Visión artificial

51 idiomas

* [Artículo](https://es.wikipedia.org/wiki/Visi%C3%B3n_artificial)
* [Discusión](https://es.wikipedia.org/wiki/Discusi%C3%B3n:Visi%C3%B3n_artificial)
* [Leer](https://es.wikipedia.org/wiki/Visi%C3%B3n_artificial)
* [Editar](https://es.wikipedia.org/w/index.php?title=Visi%C3%B3n_artificial&action=edit)
* [Ver historial](https://es.wikipedia.org/w/index.php?title=Visi%C3%B3n_artificial&action=history)

Herramientas

La **visión informática**,[[1]](https://es.wikipedia.org/wiki/Visi%C3%B3n_artificial#cite_note-1)​ también conocida como **visión artificial** o **visión por computadora** (del [inglés](https://es.wikipedia.org/wiki/Idioma_ingl%C3%A9s) *computer vision*) o **visión técnica**, es una disciplina científica que incluye métodos para adquirir, procesar, analizar y comprender las imágenes del mundo real con el fin de producir información numérica o simbólica para que puedan ser tratados por un ordenador. Tal y como los seres humanos usamos nuestros ojos y cerebros para comprender el mundo que nos rodea, la visión informática trata de producir el mismo efecto para que los ordenadores puedan percibir y comprender una imagen o secuencia de imágenes y actuar según convenga en una determinada situación. Esta comprensión se consigue gracias a distintos campos como la geometría, la estadística, la física y otras disciplinas. La adquisición de los datos se consigue por varios medios como secuencias de imágenes, vistas desde varias cámaras de video o datos multidimensionales desde un escáner médico.

[Diagrama

El contenido generado por IA puede ser incorrecto.](https://commons.wikimedia.org/wiki/File:CVoverview2-ES.jpg)Esquema de las relaciones entre la visión por ordenador y otras áreas afines.

Hay muchas tecnologías que utilizan la visión por ordenador, entre las cuales se encuentran el [reconocimiento de objetos](https://es.wikipedia.org/wiki/Reconocimiento_de_objetos), la detección de sucesos, la reconstrucción de una escena (mapping) y la [restauración de imágenes](https://es.wikipedia.org/wiki/Restauraci%C3%B3n_de_imagen).

El objetivo final de la visión informática es conseguir el desarrollo de estrategias automáticas para el [reconocimiento de patrones](https://es.wikipedia.org/wiki/Reconocimiento_de_patrones) complejos en imágenes de múltiples dominios. En la actualidad, muchos son los campos que se han visto beneficiados por este conjunto de técnicas. Uno de los más conocidos es el de la robótica, ya que los robots con cierta autonomía deben reconocer con precisión la localización de los objetos de su entorno para no colisionar contra ellos, por ejemplo. A menudo, esto lo consiguen por medio de sensores o de cámaras, siendo estos últimos dispositivos idóneos para la aplicación de las estrategias de visión por ordenador.

Sin embargo, la robótica no es el único ámbito que se ha visto beneficiado por este conjunto de técnicas. Podemos destacar el ámbito de la [imagen médica](https://es.wikipedia.org/wiki/Imagen_m%C3%A9dica), con sistemas capaces de reconocer, por ejemplo, patrones patológicos en una modalidad de imagen determinada y diagnosticar enfermedades de forma automatizada. También se emplean en otros ámbitos, como en sistemas de seguridad, seguimiento de objetos (por ejemplo, seguimiento de un futbolista en vídeo durante un partido de fútbol) o detección de anomalías en piezas fabricadas en una cadena de producción, esto último como método de control de calidad.

**Interferencias técnicas y optimizaciones**

[[editar](https://es.wikipedia.org/w/index.php?title=Visi%C3%B3n_artificial&action=edit&section=1)]

A la hora de aplicar los conceptos teóricos de la visión por ordenador encontraremos siempre interferencias y problemas relacionados con el mundo que nos rodea. Esto es por el mero hecho de que nuestro mundo no es perfecto y los aparatos de medición y captura tampoco lo son. Estos introducen siempre (a mayor o menos cantidad) una distorsión o ruido que contamina la muestra o imagen con la que deseamos trabajar.

Teniendo en cuenta estos problemas denominamos los siguientes tipos de [ruidos](https://es.wikipedia.org/wiki/Ruido_en_la_fotograf%C3%ADa_digital) (entre otros) como ruidos técnicos:

**Ruidos técnicos**

[[editar](https://es.wikipedia.org/w/index.php?title=Visi%C3%B3n_artificial&action=edit&section=2)]

**Salt and pepper**

[[editar](https://es.wikipedia.org/w/index.php?title=Visi%C3%B3n_artificial&action=edit&section=3)]

Ruido impulsivo que hace que los píxeles afectados tomen un valor extremo, es decir, máximo (blanco) o bien mínimo (negro). El efecto de este ruido en una imagen en blanco y negro, o escala de grises, es tener diversos puntos blancos y negros esparcidos aleatoriamente por la imagen. De ahí el nombre Salt and Pepper (sal y pimienta), debido a que parece que la imagen haya sido rociada por estos compuestos. Este ruido puede aparecer a causa de los canales de transmisión de las imágenes. Para solventar este ruido podremos utilizar (pese a perder definición) un filtro de promedio espacial. No es muy confiable, ya que te pueden espiar.

