

Calibración de cámara multispectral utilizando redes neuronales convolucionales

Iván A. Juárez Trujillo¹, Jonny P. Zavala de Paz²,
Omar Palillero Sandoval², Francisco A. Castillo Velásquez²

¹ Centro de Investigación en Ingeniería y Ciencias Aplicadas,
México

² Universidad Politécnica de Querétaro,
México

ivan.juarez@alumnos.uaem.mx,
{jonny.zavala, francisco.castillo}@upq.edu.mx,
omar.palillero@uaem.mx

Resumen. Se presenta una metodología para la calibración de cámara multispectral mediante redes neuronales convolucionales. Se capturaron imágenes en RGB de la cámara multispectral para cada uno de los estándares, las muestras son tomadas con las mismas condiciones de iluminación y con el mismo ángulo de captura. Las imágenes son fragmentadas en pequeños tamaños de matrices agregados a una clase específica, y guardada con una etiqueta especial para distinguirla de toda la base de datos de la clase, el mismo proceso lleva los 7 cerámicos de color restantes Lucideon Std. Uno de los cerámicos corresponderá a una clase en particular con una dimensión igual para todas las clases. Finalmente, con base a la metodología presentada se logra calibrar la cámara con respecto a las referencias.

Palabras clave. Calibración, cámara multispectral, redes neuronales convolucionales, análisis de color.

Multispectral Camera Calibration Using Convolutional Neural Networks

Abstract. A methodology for multispectral camera calibration using convolutional neural networks is presented. RGB images were captured from the multispectral camera for each of the standards, the samples are taken under the same lighting conditions and with the same capture angle. The images are fragmented into small matrix sizes added to a specific class, and saved with a special label to distinguish it from the entire class database, the same process takes the remaining 7 Lucideon Std color tiles. One of the

tiles will correspond to a particular class with an equal dimension for all classes. Finally, based on the presented methodology, it is possible to calibrate the camera with respect to the references.

Keywords. Calibration, multispectral camera, convolutional neural networks, color analysis.

1. Introducción

Las redes neuronales convoluciones (CNN) son redes complejas, el aprendizaje profundo para visión artificial cada vez se aplica para resolver problemas de visión artificial que llevan un aprendizaje de principio a fin, dando datos en bruto y una tarea a realizar en clasificación de imágenes, datos o sonidos aprendiendo a clasificar por sí solo.

En la mayoría de los casos, una red neuronal es un sistema que se adapta y cambia su estructura durante la fase de aprendizaje. Dicho aprendizaje, es una parte muy importante del proceso. A veces, estos sistemas son también asociados con un algoritmo de aprendizaje [1, 11].

Un uso común de las redes neuronales es la definición de una clase con una serie de funciones, en las que los miembros de la clase se obtienen mediante la variación de los parámetros, la conexión de los pesos o detalles específicos de

la arquitectura, tales como el número de neuronas o su conectividad [13, 14]. El aprendizaje automático trata de simular la forma que tiene de aprender el cerebro.

Ante un estímulo, ciertas neuronas se activan, evalúan la información que han recibido, reaccionan y se comunican con otras neuronas. En posteriores estímulos añaden nuevos datos a los que ya conocen, evalúan el resultado de las acciones anteriores y corrigen su funcionamiento para tener la mejor reacción posible.

Se puede decir, que es un conjunto de algoritmos de aprendizaje automático, que llevan tareas de práctica de principio a fin [1, 13]. En este trabajo se pretende normalizar las imágenes adquiridas mediante calibrar un sistema basado en redes neuronales (aprendizaje profundo).

La información de la adquisición de color mediante un sistema multiespectral no arroja información correcta, los efectos debidos a factores paramétricos (por ejemplo, tiempo de exposición, respuesta de la cámara, etc.) pueden alterar la información por lo que es necesario tener un sistema que pueda, definir con exactitud qué información estamos percibiendo con cualquier sistema de adquisición.

Los sistemas se agrupan en general en multibanda, multiespectral y sistemas hiperespectrales en función del número de bandas seleccionadas sobre un determinado espectro [2, 15]. Los instrumentos deben funcionar de manera confiable y la mejor manera de garantizar esto es verificarlos con estándares estables.

