



Facultad de Ingeniería
Escuela de Ingeniería de Sistemas y Computación

Mauricio Gaona
mauricio.gaona@correounivalle.edu.co

Profesor

2025-I

Agenda

- Introducción a la Visión Computacional
- Librerías - Servicios - Plataformas



Propósito

- Comprender los conceptos fundamentales de la visión computacional.
- Identificar las diferentes categorías y aplicaciones de la visión computacional.
- Mencionar los sectores en los que se aplica la visión computacional
- Identificar las Librerías, plataformas o servicios para desarrollar aplicaciones que integran visión computacional
- Ideas para realizar el taller 3.





¿Qué es la visión computacional?



Visión Artificial

Es un campo de la AI y la ciencia de la computación que se enfoca en entrenar a aplicaciones para interpretar y comprender contenido visual.



Visión Artificial



Es una disciplina que busca desarrollar técnicas para permitir a los computadores 'ver' y entender el contenido de las imágenes digitales o videos.



Visión Artificial

Se ocupa del análisis automático de las imágenes del mundo real para obtener información, tomar decisiones o realizar acciones



Visión Artificial



Combina aspectos de la percepción visual, el procesamiento de imágenes, el aprendizaje automático y la inteligencia artificial para procesar, analizar e interpretar datos visuales en un formato que sea comprensible y utilizable por las aplicaciones.

La Visión Computacional permite a las aplicaciones interpretar el mundo visual a través de modelos que analizan imágenes y videos.



Visión Artificial

La Visión artificial es un campo de la informática que se centra en permitir que los equipos identifiquen y comprendan objetos y personas en imágenes y videos.

Al igual que otros tipos de inteligencia artificial, el objetivo de la visión artificial es realizar y automatizar tareas que replican las capacidades humanas.

En este caso, la visión artificial busca replicar la forma en que los humanos ven y la forma en que los humanos tienen sentido de lo que ven.



Visión Artificial

Cómo se analiza una imagen con visión artificial

- Un sensor captura una imagen. A menudo, este sensor es simplemente una cámara, pero puede ser una cámara de video, un dispositivo médico de diagnóstico por imagen o cualquier otro tipo de dispositivo que capture una imagen para analizarla.
- A continuación, se envía la imagen a un dispositivo de interpretación. El dispositivo de interpretación usa el reconocimiento de patrones para descomponer la imagen, comparar los patrones que contiene con su biblioteca de patrones conocidos y determinar si hay contenido en la imagen que coincida. El patrón puede ser algo general, como la apariencia de un determinado tipo de objeto, o podría basarse en identificadores únicos, como los rasgos faciales.
- Un usuario solicita información específica sobre una imagen y el dispositivo de interpretación proporciona la información solicitada en función del análisis que ha realizado de la imagen.



Visión Artificial

Cómo se analiza una imagen con visión artificial



Visión Artificial

Aprendizaje profundo y Visión artificial

Las aplicaciones modernas de visión computacional están dejando de usar métodos estadísticos para analizar las imágenes y, cada vez más, se basan en lo que se conoce como aprendizaje profundo.

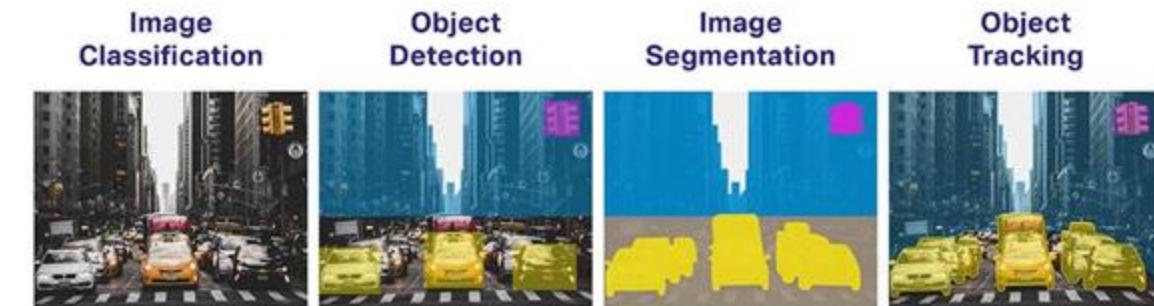
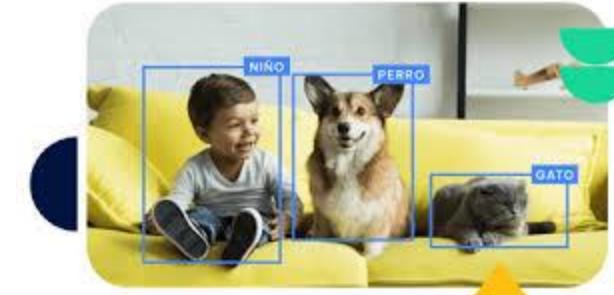
Con el aprendizaje profundo (Deep learning), utiliza redes neuronales artificiales con múltiples capas (redes profundas) y una aplicación de visión artificial ejecuta un tipo de algoritmo en una red neuronal, que le permite proporcionar análisis más precisos de las imágenes.

