ISSN 1870-4069

Sistema para clasificación de texturas en imágenes mediante aprendizaje profundo y características wavelet

Juan Manuel Fortuna-Cervantes[1], Marco Tulio Ramírez-Torres[2], Marcela Mejía-Carlos[1], José Salomé Murguia-Ibarra[1], Juan Martínez-Carranza[3]

[1] Universidad Autónoma de San Luis Potosí, Facultad de Ciencias-IICO, México

[2] Universidad Autónoma de San Luis Potosí, Coordinación Académica Región Altiplano Oeste, México

[3] Instituto Nacional de Astrofísica Óptica y Electrónica, Computer Science Department, México

a158852@alumnos.uaslp.mx, {tulio.torres, marcela.mejia, ondeleto}@uaslp.mx, carranza@inaoep.mx

Resumen. La caracterización de texturas en imágenes digitales, se ha convertido en una herramienta de análisis en el área de visión computacional. La textura dentro de la percepción visual es una propiedad física muy importante, dado que, brinda información sobre la composición estructural de las superficies y de los objetos en la imagen. En este trabajo se realiza un clasificador para las bases de datos: KTH-TIPS-2B (KT2B), Describable Textures Dataset (DTD) y Flickr Material Database (FMD). Y se estudia la adaptabilidad del aprendizaje profundo con la transformada wavelet, en particular una aproximación a la arquitectura Wavelet CNN [6]. Además, se utiliza una metodología empírica y experimental en el desarrollo e implementación de la red neuronal convolucional (CNN) y el análisis wavelet, ambas como métodos de extracción de características. La combinación de estos métodos permite lograr un rendimiento de clasificación aceptable. En la base de datos KT2B se logra un 96 %, en la base de datos DTD un 34 % y por último en FMD se obtiene un 30 % de exactitud. Las gráficas de aprendizaje reflejan que los tres conjuntos de datos muestran una generalización de aprendizaje en las primeras épocas de entrenamiento. Para pequeños conjuntos de imágenes con texturas se recomienda la fusión del aprendizaje profundo con el análisis wavelet. Debido a la limitación de aprendizaje sobre información espectral que se pierde en la CNNs convencionales. Además, es información útil que permite mejorar el rendimiento de clasificación. Los resultados muestran que es posible integrar esta metodología en el desarrollo tecnológico de aplicaciones, como tareas de clasificación o restauración de imágenes y detección de objetos.

Palabras clave: Aprendizaje profundo, clasificación de texturas, métodos de extracción de características, redes neuronales convolucionales, transformada wavelet.

System for Classification of Textures in Images Using Deep Learning and Wavelet Characteristics

Abstract. The characterization of textures in images has become an analysis tool in the area of computer vision. Texture in visual perception is a fundamental physical property since it provides information about the structural composition of surfaces and objects in the image. The aim is to develop a classifier for the databases: KTH-TIPS-2B (KT2B), Describable Textures Dataset (DTD), and Flickr Material Database (FMD). Furthermore, the adaptivity of deep learning with wavelet transform is studied using an approach of Wavelet CNN [6]. An empirical and experimental methodology is used to develop and implement convolutional neural network (CNN) and wavelet analysis, both as feature extraction methods. The combination of both methods allows achieving acceptable classification performance. In the KT2B database with 96%, in the DTD database with 34%, and finally in FMD with 30% accuracy. The learning graphs reflect that all three datasets show a generalization of learning in the early training epochs. For small sets of images with textures, the fusion of deep learning with wavelet analysis is recommended due to the limitation of learning about spectral information lost in conventional CNNs, helpful information that allows for improved classification performance. The results show that it will be possible to integrate this methodology in the technological development of applications, such as image classification or restoration tasks and object detection.

Keywords: Convolutional neural network, deep learning, feature extraction methods, wavelet transform, texture classification.

1. Introducción

La percepción visual es una capacidad humana que permite reconocer la textura de objetos a simple vista, donde interviene el sistema óptico y el sistema nervioso. Los cuales son capaces de captar la información visual, procesarla y obtener un significado, para poder interpretar y comprender de que esta compuesto el objeto. Por otro lado, el análisis de textura dentro del aprendizaje máquina juega un papel importante en tareas de clasificación, detección y localización de objetos.

Este tipo de análisis tiene algunas áreas de aplicación, como el diagnóstico médico asistido por computadora, reconocimiento de frutas utilizando inteligencia artificial, localización y detección en la navegación aérea con drones, por mencionar algunas. En el área de procesamiento de imágenes, se puede definir la textura a partir de los píxeles vecinos y de la distribución de la intensidad sobre la imagen [16].

Además, existen algunos métodos de clasificación para el análisis de la textura como estadísticos, geométricos, de modelo y espectrales. Por otro lado, los métodos espectrales describen la textura en el dominio de la frecuencia. Se basan en la descomposición de una señal en términos de funciones base y utilizan los coeficientes de expansión como elementos del vector de características.

Fig. 1. Proceso del aprendizaje profundo.

Este trabajo se centra en la clasificación de la textura en imágenes, particularmente en conocer la información estructural de las superficies y de los objetos que se encuentran en el plano de imagen. La base del sistema de clasificación es una aproximación a la arquitectura Wavelet CNN, la cual fue propuesta en [6] para clasificación de texturas y tareas sobre etiquetado múltiple en relación con el contenido de la imagen.

