¿Qué es el reconocimiento de imágenes?



[Artificial Intelligence](https://www.ibm.com/mx-es/think/artificial-intelligence)

**12 de noviembre de 2024**

* [¿Qué es el reconocimiento de imágenes?](https://www.ibm.com/mx-es/think/topics/image-recognition?mhsrc=ibmsearch_a&mhq=reconocimiento%20facial#%C2%BFQu%C3%A9+es+el+reconocimiento+de+im%C3%A1genes%3F)
* [Reconocimiento de imágenes con machine learning tradicional](https://www.ibm.com/mx-es/think/topics/image-recognition?mhsrc=ibmsearch_a&mhq=reconocimiento%20facial#Reconocimiento+de+im%C3%A1genes+con+machine+learning+tradicional)
* [Desafíos en el reconocimiento de imágenes](https://www.ibm.com/mx-es/think/topics/image-recognition?mhsrc=ibmsearch_a&mhq=reconocimiento%20facial#Desaf%C3%ADos+en+el+reconocimiento+de+im%C3%A1genes)
* [Reconocimiento de imágenes y detección de objetos](https://www.ibm.com/mx-es/think/topics/image-recognition?mhsrc=ibmsearch_a&mhq=reconocimiento%20facial#Reconocimiento+de+im%C3%A1genes+y+detecci%C3%B3n+de+objetos)
* [Usos en evolución del reconocimiento de imágenes](https://www.ibm.com/mx-es/think/topics/image-recognition?mhsrc=ibmsearch_a&mhq=reconocimiento%20facial#Usos+en+evoluci%C3%B3n+del+reconocimiento+de+im%C3%A1genes)

**Autores**



[**Tim Mucci**](https://www.ibm.com/think/author/tim-mucci)

IBM Writer

Gather

¿Qué es el reconocimiento de imágenes?

El reconocimiento de imágenes es una aplicación de machine learning (ML) (o tecnología de machine learning) que permite que el software y los dispositivos identifiquen objetos, lugares, personas, escritura y acciones en imágenes digitales o video.

Esta tecnología permite a las computadoras identificar defectos de los productos, ayuda a los profesionales médicos a detectar anomalías y es fundamental para el desarrollo de los vehículos autónomos.

El reconocimiento de imágenes es una tarea central de la [visión artificial](https://www.ibm.com/mx-es/think/topics/computer-vision), el campo más amplio de la tecnología de [inteligencia artificial](https://www.ibm.com/mx-es/think/topics/artificial-intelligence) (IA) que permite que el software y las máquinas comprendan los datos visuales y reaccionen a estos.

Los ingenieros utilizan modelos tradicionales de [machine learning](https://www.ibm.com/mx-es/topics/machine-learning) y aprendizaje profundo en el reconocimiento de imágenes. Estos enfoques suelen estar separados, y si se combinan o utilizan de forma independiente depende del problema específico y los requisitos de recursos.

Reconocimiento de imágenes con machine learning tradicional

El machine learning utiliza algoritmos que clasifican imágenes en función de características extraídas manualmente por ingenieros humanos. Los ingenieros preprocesan las imágenes y las analizan de acuerdo con el objetivo o problema específico que intentan resolver mediante el reconocimiento de imágenes.

Tal vez sea identificar rostros, detectar objetos o clasificar texturas. En cada caso, los ingenieros utilizan su conocimiento del dominio para preprocesar las imágenes y entrenar el algoritmo.



Software de reconocimiento de imágenes que utiliza detección de objetos para discernir tipos específicos de vehículos

Normalización

Los ingenieros preparan las imágenes para el análisis normalizándolas, lo que significa escalar los valores de píxeles a un rango estándar, generalmente entre 0-1 o -1-1, para que los datos sean congruentes y más manejables para que los procesen los modelos de machine learning.

