

UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

Aplicación de EasyOCR para la extracción automática de información en documentos digitalizados: estudio de caso en Banco Davivienda

---

INFORME DE PASANTÍA PARA OPTAR POR EL TÍTULO DE MATEMÁTICA
PROYECTO CURRICULAR DE MATEMÁTICAS

Manuela Cuervo Rodríguez
Dirigido por: Luis Alejandro Másmela Caita
Profesional encargado: Giovanny Rozo

Bogotá DC
Mayo de 2026

Resumen

En este trabajo se analizó el desempeño de la herramienta EasyOCR en la extracción automática de información contenida en documentos digitalizados. El estudio se desarrolló a partir de las actividades de procesamiento documental realizadas durante la pasantía en Banco Davivienda, donde se empleaban herramientas de inteligencia artificial para automatizar la captura de información.

Para la evaluación se implementó EasyOCR sobre un conjunto de 32 cédulas de ciudadanía digitalizadas. Se aplicaron técnicas de preprocesamiento de imágenes, incluyendo corrección de orientación y filtrado gaussiano, y posteriormente se calcularon métricas de precisión para diferentes campos de información presentes en los documentos.

Los resultados muestran que la configuración compuesta por GPU y preprocesamiento de imágenes alcanzó una precisión general de 90.31%, superando la obtenida mediante la configuración sin preprocesamiento. Asimismo, se observó que la calidad de las imágenes influye directamente en el desempeño del sistema, especialmente en campos con una mayor cantidad de texto.

Finalmente, el estudio permitió comprender los principales elementos de la arquitectura CNN+CTC utilizada por EasyOCR y evidenciar el potencial de este tipo de herramientas para apoyar procesos de extracción automática de información documental.

Palabras clave: Aprendizaje de máquina, Aprendizaje Profundo, Redes Neuronales Convolucionales, LSTM, CTC.

Clasificación AMS: 62Jxx, 68T07.

Índice

1. Introducción **3**

2. Objetivos **4**

2.1. Objetivo General 4

2.2. Objetivos Específicos 4

3. Banco Davivienda S.A. ***5***

3.1. Historia Banco Davivienda 5

3.2. Descripción de la pasantía 6

4. Preliminares **7**

4.1. Reconocimiento Óptico de Caracteres (OCR) 7

4.2. Aprendizaje profundo en OCR 8

4.3. EasyOCR 8

4.4. Redes Neuronales Convolucionales CNN y VGGNet 9

4.4.1. Modelado de Secuencias mediante LSTM 11

4.4.2. Decodificación y Alineación mediante CTC 12

4.4.3. Generación de Resultado 13

***5.* Metodología** **14**

5.1. Entorno de ejecución 14

5.2. Conjunto de Datos 15

5.3. Preprocesamiento de imágenes 16

5.4. Extracción de texto mediante EasyOCR 16

5.5. Validación de las extracciones 17

6. Resultados y discusión **18**

6.1. Resultados de extracción por campo 18

6.2. Limitaciones encontradas 19

7. Conclusiones **20**

1. Introducción

Durante mi pasantía en Banco Davivienda participé en actividades relacionadas con la automatización documental, donde herramientas basadas en inteligencia artificial permitían extraer información contenida en documentos digitalizados y transformarla en datos estructurados. Observar cómo estas tecnologías eran capaces de identificar automáticamente texto presente en imágenes y documentos escaneados despertó mi interés por comprender los modelos que hacen posible este proceso.

El Reconocimiento Óptico de Caracteres (OCR) se ha convertido en una herramienta ampliamente utilizada para la extracción automática de información contenida en documentos digitales. Su aplicación permite transformar imágenes o documentos escaneados en texto procesable, reduciendo la necesidad de digitación manual y facilitando el manejo de grandes volúmenes de información.

Los avances recientes en aprendizaje profundo han permitido mejorar significativamente el desempeño de estos sistemas. Entre las arquitecturas más utilizadas se encuentra la combinación de Redes Neuronales Convolucionales (CNN) y Clasificación Temporal Conexionista (CTC), la cual constituye la base de herramientas como EasyOCR. Mientras las CNN se encargan de extraer características visuales de las imágenes, la CTC permite transformar dichas representaciones en secuencias de caracteres sin requerir una alineación previa entre la imagen y el texto.

En este contexto, el presente trabajo surge del interés por comprender las tecnologías observadas durante las actividades de procesamiento documental desarrolladas en la pasantía. Para ello, se analizan los fundamentos de la arquitectura CNN+CTC utilizada por EasyOCR y se implementa esta herramienta para la extracción automática de información en documentos digitalizados, evaluando su precisión mediante métricas aplicadas a un conjunto de documentos previamente seleccionado y preparado para este propósito.

En el desarrollo del presente trabajo se utilizó la herramienta de inteligencia artificial Claude, de Anthropic, como apoyo en la redacción y corrección de estilo del texto [2]. El análisis conceptual, la implementación técnica, la evaluación de los resultados y las conclusiones presentadas corresponden en su totalidad a un desarrollo propio del autor.

2. Objetivos

2.1. Objetivo General

Analizar el funcionamiento de la herramienta EasyOCR en procesos de extracción automática de información documental, a partir de las actividades de procesamiento documental desarrolladas durante la pasantía en Banco Davivienda, evaluando su desempeño sobre documentos digitalizados y estudiando la arquitectura CNN+CTC que sustenta su funcionamiento.

