

Clasificación de Objetos de Moda usando Redes Neuronales Convolucionales

1st Santiago Andrés Rocha, 2nd Sebastián Rojas Bueno

Escuela Colombiana de Ingeniería Julio Garavito, Bogotá, Colombia

santiago.rocha-c@mail.escuelaing.edu.co, sebastian.rojas-bu@mail.escuelaing.edu.co

Resumen—La clasificación de objetos de moda, como bolsos o tenis, por medio de la tecnología, es importante porque puede automatizar el proceso de etiquetado y organización de grandes cantidades de productos, lo que a su vez puede mejorar la eficiencia de la cadena de suministro y el comercio electrónico. El uso de redes neuronales convolucionales (CNN's) en particular es importante porque permiten una alta precisión en la clasificación de objetos en imágenes, lo que es especialmente útil en la industria de la moda donde hay una gran cantidad de variaciones de colores, texturas y patrones en cada prenda. Además, la tecnología de clasificación de objetos de moda también puede ser utilizada para mejorar la experiencia del consumidor, ya que permite la recomendación de productos personalizados en línea, basados en las preferencias del usuario y su historial de compras. Este proyecto consistió en la aplicación de redes neuronales convolucionales para la clasificación de objetos de moda en imágenes. Se utilizó un conjunto de datos que contenía imágenes de prendas de vestir y se entrenó varios modelos de CNN's para clasificar cada imagen en una de las 10 categorías de ropa, para posteriormente comparar y validar los resultados y seleccionar el mejor de los modelos.

Con métricas de Exactitud de entre el **90 %** y el **95 %**, y límite de tiempo de máximo 2 horas para la ejecución del entrenamiento, validación y pruebas por modelo, usando el optimizador 'Adam', entropía cruzada categórica como Pérdida y entre 5 y 6 épocas, se obtuvo en el mejor de los resultados una Exactitud del **92 %** con un tiempo de ejecución de **30 minutos**. Cabe aclarar que más adelante en este documento ahondaremos más en los detalles de cada modelo y explicaremos a profundidad las características de las redes.

Index Terms—Problema de clasificación, redes neuronales convolucionales, objetos de moda, entrenamiento, validación, pruebas.

I. INTRODUCCIÓN

La clasificación manual de objetos de moda, como bolsos o tenis, sigue siendo importante en la industria de la moda, ya que permite una mayor precisión en la identificación y organización de productos, especialmente cuando se trata de prendas de alta gama y ediciones limitadas. Además, la clasificación manual permite a los expertos de la industria identificar y analizar las tendencias de la moda y del mercado, lo que a su vez puede influir en el diseño y la producción futura de productos de moda.

La clasificación de imágenes con la ayuda de redes neuronales convolucionales (CNN) es una tarea común en la investigación en aprendizaje profundo. Para este fin, se pueden plantear diferentes modelos de redes neuronales que aprovechen la afinidad de las CNN para el reconocimiento de imágenes.

En general, estos modelos siguen una estructura básica, que consiste en una capa de entrada que recibe la imagen, seguida

de múltiples capas de convolución que extraen características de la imagen y reducen su tamaño, y finalmente una o varias capas totalmente conectadas que realizan la clasificación.

En este proyecto planteamos el uso de redes neuronales convolucionales para el reconocimiento de imágenes y su posterior clasificación en 1 de los 10 posibles tipos de objetos de moda que podemos encontrar en el dataset 'Convolutional neural Network dataset (FashionItems)', que van desde zapatos, tenis y sandalias hasta camisas, camisetas y chaquetas, por medio de 3 modelos de CNN.

II. ESTADO DEL ARTE

II-A. Marco Teórico

La clasificación manual de objetos de moda, como bolsos o tenis, sigue siendo importante en la industria de la moda, ya que permite una mayor precisión en la identificación y organización de productos, especialmente cuando se trata de prendas de alta gama y ediciones limitadas. Además, la clasificación manual permite a los expertos de la industria identificar y analizar las tendencias de la moda y del mercado, lo que a su vez puede influir en el diseño y la producción futura de productos de moda. [1]

Entre las estrategias que se utilizan para la clasificación manual de objetos de moda se encuentran la observación visual, la identificación de patrones y características específicas de cada producto, la comparación con productos similares (características de bajo nivel) y la forma de usarse, las épocas de mayor / menor uso, junto con demás características que dejan de ser tan evidentes (Características de alto nivel), lo que se denomina en problemas de aprendizaje profundo como 'variabilidad intra-clase'.

- Características de bajo nivel: Refiriéndose a características evidentes y fácilmente mensurables, tales como la forma, el color, la textura de los objetos, así como el estilo, el diseño y los materiales. [2]
- Características de alto nivel: Refiriéndose a características ya no tan evidentes tales como la forma de usarse el objeto, las temporadas y climas en los que se usan, los estados de ánimo, así como el significado, el contexto social y el contexto religioso.