La implementación de este sistema se desarrolla con la fusión de dos enfoques; utilizando el dominio espacial, específicamente las redes neuronales convolucionales (CNN por sus siglas en ingles) y el dominio espectral, la transformada wavelet de Haar [9, 3, 11]. Internamente este sistema se divide en dos etapas: la primera corresponde a la extracción de características y la segunda a la etapa de clasificación.

Con respecto a la fase de extracción de características, el tensor creado tiene un conjunto de parámetros numéricos que describen el contenido de la imagen, como el color, la textura o la forma del objeto. Por lo tanto, la etapa de extracción de características es importante para el éxito general de cualquier sistema de clasificación y reconocimiento en imágenes. Las principales contribuciones del trabajo se resumen a continuación:

– Se propone una metodología para mejorar y evaluar el modelo de aprendizaje. Además, se valida con nuevo conjunto de imágenes.
– Se diseña e implementa un sistema de clasificación de texturas en imágenes para evaluar la adaptabilidad del aprendizaje profundo con el análisis wavelet.

En particular, se demuestra que la combinación de ambos métodos de extracción de características (CNN y la transformada wavelet de Haar), alcanzan precisiones competitivas a lo reportado en la literatura, con un número significativamente menor de parámetros entrenables que al utilizar solo un método. Como resultado, el modelo es más fácil de entrenar, generaliza su aprendizaje a la combinación de información, y tiene un menor costo computacional.

El resto del documento está organizado de la siguiente manera en la Sección 2 se muestran los trabajos relacionados, la Sección 3 introduce la metodología para abordar el problema de clasificación de la textura. Posteriormente en la sección 4 se muestran los resultados con los tres conjuntos de datos, que se han usado para probar nuestro enfoque y la parte experimental. Finalmente, en la Sección 5 se presentan las conclusiones.

Fig. 2. Imágenes de ejemplo del conjunto de datos KTH-TIPS-2B.

2. Estado del arte

En el diseño de un sistema de recuperación para imágenes, es fundamental hacer un análisis de las características sobre el contenido del plano de imagen. Para esto en [16], los autores mencionan que la calidad del sistema depende, en primer lugar, de los vectores de características utilizados, además, presentan un estudio de las técnicas más comunes de extracción y representación de las características, donde brindan una clasificación en función de las mismas características, como por color, textura o forma.

En particular, al pensar en la búsqueda por textura, se tienen diferentes métodos de extracción, que en opinión del autor se clasifican en estadísticos, geométricos, de modelo y espectrales. Por otra parte, se menciona que el uso de un enfoque espectral combinado con otros métodos de extracción, mejora los resultados en la solución de problemas sobre clasificación y reconocimiento de texturas.

En cuanto al aprendizaje profundo en la ultima década se ha posicionado como una nueva solución en áreas de la robótica [12], visión computacional [1] y el lenguaje natural [15]. En particular, las redes neuronales convolucionales son una categoría del aprendizaje profundo, ya que se adaptan al análisis de objetos mediante el aprendizaje y la extracción de características complejas.

Por otro lado, aunque la CNN es un extractor universal, en la práctica, no esta claro si la CNN puede aprender a realizar análisis espectrales. Para tener este enfoque dentro de la CNN, en [2] los autores proponen una arquitectura llamada Textura CNN. Su idea se centra en que la información extraída por las capas convolucionales es de menor importancia en el análisis de la textura.

En consecuencia, utilizan una métrica estadística de energía en la etapa de extracción de características. Esta información se concatena con la etapa de clasificación, la capa totalmente conectada. En concreto, la arquitectura muestra una mejora en el rendimiento y una reducción en el costo computacional.

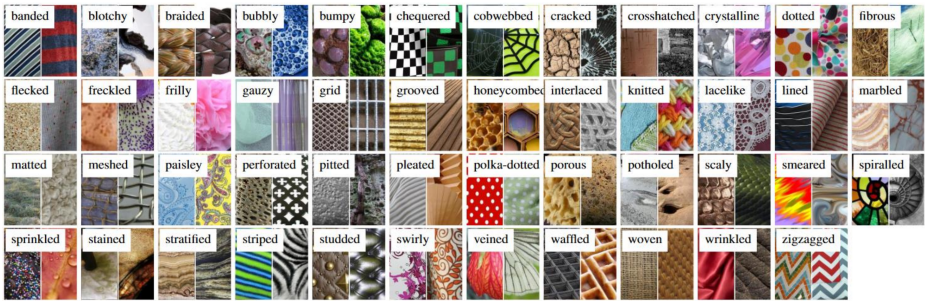


Fig. 3. Imágenes de ejemplo del conjunto de datos DTD.

Fig. 4. Imágenes de ejemplo del conjunto de datos FMD.

Los enfoques espaciales y espectrales son dos de los principales métodos para las tareas de procesamiento de imágenes. Dentro de esta linea de investigación, en [6] los autores proponen una novedosa arquitectura CNN, llamada Wavelet CNN, que combina un análisis multiresolución y el aprendizaje de la CNN en un modelo. Basándose en esta idea, complementan las partes que faltan con el análisis multiresolución mediante la transformada wavelet de Haar en su representación bidimensional, y la integran como componentes adicionales en toda la arquitectura.