2.2. Objetivos Específicos

- Describir el papel de los sistemas OCR en los procesos de extracción automática de información documental observados durante la pasantía.
- Estudiar los fundamentos de la arquitectura CNN+CTC implementada por la herramienta EasyOCR.
- Implementar EasyOCR sobre un conjunto de documentos digitalizados aplicando técnicas de preprocesamiento de imágenes.
- Evaluar la precisión de las extracciones obtenidas y analizar los factores que afectan el desempeño de la herramienta.

3. Banco Davivienda S.A.

Banco Davivienda es una entidad financiera colombiana que hace parte del Grupo Empresarial Bolívar, fundada en 1972. Cuenta con presencia internacional en Estados Unidos, El Salvador, Costa Rica, Honduras y Panamá, brindando servicios financieros a personas naturales, empresas y sectores rurales. La organización busca generar valor para sus clientes, accionistas, colaboradores y terceros a través de la innovación, la tecnología y la sinergia empresarial.

3.1. Historia Banco Davivienda

Davivienda cuenta con más de 50 años de trayectoria generando valor económico a las personas y a sus empresas en Colombia y Centroamérica. Estos son algunos de los hitos más importantes de este recorrido:

- **1972**: Nace la Corporación Colombiana de Ahorro y Vivienda con el nombre Coldeahorro. La "Casita Roja" como símbolo hasta la fecha.
- **1973**: Coldeahorro cambia de nombre por Corporación Colombiana de Ahorro y Vivienda, Davivienda.
- **1994**: Sale el producto de tarjetas de crédito.
- **1997**: Davivienda se convierte en un Banco comercial con el nombre de Banco Davivienda S.A. y se consolida en el sector financiero en Colombia.
- **2006**: Se fusiona Davivienda y el Banco Superior.
- **2007**: Davivienda adquiere a Granbanco S.A. Bancafé. Lo que permitió aumentar los servicios financieros a empresas y llegar a Panamá y Miami.
- **2010**: Davivienda emite por primera vez acciones preferenciales en la Bolsa de Valores de Colombia. Más de 80.000 personas pusieron Davivienda a su nombre.
- **2011**: Davivienda emite por segunda vez acciones en la Bolsa de Valores de Colombia. Nace DaviPlata, un servicio gratuito para hacer transacciones a través del celular.
- **2012**: Se firma un acuerdo para HSBC Costa Rica, Honduras y El Salvador ampliando su presencia en Centroamérica.
- **2013**: El Banco cierra un acuerdo con Corredores Asociados que fortalece su oferta y portafolio.

- **2016**: Crece su oferta para pymes gracias a la fusión con Leasing Bolívar S.A.
- **2019**: Nace la alianza RappiPay Davivienda potenciando los pagos digitales, bajando los costos de acceso y transaccionales.
- **2024**: Davivienda regresa al mercado de valores con una emisión de acciones ordinarias y para un total de 36 millones de acciones emitidas.

3.2. Descripción de la pasantía

Durante el periodo de prácticas desarrollé diferentes labores en el área de analítica y tecnología del banco. Las actividades abarcaron tres frentes principales, los cuales se describen a continuación.

Migración de servidores locales a Google Cloud Platform

La primera parte de la pasantía estuvo enfocada en apoyar el proceso de migración de servidores locales hacia Google Cloud Platform (GCP). Este trabajo se desarrolló desde tres frentes. El primero fue el desarrollo de códigos, donde se reescribieron consultas SQL siguiendo los nuevos lineamientos de GCP y homologando las diferentes funciones hacia BigQuery (BQ). Este proceso implicó comprender las diferencias de sintaxis y funcionamiento entre el entorno local y el entorno en la nube, así como adaptar lógicas existentes para garantizar la equivalencia en los desarrollos.

El segundo frente fue la certificación de históricos, que consistía en comparar la información almacenada en los servidores locales con la obtenida en BigQuery para validar la correcta migración de los datos. Para ello, se desarrolló una herramienta en Python que automatizaba la consulta de la información en BigQuery y realizaba la comparación contra los datos extraídos del servidor local. Este proceso era semiasistido, ya que la información del servidor local debía obtenerse manualmente mediante consultas dinámicas y cargarse en una hoja de cálculo, a partir de la cual el código ejecutaba las validaciones y generaba los resultados de la comparación. Esta tarea requería interpretar las diferencias encontradas y determinar si correspondían a errores de migración o a diferencias esperadas en la estructura de los datos.

El tercer frente fue la documentación de los desarrollos, donde por cada código se realizaba un registro que incluía las tablas utilizadas, los cruces realizados, una descripción de las variables de la tabla final y un diagrama del modelo lógico. Esta documentación era fundamental para garantizar la trazabilidad de los procesos migrados y facilitar su mantenimiento.

Extracción de información de páginas web

De manera complementaria, se realizaron labores de extracción automática de información de páginas web públicas. La información obtenida era organizada y utilizada posteriormente de acuerdo con las necesidades de los diferentes procesos apoyados durante la práctica.

Procesamiento documental con herramientas de inteligencia artificial

Otra de las actividades desarrolladas durante la pasantía estuvo relacionada con el procesamiento documental mediante herramientas de inteligencia artificial. En este proceso se trabajaba con documentos digitalizados a partir de los cuales se extraía información para reducir los tiempos asociados al procesamiento manual.

Esta actividad permitió observar cómo las herramientas utilizadas eran capaces de extraer información de manera automática y convertirla en datos que posteriormente podían ser utilizados en otros procesos. También permitió ver que factores como la calidad de la imagen o la forma en que se encontraba organizado el documento podían influir en los resultados obtenidos.

