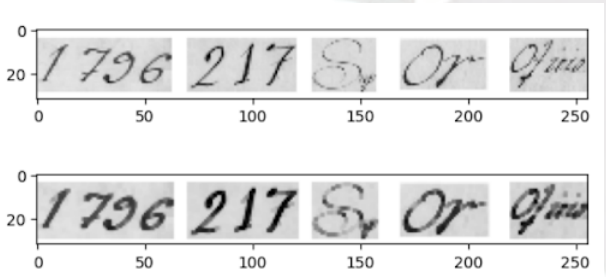


Fig. 2. En la parte superior una línea de texto original. En la parte inferior se muestra la aplicación de erosión

C. Entrenamiento

Para el entrenamiento del modelo de deep learning se ha utilizado un hardware de última generación. Se emplearon dos tarjetas de tipo NVIDIA RTX A5500 con capacidad de 24GB. Además el entrenamiento se llevó a cabo en un servidor linux con una memoria total de 250GB.

Primero se realizó el pre entrenamiento utilizando el dataset IAM para textos regulares seguidamente se aplicó un procedimiento de transferencia de aprendizaje (TA) mediante un mecanismo de fine tuning en el bloque de decodificación de la red, esto nos sirve para adaptar la salida de la red a la cantidad de caracteres asociados al idioma español.

El entrenamiento se realizó con el objetivo de minimizar la función de pérdida Clasificación Temporal Conexionista (CTC [6]).

El procedimiento de optimización se realizó con un mecanismo de Descenso de Gradiente Estocástico (DGD) utilizando el método RMSProp [7] con una tasa de aprendizaje de 0.001 considerando un tamaño de batch de 100 elementos para el pre-entrenamiento y de 5 para para el fine tuning.

El entrenamiento fue detenido cuando la métrica de evaluación no mejoró después de 25 veces consecutivas para la etapa de pre entrenamiento y 21 veces para el fine tuning.

El tamaño del beam search utilizado para la decodificación de los caracteres en ambas etapas fue de 50, en la Tabla V se detallan los hiper-parámetros utilizados en ambos procesos.

TABLE V
HIPERPARAMETROS USADOS PARA EL ENTRENAMIENTO DEL MODELO

	Pre-training	**Fine tuning**
tamaño de batch	100	5
tasa de aprendizaje	0.001	0.0005
número de épocas	163	44
early stopping	25	21

V. EXPERIMENTACIÓN Y RESULTADOS

Los experimentos se realizaron de la siguiente forma:

- Paso 1:Se realizó el entrenamiento en modo línea del modelo propuesto utilizando el dataset IAM para textos regulares obteniendo un TEC de 6,8 y un TEP de 19,64.
- Paso 2:Se realizó la Transferencia de Aprendizaje mediante un procedimiento de *fine tuning*, donde se inicializó la red con los parámetros de la red previamente entrenada, utilizando el dataset Phi para textos históricos.

A. Métricas de Evaluación

Para medir la precisión de la red fueron utilizadas las dos métricas más importantes de los sistemas de RTM: Tasa de Error de Caracteres (TEC) y Tasa de Error de Palabras (TEP).

La TEC es calculada mediante el conteo del total de errores a nivel de caracteres (inserciones, eliminaciones, sustituciones) entre el texto reconocido y el texto real dividido por el número total de caracteres. La TEP se calcula de una manera similar considerando los errores a nivel de palabras. “(1)”.

$$\text{TEC} = \frac{S + D + I}{N} \tag{1}$$

- S: número de sustituciones.
- D: número de eliminaciones.
- I: número de inserciones.
- N: número total de caracteres en el texto referenciado.

Ambas métricas se pueden medir a nivel porcentual.

B. Data

Para el entrenamiento del modelo se trató de utilizar la mayor cantidad de datos, partiendo del dataset IAM modo palabras se formaron imágenes de líneas tal como se muestra en la Figura 3.

La cantidad de palabras utilizadas en el pre-entrenamiento fue obtenida de forma aleatoria entre 1 a 8, tal como muestra en el Tabla VI. Además se realizó el mismo procedimiento de conformación de líneas de texto para el entrenamiento con el dataset Phi, en la Figura 3 se muestra una imagen de ejemplo.

TABLE VI
DETALLE DE FORMACIÓN DE LÍNEAS EN BASE A PALABRAS PARA EL ENTRENAMIENTO.