La arquitectura Wavelet CNN permite utilizar información espectral que se pierde en su mayoría en las CNN convencionales, pero que es información útil en la mayoría de las tareas de procesamiento de imágenes. El rendimiento alcanzado en la clasificación de texturas y etiquetado de imágenes, muestran una mayor exactitud de clasificación que al utilizar AlexNet, donde unicamente utilizan las CNN convencionales. Así como, reducir el numero de parámetros a entrenar.

3. Materiales y métodos

El aprendizaje profundo es un subcampo del aprendizaje automático, una nueva manera de aprender características a partir de los datos, como texto, audio e imágenes [8]. El término profundo en esta área no hace referencia a ningún tipo de arquitectura; más bien, representa la idea de capas sucesivas de representaciones en diferentes niveles.

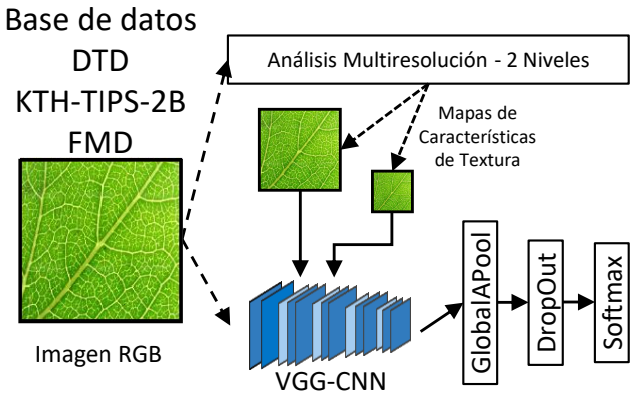
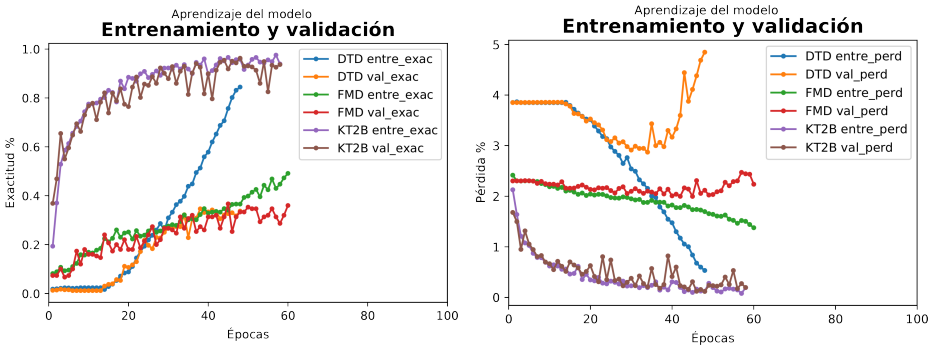


Fig. 5. Arquitectura para el sistema de clasificación de imágenes con textura.



(a) Exactitud del modelo.

(b) Pérdida del modelo.

Fig. 6. Evaluación de las métricas de exactitud y pérdida para los conjuntos de entrenamiento y validación.

Estas nuevas representaciones son cada vez más significativas. Por otro lado, dentro del aprendizaje profundo están la redes neuronales convolucionales o CNN, que son un tipo especializado de red neuronal para el procesamiento de datos. El nombre red neuronal convolucional indica que la red emplea una operación matemática llamada convolución.

La convolución es un tipo especializado de operación lineal, que es utilizada en lugar de la multiplicación general de matrices dentro de la red. La metodología propuesta para mejorar el rendimiento sobre el modelo de aprendizaje, se resume en tres etapas, las cuales se muestran en la Fig. 1. Cada etapa se describe a continuación.

Para la primer etapa, se necesita seleccionar los datos con los cuales vamos a generalizar el conocimiento del modelo, en nuestro caso se utilizan imágenes de tres conjuntos de datos que existen en la literatura, KTH-TIPS-2B, DTD y FMD. Los cuales contienen imágenes con diferentes texturas y materiales en condiciones naturales. Estos conjuntos de datos serán descritos a continuación.

Tabla 1. Capacidad de aprendizaje en las tres etapas de evaluación.

	Entrenamiento	Validación	Prueba
Exactitud DTD	0.3974	0.3221	0.3451
Exactitud FMD	0.3414	0.3667	0.3000
Exactitud KT2B	0.9580	0.9629	0.9678