El preprocesamiento también incluye cambiar el tamaño de las imágenes, convertirlas a escala de grises para reducir la complejidad computacional o eliminar el ruido mediante el uso de técnicas de filtrado gaussiano. En el reconocimiento de imágenes, “ruido” se refiere a cualquier variación no deseada o aleatoria en píxeles; por ejemplo, una imagen moteada, granulada, borrosa o distorsionada.

Extracción de características

A continuación, los ingenieros deben seleccionar las características que proporcionen la información más significativa. Puede ser bordes al detectar formas o la intensidad de color si el resultado es distinguir objetos por el tono. Debido a que los modelos de machine learning se basan en características extraídas manualmente, la anotación de datos etiqueta información esencial.

Al anotar objetos de interés dentro de las imágenes, los modelos pueden reconocer y clasificar más fácilmente objetos específicos como "gato" o "perro". Los datos anotados con precisión permiten a los algoritmos de machine learning aprender con exactitud las características visuales de cada categoría.

Codificación para machine learning

Los ingenieros extraen estas características y las formatean en vectores numéricos, lo que facilita que los modelos de machine learning procesen y comparen imágenes. Los ingenieros traducen cada imagen en un vector de características de longitud fija, una lista de números que resumen su importancia.

Reconocimiento de imágenes con aprendizaje profundo

Por el contrario, los modelos de aprendizaje profundo pueden aprender directamente de la imagen. El aprendizaje profundo, un subconjunto del machine learning, emplea redes neuronales para realizar tareas complejas de preprocesamiento y reconocimiento de imágenes, pero a costa de mayores requisitos computacionales y de datos.

Las redes neuronales convolucionales (CNN)son arquitecturas de [aprendizaje profundo](https://www.ibm.com/mx-es/topics/deep-learning) con capas convolucionales que analizan y aprenden la naturaleza estructurada de los datos de imagen.

Capa de entrada

[La red neuronal profunda](https://www.ibm.com/mx-es/topics/neural-networks) de CNN detecta automáticamente el valor de píxel sin procesar de la imagen. La CNN pasa esa información a través de capas de la red profunda para extraer patrones y, en última instancia, hacer predicciones sobre la imagen.

Las capas de la red comienzan con la capa de entrada. La capa de entrada procesa los valores de píxeles sin procesar de la imagen, tratándolos como una cuadrícula de intensidades numéricas y los pasa a capas posteriores para la extracción de patrones.

Extracción de características

A continuación, la capa convolucional aplica pequeños filtros o kernels sobre la imagen para detectar patrones locales, como bordes o textura. La convolución reduce la necesidad de extracción manual de características porque la red puede aprender los patrones directamente de los datos.

Después de cada convolución, una función de activación introduce no linealidad en el modelo, lo que permite a la red aprender patrones, formas y objetos complejos apilando varias capas.

Agrupación y aplanamiento

La agrupación de capas reduce la muestra de la imagen para reducir su tamaño, al tiempo que conserva características importantes para garantizar que el modelo sea computacionalmente eficiente en el manejo de variaciones, como ligeras rotaciones o cambios en la imagen.

Después de que la red extrae características, aplana los datos en un vector unidimensional y los pasa a través de capas completamente conectadas. Estas capas integran los patrones aprendidos en etapas anteriores para identificar relaciones complejas y refinar el proceso de clasificación.

Capa de salida

Finalmente, los datos llegan a la capa de salida, que consolida las características extraídas y produce una predicción final. Esta predicción se compara con el conjunto de datos de entrenamiento anotado para calcular errores y ajustar los pesos de la red para mejorar la precisión.

Por ejemplo, para entrenar a un modelo para que reconozca imágenes de gatos, los ingenieros pueden usar el [aprendizaje supervisado](https://www.ibm.com/mx-es/topics/supervised-learning), etiquetando miles de imágenes con etiquetas, como "gato" o "no gato", para que el modelo pueda aprender características clave, como la textura del pelaje, los bigotes y la forma de las orejas.