Ahora, hablemos de las redes neuronales y el uso de las redes neuronales convolucionales. Las redes neuronales son un tipo de modelo de aprendizaje automático inspirado en el funcionamiento del cerebro humano. Están compuestas por múltiples nodos (neuronas) interconectados que procesan

y transmiten información a través de capas, con el fin de realizar tareas de clasificación, predicción o reconocimiento de patrones en los datos de entrada. [3]

Por otro lado, las redes neuronales convolucionales (CNN, por sus siglas en inglés) son un tipo de red neuronal particularmente útil en procesamiento de imágenes. Se utilizan para tareas como la clasificación, segmentación y detección de objetos en imágenes.

Las CNN tienen la capacidad de identificar características relevantes en una imagen, como bordes, formas, texturas y patrones. Esto se logra mediante el uso de capas de convolución, que aplican filtros para extraer características de las imágenes, seguido de capas de pooling, que reducen la dimensionalidad de la salida de las capas de convolución. [4]

Las entradas de una red convolucional se representan como matrices tridimensionales (ancho x alto x canales). El ancho y la altura se refieren a las dimensiones espaciales de la imagen, mientras que los canales se refieren a la cantidad de colores o características que se están procesando. Por ejemplo, una imagen en escala de grises tendría un solo canal, mientras que una imagen en color tendría tres canales (rojo, verde y azul). [5]

La topología de las entradas de una red convolucional se relaciona con la topología de las imágenes de la siguiente manera: las capas iniciales de la red convolucional suelen tener filtros de tamaño pequeño que se aplican a regiones pequeñas de la imagen. Estos filtros se mueven por toda la imagen, extrayendo características locales como bordes y texturas. A medida que se avanza en la red convolucional, las capas tienen filtros de tamaño más grande que cubren regiones más grandes de la imagen. Estos filtros extraen características más complejas, como formas y objetos. [5]

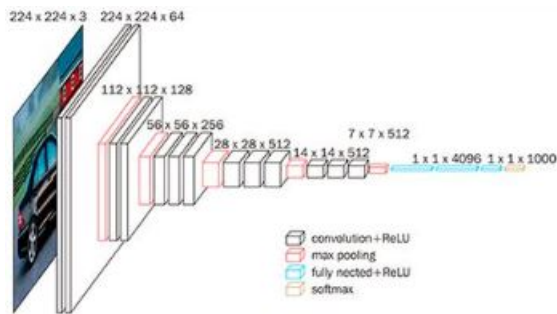


Figura 1. Ejemplo de una Red Neuronal Convolucional [9]

Más adelante, necesitaremos del uso del optimizador Adam presente en Keras, framework especializado para el desarrollo de redes neuronales, por lo que describiremos a continuación el funcionamiento del optimizador. Adam utiliza la estimación de momentos de primer y segundo orden de los gradientes de los pesos para adaptar la tasa de aprendizaje de cada peso de forma individual, lo que ayuda a ajustar la tasa de aprendizaje para cada peso en función de su contribución a la red. Esto permite una convergencia más rápida y una mayor estabilidad en comparación con otros optimizadores. [6]

Por último, en cuanto a los Datasets relacionados con objetos de moda, o fashion items en inglés, son conjuntos de datos

que contienen imágenes de ropa en diferentes categorías, como camisetas, zapatos, bolsos, etc. Estos datasets son similares al conjunto de datos MNIST (Modified National Institute of Standards and Technology), que se utiliza comúnmente para el reconocimiento de dígitos escritos a mano. Sin embargo, mientras que MNIST solo contiene imágenes de dígitos escritos a mano en escala de grises, el dataset de fashion items tiene imágenes en color y en grises de diferentes tipos de ropa. [7]

Aunque los datasets de fashion items y MNIST tienen diferentes características, ambos son útiles para la evaluación de modelos de aprendizaje automático. Ambos tienen un conjunto de entrenamiento y un conjunto de pruebas etiquetados, lo que permite a los modelos ser entrenados y evaluados en los mismos datos. [8]

II-B. Trabajo Relacionado

Para los datasets relacionados / similares a MNIST, existen enormes cantidades de trabajos relacionados, entre implementaciones de sistemas de clasificación simples hasta implementación de redes neuronales profundas. A continuación, hablaremos sobre 3 trabajos destacados que por sus resultados o por sus formas de ser implementados, nos llamaron la atención desde un principio, claro está, no sin antes hablar del dataset seleccionado:

1. Dataset 'Convolutional neural Network dataset(FashionItems)' [10]: El dataset se compone de dos archivos, fashion-mnist_test.csv y fashion-mnist_train.csv, conteniendo 60000 y 10000 ejemplares de objetos / artículos de moda respectivamente. La primera de sus columnas especifica el tipo de objeto al que corresponden (i.e. zapatos, sandalias, etc), y el resto de las 784 columnas corresponden a cada píxel de la imagen que representa al objeto, siendo 'imágenes' de 28x28 píxeles. (Ver figuras 2 y 3). Los 10 tipos de objetos posibles, junto con la etiqueta que los representa, son los siguientes:

- a) Camisas (0)
- b) Pantalones (1)
- c) Buzos (2)
- d) Vestidos (3)
- e) Chaquetas (4)
- f) Sandalias (5)
- g) Camisetas (6)
- h) Tenis (7)
- i) Bolsos (8)
- j) Zapatos (9)

2. Inception-v3 [11]: En un estudio realizado por el equipo de investigación de Google en 2016, se presentó la red neuronal que fue entrenada con una base de datos de más de un millón de imágenes de ropa. La red neuronal logró una precisión del **90,8 %** en la clasificación de las prendas en una colección de prueba, superando a los métodos de clasificación manual tradicionales. Presentamos su arquitectura en la figura 4.
3. Otro ejemplo es el trabajo realizado por investigadores de la Universidad de Pekín en 2018 [12], donde presentaron

una red neuronal basada en una arquitectura de red ResNet”. La red fue entrenada con más de 120,000 imágenes de ropa y logró una precisión del **92,5 %** en la clasificación de la ropa en una colección de prueba. Además, los investigadores utilizaron un enfoque de transferencia de aprendizaje para mejorar la capacidad de la red para clasificar imágenes de ropa con fondos complejos. Presentamos su arquitectura en la figura 5.

4. 'Sample CNN code to get you started' [13], que es utilizado como base y comienzo de nuestro proyecto, fue realizado por Sujay Kapadnis [14] donde de forma detallada proporciona un modelo de CNN funcional para la clasificación de los objetos del mismo dataset que estamos trabajando. Sujay proporciona una manera de visualización y transformación de los datos del dataset para que puedan ser usados a modo de entrada de la CNN, para posteriormente construirla y especificar hiper parámetros, capas, épocas, etc. resultando en un modelo para la clasificación de imágenes de 5 épocas que alcanza una pérdida por entropía cruzada categórica de **0.1299** en entrenamiento y con una métrica de Exactitud alcanza un **95.3213 %** en entrenamiento junto con un **91.867 %** en validación. En la figura 6 podemos evidenciar que época tras época, la pérdida disminuye significativamente alcanzando finalmente un valor bajo, indicando que la diferencia entre los resultados obtenido vs los esperados en el conjunto de entrenamiento es muy baja, resultando favorable la implementación del modelo.

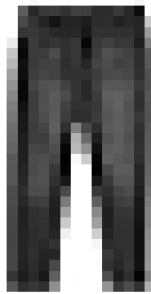


Figura 2. Unos pantalones, ejemplo del dataset [10]

III. METODOLOGÍA

La metodología para el desarrollo de la solución se planteó de la siguiente forma:

1. Seleccionar el dataset a trabajar.
2. Pensar en el tipo de problema.
3. Analizar el tipo de dato a utilizar como fuente.
4. Considerar una solución con redes neuronales.
5. Considerar qué tipo de red neuronal convendría más.
6. Buscar trabajo relacionado para ayudarnos a comenzar.
7. Empezar a desarrollar la solución.

A continuación, describiremos nuestro desarrollo para cada una de estas etapas en nuestra metodología:

1. Seleccionar el dataset a trabajar:
En esta etapa, se eligió el conjunto de datos que se utilizaría para entrenar el modelo de clasificación de



Figura 3. 100/60000 ejemplos del dataset [10]

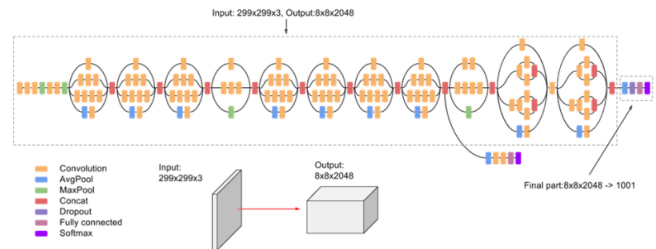
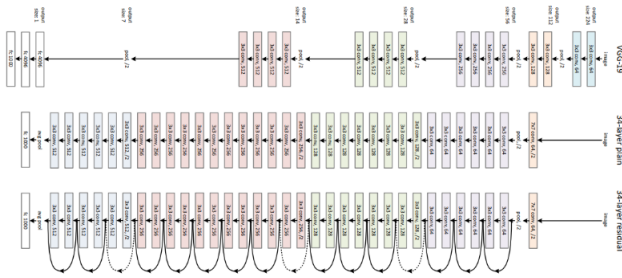


Figura 4. Arquitectura de Inception-v3 [11]

objetos de moda. Se seleccionó un conjunto de datos lo suficientemente grande y diverso para permitir que el modelo aprendiera a reconocer diferentes tipos de objetos de moda.