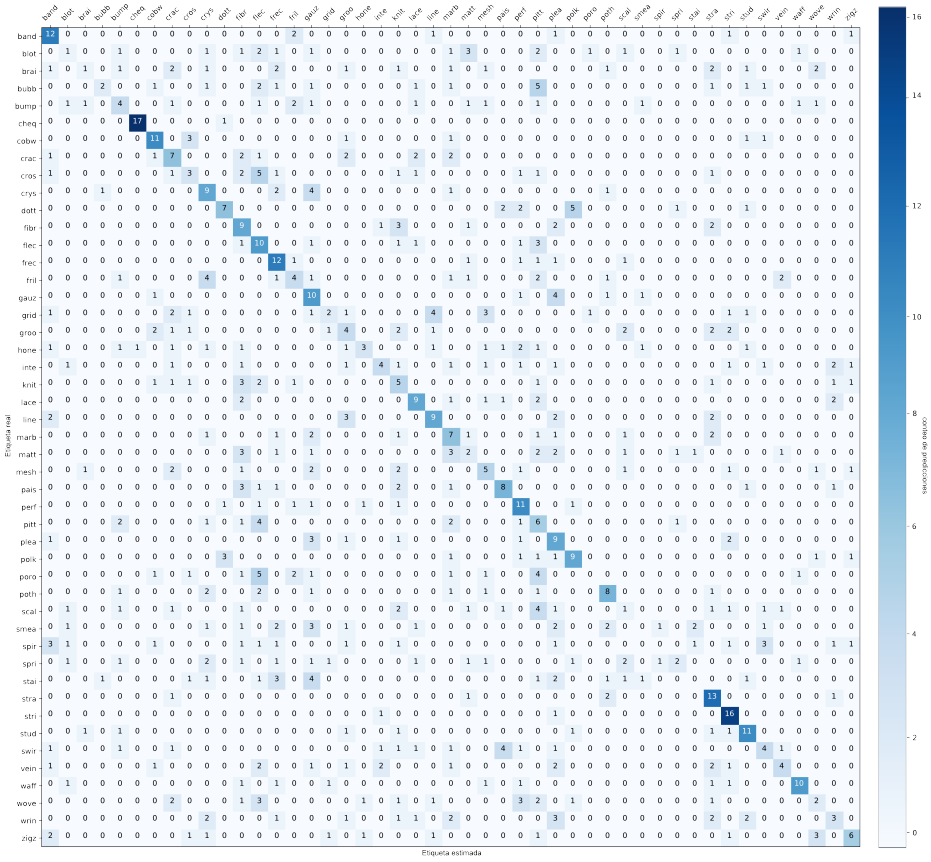


Fig. 7. Matriz de confusión para el conjunto de datos DTD.

3.1. Conjunto de datos

El primer conjunto de datos seleccionado es KTH-TIPS-2B (KT2B) el cual contiene 432 imágenes clasificadas en 11 clases. Cada clase consta de cuatro muestras y cada muestra tiene 108 imágenes [7]. La Fig. 2 ilustra las 11 clases con 4 muestras aleatorias. El segundo conjunto de datos es Describable Textures Dataset (DTD) el cual contiene 47 clases de 120 imágenes en la naturaleza, esto significa que las imágenes fueron adquiridas en condiciones no controladas [5]. Este conjunto de datos incluye 10 divisiones disponibles con 40 imágenes de entrenamiento, 40 imágenes de validación, y 40 imágenes de prueba para cada clase.

Tabla 2. Resultados de clasificación y comparación con otras arquitecturas del estado del arte, en términos de exactitud (%).

	AlexNet	T-CNN	Wavelet CNN	Propuesta
DTD	22.7	27.8	35.6	34.5
KT2B	48.3	49.6	63.7	96.7
FMD	–	–	–	30.0

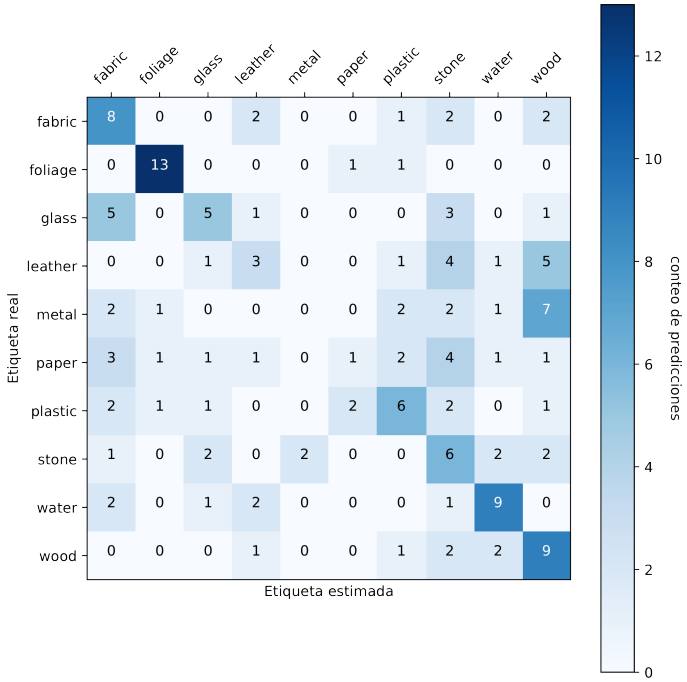


Fig. 8. Matriz de confusión para el conjunto de datos FMD.

En la Fig. 3 se muestra algunas imágenes de este conjunto. Por último, el tercer conjunto de datos seleccionado es Flickr Material Database (FMD). El cual esta construido con una gama de materiales comunes (por ejemplo vidrio, plástico, etc.). Cada imagen de esta base de datos (100 imágenes por categoría, 10 categorías) está seleccionada manualmente de Flickr.com (bajo licencia Creative Commons) para garantizar una variedad de condiciones de iluminación, composiciones, colores, textura y subtipos de materiales [13]. Se muestran algunas imágenes de esta base de datos en la Fig. 4.