En 1978 introducen el termino de imágenes intrínsecas, esto hace referencia a una descomposición de nivel medio de una imagen separada en dos partes importantes: reflectancia y luminancia. Encontrar tal descomposición es un problema de investigación abierto e importante para el desarrollo de aplicaciones de visión por computadora.

Debido a que el problema fundamental de obtener las imágenes de reflectancia y luminancia es mal condicionado, algunos investigadores se han centrado en una solución al problema en base a algunas consideraciones iniciales [3, 16].

El aprendizaje automático es una rama de la inteligencia artificial que tiene como finalidad el desarrollo de técnicas que permiten a un sistema

la toma de decisiones de manera autónoma mediante aprendizaje, de forma más concreta, se trata de crear algoritmos con capacidad de generar comportamientos a partir de una información suministrada en forma de ejemplos [6, 7]. Las redes neuronales convolucionales no solo pueden descubrir automáticamente características útiles sin extracción manual, sino también procesar los datos en múltiples matrices (imágenes).

Por lo tanto, codificamos series de tiempo de aceleraciones triaxiales como imágenes con la relevancia en diferentes momentos y usamos el algoritmo de detección de caídas basado en CNN para clasificar los patrones de caída sobre la base de las imágenes [4].

Uno de los tipos de clasificadores que dan mejor resultado son las redes neuronales convolutivas. Este tipo de algoritmo de aprendizaje tiene la facilidad de que no depende tanto del conocimiento previo que se tiene de la materia, sino que solo requieren saber para cada muestra cual es el resultado esperado, y es el propio algoritmo el que extrae las características que usara después para clasificar.

Esto facilita la tarea de programación de la red neuronal, pero también tienen la desventaja que necesitan muchas muestras y mucho tiempo de aprendizaje [13, 14]. Los enfoques actuales para el reconocimiento de objetos hacen un uso esencial de los métodos de aprendizaje automático con densidad de información [10].

Para mejorar su rendimiento, podemos recopilar conjuntos de datos más grandes, aprender modelos más potentes y utilizar mejores técnicas para prevenir el sobreajuste. Hasta hace poco, los conjuntos de datos de imágenes etiquetadas eran relativamente pequeño del orden de decenas de miles de imágenes [14, 17].

2. Metodologías y materiales

2.1. Conversión de estímulos de color

Dado un color XYZ cuyos componentes están en el rango nominal [0.0, 1.0] y cuyo blanco de referencia es el mismo que el del sistema RGB, la conversión a RGB complementario se realiza en dos pasos.

Tabla 1. Especificaciones técnicas de la cámara multiespectral

Unidad principal	Dimensiones	59×41×28 mm
	Peso	72 g
Cámara multiespectral	Resolución	1,2 MP
	Tamaño de la imagen	1280×960 píxeles
	Obturador global	
	Disparador	3,98 mm
	Distancia focal	Verde 550 nm (ancho de 440 nm)
	Longitudes	Rojo 660 nm (ancho de 40 nm)
		Borde rojo 735 nm (ancho de 10 nm)
		Infrarrojo cercano 790 nm (ancho de 40 nm)

Dadas las coordenadas de cromaticidad de un sistema RGB (x_r, y_r), (x_g, y_g) y (x_b, y_b) y su referencia blanca (X_w, Y_w, Z_w). Aquí está el método para calcular la matriz 3×3 para convertir XYZ a RGB [5, 18]:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = (M) \begin{pmatrix} R \\ G \\ B \end{pmatrix}. \quad (1)$$

Para utilizar esta matriz correctamente, los valores RGB deben ser lineales y estar en el rango nominal [0.0, 1.0]. En muchos casos, los valores RGB pueden necesitar primero conversión (por ejemplo, dividir entre 255 y luego elevarlos a una potencia).

Tenga cuidado de que los blancos de referencia se usen de manera consistente. Por ejemplo, RGB se define en relación con un blanco de referencia D65 y los perfiles ICC se definen en relación con un blanco de referencia D50.

Los blancos de referencia que no coinciden deben tenerse en cuenta en otros lugares, generalmente mediante el uso de un algoritmo de adaptación cromática. La matriz inversa (es decir, la matriz que convierte XYZ a RGB) se calcula invirtiendo la matriz [M] anterior.

2.2. Estándares de color

Los estándares de cerámica Lucideon Ltd., son estándares populares para la calibración de color en los Estados Unidos y en el extranjero.