2. Pensar en el tipo de problema:
Se definió el problema de clasificación de objetos de moda, es decir, el objetivo del proyecto era desarrollar un modelo de clasificación que pudiera identificar diferentes tipos de objetos de moda en imágenes.
3. Analizar el tipo de dato a utilizar como fuente:
Se analizó el tipo de datos que se utilizarían como fuente para el modelo de clasificación. En este caso, se utilizaron imágenes de objetos de moda.
4. Considerar una solución con redes neuronales:
Se decidió utilizar una solución basada en redes neuronales para abordar el problema de clasificación de objetos de moda. Esto se debió a que las redes neuronales han demostrado ser eficaces en la clasificación de imágenes.
5. Considerar qué tipo de red neuronal convendría más:
Se analizaron diferentes tipos de redes neuronales y se decidió utilizar una red neuronal convolucional (CNN). Las CNN son especialmente adecuadas para la clasificación de imágenes debido a su capacidad para detectar patrones en los datos de entrada.
6. Buscar trabajo relacionado para ayudarnos a comenzar:
Se investigaron trabajos relacionados para obtener ideas



- Capa de Convolución: 1 kernel de 3x3, Padding: Same, Función de Activación: Relu
- Pooling: 2x2, Strides: 2
- Capa de Convolución: 1 kernel de 3x3, Padding: Same, Función de Activación: Relu
- Capa de Convolución: 1 kernel de 3x3, Padding: Same, Función de Activación: Relu
- Pooling: 2x2, Strides: 2
- Capa de Convolución: 1 kernel de 3x3, Padding: Same, Función de Activación: Relu
- Pooling: 2x2, Strides: 2
- Capa Densa: 10 neuronas - Función de Activación: Sigmoid
- Función de activación Capas de Salida: Sigmoide
- Épocas: 5
- Costo: Entropía cruzada Categórica
- Optimizador: Adam
- Métricas: Exactitud 90 %-95 %
- Métricas: Tiempo de Ejecución: 2 horas

Segundo modelo:

- Entradas: 28x28x1 (Imágenes de 28x28 con un solo canal)
- Capa de Convolución: 1 kernel de 3x3, Padding: Same, Función de Activación: Relu
- Capa de Convolución: 1 kernel de 3x3, Padding: Same, Función de Activación: Relu
- Pooling: 2x2, Strides: 2
- Capa de Convolución: 1 kernel de 3x3, Padding: Same, Función de Activación: Relu
- Capa de Convolución: 1 kernel de 3x3, Padding: Same, Función de Activación: Relu
- Pooling: 2x2, Strides: 2
- Capa de Convolución: 1 kernel de 3x3, Padding: Same, Función de Activación: Relu
- Pooling: 2x2, Strides: 2
- Capa Densa: 10 neuronas - Función de Activación: Sigmoid
- **Función de activación Capas de Salida: Softmax**
- Épocas: 5
- Costo: Entropía cruzada Categórica
- Optimizador: Adam
- Métricas: Exactitud 90 %-95 %
- Métricas: Tiempo de Ejecución: 2 horas

Tercer modelo:

- Entradas: 28x28x1 (Imágenes de 28x28 con un solo canal)
- Capa de Convolución: 1 kernel de 3x3, Padding: Same, Función de Activación: Relu
- Capa de Convolución: 1 kernel de 3x3, Padding: Same, Función de Activación: Relu
- Pooling: 2x2, Strides: 2
- Capa de Convolución: 1 kernel de 3x3, Padding: Same, Función de Activación: Relu
- Capa de Convolución: 1 kernel de 3x3, Padding: Same, Función de Activación: Relu
- Pooling: 2x2, Strides: 2
- Capa de Convolución: 1 kernel de 3x3, Padding: Same, Función de Activación: Relu
- Pooling: 2x2, Strides: 2
- Capa Densa: 10 neuronas - Función de Activación: Sigmoid
- **Función de activación Capas de Salida: Softmax**
- Épocas: 5
- Costo: Entropía cruzada Categórica

Función de Activación: Relu

- Pooling: 2x2, Strides: 2
- Capa Densa: 10 neuronas - Función de Activación: Sigmoid
- **Función de activación Capas de Salida: Softmax**
- **Épocas: 7**
- Costo: Entropía cruzada Categórica
- Optimizador: Adam
- Métricas: Exactitud 90 %-95 %
- Métricas: Tiempo de Ejecución: 2 horas

Cuadro II
DIFERENCIAS ENTRE LOS MODELOS

Modelos	F. Act. Densa	Épocas
Modelo 1	Sigmoid	5
Modelo 2	Softmax	5
Modelo 3	Softmax	7

El objetivo de estos cambios fue verificar dos aspectos importantes: Primero, queríamos revisar por qué quien proporcionó el código base no utilizó una función de activación Softmax en la capa densa, sabiendo que el problema de clasificación 'excluyente' y cada objeto solo puede llegar a ser clasificado como 1 tipo. Segundo, queríamos visualizar el porqué se decidió utilizar solo 5 épocas para el modelo, aumentándolas a 7 para ver cómo se comportan los resultados.