3.2. Preparación de datos

Dado que inicialmente las imágenes dentro de los conjuntos tienen diferente tamaño. En la etapa de preprocesamiento, las imágenes son redimensionadas a un tamaño de 300×300 para los tres conjuntos, KT2B, FMD y DTD. También, al utilizar una arquitectura CNN se debe procesar las entradas.

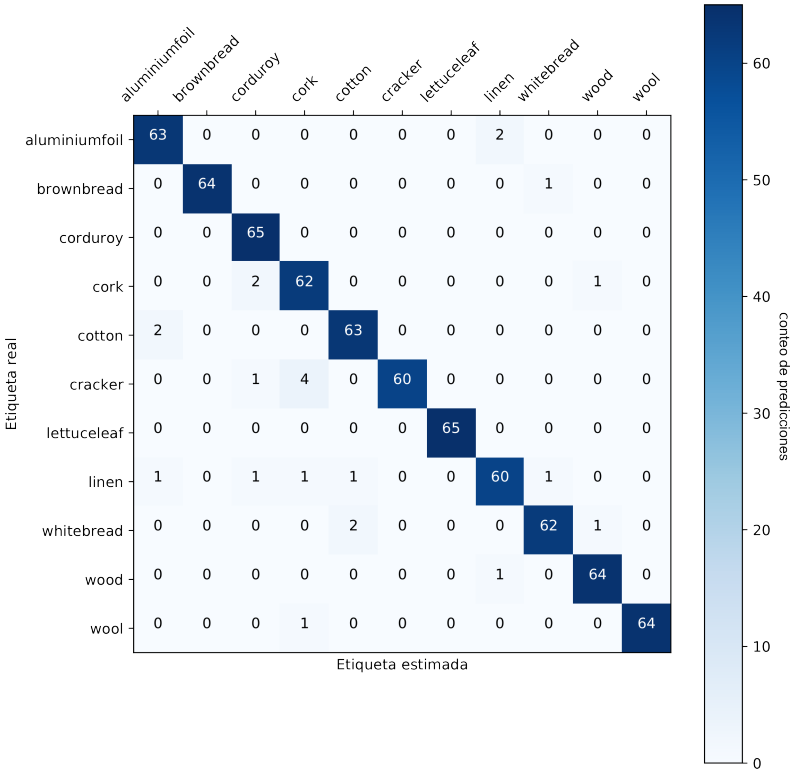


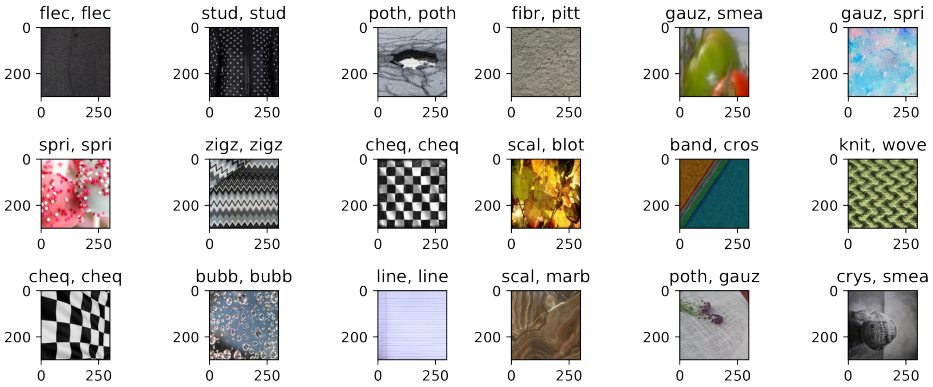
Fig. 9. Matriz de confusión para el conjunto de datos KT2B.

Por esa razón, al trabajar con imágenes, es conveniente normalizar los valores de los píxeles en un rango de 0 a 1, con la finalidad de que nuestro modelo converja rápidamente a un mínimo local, ya que las entradas con valores enteros grandes pueden ralentizar el proceso de aprendizaje.

3.3. Evaluación del modelo

En esta etapa se genera un conjunto de datos de manera aleatoria, el cual se divide en tres subconjuntos, uno de entrenamiento con el 75 % de las imágenes, otro de validación con el 15 % y el 15 % restante se utiliza para la etapa de prueba. Esta nueva distribución de imágenes para la evaluación del modelo durante el entrenamiento y validación, se aplicará a los tres conjuntos de datos, KT2B, FMD y DTD.

Después, se seleccionan los parámetros de entrenamiento, con la finalidad de ir evaluando el modelo durante cada época o iteración de aprendizaje. En la búsqueda de los filtros kernel los cuales establecen los rasgos característicos de cada textura, se debe seleccionar el que tenga un mejor desempeño de manera automática. Por lo tanto, este proceso permite ajustar y actualizar el modelo, conforme se esté entrenando la arquitectura.



(a) 292 Clasificadas correctamente. (b) 554 Clasificadas incorrectamente.

Fig. 10. Clasificación de texturas de manera aleatoria (de un total de 846 imágenes) utilizando el modelo de predicción DTD.