Idioma	Dataset	Cant. Palabras por línea
Inglés	IAM	Aleatorio de 1 a 8
Español	Phi	Aleatorio de 2 a 7

Cantidad mínima y máxima de palabras usadas obedecen a una asignación experimenta.

Para la evaluación del modelo se utilizaron 5 palabras fijas por línea para ambos datasets.

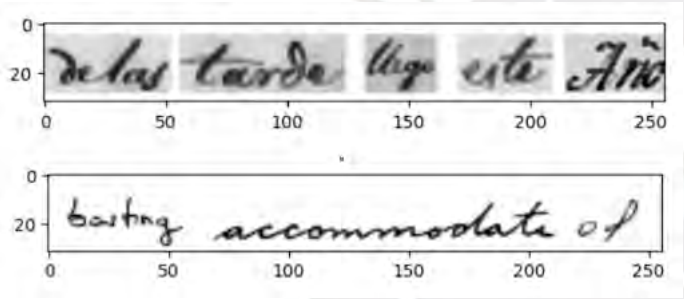


Fig. 3. En la parte superior e inferior de la imagen se muestra un ejemplo de creación de líneas para el dataset Phi y el dataset IAM respectivamente.

C. Resultados del Entrenamiento

La Figura 4 muestra el comportamiento de la función de pérdida CTC.

En la Figura 4 se puede observar que el modelo converge a una solución óptima, se aprecia además que después de la época 80 la disminución en la función de pérdida ya no es significativa. En este experimento, se realizaron un total de 145 épocas.

En la Figura 5 muestra el comportamiento de la TEC (Tasa de Error de Caracteres) , se observa que después de la época 40 no hay una disminución significativa en la métrica, alcanzando un valor final de 6,8. A pesar de ello, la aplicación del early stopping prolongó el entrenamiento, ya que cada decimal de mejora en esta métrica es crucial para mejorar la precisión del modelo.

Como siguiente paso se procedió a entrenar el modelo con el dataset Phi generando los resultados de línea base mostrados en la Tabla VII, posteriormente

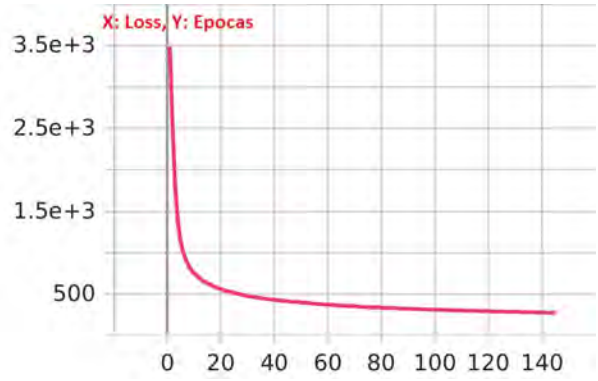


Fig. 4. Comportamiento de la función de pérdida en el pre-entrenamiento.

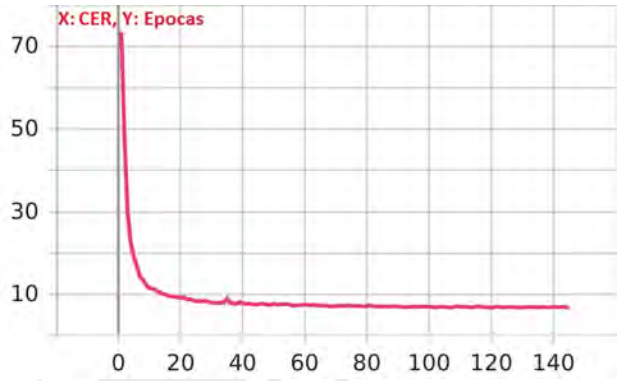


Fig. 5. Comportamiento de la TEC en el pre-entrenamiento.

se aplicó Transferencia de Aprendizaje utilizando los pesos obtenidos en el pre-entrenamiento y finalmente se aplicaron diversas técnicas de Aumento de Datos y finalmente se agregó la utilización del Modelo de Lenguaje previamente estimado. El resumen de los resultados se muestran en la Tabla VII.