Aunque las actividades desarrolladas durante la práctica abarcaron diferentes áreas, el procesamiento documental fue una de las que más despertó mi interés. Observar cómo estas herramientas eran capaces de identificar información contenida en documentos escaneados y transformarla en datos estructurados generó inquietudes sobre los modelos que hacen posible este proceso. A partir de esta experiencia surgió el interés por estudiar los sistemas de reconocimiento óptico de caracteres (OCR) y, en particular, las arquitecturas CNN+CTC utilizadas por herramientas como EasyOCR, tema central del presente trabajo.

4. Preliminares

Las actividades de procesamiento documental desarrolladas durante la pasantía despertaron el interés por comprender los modelos utilizados para la extracción automática de información. Por esta razón, en este capítulo se presentan algunos conceptos necesarios para entender el funcionamiento de los sistemas OCR y de la arquitectura utilizada por EasyOCR.

4.1. Reconocimiento Óptico de Caracteres (OCR)

El Reconocimiento Óptico de Caracteres (OCR) es una tecnología que permite convertir texto contenido en imágenes o documentos escaneados en información digital que puede ser procesada por un computador. Su principal objetivo es automatizar tareas que tradicionalmente requerían la lectura y digitación manual de información.

Los sistemas OCR tienen aplicaciones en diferentes sectores, donde son utilizados para digitalizar documentos, automatizar procesos administrativos y facilitar la consulta de información almacenada en formatos físicos. En el sector financiero, por ejemplo, permiten extraer datos de formularios, documentos de identificación y otros soportes documentales que hacen parte de los procesos operativos de las organizaciones.

Los primeros sistemas OCR se basaban en reglas diseñadas para reconocer patrones específicos y requerían documentos con formatos muy controlados para obtener buenos resultados. Con el avance del aprendizaje profundo, surgieron modelos capaces de aprender automáticamente las características relevantes de los caracteres a partir de grandes conjuntos de datos, mejorando significativamente la precisión y la capacidad de adaptación a diferentes tipos de documentos.

4.2. Aprendizaje profundo en OCR

El aprendizaje profundo ha transformado la forma en que se desarrollan los sistemas modernos de reconocimiento óptico de caracteres. A diferencia de los métodos tradicionales, que dependían de reglas definidas manualmente, estos modelos aprenden directamente a partir de ejemplos, identificando patrones presentes en las imágenes y relacionándolos con los caracteres correspondientes.

En la actualidad, muchas herramientas OCR utilizan arquitecturas compuestas por diferentes tipos de redes neuronales que trabajan de manera conjunta. Estas arquitecturas permiten extraer características visuales de las imágenes, interpretar la relación secuencial entre los caracteres y generar una transcripción final del texto detectado.

Entre las arquitecturas más utilizadas se encuentran aquellas basadas en Redes Neuronales Convolucionales (CNN), redes recurrentes tipo LSTM y mecanismos de decodificación mediante Clasificación Temporal Conexionista (CTC). Esta combinación constituye la base de funcionamiento de herramientas como EasyOCR, utilizada en el desarrollo del presente trabajo.

4.3. EasyOCR

EasyOCR es una librería de Python especializada en reconocimiento óptico de caracteres (OCR), diseñada para detectar y extraer texto a partir de imágenes y documentos digitalizados mediante técnicas de aprendizaje profundo. Su implementación permite automatizar procesos de extracción de información sin necesidad de entrenar modelos desde cero, lo que facilita su integración en aplicaciones de procesamiento documental.

Para el desarrollo de este trabajo se seleccionó EasyOCR debido a que es una herramienta de código abierto, cuenta con modelos preentrenados para múltiples idiomas y puede implementarse fácilmente en Python. Además, su arquitectura incorpora técnicas ampliamente utilizadas en sistemas modernos de reconocimiento de texto, lo que permite estudiar tanto su funcionamiento teórico como su aplicación práctica.

De manera general, EasyOCR combina tres componentes principales. En primer lugar, una red neuronal convolucional (CNN) se encarga de extraer características visuales de la imagen. Posteriormente, redes recurrentes tipo LSTM procesan estas características considerando la naturaleza secuencial del

texto. Finalmente, el algoritmo de Clasificación Temporal Conexionista (CTC) transforma las predicciones generadas por la red en una secuencia de caracteres legible.

Una de las ventajas de EasyOCR es su capacidad para trabajar con múltiples idiomas y distintos estilos tipográficos, manteniendo un desempeño adecuado en una amplia variedad de documentos. Sin embargo, factores como la calidad de la imagen, la presencia de ruido o una orientación incorrecta del documento pueden afectar la precisión de los resultados, por lo que en muchos casos es necesario aplicar técnicas de preprocesamiento antes de la extracción de texto.

4.4. Redes Neuronales Convolucionales CNN y VGGNet

Una CNN es un tipo de arquitectura inspirada en el sistema visual humano, diseñada específicamente para procesar datos con estructura de rejilla, como las imágenes.

Funcionamiento sencillo: En lugar de analizar cada píxel de forma aislada, la CNN utiliza filtros que se deslizan sobre la imagen para detectar patrones [9].

Jerarquía de aprendizaje: Las primeras capas de la red aprenden a reconocer rasgos básicos como bordes y colores; a medida que la red se vuelve más profunda, combina esos rasgos para identificar formas complejas como letras o caracteres específicos.