Por otro lado, al tener clases con una misma cantidad de imágenes, es posible utilizar una métrica de desempeño. En este caso, se puede calcular la exactitud, una métrica de desempeño muy relevante en tareas de clasificación. La exactitud (*accuracy*) es calculada como un porcentaje de imágenes que están correctamente etiquetadas por el modelo creado.

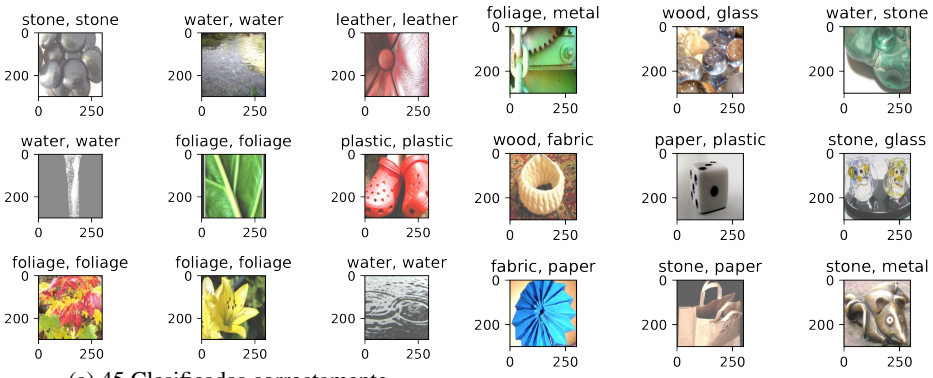
En relación con el rendimiento del modelo, una manera de determinar algunos patrones de error en la predicción o clasificación de las texturas, es utilizando la matriz de confusión múltiple. La cual es una tabla de N×N, que resume el nivel de éxito de las predicciones de un modelo de clasificación; es decir, la correlación entre la etiqueta y la clasificación del modelo. En este caso N representa el número de clases, un eje de la matriz de confusión es la etiqueta que el modelo predijo, y el otro es la etiqueta real.

3.4. Método para extracción de características

Por lo general, los diferentes métodos de extracción de características conducen a distintos elementos de información sobre la textura dentro de una imagen. Por lo que, retomando la idea central de la arquitectura Wavelet CNN [6], se decide diseñar una aproximación de la arquitectura. Dando como resultado un sistema híbrido para combinar los rasgos que genera la CNN a través de los filtros o kernel, junto con los atributos o mapas de características generados de manera manual con la transformada wavelet de Haar, mediante el análisis multiresolución a dos niveles de descomposición.

El diseño de la red incluye dos procesos separados (extracción de características mediante CNN y extracción de características usando el análisis wavelet) que posteriormente son fusionados en el entrenamiento del modelo. En el primer proceso, la imagen RGB preprocesada en la etapa de preparación de datos, entra como tensor a la arquitectura base VGG para conseguir ciertos patrones de manera automática.

Esto permitirá clasificar las texturas y diferenciar una de otra. El segundo proceso se realiza de manera adicional, dicho de otra manera, antes de ingresar la información espacial a la arquitectura CNN, debemos de generar nuevos datos sintéticos, que serán los atributos (coeficientes wavelet) o mapas de características en el dominio espectral.



(a) 45 Clasificadas correctamente. (b) Textura no detectada.

Fig. 11. Clasificación de texturas de manera aleatoria (de un total de 150 imágenes) utilizando el modelo de predicción FMD.

Conviene subrayar que, el proceso de análisis multiresolución en la etapa de descomposición, permite generar mapas de características que pueden ser adaptadas a los bloques convolucionales de la CNN base. Por lo tanto, con la fusión de estos procesos en la etapa de extracción de características se puede pasar a la siguiente etapa del aprendizaje profundo para la clasificación. En la Fig. 5, se muestra el sistema de clasificación con un enfoque espacial y con la fusión del enfoque espectral.

4. Resultados experimentales

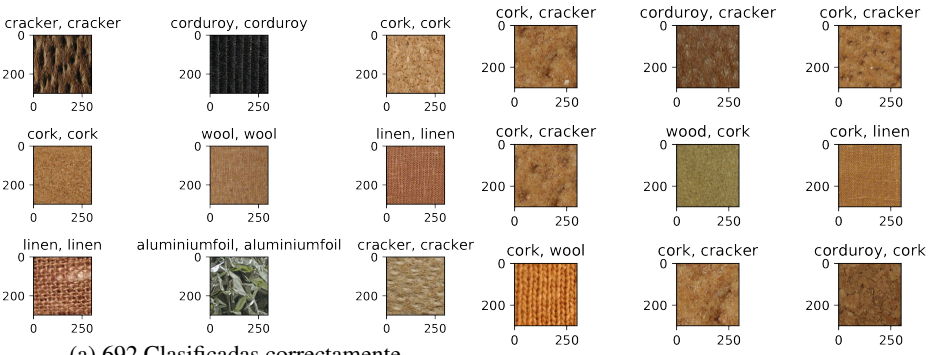
Para validar nuestro enfoque se ha utilizado tres conjuntos de datos KT2B, DTD y FMD. Dos de ellos (KT2B y DTD) son un caso especial de bases de datos de texturas porque contienen imágenes capturadas bajo condiciones no controladas.