Estos estándares se pueden usar para calibrar espectrofotómetros de color en geometría tanto hemisférica como direccional/direccional.

Los juegos Lucideon BCRA de Avian Technologies constan de catorce azulejos (el conjunto estándar Lucideon BCRA más un azulejo blanco y negro), montados y sellados para evitar los efectos de la humedad [6].

3. Desarrollo experimental

3.1. Imágenes multiespectrales

Se toma la imagen en RGB de la cámara multiespectral (ver Tabla 1) para cada uno de los estándares, las muestras son tomadas con las mismas condiciones de iluminación y con el mismo ángulo de captura, es ingresado cada estándar de color a una cabina oscura, la fuente de iluminación a 90° con respecto a la base, como se muestra en la fig. 1.

La cámara multiespectral Parrot sequoia es un equipo que proporciona imágenes espectrales con anchos de banda establecidos por el fabricante, dichas imágenes entregan en espacios de colores CIE y componentes RGB. Además de que entregan imágenes con dimensiones en píxeles de 4608 x 3456 como se muestra en la fig. 1.

Las imágenes para manipular deberán contener una dimensión de solo el estándar, por lo que es necesario obtener una imagen recortada como se muestra en la fig. 1, se muestran las 8 imágenes capturadas por la cámara multiespectral Parrot Sequoia [9].

Con la ayuda del software MATLAB R2019 [4], se procede a separar los componentes de la imagen, la imagen multiespectral contiene los dos componentes.

3.2. Adquisición de muestras

La base de datos o imágenes generadas virtualmente de estándares de color definidas por fabricantes de calibración, nos permiten manipular los datos físicos, para sistemas virtuales de imágenes.

Se inició con la toma de lectura de 8 estándares con un espectrofotómetro que mide

longitudes de onda de 400 a 700nm, el barrido de lectura inicia con 400nm y una resolución de 10nm de separación para cada lectura, teniendo un total de 30 lecturas, las lecturas son imágenes se encuentran en el espacio de color LAB, con el objetivo de tener un estándar de color se realiza el ajuste de espacio de color LAB a CIE ya que nos permite utilizar cualquier equipo o dispositivo que capture imágenes en espacios CIE con componentes RGB.

El resultado de dicha conversión utilizando las fórmulas y generación de patrones virtuales de 30 a 3 componentes de color. En la fig. 1 observamos los colores creados a partir de estándares físicos, se puede observar a detalle comparando los resultados y los estándares físicos. Los lotes de entrenamiento contienen las imágenes restantes en orden aleatorio, pero algunos lotes de entrenamiento pueden contener más imágenes de una clase que de otra [10, 12].

3.3. Categorías y clases

La base de datos se basa en 8 categorías de 8 distintas clases de test de color De la marca Lucideon STD, con una dimensión en píxeles de 32 x 32 en RGB, con una dimensión de la base de datos de 8,000 imágenes de 8 categorías y 8 clases diferentes.

La base de datos se inicia con la toma de imágenes con la cámara multispectral Parrot Sequoia (ver Fig. 2), tomando una imagen a 8 cerámicos de color de la marca Lucideon STD.

Con una dimensión en píxeles de 3000 x 3000 en RGB, a través de la separación y segmentación de cada una de las imágenes para formar pequeñas matrices con dimensiones en píxeles de 32 x 32 en RGB y con una etiqueta distinta una de otra, con tal método se forma las 8 categorías de las 8 distintas clases de cerámicos de color de la marca Lucideon STD [6, 9].

3.4 Calibración con redes neuronales

Observar una red típica de aprendizaje con n cantidad de capas de las cuales estarán detectando una serie de características, mientras que las últimas clasificarán las imágenes. Al inicio tendremos una serie de imágenes con una dimensión largo y ancho, y con las tres