Además, el aprendizaje profundo permite a un programa de visión computacional conservar la información de cada imagen que analiza, de modo que, cuanto más se usa, mejor es su precisión.



Categorías

- Detección y Reconocimiento de Objetos
- Segmentación de Imágenes
- Reconocimiento de Patrones y Clasificación
- Estimación de Movimiento y Seguimiento de Objetos
- Reconstrucción 3D
- Análisis de Escenas y Contexto
- Procesamiento de Imágenes y Mejora



Fundamentos

El procesamiento de imágenes con IA consiste en analizar y transformar imágenes digitales usando algoritmos de aprendizaje profundo. A partir de los valores de los píxeles, una red neuronal convolucional (CNN) extrae automáticamente patrones y características relevantes. El proceso comienza con el preprocesamiento de la imagen, como normalización o reducción de ruido, seguido del paso por múltiples capas que resaltan información cada vez más compleja. Así, la IA interpreta el contenido visual sin necesidad de intervención manual directa.

Durante el entrenamiento, la IA aprende de grandes cantidades de imágenes etiquetadas, ajustando sus parámetros para mejorar su precisión. Una vez entrenada, puede realizar tareas como clasificación, detección, segmentación o generación de nuevas imágenes. Cada predicción se basa en las representaciones internas que la red ha construido a partir de los datos. De este modo, la IA convierte patrones numéricos en información útil para resolver problemas del mundo real.



Segmentos



- Automotriz y Transporte
- Salud y Medicina
- Seguridad y Vigilancia
- Retail y Comercio Electrónico
- Agricultura y Medio Ambiente
- Manufactura y Automatización Industrial
- Entretenimiento y Medios de Comunicación
- Deportes y Fitness
- Ciencia e Investigación
- Robótica y Drones
- Educación y Capacitación
- Banca y Finanzas



Segmentos - Automotriz y Transporte



En el desarrollo de vehículos autónomos y sistemas de asistencia al conductor, como el reconocimiento de señales de tráfico, detección de peatones y otros vehículos, y la navegación



Segmentos - Salud y Medicina



Para análisis de imágenes médicas, como la detección de tumores en radiografías o resonancias magnéticas y análisis de tejidos



Segmentos - Seguridad y Vigilancia



Utilizada en el reconocimiento facial, análisis de comportamiento anómalo, y monitoreo de áreas públicas o privadas para garantizar la seguridad.



Segmentos - Retail y Comercio Electrónico



En aplicaciones como el análisis de comportamiento del consumidor, gestión de inventarios mediante reconocimiento de productos (códigos), y experiencias de compra interactivas.



Segmentos - Agricultura y Medio Ambiente



Para la detección de plagas, análisis de cultivos, monitoreo de la salud de las plantas, y gestión de recursos naturales.



Segmentos - Manufactura y Automatización Industrial



En la inspección de calidad, control de procesos, automatización de tareas, y en el mantenimiento predictivo de maquinaria.