La primera etapa del reconocimiento de texto en escenas (STR) consiste en la extracción de características, donde una red neuronal convolucional (CNN) mapea la imagen de entrada a una representación abstracta que resalta los atributos relevantes para el reconocimiento de caracteres [12].

La Figura 1 ejemplifica cómo una imagen se representa como matriz binaria antes de ser procesada por la red.

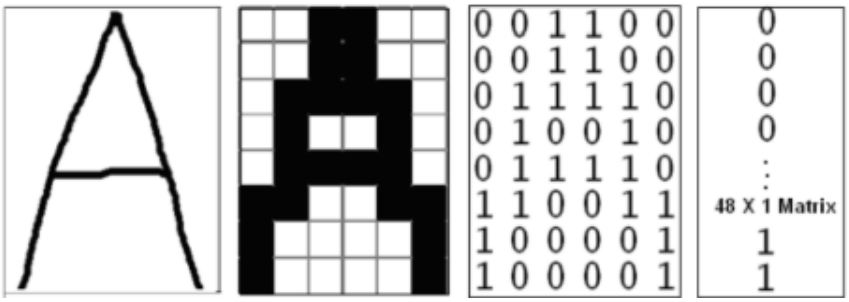


Figura 1: Representación de una imagen en escala de grises y su matriz binaria correspondiente para el carácter “A”, donde cada celda indica la presencia (1) o ausencia (0) de píxel activo.

Nota. Tomada de [4].

Arquitectura VGG: Históricamente, modelos como VGGNet introdujeron el uso de filtros pequeños de 3×3 aplicados de forma modular y profunda. Este diseño demostró que apilar múltiples capas de filtrado fino permite capturar detalles complejos con un menor número de parámetros en comparación con kernels de mayor tamaño [11, 3] . En OCR, estas capas aprenden a detectar desde bordes simples hasta formas abstractas como los arcos y trazos de los caracteres.

La Figura 2 ilustra el flujo de procesamiento de VGGNet, desde la entrada de la imagen hasta la generación de las coordenadas y puntajes de confianza.

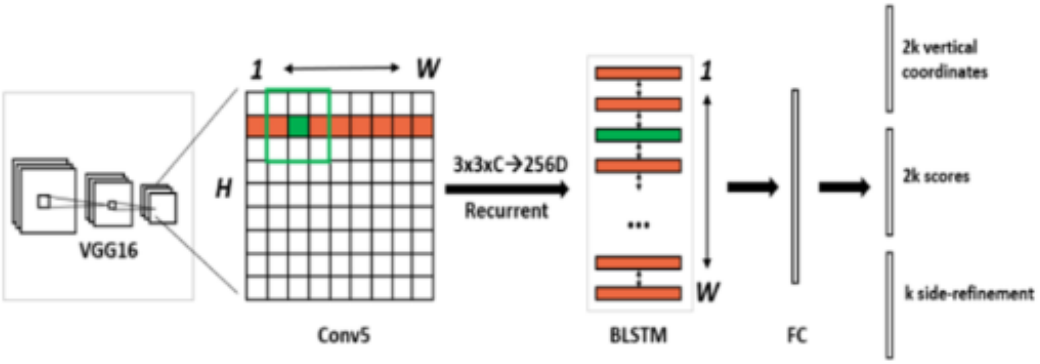


Figura 2: Flujo de procesamiento de la arquitectura VGG16 combinada con una red BLSTM, desde la extracción de características convolucionales hasta la generación de coordenadas verticales, puntajes de confianza y refinamiento lateral.

Nota. Tomado de [14].

4.4.1. Modelado de Secuencias mediante LSTM

Una vez extraídas las características visuales, estas se reformatean como una secuencia de columnas donde cada elemento representa una franja horizontal de la imagen. Sin embargo, el análisis independiente de estas franjas carece de información contextual, problema que se resuelve mediante el modelado de secuencias. La LSTM es un tipo especializado de Red Neuronal Recurrente (RNN) diseñada para trabajar con secuencias de datos donde el orden es fundamental, como el habla o el texto.

- Long Short-Term Memory (LSTM): Las redes recurrentes tradicionales sufren del problema del gradiente desvaneciente, lo que les impide retener información de pasos de tiempo lejanos. La arquitectura LSTM soluciona esto mediante bloques de memoria que contienen una celda autoconectada y tres compuertas multiplicativas: entrada (input), olvido (forget) y salida (output) [7].

  Una vez extraídas las características visuales, se procesan como una secuencia. La LSTM actúa como el motor contextual, analizando franja por franja de la imagen. Gracias a su capacidad para gestionar dependencias a largo plazo, el sistema no reconoce caracteres aislados, sino que entiende la relación espacial y lingüística entre ellos.

- Funcionamiento de las Compuertas: Utiliza tres compuertas inteligentes. Compuerta de Olvido: Decide qué información antigua ya no es útil y debe borrarse. Compuerta de Entrada: Selecciona qué nueva información de la secuencia actual vale la pena guardar. Compuerta de Salida: Determina qué parte de la memoria se usará para realizar la predicción en ese momento. En el OCR, esto permite que el sistema entienda que si ya leyó "H", O y "L", existe una alta probabilidad de que la siguiente letra sea una A.

- Bidireccionalidad (BiLSTM): Para maximizar el contexto, se utilizan capas BiLSTM que procesan la secuencia tanto hacia adelante como hacia atrás, permitiendo que cada predicción se beneficie de la información pasada y futura de la cadena de texto [13] .