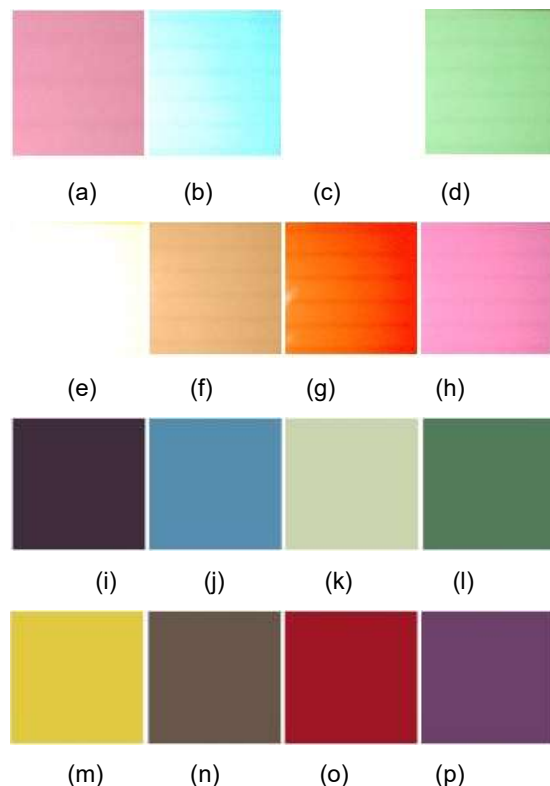


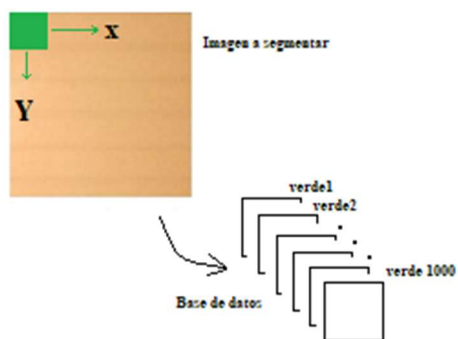
Fig. 1. Imágenes de cerámico de color Lucideon Ltd . (a) B5-G, (c)B6-G, (e)B7-G, (g)B8-G, (i)B9-G, (k)B10-G, (m)B11-G y (o)B12-G imágenes capturadas por la cámara multispectral parrot sequoia, (b)B5-G1, (d)B6-G1, (f)B7-G1, (h)B8-G1, (j)B9-G1, (l)B10-G1, (n)B11-G1 y (p)B12-G1 imágenes creadas por medio del espectrofotómetro CM-700D de la marca Konica Minolta (ver Tabla 2)

componentes de color RGB, las capas internas como convolución y pooling estarán encargadas de las caracterizas de las imágenes con dimensiones en 2D.

Es un tipo de red neuronal profunda que puede trabajar directamente con datos estructurados como imágenes con el fin de clasificarlas, al pilar múltiples capas para formar la arquitectura se crean las redes capaz de llevar a cabo el aprendizaje de las propias características eliminando la necesidad de llevar a cabo una selección manual de las mismas, a diferencia del enfoque clásico en redes neuronales donde se utiliza un número de capas reducido, utilizando CNN estaremos trabajando fácilmente con un número de capas elevado

Tabla 2. Especificaciones técnicas del espectrofotómetro Konica Minolta CM700D

Características	Especificaciones
Sistema de iluminación / visualización	Iluminación difusa. 8° ángulo de visión. Componente especular incluido/excluido. Cumple con CIE No. 15, ISO 7724/1, DIN5033 Teil7, ASTM E 1164 y JIS Z 8722.
Tamaño de la esfera integradora	Diámetro de 40 mm.
Dispositivo de separación espectral	Rejilla de difracción.
Rango de longitud de onda	400 a 700 nm.
Tono de longitud de onda	10 nm.
Medio ancho de banda	Aproximadamente 10 nm.
Rango de reflectancia	de 0 a 175%, resolución de pantalla: 0.01%.
Fuente de luz	Lámpara de xenón pulsada (con filtro de corte UV).
Tiempo de medición	Aproximadamente 1 segundo.
Intervalo mínimo de medición	Aprox. 2 segundos (en modo SCI o SCE).

**Fig. 2.** Creación de la base de datos a partir de las imágenes multiespectral. La imagen se secciona dependiendo de tamaño de los filtros y se crea una base de datos a partir de una imagen

posiblemente entre 5 y 50 capas o incluso más, por este motivo las GPU van a jugar un papel clave en el entrenamiento de estas redes neuronales convolucionales, en la fig. 3 podemos ver una arquitectura CNN como una

representación de los datos, a medida que viajan por la red, a muy alto nivel el primer conjunto de capas estarán detectando características, mientras que las capas finales tendrán como objetivo resolver la clasificación de la imagen [4].