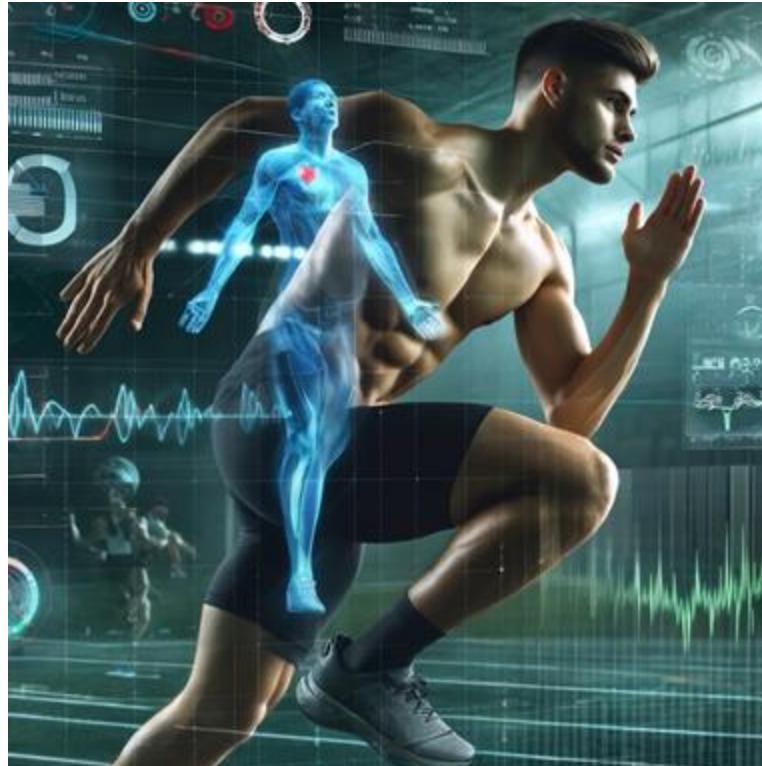
Segmentos - Entretenimiento y Medios de Comunicación



En efectos visuales, realidad aumentada y virtual, y en la creación de experiencias interactivas y personalizadas para el usuario.



Segmentos - Deportes y Fitness



Para el análisis de rendimiento, seguimiento de movimientos, y en aplicaciones de entrenamiento personalizado.



Segmentos - Robótica y Drones



En la navegación, mapeo de terrenos, y en la realización de tareas específicas en entornos complejos o peligrosos



Segmentos - Educación y Capacitación



Como herramienta para la creación de contenido educativo interactivo, calificación de exámenes (texto, pdf, imágenes) y en sistemas de enseñanza virtual



Segmentos - Banca y Finanzas



Para la verificación de identidad, prevención de fraudes, y en el análisis de documentos financieros.



Algunas conclusiones

- La visión computacional ha revolucionado múltiples sectores, desde la robótica hasta la medicina.
- Las CNNs y otros modelos de IA son fundamentales para implementar aplicaciones de visión computacional.
- A medida que la tecnología avanza, las capacidades de los modelos para interpretar el mundo visual se volverán cada vez más precisas.



<https://ai.google.dev/gemini-api/docs/files?hl=es-419#prompt-guide>

Ideas: Modelo multimodal de Gemini

Instrucción	Respuesta
<p>Escribe una entrada de blog breve y atractiva basada en esta foto. Debe incluir una descripción de la comida que aparece en la foto y hablar sobre mi proceso de preparación de comidas.</p> 	<p>Preparar las comidas es una excelente manera de ahorrar tiempo y dinero, y también puede ayudarte a comer de forma más saludable. Esta comida es un excelente ejemplo de una comida saludable y deliciosa que se puede preparar con anticipación con facilidad.</p> <p>Esta comida incluye arroz integral, verduras asadas y pollo teriyaki. El arroz integral es un carbohidrato saludable y complejo que te brindará energía sostenida. Las verduras asadas son una excelente manera de obtener tu dosis diaria de vitaminas y minerales, y el pollo teriyaki es una opción deliciosa y rica en proteínas.</p> <p>Esta comida también es muy fácil de preparar. Solo tienes que cocinar el arroz integral, asar los vegetales y cocinar el pollo teriyaki. Una vez que todo esté cocido, divídalo en recipientes para preparar comidas y guárdalos en el refrigerador. Luego, puedes tomar un contenedor y salir en las mañanas ocupadas.</p> <p>Si buscas una comida saludable y deliciosa que se pueda preparar con facilidad con anticipación, esta es una excelente opción. Está repleto de nutrientes y sabor, y te mantendrá satisfecho.</p> <p>¡Salud y delicias para la preparación de comidas!</p>

<https://platform.openai.com/docs/guides/image-generation?image-generation-model=gpt-image-1>

Ideas: Modelo multimodal de openAI

Generación de imágenes: En este ejemplo, se usan 4 imágenes de entrada para generar una nueva imagen de una canasta de regalo que contenga los elementos de las imágenes de referencia.