Alternativamente, en el [aprendizaje no supervisado](https://www.ibm.com/mx-es/topics/unsupervised-learning), el modelo funciona con datos sin etiquetar para descubrir patrones de forma independiente. El modelo identifica relaciones sin categorías predefinidas agrupando imágenes en función de características compartidas (por ejemplo, formas o texturas similares).

Este enfoque es útil para tareas, como la detección de fraudes, el control de calidad y el análisis de patrones cuando los datos etiquetados no están disponibles. En el aprendizaje no supervisado, el modelo agruparía imágenes de forma independiente en función de patrones compartidos, agrupando todas las imágenes de gatos sin saber explícitamente que son gatos.

Un tercer enfoque, el [aprendizaje autosupervisado](https://www.ibm.com/mx-es/topics/self-supervised-learning), combina aspectos del aprendizaje no supervisado comenzando con datos sin etiquetar, pero generando pseudoetiquetas a partir de la estructura inherente de los datos, lo que permite que los modelos aprendan representaciones significativas sin etiquetas tradicionales, lo que los vuelve eficaces para tareas con conjuntos de datos etiquetados limitados.

Con el aprendizaje autosupervisado, el modelo podría analizar partes de una imagen, como la reconstrucción de la cara de un gato parcialmente oculta, para identificar patrones y características. En última instancia, el modelo entrenado (ya sea mediante machine learning o aprendizaje profundo) podría identificar y clasificar con precisión imágenes nuevas e inéditas de gatos, distinguiéndolos de otros animales u objetos.



Reconocimiento de imágenes mediante cuadros delimitadores para clasificar vehículos

Desafíos en el reconocimiento de imágenes

Si bien las tecnologías de reconocimiento de imágenes han avanzado, aún enfrentan desafíos que afectan la precisión y la confiabilidad. Los ingenieros mitigan estos problemas combinando arquitecturas de modelos mejoradas, diversos conjuntos de datos de entrenamiento y técnicas de preprocesamiento.

Imágenes desordenadas u oscurecidas

[El aprendizaje supervisado](https://www.ibm.com/mx-es/topics/supervised-learning)utiliza datos etiquetados, con cada imagen etiquetada con su categoría correcta para guiar el algoritmo a través de ejemplos claros. Por ejemplo, entrenar un sistema para reconocer automóviles implica un conjunto de datos etiquetado como "gato" y "no gato". Luego, el modelo aprende a diferenciar en función de patrones visuales dentro de estos ejemplos etiquetados.

Variaciones de ángulo y perspectiva

En el [aprendizaje no supervisado](https://www.ibm.com/mx-es/topics/unsupervised-learning), el algoritmo emplea datos no etiquetados, descubriendo patrones de forma independiente. Es similar a darle a un niño una caja de juguetes para clasificarlos por similitud; los algoritmos no supervisados agrupan imágenes según características compartidas (por ejemplo, bigotes, pelaje, cuatro patas y cola) sin conocer las categorías explícitas.

Condiciones de iluminación

Los cambios en la iluminación, como sombras, variaciones de brillo o entornos con poca luz, pueden afectar el rendimiento de los sistemas de reconocimiento de imágenes. Los puntos brillantes pueden borrar detalles, mientras que las sombras pueden oscurecer características críticas, lo que hace que el modelo malinterprete la forma o el color de un objeto.

Los métodos avanzados, como la ecualización adaptativa de histogramas o la incorporación de datos de entrenamiento con diferentes condiciones de iluminación, ayudan a que los modelos funcionen mejor en diferentes escenarios de iluminación.

Limitaciones en los datos de entrenamiento

El rendimiento de un modelo de reconocimiento de imágenes depende de la diversidad y calidad de sus datos de entrenamiento. Los modelos entrenados en conjuntos de datos que presentan principalmente imágenes idealizadas de alta resolución pueden tener dificultades al encontrar variaciones de menor calidad o del mundo real.