La clave es que todas estas capas van a ser entrenadas conjuntamente es decir el proceso de entrenamiento va a consistir en ajustar correctamente los pesos en todas estas capas para que las imágenes sean clasificadas correctamente, nos fijamos en la escritura de la CNN en algo más de detalle vamos a comenzar con una capa de entrada, cuyos parámetros van a venir determinados por el conjunto de imágenes en este caso vamos a trabajar con imágenes que han terminado ancho y alto y 3 canales, rojo, verde y azul.

A continuación, podemos tener varias capas convolución, ReLu y Pooling, en qué consisten, estas capas fundamentalmente van a estar involucradas en el aprendizaje de características de las imágenes mediante la aplicación de convoluciones 2D.

Posteriormente, una vez que llegamos al último volumen, lo que vamos a hacer simplemente es aplanar esta capa está bien estirla en un vector de modo que podemos trabajar con redes totalmente conectadas y finalmente una capa Softmax que nos permite determinar a qué etiqueta corresponde a nuestra imagen.

3.5 Capa de convolución

En la fig. 3, vemos una representación o arquitectura de la CNN y es interesante ver aquí como los filtros que vamos a utilizar en las capas que llevan a cabo convoluciones estarán formando formas o estructuras muy simples en capas tempranas o en las primeras capas posiblemente segmentos verticales, horizontales quizá pequeños cubos y estructuras más complejas a medida que nos adentramos en capas más y más profundas, de modo que podemos utilizar estas características para distinguir una flor de otros objetos [8].

El modo en que esta capa de convolución va a funcionar es que vamos a llevar a cabo una serie de convolución es para el volumen 3D de datos de entrada que se ve a la izquierda, con una serie

de filtros para producir el volumen 3D que vemos en la derecha, vamos a definir el tamaño de estos filtros como parte de lo que se conocerán como hiper parámetros de la capa de convolución estando en la profundidad del volumen de salida determinada por el número de filtros a considerar, en este caso vemos en el ejemplo que estaremos trabajando con dos filtros.

Si empezamos con 1 de estos filtros vamos a realizar la convolución utilizando el filtro a lo largo del volumen de entrada para cada una de las posiciones que vamos cubriendo calcularemos el producto escalar y sumaremos el resultado tal y como se puede apreciar en el volumen de salida a la derecha con cada aplicación del filtro, sobre el volumen de entrada estaremos obteniendo un único valor escalar.

El proceso va a repetirse para todos los filtros deslizándolos con el volumen de entrada calculando ese producto escalar con la región que corresponde sumando el resultado de modo que obtenemos diferentes valores escalares para el volumen de salida y calculamos así diferentes mapas de activación para cada 1 de los filtros.

En este caso puesto que habíamos considerado los filtros obtenemos dos mapas de activación, aspecto clave para tener en cuenta aquí es que los valores de estos filtros están siendo apreñados por de modo que se activarán cuando la red ve a determinadas características específicas.

En la fig. 5 de presentan los distintos hiper parámetros que podemos tener en cuenta para la capa de convolución, aquí vemos el aspecto que tiene el volumen de entrada y de salida para la capa de convolución y como la selección de parámetros afecta al resultado.

Los hiper parámetros más relevantes serán, por un lado, el número de filtros aquí vemos 1 de los filtros 3D en la capa de entrada vamos a tener cada estos filtros 3D y para entender mejor el diagrama esta es una vista frontal de la entrada a la capa de convolución en realidad el número de filtros determinará la profundidad del volumen de salida, esto es de 2 será igual K.

Por otro lado, otro de los parámetros era el tamaño del filtro, aquí en la transparencia nos hemos permitido simplificar y considerar que la altura y anchura del filtro coinciden obviamente la profundidad del filtro vendrá determinada por D1.