**Ruido uniforme**

[[editar](https://es.wikipedia.org/w/index.php?title=Visi%C3%B3n_artificial&action=edit&section=4)]

El valor original del píxel distorsionado es sustituido por otro siguiendo una distribución uniforme en el intervalo de valores posibles, esto es, desde el blanco al negro. El efecto a simple vista de este ruido es percibir interferencias en la imagen, como si esta estuviese codificada. Observamos una pantalla por encima de la imagen llena de pixeles de valores aleatorios y uniformemente expandidos. Este ruido puede aparecer en el procesado de cuantificación de una imagen. Para solventar este ruido podremos utilizar (pese a perder definición) un filtro de promedio espacial.

**Ruido Gaussiano**

[[editar](https://es.wikipedia.org/w/index.php?title=Visi%C3%B3n_artificial&action=edit&section=5)]

Ruido derivado de los equipos de captura que sigue la fórmula (falta fórmula). El efecto en la imagen será parecido al uniforme solo que los valores del ruido no son tan abruptos, tenderán más a grises que a negros y blancos. Para solventar el problema podríamos utilizar un filtro de promedio espacial con coeficientes Gaussianos.

En todos los casos, para solventar cualquier tipo de ruido deberemos aplicar algún tipo de [filtro](https://es.wikipedia.org/wiki/Procesamiento_digital_de_im%C3%A1genes) compatible con nuestro ruido.

Existen otro tipo de interferencias debidas al contexto e interpretación de la imagen. Pasamos a enumerar algunas:

**Interferencias debidas al contexto**

[[editar](https://es.wikipedia.org/w/index.php?title=Visi%C3%B3n_artificial&action=edit&section=6)]

**Punto de vista**

[[editar](https://es.wikipedia.org/w/index.php?title=Visi%C3%B3n_artificial&action=edit&section=7)]

Cómo está la imagen orientada respecto al observador. Un animal no se ve igual de frente que de espaldas, aunque este sigue siendo un animal. Iluminación: La cantidad de luz que recibe el objeto. Un objeto deberá ser distinguible independientemente de si tiene alguna cara oscura debido a la iluminación.

**Oclusión**

[[editar](https://es.wikipedia.org/w/index.php?title=Visi%C3%B3n_artificial&action=edit&section=8)]

Un objeto puede estar en segundo plano, es decir, que otro objeto tape parcialmente nuestro objetivo a la hora de extraerlo o analizarlo. Deberemos ser capaces de saber interpretar que un objeto diferente está entre nuestro objeto a extraer y el observador.

**Escala**

[[editar](https://es.wikipedia.org/w/index.php?title=Visi%C3%B3n_artificial&action=edit&section=9)]

Factor que determina el tamaño de la imagen respecto al real. Un edificio no parecerá tener el mismo tamaño dependiendo de la imagen capturada (ya sea por distancia, ángulo…). Deberemos ser capaces de interpretar que un objeto puede ser el mismo pese a que el tamaño en las fotografías pueda ser diferente.

**Deformación**

[[editar](https://es.wikipedia.org/w/index.php?title=Visi%C3%B3n_artificial&action=edit&section=10)]

Un objeto puede estar deformado debido a múltiples factores (véase el ejemplo de la carretera en pleno verano) o simplemente debido a errores de captura, o posicionamiento y ángulo de la captura. Deberemos interpretar que el objeto pertenece a una categoría pese a sus deformaciones.

**Fondo desordenado**

[[editar](https://es.wikipedia.org/w/index.php?title=Visi%C3%B3n_artificial&action=edit&section=11)]

Un objeto puede estar en un contexto desordenado y caótico. Como por ejemplo un mosaico. Deberemos saber distinguir el objeto entre el caos que le envuelve.

**Variaciones dentro de una misma clase**

[[editar](https://es.wikipedia.org/w/index.php?title=Visi%C3%B3n_artificial&action=edit&section=12)]

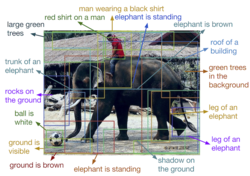
Un tipo de objeto puede ser muy dispar a otro de su misma categoría. Si tomamos el ejemplo de una silla observamos que hay multitud de sillas diferentes pero a todas les unen los mismos rasgos característicos: 4 patas, una placa donde sentarse, un reposo para la espalda…

Dependiendo de cómo solucionemos estos problemas (qué tipo de algoritmos y procesos aplicamos) nuestro programa/aplicación será más o menos eficiente y más o menos fiable.