- Entrenamiento de los modelos: Para entrenar los modelos, fueron usados los datos destinados a Entrenamiento, y con las variaciones previamente descritas, se obtuvo distintos resultados a tener en cuenta.
- Evaluación del modelo: De acuerdo con los siguientes parámetros propuestos:
 - Función de pérdida/costo: Entropía Cruzada Categórica
 - Métrica: Precisión entre el **90 %** y el **95 %** .
 - Métrica: Tiempo de ejecución no mayor a **2 horas** .

Modelo Seleccionado:

Segundo modelo:

- Entradas: 28x28x1 (Imágenes de 28x28 con un solo canal)
- Capa de Convolución: 1 kernel de 3x3, Padding: Same, Función de Activación: Relu
- Capa de Convolución: 1 kernel de 3x3, Padding: Same, Función de Activación: Relu
- Pooling: 2x2, Strides: 2
- Capa de Convolución: 1 kernel de 3x3, Padding: Same, Función de Activación: Relu
- Capa de Convolución: 1 kernel de 3x3, Padding: Same, Función de Activación: Relu
- Pooling: 2x2, Strides: 2
- Capa de Convolución: 1 kernel de 3x3, Padding: Same, Función de Activación: Relu
- Pooling: 2x2, Strides: 2
- Capa Densa: 10 neuronas - Función de Activación: Sigmoid
- **Función de activación Capas de Salida: Softmax**
- Épocas: 5
- Costo: Entropía cruzada Categórica

- Optimizador: Adam
- Métricas: Exactitud 90 %-95 %
- Métricas: Tiempo de Ejecución: 2 horas

Con los siguientes resultados:

Modelo 2, Entrenamiento:

- Resultados en las pérdidas: Figura 8
- Resultados de Exactitud: **95.492 %**

Modelo 2, Pruebas:

- Resultados de Exactitud: **91.6667 %**

Modelo 2, Tiempo:

- Aproximadamente, **30 minutos**

Más adelante enseñaremos todos los resultados obtenidos, y podremos justificar finalmente la selección de este modelo, pero adelantamos que cumple con las métricas establecidas, teniendo mejores resultados que los demás, y tiene comportamientos adecuados a lo que se esperaría de una red neuronal convolucional para un problema de clasificación.

5. Pruebas y análisis de resultados: Enseñaremos a continuación los resultados que obtuvimos, teniendo como prioridad enseñar la validez de la construcción y su eficacia para la entrada de los resultados.

V. RESULTADOS

Los resultados obtenidos fueron los siguientes: Se evaluaron los modelos obteniendo como resultados en cuanto a Exactitud y Pérdida:

- Modelo 1, Entrenamiento:
 - Resultados en las pérdidas: Figura 7
 - Resultados de Exactitud: **95.1212 %**
- Modelo 1, Pruebas:
 - Resultados de Exactitud: **91.1477 %**
- Modelo 1, Tiempo:
 - Aproximadamente, **35 minutos**
- Modelo 2, Entrenamiento:
 - Resultados en las pérdidas: Figura 8
 - Resultados de Exactitud: **95.492 %**
- Modelo 2, Pruebas:
 - Resultados de Exactitud: **91.6667 %**
- Modelo 2, Tiempo:
 - Aproximadamente, **30 minutos**
- Modelo 3, Entrenamiento:
 - Resultados en las pérdidas: Figura 9
 - Resultados de Exactitud: **96.6082 %**
- Modelo 3, Pruebas:
 - Resultados de Exactitud: **91.0587 %**
- Modelo 3, Tiempo:
 - Aproximadamente, **57 minutos**

Nota: Para los gráficos enseñados, se tiene como valores en el eje X las épocas de ejecución de los modelos, mientras en el eje Y tenemos los valores de pérdida, para cada uno de los modelos.

Cuadro III
COMPARACIÓN RESULTADOS

Modelos	Train Loss	Train Prec	Test Prec	T. de Ejecución
Modelo 1	0.1374	95.1212 %	91.1477 %	35 min
Modelo 2	0.2943	95.492 %	91.6667 %	30 min
Modelo 3	0.0929	96.6082 %	91.0587 %	57 min

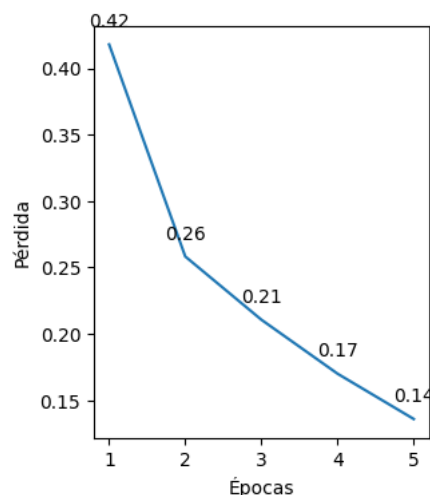


Figura 7. Pérdida en entrenamiento del modelo de Sujay [13]

Teniendo en cuenta que los resultados arrojados por la red no son enteros, decidimos aproximar los resultados con 4 cifras decimales, pues consideramos que son suficientes para visualizar las diferencias entre las evaluaciones.