Ideas: Verificación de identidad

Suplantación de identidad: la foto tomada de la persona coincide con la de la cédula



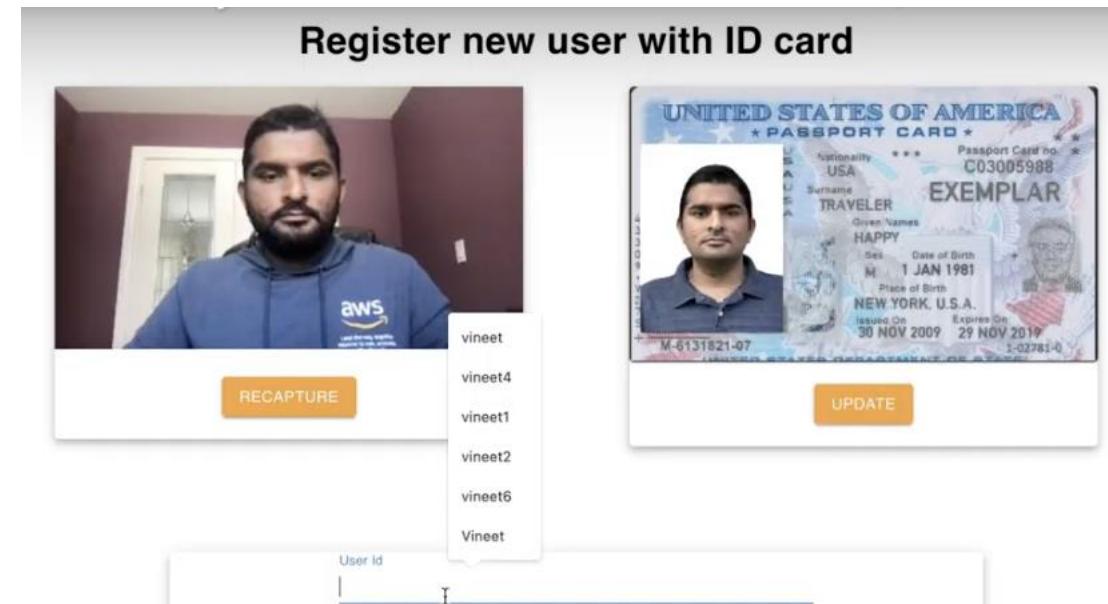
Se podría usar la huella de una persona !





Ideas: Verificación de identidad

Login usando una foto



Ideas: Generación de interfaces con Python

Tools For Building AI Chat UIs in Python (Gradio)

<https://www.gradio.app>

<https://medium.com/@amosgyamfi/the-3-best-python-frameworks-to-build-uis-for-ai-apps-08619d53d281>



The screenshot shows a Gradio AI Chat interface. At the top left is a "Chatbot" button. To its right is a search bar containing the text "que es inteligencia artificial". On the far right is a "Configuración del Modelo" sidebar with sections for "Model" (set to "gemini-2.0-flash-thinking-exp-01-21") and "Mensajes del sistema" (containing the message "Eres un asistente de AI útil..."). The main content area has a dark background. It displays a text block about Artificial Intelligence (IA) and a numbered list of IA capabilities. Below this is a text input field with placeholder "Enviar un mensaje...", and three buttons: "Enviar" (with a rocket icon), "Limpiar" (with a checkmark icon), and "Guardar Chat" (with a disk icon). At the bottom left is a "File" button, and at the bottom center is a "Usar vía API" link.

Librerías, Servicios y Plataformas

1. Gradio (UI con Python)
2. Amazon Rekognition (Amazon Web Services)
3. Google Cloud Vision (Google Cloud Platform)
4. Azure Computer Vision (Microsoft Azure)
5. IBM Watson Visual Recognition (IBM Cloud)
6. OpenCV (Open Source)
7. Roboflow
8. OpenAI
9. Ximiliar
10. [opcional] Google API de Cloud Healthcare