La Figura 3 esquematiza este procesamiento, mostrando cómo la secuencia de características atraviesa las capas recurrentes bidireccionales hasta generar las predicciones por cada franja de la imagen.

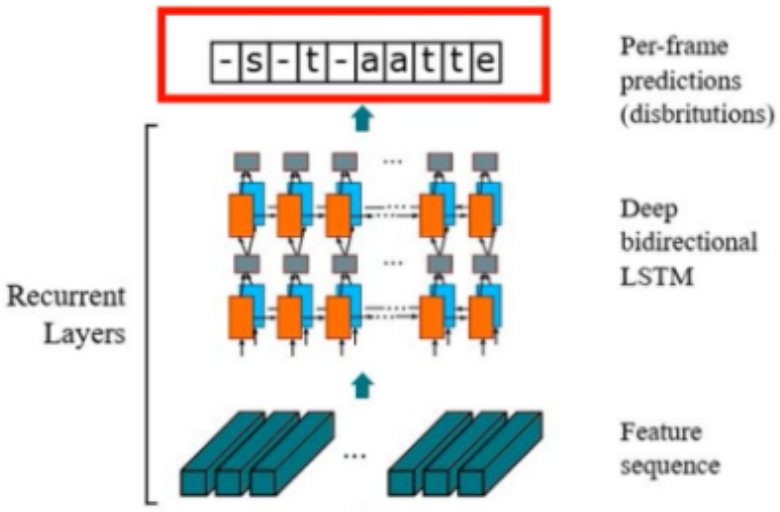


Figura 3: Procesamiento secuencial de las características visuales mediante una red LSTM bidireccional (BiLSTM), donde cada franja de la secuencia de entrada genera una predicción por cuadro.

Nota. Adaptada de [1].

4.4.2. Decodificación y Alineación mediante CTC

El mayor desafío en el reconocimiento de texto es que la alineación entre la imagen (secuencia de entrada) y el texto (secuencia de etiquetas) es desconocida; no sabemos exactamente cuántas franjas de la imagen corresponden a una letra específica.

- Símbolo Blank y Colapso de Secuencias: El método Connectionist Temporal Classification (CTC) resuelve esto interpretando las salidas de la red como una distribución de probabilidad sobre todas las posibles secuencias de etiquetas, incluyendo un símbolo especial llamado "blank"(-) [5].
- Mapeo Muchos-a-Uno: CTC define una función de mapeo B que transforma una ruta de predicciones de la red en la transcripción final eliminando primero las etiquetas repetidas y luego los espacios en blanco. Por ejemplo, una secuencia predicha como HH-EE-LL-LL-OO se colapsa

legalmente en HELLO. Este proceso permite que la red se entrene directamente con pares de imágenes y textos sin necesidad de que los datos estén pre-segmentados manualmente.

- Entrenamiento de extremo a extremo: Esto permite que la red aprenda directamente de imágenes de palabras completas sin necesidad de que un humano le indique dónde empieza y termina cada letra individualmente

La Figura 4 ilustra este proceso de alineación, mostrando las distintas rutas posibles y sus probabilidades asociadas para una secuencia de entrada.

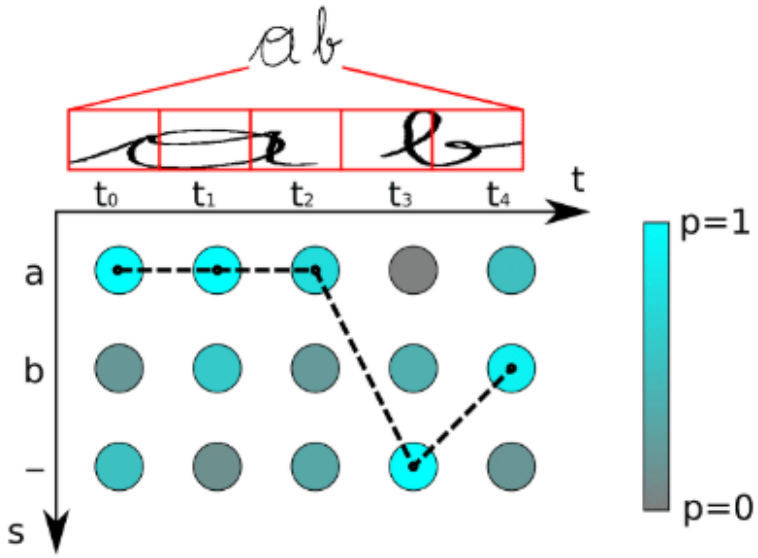


Figura 4: Matriz de probabilidades de salida del algoritmo de Clasificación Temporal Conexionista (CTC), donde cada celda representa la probabilidad asignada a un símbolo en un instante de tiempo determinado, para la secuencia manuscrita “ab”.

Nota. Tomada de [10].

4.4.3. Generación de Resultado

Finalmente, se agrupa la información procesada y entrega una salida estructurada para el usuario que incluye la ubicación exacta (coordenadas del bounding box), el texto final decodificado y un índice de confianza que representa qué tan segura está la red de su predicción.

La Figura 5 presenta un ejemplo de esta salida, donde se observa el texto detectado junto con su respectivo índice de confianza.

```
Text: REPUBLICA DE COLOMBIA, Probability: 0.5870270871025096
Text: IDENTIFICACION PERSONAL, Probability: 0.8279591088904708
Text: CEDULA DE CIUDADANIA, Probability: 0.6727226252358358
Text: NUMERO, Probability: 0.9591163729017688
```

Figura 5: Salida generada por EasyOCR sobre un documento de prueba, mostrando el texto detectado y su respectivo índice de confianza (probability) para cada campo reconocido.