4.1. Configuración experimental

En la Fig. 5 se ilustra el sistema de clasificación propuesto. La arquitectura esta diseñada de acuerdo a la red VGG16 [5], internamente tiene 5 bloques convolucionales con kernels de tamaño 3×3 y un padding (same) de manera que la salida tenga el mismo tamaño que la entrada. Además, cada bloque convolucional contiene dos métodos de convolución Conv2D, el primero con un stride de 1 y el segundo con un stride de 2 para asegurar que la salida sea la mitad de tamaño que la entrada.

Esto permite que los bloques convolucionales puedan extraer las características de las texturas en el dominio espacial. Por otro lado, dado que la arquitectura VGG y el análisis multiresolución (etapa de descomposición) tienen la misma característica de reducción, es posible concatenar cada nivel de descomposición (información en el dominio espectral) con los mapas de características de cada bloque convolucional.

En cuanto a determinar los atributos característicos de cada textura, se decide utilizar la transformada wavelet de Haar a 2 niveles. Este factor de 2, depende de la posibilidad de disminuir la imagen en cuanto a su tamaño.



(a) 692 Clasificadas correctamente. (b) Textura no detectada.

Fig. 12. Clasificación de texturas de manera aleatoria (de un total de 150 imágenes) utilizando el modelo de predicción FMD.

También, es importante mencionar que este proceso se aplica para cada uno de los canales RGB de la imagen, realizando la descomposición individual por canal y al final los mapas encontrados son concatenados en un vector.

La nueva información es concatenada en un cubo de características (espectral y espacial), y esté debe ser transformado para cambiar su representación mediante el método GlobalAveragePooling2D. Este nuevo vector a su vez alimenta el método de regularización DropOut para evitar el sobreajuste antes de pasar por la ultima capa de predicción, Softmax.

4.2. Implementación

En cuanto a la implementación del sistema, se utiliza el lenguaje Python y el API de Keras con Tensorflow como Backend [4]. En resumen, el sistema de clasificación tiene un total de 5,441,866 parámetros entrenables o pesos sinápticos de aprendizaje.

También, en la selección de los hyperparámetros se establece un indice de aprendizaje de 0.001, un minibatch de 30, 500 épocas de entrenamiento para el aprendizaje, cuatro Callbacks API para mejorar el rendimiento del modelo (ModelChekpoint, EarlyStopping, CVLogger, ReduceLROnPlateau), además del optimizador Adam, que es una variante de Gradiente Descendente.

Por otro lado, para el procesamiento de imágenes se utiliza las librerías OpenCV, por su fácil manejo y adaptabilidad dentro de la programación. En el caso del método adicional, la transformada wavelet de Haar, usamos la librería Pywt [10].

4.3. Resultados y discusión

La Fig. 6 muestra el comportamiento de aprendizaje para los tres conjuntos de datos. El máximo local alcanzado para la métrica de exactitud se muestra en la Fig. 6(a). Para el conjunto DTD el valor máximo alcanzado se dá alrededor de la época 21, con un valor de 39.74 % en entrenamiento (DTD entre_exac). Y para validación (DTD val_exac) de un 32.21 %.

En el caso del conjunto FMD el máximo local se dá en la época 45, con una exactitud de 34.14 % en entrenamiento (FMD entre_exac) y un 36.67 % para validación (FMD val_exac). Por último, para el conjunto de KT2B el máximo local se dá en la época 48, con un mejor rendimiento, un 95.80 % de exactitud para entrenamiento (KT2B entre_exac) y 96.29 % (KT2B val_exac) para validación.

En la Fig. 6 (b) se ilustra la pérdida del modelo en cada uno de los tres conjuntos de datos. También, se muestra que existe un punto de divergencia entre los dos conjuntos que se están entrenando (entrenamiento y validación). Este punto coincide con el número de época donde se encontró el máximo local de exactitud.

La tabla 1 resume el rendimiento alcanzado por cada modelo para los conjuntos de datos DTD, FMD y KT2B. La exactitud alcanzada para el conjunto de prueba en KT2B es del 96 %, FMD con un 30 % y para DTD es del 34 %. Estos resultados son muy importantes, dado que son imágenes que nunca ha visto el modelo. Por otro lado, a pesar de que las métricas en entrenamiento y validación se dan en diferentes tiempos, se muestra un comportamiento homogéneo con los datos de prueba.

Así que, existe una generalización del conocimiento entre los tres conjuntos, entrenamiento, validación y prueba. Con el fin de evaluar el desempeño de clasificación de nuestro modelo, se determina utilizar la matriz de confusión múltiple. Conviene subrayar, que existe una relación del conjunto de prueba con los resultados de la matriz de confusión. Para DTD en la Fig. 7, muestra una diagonal azul difícil de percibir y el mapa de calor aún distribuido en algunas zonas, esto indica que el modelo desarrollado aún encuentra semejanza en la mayoría de las clases.

En particular, para el conjunto de imágenes FMD, ver Fig. 8, la diagonal azul también es muy difícil de percibir. Se observa que en el centro de la matriz no existe una predicción correcta, ni con la etiqueta real y con ninguna otra clase, por lo que no se completa la diagonal. También, se encuentra que la mayor parte del mapa de calor esta distribuida a la derecha.