Ideas y ejemplos de visión computacional

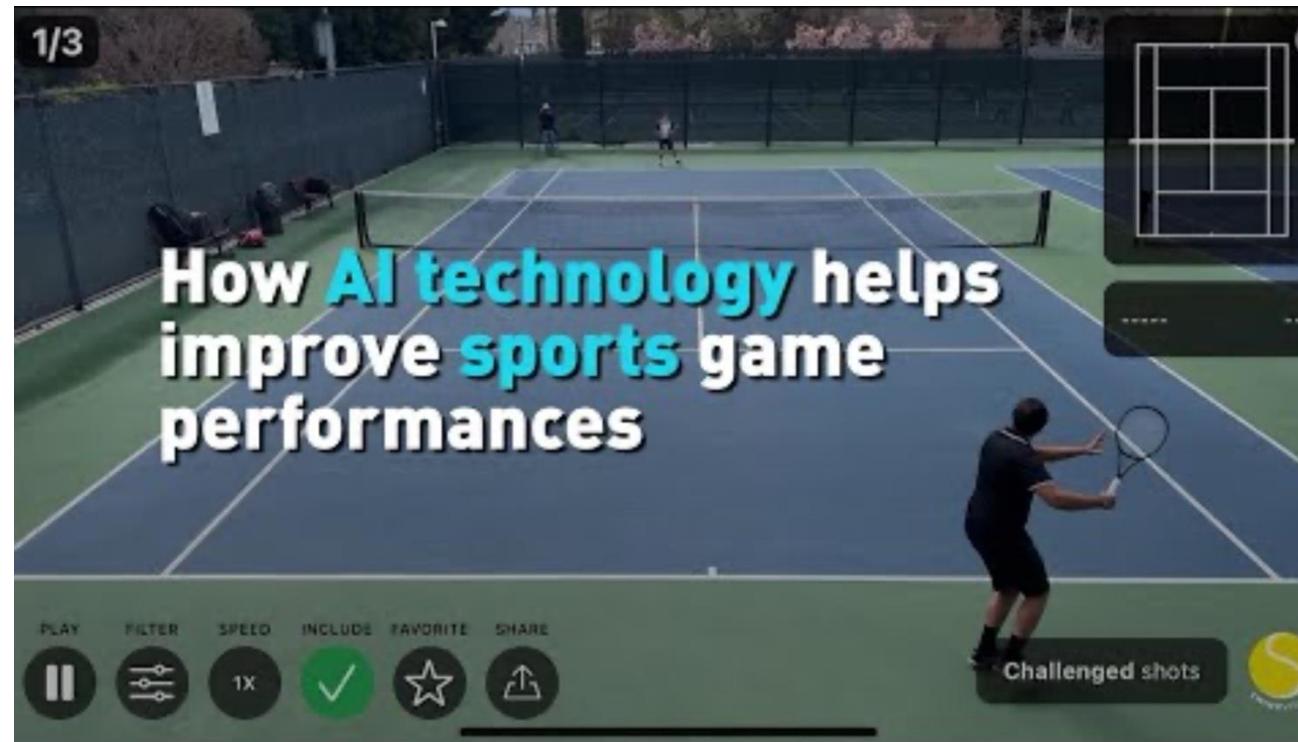
<https://www.youtube.com/watch?v=tGdHi4NIXG4>



Otras fuentes recomendadas: Ideas y ejemplos de visión computacional

Análisis del juego de tenis:

<https://youtu.be/B99egqUoP1s?t=119>



Otras fuentes recomendadas: Ideas y ejemplos de visión computacional

Control de Aforo con Vision Computacional

<https://www.youtube.com/watch?v=tGdHi4NIXG4>

Tecnología que permite a las máquinas ver y guiar a los humanos en la resolución de problemas técnicos complejos.

<https://techsee.com/computer-vision/>



Reconocimiento Facial | Python – OpenCV

<https://www.youtube.com/watch?v=cZkpaL36fW4>

DETECTOR DE OBJETOS con Haar Cascade | Python y OpenCV

<https://amin-ahmadi.com/cascade-trainer-gui/>

https://www.youtube.com/watch?v=v_cwOq06g9E

OpenMV se trata de crear módulos de visión artificial extensibles y extensibles de bajo costo y con Python y tiene como objetivo convertirse en el "Arduino de la Visión Artificial".

https://openmv.io/?gad_source=1&gbraid=0AAAAADOSICrthpPU38YsJtHI1ALzDSvYm&gclid=Cj0KCQjwsJO4BhDoARIsADDv4vBL-FYIXAeM8Oe7v00NMICejkvXJSTRM-I6pUhxaJXK1s63re3NHdAaAs6PEALw_wcB



Asignación de Servicio



Librerías, Servicios y Plataformas

1. Gradio (UI con Python)	Grupo 3	Mayo 6
2. Amazon Rekognition (Amazon Web Services)	Grupo 7	Mayo 6
3. Google Cloud Vision (Google Cloud Platform)	Grupo 6	Mayo 6
4. Azure Computer Vision (Microsoft Azure)	Grupo 9	Mayo 6
5. IBM Watson Visual Recognition (IBM Cloud)	Grupo 8	Mayo 7
6. OpenCV (Open Source)	Grupo 4	Mayo 7
7. Roboflow	Grupo 5	Mayo 7
8. OpenAI	Grupo 1	Mayo 7
9. Ximiliar	Grupo 2	Mayo 7
10.[opcional] Google API de Cloud Healthcare		

Taller 3



¿Preguntas?



Proyecto Integrador II



Gracias