Nota. Elaboración propia.

5. Metodología

La siguiente sección describe la metodología implementada para la aplicación de EasyOCR en la extracción de texto por campo sobre un conjunto de cédulas de ciudadanía digitalizadas mediante escáner.

5.1. Entorno de ejecución

Las pruebas realizadas en este trabajo se ejecutaron utilizando dos configuraciones de hardware CPU (Central Processing Unit) y GPU (Graphics Processing Unit). La CPU es el componente encargado de ejecutar las operaciones generales de un computador y se caracteriza por procesar tareas de forma principalmente secuencial. Por su parte, la GPU fue desarrollada inicialmente para el procesamiento gráfico, pero su arquitectura basada en miles de núcleos de procesamiento permite ejecutar un gran número de operaciones de manera paralela.[8]

Esta capacidad resulta especialmente útil en aplicaciones de aprendizaje profundo, ya que los modelos de redes neuronales requieren una gran cantidad de operaciones matriciales durante la etapa de inferencia. Debido a ello, las GPU suelen ofrecer tiempos de ejecución considerablemente menores que las CPU en tareas relacionadas con visión por computador y reconocimiento óptico de caracteres.[6]

Con el fin de evaluar el impacto del hardware sobre el desempeño del modelo, EasyOCR se ejecutó utilizando ambas configuraciones. Los tiempos obtenidos se presentan en la Tabla 1.

Configuración	Tiempo de ejecución
CPU	20 min
GPU	1 min

Tabla 1: Comparación de los tiempos de ejecución de EasyOCR sobre el conjunto de 32 cédulas de ciudadanía, utilizando configuración de hardware basada en CPU y en GPU, sin aplicar preprocesamiento de imágenes.

Nota. Elaboración propia.

Los resultados muestran una diferencia significativa en los tiempos de procesamiento. Mientras que la ejecución completa del conjunto de documentos utilizando CPU requirió aproximadamente 20 minutos, la misma tarea ejecutada sobre GPU se completó en cerca de 1 minuto. Esto representa una aceleración aproximada de 20 veces respecto a la ejecución sobre CPU.

La reducción observada se atribuye a la capacidad de las GPU para ejecutar múltiples operaciones de manera paralela, característica especialmente útil en modelos basados en redes neuronales profundas. Estos resultados evidencian la importancia del hardware utilizado en aplicaciones de reconocimiento óptico de caracteres, particularmente cuando se trabaja con grandes volúmenes de documentos o cuando se requieren tiempos de respuesta reducidos.

5.2. Conjunto de Datos

Con el fin de evaluar la precisión del modelo, se utilizaron 32 cédulas de ciudadanía colombianas de modelo antiguo, digitalizadas mediante escáner. Los campos sobre los cuales se van a calcular las métricas de extracción son número de cédula, apellidos, nombres, fecha de nacimiento, lugar de nacimiento, estatura, grupo sanguíneo (RH), sexo y fecha y lugar de expedición.

El conjunto de imágenes presenta una variación significativa en la calidad del escaneo. Algunos documentos presentan baja resolución, otros llegaron rotados o con la hoja desplazada durante el escaneo, lo que generó solapamiento entre campos. También se encuentran documentos en buenas condiciones. Esta variedad en la calidad permite evaluar el comportamiento del sistema ante condiciones reales de uso.

La selección de las cédulas se realizó debido a que se contaba con este tipo de documentos para realizar las pruebas. Además, su estructura definida y la cantidad limitada de campos facilitaron la validación de los resultados obtenidos y el cálculo de las métricas utilizadas en la evaluación.

Para cada documento se construyó una tabla con los valores reales de cada campo, utilizada como referencia para el cálculo de las métricas de evaluación y la comparación con los resultados obtenidos por el modelo.

5.3. Preprocesamiento de imágenes

Antes de realizar la extracción de texto, se aplicó una etapa de preprocesamiento con el objetivo de mejorar la calidad de las imágenes de entrada y facilitar el reconocimiento de caracteres. Debido a que algunos documentos presentaban problemas de orientación, se realizó una corrección de rotación cuando fue necesario, garantizando que el texto se encontrara en una posición adecuada para su procesamiento.

Posteriormente, se aplicó un filtro gaussiano utilizando la función `GaussianBlur` de OpenCV con una ventana de tamaño (5×5). Este filtro permite suavizar la imagen y reducir pequeñas variaciones de intensidad asociadas al ruido presente en el proceso de digitalización.

Una vez realizada esta etapa, la imagen preprocesada fue utilizada como entrada para EasyOCR, encargado de detectar las regiones de texto y generar la transcripción correspondiente junto con un nivel de confianza asociado a cada predicción.

5.4. Extracción de texto mediante EasyOCR

Después de aplicar el preprocesamiento descrito en la sección anterior, se realizó la extracción de texto mediante EasyOCR. Como resultado, se obtuvo el texto identificado en cada documento junto con un valor de confianza para cada extracción.

Dado que EasyOCR extrae todo el texto presente en la imagen, fue necesario revisar los resultados obtenidos para identificar la información correspondiente a los campos evaluados. Esta revisión se realizó de forma manual, seleccionando únicamente los resultados asociados a dichos campos y descartando la información que no era necesaria para el cálculo de las métricas.