En cambio para el conjunto de datos KT2B, ver Fig. 9, muestra que existe una diagonal de color azul completa, esto resume que hay una tendencia positiva de predicción con respecto a la etiqueta original de las texturas KT2B. Las figuras, Fig. 10, 11 y 12 muestran algunas imágenes del conjunto de pruebas.

Estas imágenes son evaluadas por cada modelo, en la parte superior, se puede ver la etiqueta real y la predicción. En concreto, permite tener una idea visual sobre los rasgos de clasificación, y la correlación entre clases. La tabla 2, resume el rendimiento alcanzado de nuestro sistema de clasificación, así como una comparativa contra AlexNet, Textura CNN y Wavelet CNN [6][2].

Adicionalmente, se propone el conjunto de datos FMD para validar el modelo de aprendizaje. Por lo tanto, acorde a los resultados se puede ver que el modelo brinda una generalización de aprendizaje hacia otros conjuntos de datos.

5. Conclusión

En este trabajo se estudia la posibilidad de incorporar el análisis espectral a la arquitectura CNN, de acuerdo con algunas arquitecturas reportadas en la literatura.

Este novedoso sistema de clasificación, brinda un nuevo enfoque en la reestructuración de las capas convolucionales, la forma de generalizar el aprendizaje y la reducción del mapa de características.

También, se representa como una arquitectura de tres entradas - un modelo, dado que se alimenta con la información espacial y los dos mapas de características en el dominio espectral. En particular, con la fusión de los mapas de características creados de manera adicional, no limitamos el aprendizaje a las características espectrales. También, se demuestra que el modelo utilizado logra una mejor exactitud para la clasificación de la textura, con un número menor de parámetros a entrenar que los modelos existentes.

Además, los resultados mostraron que las características de textura creadas de manera adicional, podrán ser de ayuda para las arquitecturas CNN, especialmente en el caso de pequeños conjuntos de datos. Este enfoque, permitirá en un futuro probar otras características de textura y de arquitecturas CNN, en aplicaciones de reconocimiento de patrones en la restauración de imágenes, tareas de clasificación y detección de objetos.

Referencias

1. Alcalá-Rmz, V., Maeda-Gutiérrez, V., Zanella-Calzada, L. A., Valladares-Salgado, A., Celaya-Padilla, J. M., Galván-Tejada, C. E.: Convolutional Neural Network for Classification of Diabetic Retinopathy Grade. Advances in Soft Computing, Mexican International Conference on Artificial Intelligence, vol. 12468, pp. 104–118 (2020) doi: 10.1007/978-3-030-60884-2_8
2. Andrearczyk, V., Whelan, P. F.: Using filter banks in convolutional neural networks for texture classification. Computer Science, Computer Vision and Pattern Recognition, vol. 84, pp. 63–69 (2016) doi: 10.48550/ARXIV.1601.02919
3. Bengio, Y., Goodfellow, I., Courville, A.: Deep learning. MIT press, vol. 1 (2017)
4. Chollet, F.: Deep learning with Python. vol. 361 (2018)
5. Cimpoi, M., Maji, S., Kokkinos, I., Mohamed, S., Vedaldi, A.: Describing textures in the wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3606–3613 (2014)
6. Fujieda, S., Takayama, K., Hachisuka, T.: Wavelet convolutional neural networks (2018)
7. Hayman, E., Caputo, B., Fritz, M., Eklundh, J. O.: On the significance of real-world conditions for material classification. Computer Vision, European conference on computer vision, vol 3024, pp. 253–266 (2004) doi: 10.1007/978-3-540-24673-2_21
8. LeCun, Y.: Generalization and network design strategies. Connectionism in perspective, vol. 19, pp. 143–155 (1989)
9. LeCun, Y., Bengio, J., Hinton, G.: Deep learning. Nature, vol. 521, pp. 436–444 (2015)
10. Lee, G., Gommers, R., Waselewski, F., Wohlfahrt, K., O'Leary, A.: PyWavelets: A Python package for wavelet analysis. Journal of Open Source Software, vol. 4, no. 36, pp. 1237 (2019) doi: 10.21105/joss.01237.36
11. Mallat, S. G.: A theory for multiresolution signal decomposition: The wavelet representation. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11, no. 7, pp. 674–693 (1989) doi: 10.1109/34.192463
12. Rojas-Perez, L. O., Martinez-Carranza, J.: Autonomous Drone Racing with an Opponent: A First Approach. Computación y Sistemas, vol. 24, no. 3 (2020)
13. Sharan, L., Rosenholtz, R., Adelson, E.: Material perception: What can you see in a brief glance? Journal of Vision, vol. 9, no 8, pp. 784 (2009) doi: 10.1167/9.8.784

14. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition, pp. 1409–1556 (2014) doi: 10.48550/arXiv.1409.1556
15. Tapia-Téllez, J. M., Escalante, H. J.: Data augmentation with transformers for text classification. Lecture Notes in Computer Science, Mexican International Conference on Artificial Intelligence, vol. 12469, pp. 247–259 (2020) doi: 10.1007/978-3-030-60887-3_22
16. Vassilieva, N. S.: Content-based image retrieval methods. Programming and Computer Software, vol. 35, no. 3, pp. 158–180 (2009) doi: 10.1134/s036 1768809030049