Comparamos en la figura 10 los resultados obtenidos en el entrenamiento de los modelos, así como en la figura 11 distintos valores de Exactitud obtenidos en entrenamiento, y en la figura 12 distintos valores de Exactitud obtenidos en pruebas. En el caso de la comparación de los entrenamientos, podemos ver que se comportan de forma similar tanto el modelo 2 y 3 en el que se usó la función de activación Softmax para la capa densa en lugar del Sigmoid del modelo original. En el caso de la comparación de las pruebas, vemos como en efecto la Exactitud del modelo 3 es mejor que la de los demás modelos, sin embargo, el modelo que posee mejores resultados para Pruebas, es el Modelo 2. De paso, evidenciamos que tener más épocas para el modelo resulta inoficioso.

En síntesis, el Modelo 2 posee buenos resultados en cuanto a pérdidas, tiene el menor tiempo de ejecución y tiene el mejor porcentaje de Exactitud de entre todos los modelos.

De esta forma, se construyó y evaluó varios modelos de redes neuronales convolucionales aplicados a problemas de clasificación, que logran cumplir con las métricas establecidas y arrojan resultados favorables para el desarrollo de sistemas inteligentes, y resultados favorables para el posible uso de herramientas de ayuda para profesionales en la clasificación de objetos de moda.

VI. CONCLUSIONES

Específicamente, la red neuronal convolucional definitiva posee la siguiente configuración:

Segundo modelo:

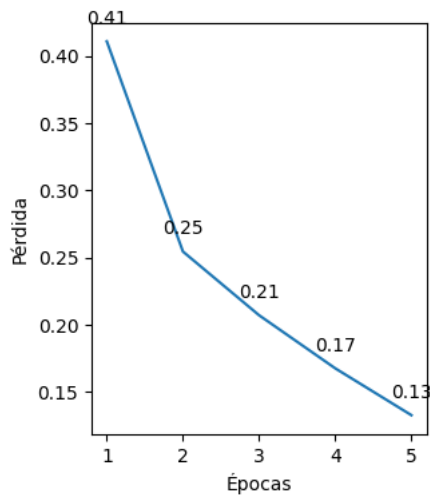


Figura 8. Pérdida en entrenamiento del modelo 2

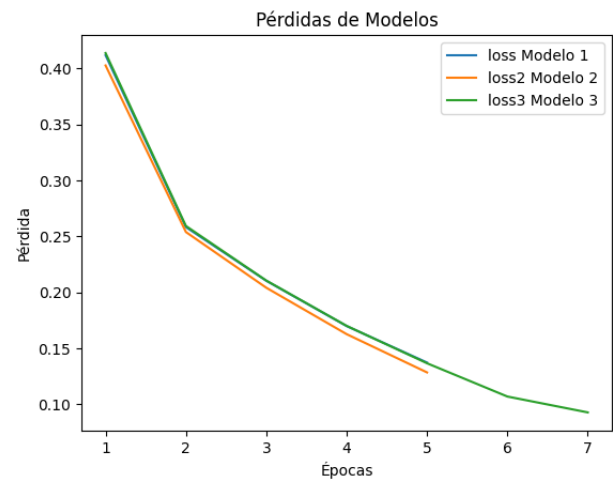


Figura 10. Pérdida en entrenamiento de los 3 modelos

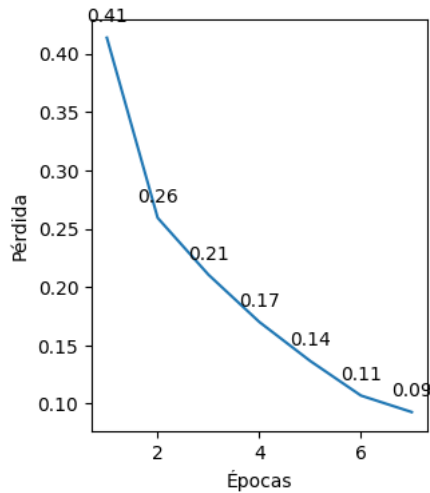


Figura 9. Pérdida en entrenamiento del modelo 3

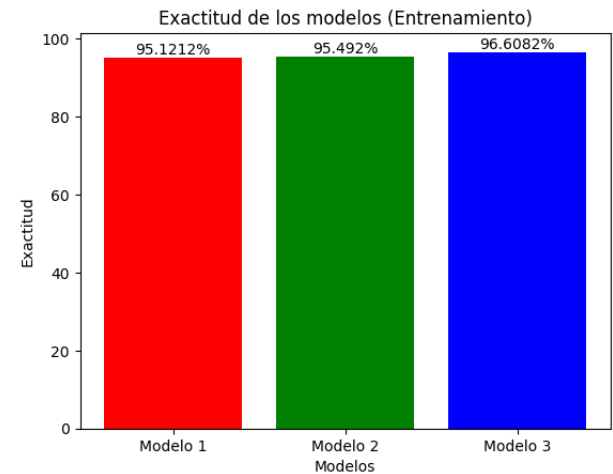


Figura 11. Exactitud en entrenamiento de los 3 modelos

- Entradas: 28x28x1 (Imágenes de 28x28 con un solo canal)
- Capa de Convolución: 1 kernel de 3x3, Padding: Same, Función de Activación: Relu
- Capa de Convolución: 1 kernel de 3x3, Padding: Same, Función de Activación: Relu
- Pooling: 2x2, Strides: 2
- Capa de Convolución: 1 kernel de 3x3, Padding: Same, Función de Activación: Relu
- Capa de Convolución: 1 kernel de 3x3, Padding: Same, Función de Activación: Relu
- Pooling: 2x2, Strides: 2
- Capa Densa: 10 neuronas - Función de Activación: Sigmoid
- Función de activación Capas de Salida: Softmax
- Épocas: 5
- Costo: Entropía cruzada Categórica

- Optimizador: Adam
- Métricas: Exactitud 90 %-95 %
- Métricas: Tiempo de Ejecución: 2 horas

Con los siguientes resultados:

Modelo 2, Entrenamiento:

- Resultados en las pérdidas: Figura 8
- Resultados de Exactitud: **95.492 %**

Modelo 2, Pruebas:

- Resultados de Exactitud: **91.6667 %**

Modelo 2, Tiempo:

- Aproximadamente, **30 minutos**

1. Las redes neuronales convolucionales son una herramienta muy útil para la clasificación de objetos de moda: Las redes neuronales convolucionales (CNN) son una arquitectura de red neuronal que se utiliza comúnmente para tareas de visión por computadora, como la clasificación de imágenes. En este proyecto, se demostró que las CNN son una herramienta muy útil para la clasificación de objetos de moda, lo que indica que esta tecnología puede ser aplicada para la automatización de procesos de clasificación de objetos de moda en la industria.

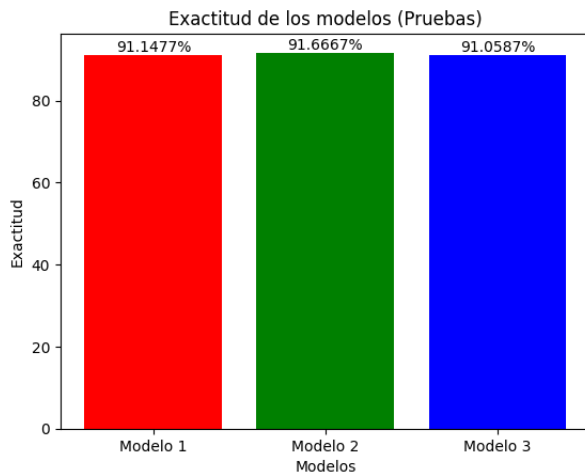


Figura 12. Exactitud en pruebas de los 3 modelos

- La exploración de diferentes modelos, funciones de activación de las capas densas y épocas es fundamental para obtener un modelo óptimo: Durante el proyecto se evaluaron diferentes modelos, variando la cantidad y tipos de capas, funciones de activación y épocas de entrenamiento. Esta exploración permitió encontrar el modelo óptimo que logró la mejor métrica de precisión. Es importante destacar que esta exploración requiere de un enfoque sistemático, ya que implica el entrenamiento y la evaluación de múltiples modelos.
- Se logró cumplir con las métricas de exactitud entre el 90% y el 95%, lo que indica que el modelo es muy preciso para la clasificación de objetos de moda: Las métricas de precisión obtenidas durante el proyecto indican que el modelo de CNN desarrollado es muy preciso para la clasificación de objetos de moda. Al lograr métricas de precisión tan altas, se puede considerar que el modelo es efectivo para su aplicación en la industria de la moda.
- El tiempo de ejecución menor a 2 horas es una gran ventaja, ya que permite obtener los resultados de manera rápida y eficiente: El tiempo de ejecución de un modelo de CNN es un factor crítico para su aplicación en la industria. En este proyecto, se logró desarrollar un modelo que cumple con altas métricas de precisión y, al mismo tiempo, su tiempo de ejecución es menor a 2 horas. Esto es una gran ventaja ya que permite obtener los resultados de manera rápida y eficiente, lo que puede ser muy útil en aplicaciones de la industria de la moda en tiempo real.
- El modelo mejorado que se obtuvo puede ser utilizado en la industria de la moda para mejorar la eficiencia en la clasificación de objetos, lo que puede llevar a una mayor satisfacción del cliente y una mejora en la rentabilidad: El modelo de CNN desarrollado en este proyecto tiene un alto potencial para su aplicación en la industria de la moda. Al utilizar este modelo para la clasificación de objetos de moda, se puede mejorar la eficiencia en los procesos de clasificación, lo que puede llevar a una mayor satisfacción del cliente y una mejora en la rentabilidad de

las empresas de la industria de la moda.