TABLE VII
COMPARACIÓN DE RESULTADOS DE TEC Y TEP DEL MODELO INCORPORANDO DIFERENTES TÉCNICAS

Técnicas aplicadas	TEC	TEP
Línea base	52.8	76.9
Línea base + Transfer Learn.	17.7	39.0
Línea base + Transfer Learn. + Data Aug.	16.1	44.0
Línea base + Transfer Learn. + Data Aug. + LM	13.0	39.8

En la tabla VII podemos apreciar que cuando la red es entrenada desde cero las métricas de TEC y TEP muestran valores significativamente altos lo que indica un bajo nivel de precisión. Esto se debe a que el conjunto de datos Phi es relativamente pequeño, lo que limita la capacidad del modelo para aprender la amplia variedad de estilos de escritura de diferentes letras y símbolos.

Por esta razón se considera beneficioso emplear el

mecanismo de Transferencia de Aprendizaje para poder aprovechar lo aprendido previamente por el modelo a través de un conjunto de datos más extenso, como lo es dataset IAM.

Los resultados realizando fine tuning son alentadores debido a que reducen el TEC a 17.7%, además aplicando un procedimiento de Aumento de Datos, detallado en la sección Procesamiento y Aumento de Datos, se observa una mejora en la métrica TEC de 1.6 puntos porcentuales y finalmente la aplicación del modelo de lenguaje tuvo un efecto importante en la disminución de la tasa de error de 3 puntos.

Para poder medir la tasa de precisión del modelo calculamos:

$$\text{precisión} = 100 - TEC \tag{2}$$

- TEC: Tasa de Error de Caracteres.

En la Figura 6 se muestran las mejoras en la tasa de precisión del modelo aplicando todos los mecanismos antes mencionados, se observa que la aplicación de Transfer Learning tuvo un mayor impacto en la mejora de la precisión del modelo.

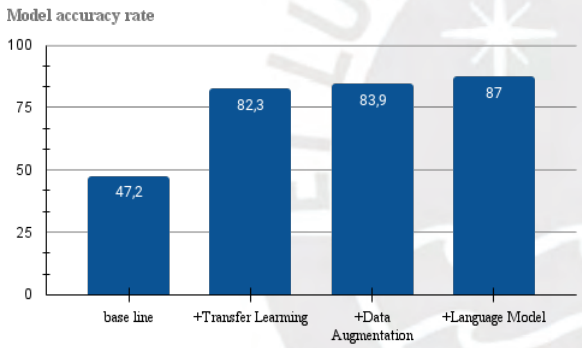


Fig. 6. Evolución de la precisión del modelo a nivel de CER

En la Figura 7. Se muestran resultados de la aplicación del modelo, donde partiendo de una línea de texto se extiende la imagen con nuevas palabras. Dado que el modelo fue entrenado con líneas formadas con un máximo de 8 palabras, lo cual se detalla en la tabla VI, se observa que al aumentar la longitud de la línea de texto se produce una disminución en la precisión del modelo.

VI. Conclusión y Trabajo Futuro

En este trabajo de investigación se ha presentado una aplicación de Aprendizaje Profundo para el Reconocimiento de Textos Manuscritos Históricos en Español mediante un procedimiento de Transferencia de Aprendizaje sobre un modelo pre-entrenado. Los resultados obtenidos han sido alentadores a nivel de la métrica TEC y TEP en contraste con un enfoque que no considera parámetros pre-entrenados. Así también se explora el aporte de otras técnicas como Aumento de Datos y Modelo de Lenguaje n-gram para la mejora de los resultados.

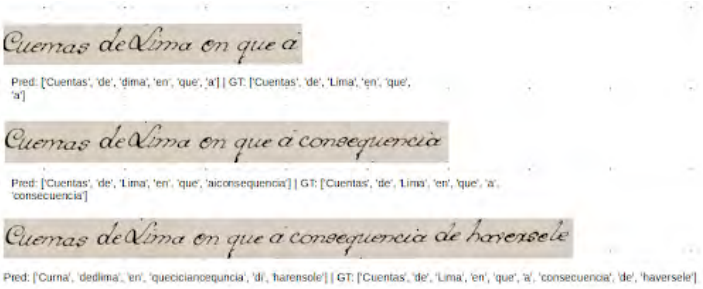


Fig. 7. Ejemplo de reconocimiento de líneas de texto y decaimiento de la precisión al aumentar el tamaño de la línea de texto.

Este avance se ha alcanzado explorando la aplicación de un nuevo conjunto de datos, propuesto en este trabajo de investigación, de textos históricos denominado *Dataset Phi*. Para un trabajo futuro se pretende enriquecer el conjunto de datos con un mayor número de elementos de lenguaje como abreviaturas y signos de puntuación lo cual impacta directamente en la precisión del modelo.