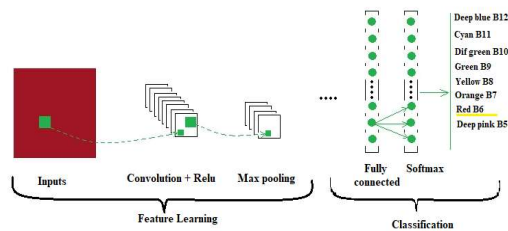


Fig. 3. Arquitectura general de una red neuronal convolucional

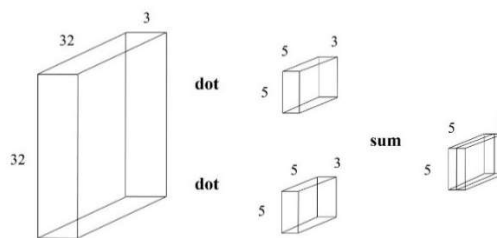


Fig. 4. Arquitectura de las capas de convolución, imágenes de entrada de 32 x 32 píxeles en RGB con filtros de 5 x 5 píxeles para formas en la salida de la capa de convolución de estructuras de 5 x 5 en 2 dimensiones

Sin embargo, el stride que especifica el intervalo desde el cual aplicar los filtros al volumen de entrada y por último el relleno con ceros a cero padding, que permite extender implícitamente los lados de la entrada con ceros este relleno con ceros es especialmente relevante cuando queremos preservar el tamaño del volumen de entrada o para evitar una reducción muy temprana de las dimensiones de la red que nos llevaría a unos malos resultados.

En la fig. 5 de presenta el diseño de los filtros de las capas de convolución, los filtros son estructuras con características y tamaños previamente calculados como se muestra en las ecuaciones 1, 2 y 3, donde deben diseñar el tamaño del filtro, el cual se desplazará en la imagen de entrada para formar un nuevo dato de salida con dimensiones cada más pequeñas.

Una vez hayamos dado valores a estos parámetros los valores para las dimensiones del volumen de salida pueden calcularse fácilmente con las siguientes fórmulas:

$$W_2 = (W_1 - F + 2P)/S + 1. \tag{2}$$

$$H_2 = (H_1 - F + 2P)/S + 1. \quad (3)$$

$$D_2 = K. \quad (4)$$

4. Resultados

Comenzamos con imágenes, de un determinado tamaño y en este caso 3 canales RGB. El cerámico de color Lucideon Std, Es fragmentado en pequeños tamaños de matrices agregados a una clase específica, y guardada con una etiqueta especial para distinguirla de toda la base de datos de la clase, el mismo proceso llevarán los 7 cerámicos de color restantes Lucideon Std.

Uno de los cerámicos corresponderá a una clase en particular con una dimensión igual para todas las clases. En la fig. 6 se presentan las 10 muestras de 10 clases distintas, en la cual se tomaron al azar 10 imágenes de cada categoría de las 5000 imágenes. Estas fueron acomodadas conforme se almacenaron en la base de datos para formar cada una de las imágenes.

Vamos a hablar de una red neuronal profunda, las redes neuronales convolucionales. Es un tipo de red profunda que puede trabajar directamente con datos estructurados, con el objetivo de clasificar, El objetivo o, lo que vamos a hacer es afilar múltiples capas. Para formar la arquitectura de la red.

De modo que seamos capaces de que la red pueda llevar a cabo el aprendizaje de las propias características presentes en los datos. Así como la clasificación, y eliminaremos así la necesidad. Manualmente elegir cuáles son las características importantes.

En la fig. 7 se presenta la primera capa de convolución con los 32 filtros. La información de la imagen es el resultado de 32 filtros de tamaño 5 x 5 que había en la primera capa de convolución de modo que esta red en esa primera capa debe ser capaz de detectar bordes verticales, horizontales pequeñas regiones, pequeños blogs que haya en la imagen, otras capas más profundas van a estar desempeñando otras tareas como entender o interpretar estructuras más complejas.