Para mitigar esto, los ingenieros curan diversos conjuntos de datos que representan las condiciones del mundo real. Algunas técnicas, como el aprendizaje por transferencia, permiten que los modelos empleen conocimientos previamente entrenados a partir de conjuntos de datos grandes y sólidos, y mejoren el rendimiento incluso con datos limitados.

Tamaño y proximidad del objeto

El tamaño de un objeto en una imagen, influenciado por su proximidad a la cámara, puede afectar la capacidad del modelo para identificarlo con precisión. Los objetos pequeños pueden no tener suficientes detalles para el reconocimiento, mientras que los objetos demasiado cercanos pueden parecer distorsionados o demasiado grandes para que el modelo se clasifique correctamente.

Los ingenieros entrenan modelos en conjuntos de datos que incluyen objetos de diferentes tamaños y distancias para manejar esto. También se emplean técnicas de procesamiento de imágenes multiescala y pirámides de características para ayudar a garantizar que el modelo pueda manejar objetos en una amplia gama de tamaños.

Reconocimiento de imágenes y detección de objetos

[La detección de objetos](https://www.ibm.com/mx-es/topics/object-detection" \t "_self) amplía el reconocimiento de imágenes mediante la identificación de objetos y la localización de sus ubicaciones dentro de una imagen. Esta técnica permite que el sistema responda preguntas como "¿Dónde está el gato en esta imagen?" o "¿Cuántos gatos hay en la escena?" La detección de objetos proporciona más contexto, ya que reconoce los objetos y sus posiciones, tamaños y orientaciones dentro del marco.

Por ejemplo, en lugar de identificar "un gato" en una imagen, la detección de objetos permite a la computadora especificar: "Hay un gato sentado en el sofá de la esquina izquierda de la imagen", lo que proporciona una comprensión espacial de la escena y de las relaciones entre objetos.

Las tareas de reconocimiento de imágenes también pueden variar en complejidad. La categorización o clasificación de imágenes asigna una sola etiqueta a toda una imagen en función de su contenido, respondiendo a la pregunta “¿Qué hay en esta imagen?”

Por ejemplo, un modelo entrenado con conjuntos de datos etiquetados de gatos y perros aprende a distinguir entre ambos identificando sus características únicas. Cuando se le presenta una nueva imagen, el modelo analiza estas características para predecir si muestra un gato o un perro.

Los modelos utilizan cuadros delimitadores para delinear estos objetos individuales, separándolos del fondo y marcando dónde comienza y termina cada objeto. Esta precisión es crítica para aplicaciones, como la conducción autónoma, donde la detección precisa de objetos, como vehículos, peatones y señales de tráfico, es esencial para la seguridad.



Las últimas novedades e insights sobre IA

Descubra insights y noticias de expertos sobre IA, la nube y mucho más en el boletín semanal Think.

[Suscríbase hoy](https://www.ibm.com/account/reg/signup?formid=news-urx-52954)

Usos en evolución del reconocimiento de imágenes

El reconocimiento de imágenes avanza con rapidez y abre el camino a aplicaciones más sofisticadas en numerosas industrias y casos de uso. Estas son algunas aplicaciones dominantes del reconocimiento de imágenes en el mundo real:

Vehículos autónomos

Muchos teléfonos inteligentes están equipados con tecnología de reconocimiento facial que permite a los usuarios desbloquear sus dispositivos mirando la pantalla. Esta aplicación de reconocimiento de imágenes se ha vuelto común, con sistemas que reconocen rasgos faciales individuales para verificar la identidad.

Reconocimiento facial

El reconocimiento facial también se utiliza ampliamente en seguridad y vigilancia para identificar personas a partir de transmisiones de video. Esta tecnología ayuda a los organismos encargados de hacer cumplir la ley a rastrear a los sospechosos en áreas públicas, mientras que las empresas la utilizan en la seguridad de los edificios para controlar el acceso.