En relación con la red neuronal, la integración de un modelo de Transformaciones Espaciales [20] durante el proceso de entrenamiento podría conllevar a obtener mejores resultados. Esta técnica podría ser una alternativa eficaz para lidiar con la considerable variabilidad de trazos característicos presentes en los textos históricos manuscritos.

References

[1] J. Puigcerver, "Are Multidimensional Recurrent Layers Really Necessary for Handwritten Text Recognition?", November 2017.

[2] T. Bluche and R. Messina, "Gated Convolutional Recurrent Neuronal Networks for Multilingual Handwriting Recognition", in Proceedings of the 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), November 2017.

[3] D. Kass and E. Vats, "AttentionHTR: Handwritten Text Recognition Based on Attention Encoder-Decoder Networks.", May 2022.

[4] A. Mondal and C. Jawahar, "Enhancing Indic Handwritten Text Recognition using Global Semantic Information.", December 2022.

[5] A. Sousa, B. Dantas, A. Toselli, and E. Baptista, "HTR-Flor: A Deep Learning System for Offline Handwritten Text Recognition", in Proceedings of the 35th Conference on Graphics, Patterns and Images (SIBGRAPI), 2020.

[6] A. Graves, S. Fernández, F. Gomez, and J. Schmidhuber, "Connectionist Temporal Classification: Labeling Unsegmented Sequence Data with Recurrent Neural Networks", in Proceedings of the International Conference on Machine Learning, pp. 369-–376, 2006.

[7] G. Tieleman and G. Hinton, "Lecture 6.5-RMSProp: Divide the gradient by a running average of its recent magnitude", in COURSERA: Neural Networks for Machine Learning, vol. 4, no. 2, 2012.

[8] U.-V. Marti and H. Bunke, "The IAM-database: an English sentence database for offline handwriting recognition", International Journal on Document Analysis and Recognition, vol. 5, no. 1, pp. 39–46, 2002.

[9] "Evento Transcripton BNP", [Online]. Available: <https://memoriamanuscrita.bnp.gob.pe/ >.

[10] T. Plötz and G. A. Fink, "Markov models for offline handwriting recognition: a survey", International Journal on Document Analysis and Recognition, vol. 12, pp. 269–298, 2009. [Online]. Available: <https://doi.org/10.1007/s10032-009-0098-4 >
[11] U.-V. Marti and H. Bunke, "The IAM-database: An English sentence database for offline handwriting recognition", International Journal on Document Analysis and Recognition, vol. 5, no. 1, pp. 39-46, 2002.
[12] D. Coquenet, C. Chatelain, and T. Paquet, "End-to-End Handwritten Paragraph Text Recognition Using a Vertical Attention Network", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 1, pp. 508–524, 1 Jan. 2023, doi: 10.1109/TPAMI.2022.3144899.
[13] S. Cascianelli, M. Cornia, L. Baraldi, et al., "Boosting modern and historical handwritten text recognition with deformable convolutions", International Journal on Document Analysis and Recognition, vol. 25, pp. 207-217, 2022. [Online]. Available: <https://doi.org/10.1007/s10032-022-00401 >
[14] A. Graves and J. Schmidhuber, "Offline handwriting recognition with multidimensional recurrent neural networks", in NIPS, 2008.
[15] Z. Xie, Z. Sun, L. Jin, Z. Feng, and S. Zhang, "Fully convolutional recurrent network for handwritten Chinese text recognition" in ICPR, 2016.
[16] R. Maalej and M. Kherallah, "Improving the DBLSTM for on-line Arabic handwriting recognition", Multimedia Tools and Applications, 2020.
[17] K. C. Nguyen, C. T. Nguyen, and M. Nakagawa, "A semantic segmentation-based method for handwritten Japanese text recognition", in ICFHR, 2020.
[18] A. Hannun, "Sequence Modeling with CTC", Distill, 2017.
[19] F. Andreas, "Handwriting recognition in historical documents", 2012.
[20] M. Jaderberg, K. Simonyan, A.Zisserman, K.Kavukcuoglu, "Spatial Transformer Networks", 2012.
[21] OpenCV, "Erosion and Dilation," OpenCV Documentation. Available: https://docs.opencv.org/3.4/db/df6/tutorial_erosion_dilatation.html. Accessed: August 2023.