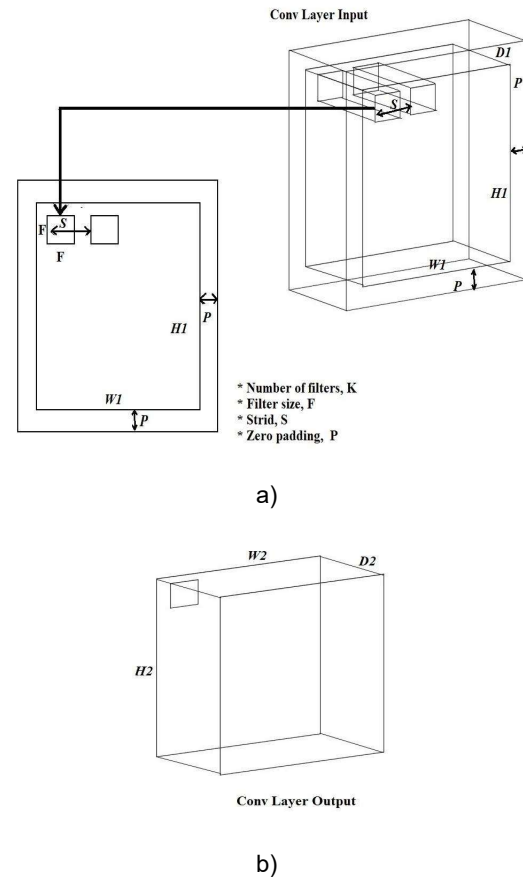


Fig. 5. a) Imagen de entrada implementando filtro de convolución, b) Resultado de la aplicación del filtro de convolución



Fig. 6. Imagen de 10 muestras de 10 clases distintas

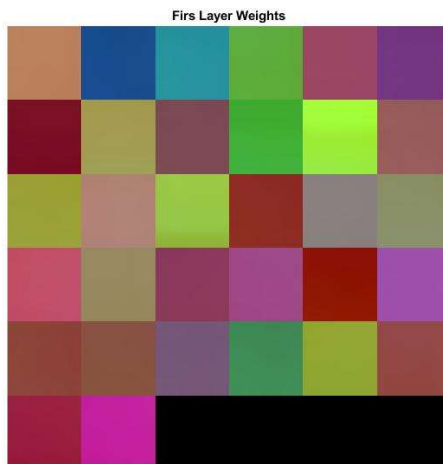


Fig.7. Primera capa de convolución con los 32 filtros

Tabla 3. Comparación de resultados y técnica de los trabajos relacionados.

Método	Aplicación	Referencia	Cuantificador de calibración
Redes neuronales convolucionales	Calibración de cámara multiespectral	Patrones de color Mediciones con espectrofotómetro	Calibración por reflectancia Espacio de color CIE RGB
Parámetros intrínsecos relacionados con el enfoque basado en el modelo de lente delgada [17]	Objetos pequeños con profundidad de campo ampliada	Modelos matemáticos	Calibración de una cámara con parámetros intrínsecos relacionados con el enfoque basado en el modelo de lente delgada para realizar mediciones de alta precisión para objetos pequeños con profundidad de campo ampliada
Integración de una cámara multiespectral y aprendizaje máquina [14]	Calibración de cámara multiespectral	Modelos matemáticos Algoritmos de aprendizaje máquina	Distintas arquitecturas de redes neuronales
Transformación usando Redes neuronales convolucionales [13]	Compresión de imágenes multiespectrales de detección remota	Imágenes multiespectrales	Rendimiento de compresión del algoritmo propuesto con la transformada espectral usando redes neuronales convolucionales, realizando experimentos con Matlab
Clasificación de imágenes de alta resolución [11]	Imágenes recopiladas de la web y etiquetadas por humanos	Base de datos de ImageNet Imágenes RGB	Red neuronal convolucional, la cual se tuvo una pérdida de 2% de rendimiento

5. Conclusiones

A medida que los algoritmos relacionados con el aprendizaje profundo han hecho avanzar los resultados de vanguardia de varias tareas de visión por computadora por un amplio margen, se

vuelve más difícil progresar en la parte superior de eso.

Puede haber varias direcciones para modelos más potentes: La primera dirección es aumentar la capacidad de generalización aumentando el tamaño de las redes.

Las redes más grandes normalmente podrían brindar un rendimiento de mayor calidad, pero se debe tener cuidado para abordar los problemas que esto puede causar, como el exceso de información y la necesidad de una gran cantidad de recursos computacionales.

Una segunda dirección es combinar la información de múltiples fuentes. La fusión de características ha sido popular y atractiva durante mucho tiempo, y esta fusión se puede clasificar en dos tipos. 1) Combine las características de cada capa de la red.

Diferentes capas pueden aprender diferentes características. Es prometedor si pudiéramos desarrollar un algoritmo para hacer que las entidades de cada capa sean complementarias.

La base teórica subyacente aún no explica en qué condiciones se desempeñarán bien o superarán a otros enfoques, y cómo determinar la estructura óptima para una determinada tarea.

Este documento describe estos desafíos y resume las nuevas tendencias en el diseño y entrenamiento de redes neuronales profundas, junto con varias direcciones que pueden explorarse más a fondo en el futuro. Los resultados de reflectancia multispectral se presentan de manera cualitativa.