Gestión y moderación de redes sociales

Las plataformas de redes sociales utilizan el reconocimiento de imágenes para sugerir etiquetas en las fotos, identificando y reconociendo los rostros de amigos y familiares. Las redes sociales también utilizan filtros de RA que detectan puntos de referencia faciales para colocar elementos virtuales, como gafas u orejas de animales, de forma que se alineen con los movimientos faciales.

Además, estas plataformas utilizan el reconocimiento de imágenes para moderar el contenido filtrando imágenes inapropiadas, manteniendo la seguridad de la plataforma y mejorando la experiencia del usuario.

Gafas inteligentes e información en tiempo real

Sobre la base de las actuales aplicaciones de [realidad aumentada](https://www.ibm.com/mx-es/topics/augmented-reality) (RA) en dispositivos móviles, unas gafas inteligentes equipadas con software de reconocimiento de imágenes pueden ofrecer a los usuarios vistas aumentadas de su entorno, superponiendo información en tiempo real sobre objetos y ubicaciones.

La tecnología de RA proporciona datos contextuales sobre cualquier cosa que el usuario mire, desde identificar puntos de referencia hasta recuperar detalles de productos en las tiendas.

Electrodomésticos

El reconocimiento de imágenes en electrodomésticos permite funciones, como el seguimiento de inventario en refrigeradores inteligentes, la detección de obstáculos en aspiradoras robóticas y el reconocimiento de personas u objetos en cámaras de seguridad.

También impulsa funciones, como la detección del tipo de tejido en lavadoras, el reconocimiento de alimentos en hornos inteligentes y el análisis facial en espejos inteligentes o monitores de bebés.

Los robots de reparto se basan en el reconocimiento de imágenes para navegar por entornos, detectar obstáculos e identificar ubicaciones de reparto para realizar entregas autónomas precisas y eficientes.

En contraste, los robots en almacenes y entornos industriales utilizan la misma tecnología para escanear y recuperar artículos, realizar controles de calidad, ensamblar piezas y clasificar materiales.

Imágenes médicas

El análisis de imágenes médicas ayuda a los profesionales de la salud a analizar radiografías, resonancias magnéticas y tomografías computarizadas. Estos sistemas pueden detectar anomalías que el ojo humano podría pasar por alto, como signos tempranos de cáncer de pulmón, accidentes cerebrovasculares o tumores cerebrales, lo que lleva a diagnósticos más oportunos.

Merative, anteriormente IBM® Watson Health, aplica el reconocimiento de imágenes para analizar datos de imágenes complejos, ayudando a los radiólogos a identificar hallazgos críticos.

El reconocimiento de imágenes médicas está avanzando con diagnósticos impulsados por IA, por lo que los sistemas de reconocimiento de imágenes pueden ayudar a detectar enfermedades en etapa temprana con mayor precisión.

La tecnología, que ya mejora áreas como la detección de tumores, apoya a los especialistas con un "segundo par de ojos" altamente entrenado para diagnósticos avanzados, particularmente en áreas donde los detalles minuciosos son críticos.

Reconocimiento óptico de caracteres (OCR)

La tecnología [OCR](https://www.ibm.com/mx-es/think/topics/optical-character-recognition) digitaliza el texto impreso mediante el escaneo de documentos, libros y recibos. Las aplicaciones utilizan OCR para reconocer y convertir texto impreso en formatos digitales que los usuarios pueden editar o buscar. El OCR fue uno de los primeros casos de uso críticos para el reconocimiento de imágenes, lo que ayudó a sentar las bases para la digitalización generalizada en todas las industrias.

Procesamiento de documentos

Los bancos y las instituciones financieras emplean el reconocimiento de imágenes para automatizar controles de verificación, identificaciones y otros documentos, reduciendo el fraude y agilizando la incorporación de clientes. La tecnología escanea imágenes de documentos en busca de detalles cruciales, los autentica y marca cualquier anomalía para su revisión.