La Figura 6 muestra un ejemplo de los resultados obtenidos por EasyOCR sobre uno de los documentos evaluados, incluyendo los distintos campos detectados y su nivel de confianza asociado.

```
Text: NÚMERO, Probability: 0.9989755879324125
Text: 1.010.XXX.XXX, Probability: 0.9893924964133678
Text: CUERVO RODRIGUEZ, Probability: 0.9213849129515671
Text: APELLIDOS, Probability: 0.9988498558945661
Text: MANUELA, Probability: 0.996914226127606
Text: NOMBRES, Probability: 0.7778512189930988
Text: 0, Probability: 0.5677965403670129
Text: Quofuo, Probability: 0.2699236413862778
Text: IFIRMA, Probability: 0.786953889327222
Text: FECHA DE NACIMIENTO, Probability: 0.9971597224319699
Text: 20-OCT-2003, Probability: 0.9965766771228374
Text: BOGOTA D.C, Probability: 0.9258068364588395
Text: (CUNDINAMARCA), Probability: 0.8401219739250592
Text: LUGAR DE NACIMIENTO, Probability: 0.8800268745152027
Text: 1.54, Probability: 0.7424960444198102
Text: 0-, Probability: 0.9942842886855187
Text: F, Probability: 0.9971200727022733
Text: ESTATURA, Probability: 0.9951394905660906
Text: G.S. RH, Probability: 0.8406448027725288
Text: SEXO, Probability: 0.9991546273231506
Text: 28-XXX-20XX BOGOTA D.C., Probability: 0.9539934325122983
Text: FECHA, Probability: 0.9999379445872724
Text: Y LUGAR DE EXPEDICIÓN, Probability: 0.9421244051187491
Text: INDICE DERECHO, Probability: 0.9871307213815127
Text: REGISTRADOA NACIONAL, Probability: 0.5024426234058187
```

Figura 6: Resultado de la extracción de EasyOCR sobre una cédula de ciudadanía del conjunto de datos evaluado, mostrando los distintos campos detectados (número de identificación, apellidos, nombres, fecha y lugar de nacimiento, estatura, grupo sanguíneo, sexo, y fecha y lugar de expedición) junto con su índice de confianza asociado.

Nota. Elaboración propia.

5.5. Validación de las extracciones

Con el fin de evaluar el desempeño del modelo, se compararon los textos extraídos por EasyOCR con los valores de referencia definidos para cada documento. La evaluación se realizó de manera independiente para cada uno de los campos analizados: número de cédula, apellidos, nombres, fecha de nacimiento, lugar de nacimiento, estatura, grupo sanguíneo (RH), sexo y fecha y lugar de expedición.

Para cada campo se contabilizó el número de extracciones correctas y se calculó el porcentaje de aciertos respecto al total de documentos evaluados. Se consideró una extracción correcta cuando el valor obtenido por el modelo coincidía con el valor registrado en la tabla de referencia correspondiente.

La precisión por campo se calculó mediante la siguiente expresión:

$$\text{precisión campo} = \frac{\text{Campos correctos}}{\text{Total de documentos}} \times 100 \tag{1}$$

donde las extracciones correctas corresponden al número de documentos en los que el campo fue reconocido adecuadamente y el total de documentos corresponde al tamaño completo del conjunto de datos evaluado.

Esta metodología permitió analizar el comportamiento del modelo para cada tipo de información presente en la cédula y determinar cuáles campos presentan mayor o menor dificultad de reconocimiento.

6. Resultados y discusión

6.1. Resultados de extracción por campo

La Tabla 2 presenta los porcentajes de precisión obtenidos para cada uno de los campos evaluados bajo las configuraciones CPU sin filtro y GPU con filtro gaussiano.

Campo	**GPU + Filtro (%)**	**CPU sin filtro (%)**
Número de cédula	93.75	78.125
Apellidos	90.625	75.000
Nombres	75.000	62.500
Lugar de nacimiento 1	96.875	90.625
Lugar de nacimiento 2	87.500	84.375
Estatura	84.375	81.250
RH	96.875	84.375
Sexo	87.500	84.375
Fecha y lugar expedición	96.875	75.000
Métrica general	**90.3125**	**77.8125**

Tabla 2: Precisión de extracción por campo obtenida con EasyOCR sobre el conjunto de 32 cédulas de ciudadanía, comparando la configuración GPU con preprocesamiento (corrección de rotación y filtro gaussiano) frente a la configuración CPU sin preprocesamiento.

Nota. Elaboración propia.

Al comparar ambas configuraciones, se observa una mejora en todos los campos analizados. La precisión general pasó de 77.81% a 90.31%, lo que indica que la configuración con GPU y preprocesamiento permitió obtener resultados más precisos en la extracción de la información contenida en las cédulas.

Los mayores incrementos se presentaron en los campos número de cédula, apellidos y fecha y lugar de expedición. Estos campos contienen una cantidad considerable de caracteres y suelen verse más afectados por problemas de calidad en la imagen, por lo que el preprocesamiento aplicado parece haber contribuido a mejorar su reconocimiento.

Por otra parte, campos como estatura, sexo y lugar de nacimiento presentaron diferencias menores entre ambas configuraciones. Esto puede deberse a que son campos cortos o que, en la mayoría de los documentos, se encontraban en condiciones que facilitaron su reconocimiento desde la primera prueba.

También es importante resaltar que algunos campos alcanzaron precisiones superiores al 95%, como el grupo sanguíneo (RH), el lugar de nacimiento y la fecha y lugar de expedición. En contraste, el campo nombres obtuvo el porcentaje más bajo de reconocimiento, incluso en la mejor configuración evaluada. Una posible explicación es la mayor variabilidad presente en este campo, tanto en longitud como en calidad de impresión y escaneo.