VII. TRABAJO FUTURO

Como trabajo futuro, nos gustaría abordar los siguientes tópicos que consideramos altamente relevantes:

- Experimentar con la regularización de los datos: Una técnica de regularización consiste en agregar penalizaciones a la función de costo de la red neuronal para evitar el sobreajuste. En futuras investigaciones, se pueden experimentar diferentes técnicas de regularización, como la regularización L1 o L2, la eliminación de neuronas aleatorias o la eliminación de capas, para evaluar si estas técnicas mejoran el rendimiento del modelo.
- Evaluar modelos más complejos ampliando el límite de 2 horas siempre y cuando los resultados obtenidos lo ameriten: En la investigación actual, se limitó el tiempo de ejecución del modelo a 2 horas. En futuras investigaciones, se puede extender este límite para evaluar modelos más complejos, como redes neuronales recurrentes o modelos de atención, siempre y cuando los resultados obtenidos lo ameriten.
- Expandir el alcance del modelo permitiendo mejorar la calidad de las imágenes de entrada (más de 28x28 píxeles): En la investigación actual, las imágenes de entrada del modelo tenían una resolución de 28x28 píxeles. En futuras investigaciones, se pueden utilizar imágenes de mayor resolución, lo que permitiría una mejor identificación de los objetos de moda.
- Incluir más canales permitiendo imágenes a color y no solo en blanco y negro: En la investigación actual, se utilizaron imágenes en blanco y negro para la clasificación de objetos de moda. En futuras investigaciones, se pueden utilizar imágenes a color, lo que permitiría una mejor identificación de los objetos de moda. Para ello, se pueden utilizar redes neuronales convolucionales tridimensionales (3D-CNN), que permiten procesar imágenes en color o en múltiples canales.

REFERENCIAS

- (N.d.). Ulpgec.Es. Retrieved April 25, 2023, from Desarrollo de una red neuronal profunda para la clasificación y caracterización de prendas de vestir.
- Belmartino, A., Liseras, N., & Berges, M. (2016). ¿Qué atributos busca el consumidor en prendas de vestir? Un análisis aplicado al mercado de la indumentaria en Argentina. What does the consumer look for? An analysis of clothing market in Argentina. *Tec empresarial*, 10(1), 7. <https://doi.org/10.18845/te.v10i1.2530>
- (N.d.-b). Amazon.com. Retrieved April 25, 2023, from <https://aws.amazon.com/es/what-is/neural-network/>
- ¿Qué son las Redes Neuronales Convolucionales? (2020, November 11). KeepCoding Bootcamps. <https://keepcoding.io/blog/redes-neuronales-convolucionales/>
- Mishra, M. (2020, August 26). Convolutional neural networks, explained. *Towards Data Science*. <https://towardsdatascience.com/convolutional-neural-networks-explained-9cc5188c4939>
- (N.d.-c). Machinelearningmastery.com. Retrieved April 25, 2023, from <https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/>
- Olabiya, K. (2022, December 8). Looking for fashion datasets for your data science projects? DATA4FASHION. <https://medium.com/data4fashion/need-some-datasets-for-your-data-science-projects-1f7b0d28e9df>

- [8] Data.World. (n.d.). Data.World. Retrieved April 25, 2023, from <https://data.world/datasets/fashion>
- [9] Torralba, P. P. (2022). Qué son las Redes Neuronales Convolucionales. Thinking for Innovation. <https://www.iebschool.com/blog/redes-neuronales-convolucionales-big-data/>
- [10] Kapadnis, S. (2023). Convolutional neural Network dataset(FashionItems) [Data set].
- [11] Guía avanzada de Inception v3. (n.d.). Google Cloud. Retrieved April 25, 2023, from <https://cloud.google.com/tpu/docs/inception-v3-advanced?hl=es-419>
- [12] Residual networks (ResNet) - deep learning. (2020, June 3). GeeksforGeeks. <https://www.geeksforgeeks.org/residual-networks-resnet-deep-learning/>
- [13] Sample CNN code to get you started. (2023, January 14). Kaggle.com; Kaggle. <https://www.kaggle.com/code/sujaykapadnis/sample-cnn-code-to-get-you-started>
- [14] Kapadnis, S. (n.d.). Sujay Kapadnis -. Medium. Retrieved April 25, 2023, from <https://medium.com/@Sujaykapadnis>