En la Tabla 3, se presenta una comparación del método presentado con otros autores, el cual proporciona una cantidad adicional entre las posibles aplicaciones del método propuesto para reducir tiempo y complejidad de los procedimientos, que pueden ser aplicados a diferentes tipos de procesos industriales.

Referencias

1. **César, A. (2017)**. Deep learning aplicado al diagnóstico de soldadura mediante espectroscopia óptica de plasma. Universidad de Cantabria, pp. 28–31.
2. **Barrow, H., Tenenbaum, J. (1978)**. Recovering intrinsic scene characteristics from images. *Computer Vision Systems*, pp. 3–26.
3. **Simone, B., Alain, T., Raimondo, S. (2016)**. RGB-D-λ: 3D multispectral acquisition with stereo RGB cameras. *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, Vol. 4, pp. 689–694. DOI: 10.5220/0005845406890696.
4. **International Commission on Illumination**. CIE 15: Technical report: Colorimetry. 3rd edition, Incorporated into U.S. Law in 10 CFR.
5. **AvianTechnologies**, Lucideon BCRA color standard tile sets. <https://aviantechnologies.com/product/brca-color-tile-sets/>.
6. **Villaluenga-Morán, J., L. (2019)**. OpenStreetCam: reconocimiento automático de objetos en imágenes mediante machine learning. Universitat Oberta de Catalunya, <http://hdl.handle.net/10609/88287>.
7. **LeCun, Y., Bottou, L., Bengio, Y., Haffner, P. (1998)**. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, Vol. 86, No. 11, pp. 2278–2324. DOI: 10.1109/5.726791.
8. **Krizhevsky, A. (2009)**. The CIFAR-10 dataset, learning multiple layers of features from tiny images. <https://www.cs.toronto.edu/~kriz/cifar.html>.
9. **Zeiler, M. (2012)**. Hierarchical convolutional deep learning in computer vision. Ph.D. tesis, New York University.
10. **Bordes, A., Glorot, X., Weston, J., Bengio, Y. (2012)**. Joint learning of words and meaning representations for open-text semantic parsing. *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, Vol. 22, pp.127–35.
11. **Krizhevsky, A., Sutskever, I., Hinton, G. (2012)**. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, Vol. 60, pp. 84–9. DOI: 10.1145/3065386.
12. **Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J. (2013)**. Distributed representations of words and phrases and their compositionality. *Proceedings of the 26th International Conference on Neural Information Processing Systems*, Vol. 2, pp. 3111–3119.
13. **Jin, L., Zilong, L. (2019)**. Multispectral transforms using convolution neural networks for remote sensing multispectral image

- compression. *Remote sensing*, Vol. 11, No. 7, pp. 1–20. DOI: 10.3390/rs11070759.
14. **Gante, S. A., Lozano, H. Y., Maldonado, M. A., Flores, G. R., Villegas, D. (2022).** Integración de una cámara multispectral y aprendizaje automático para clasificación de manzanas. *Ingeniería Investigación y Tecnología*, Vol. 23, No. 4, pp. 1–13. DOI: 10.22201/ii.25940732e.2022.23.4.031.
15. **Ashraf, S., Kadery, I., Chowdhury, M. A. A., Mahbub, T. Z., Rahman, R. M. (2019).** Fruit image classification using convolutional neural networks. *International Journal of Software Innovation (IJSI)*, Vol. 7, No. 4, pp. 51-70. DOI: 10.4018/IJSI.2019100103.
16. **Chollet, F. (2017).** Xception: Deep learning with depthwise separable convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1800–1807. DOI: 10.1109/CVPR.2017.195.
17. **Lu, Z., Cai, L. (2020).** Camera calibration method with focus related intrinsic parameters based on the thin-lens model. *Optics Express*, Vol. 28, No. 14, pp. 20858–20878. DOI: 10.1364/OE.392731.
18. **Hameed, K., Chai, D., Rassau, A. (2018).** A comprehensive review of fruit and vegetable classification techniques. *Image and Vision Computing*, Vol. 80, pp. 24–44. DOI: 10.1016/j.imavis.2018.09.016.

*Article received on 08/05/2023; accepted on 01/09/2023.
Corresponding author is Jonny P. Zavala De Paz.*