En general, los resultados muestran que la combinación de preprocesamiento y ejecución sobre GPU permitió mejorar el desempeño de EasyOCR en la extracción de información de documentos de identidad, alcanzando niveles de precisión cercanos al 90% para el conjunto de datos analizado.

6.2. Limitaciones encontradas

Durante el desarrollo de las pruebas se identificaron diferentes factores que afectaron el desempeño del modelo. La principal limitación estuvo relacionada con la calidad de los documentos digitalizados. Algunas imágenes presentaban baja resolución, rotación o desplazamientos durante el proceso de escaneo, dificultando la correcta identificación de ciertos caracteres.

También se observó que los errores de reconocimiento no afectaban de la misma manera a todos los campos. Los campos con mayor cantidad de texto, como nombres o fecha y lugar de expedición, tendían a presentar más errores cuando la calidad de la imagen era deficiente. Por el contrario, campos más cortos, como sexo o grupo sanguíneo, generalmente fueron reconocidos con mayor facilidad.

Otra limitación corresponde al tamaño del conjunto de datos utilizado. Aunque las 32 cédulas permitieron realizar una evaluación inicial del desempeño de EasyOCR, un conjunto de documentos más amplio podría proporcionar una caracterización más completa del comportamiento del modelo frente a diferentes condiciones de digitalización.

7. Conclusiones

Las actividades de procesamiento documental desarrolladas durante la pasantía permitieron conocer de cerca el uso de herramientas para la extracción automática de información contenida en documentos digitalizados. A partir de esta experiencia surgió el interés por estudiar los sistemas de reconocimiento óptico de caracteres (OCR) y, en particular, la arquitectura utilizada por herramientas como EasyOCR.

El estudio realizado permitió comprender los principales componentes de la arquitectura CNN+CTC empleada por EasyOCR. Se revisó el papel de las redes neuronales convolucionales en la extracción de características de la imagen, el uso de redes LSTM para el modelado de secuencias y la función de la Clasificación Temporal Conexionista (CTC) en la generación de la transcripción final del texto.

La implementación de EasyOCR sobre el conjunto de documentos seleccionado permitió evaluar su desempeño en la extracción automática de información contenida en documentos de identidad. Los resultados obtenidos muestran que la herramienta alcanzó una precisión general del 90.31% utilizando la configuración con GPU y preprocesamiento de imágenes, obteniendo los mejores resultados entre las configuraciones evaluadas.

También se observó que el preprocesamiento influye de manera importante en la calidad de las extracciones. La corrección de orientación y la aplicación del filtro gaussiano contribuyeron a mejorar el reconocimiento de varios de los campos analizados, especialmente en documentos que presentaban problemas asociados al proceso de digitalización.

Finalmente, se identificó que la calidad de las imágenes continúa siendo uno de los factores que más afecta el desempeño de los sistemas OCR. Problemas como baja resolución, rotación o desplazamientos durante el escaneo pueden generar errores en la extracción de información, particularmente en campos que contienen una mayor cantidad de texto. A pesar de estas limitaciones, los resultados obtenidos muestran el potencial de este tipo de herramientas para apoyar procesos de extracción automática de información documental similares a los observados durante la pasantía.

Referencias

[1] Mahya Ameryan and Lambert Schomaker. A limited-size ensemble of homogeneous cnn/lstms for high-performance word classification. *Neural Computing and Applications*, 33:8615–8634, 2021.

[2] Anthropic. Claude. `https://claude.ai`, 2026. Modelo de lenguaje de gran escala.

[3] Xiang Bai Baoguang Shi and Cong Yao. An end-to-end trainable neural network for image-based sequence recogni-tion and its application to scene text recognition. IEEE.

[4] Amit Choudhary, Rahul Rishi, Vijaypal Singh Dhaka, and Savita Ahlawat. Influence of introducing an additional hidden layer on the character recognition capability of a bp neural network having one hidden layer. *International Journal of Engineering and Technology*, 2(1):24–28, 2010.

[5] Alex Graves, Santiago Fernández, Faustino Gomez, and Jürgen Schmidhuber. Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural 'networks. volume 2006, pages 369–376, 01 2006.

[6] Dipesh Gyawali. Comparative analysis of cpu and gpu profiling for deep learning models. *Research Report*.

[7] J. Hochreiter, S. Schmidhuber. Long short-term memory. 1997.

[8] Adrian Ilie. Optical character recognition on graphics hardware. Technical report, Department of Computer Science, University of North Carolina at Chapel Hill.

[9] Keiron O'Shea and Ryan Nash. An introduction to convolutional neural networks. *CoRR*, abs/1511.08458, 2015.

[10] Basem Rizk. Evaluation of state of art open-source asr engines with local inferencing, August 2019.

[11] A. Simonyan, K. Zisserman. Very deep convolutional networks for large-scale image recognition. 2015.

[12] Peijie Li Liangju He Xianye Liang, Bocheng Zhuo. Cnn based texture synthesize with semantic segment.

[13] Peijie Li Liangju He Xianye Liang, Bocheng Zhuo. Bidirectional recurrent neural networks. 1997.

[14] Wenxian Zeng, Qinglin Meng, and Shuqing Zhang. Natural scene chinese character text detection method based on improved ctpn. In *IOP Conference Series: Journal of Physics: Conference Series*, volume 1314, page 012200. IOP Publishing, 2019.