**Aprendizaje automático**

[[editar](https://es.wikipedia.org/w/index.php?title=Visi%C3%B3n_artificial&action=edit&section=13)]

*Artículo principal:*[Aprendizaje automático](https://es.wikipedia.org/wiki/Aprendizaje_autom%C3%A1tico)

[](https://commons.wikimedia.org/wiki/File:DenseCap_(Johnson_et_al.,_2016)_(cropped).png)Image Captioningと物体認識を組み合わせた手法であるDenseCap (Johnson et al., 2016)というモデルの出力例

Las técnicas de [aprendizaje automático](https://es.wikipedia.org/wiki/Aprendizaje_autom%C3%A1tico) tienen como objetivo conseguir diferenciar automáticamente patrones usando algoritmos matemáticos. Estas técnicas son comúnmente usadas para clasificar imágenes, para tomar decisiones dentro del mundo empresarial (por ejemplo, para decidir qué clientes de un banco pueden recibir un préstamo o cuánto ha de pagar cada cliente por un seguro dependiendo de sus antecedentes), así como dentro de muchos otros ámbitos de la ciencia y la tecnología. Principalmente se pueden distinguir dos tipos de técnicas: supervisadas y no supervisadas.

En el aprendizaje supervisado se entrena al ordenador proporcionando patrones previamente etiquetados, de forma que algoritmo usado debe encontrar las fronteras que separan los posibles diferentes tipos de patrones. Adaboost y algunas redes neuronales forman parte de este grupo.

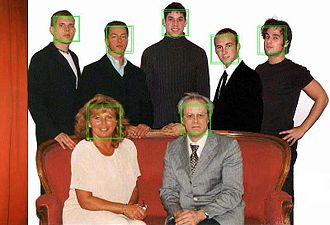
En el aprendizaje no supervisado se entrena al ordenador con patrones que no han sido previamente clasificados y es el propio ordenador el que debe agrupar los distintos patrones en diferentes clases. [K-means](https://es.wikipedia.org/wiki/K-means) y algunas [redes neuronales](https://es.wikipedia.org/wiki/Red_neuronal_artificial) forman parte de este grupo.

Ambas técnicas son muy utilizadas en la visión informática, sobre todo en clasificación y segmentación de imágenes.[[2]](https://es.wikipedia.org/wiki/Visi%C3%B3n_artificial#cite_note-2)​

**Detección de objetos**

[[editar](https://es.wikipedia.org/w/index.php?title=Visi%C3%B3n_artificial&action=edit&section=14)]

*Artículo principal:*[Reconocimiento de objetos](https://es.wikipedia.org/wiki/Reconocimiento_de_objetos)

[](https://commons.wikimedia.org/wiki/File:Face_detection_example_openCV.jpg)Detección de caras con el software [OpenCV](https://es.wikipedia.org/wiki/OpenCV" \o "OpenCV)

La detección de objetos es la parte de la visión informática que estudia cómo detectar la presencia de objetos en una imagen sobre la base de su apariencia visual, bien sea atendiendo al tipo de objeto (una persona, un coche) o a la instancia del objeto (mi coche, el coche del vecino). Generalmente se pueden distinguir dos partes en el proceso de detección: la extracción de características del contenido de una imagen y la búsqueda de objetos basada en dichas características.

La extracción de características consiste en la obtención de modelos matemáticos compactos que "resuman" el contenido de la imagen con el fin de simplificar el proceso de aprendizaje de los objetos a reconocer. Dichas características son comúnmente llamadas descriptores.

Existen diversos tipos de descriptores, que tendrán mejor o peor rendimiento en función al tipo de objeto a reconocer y a las condiciones del proceso de reconocimiento (la luz controlada o no, distancia al objeto a reconocer conocida o no). Se pueden usar desde básicos [histogramas](https://es.wikipedia.org/wiki/Histograma) de color o intensidad de luz, descriptores LBP (Local Binary Pattern, usado sobre todo para texturas) o más avanzados como el HOG (Histogram of Oriented Gradientes) o [SIFT](https://es.wikipedia.org/wiki/Scale-invariant_feature_transform).

Para el proceso de clasificación se pueden usar diferentes técnicas de aprendizaje máquina. Existen diferentes métodos, como la [regresión logística](https://es.wikipedia.org/wiki/Regresi%C3%B3n_log%C3%ADstica), o más avanzados basados en técnicas de aprendizaje automático como el [SVM](https://es.wikipedia.org/wiki/M%C3%A1quinas_de_vectores_de_soporte) o AdaBoost (Adaptative Boost).

Los mayores retos tanto de la extracción de características como la clasificación es encontrar descriptores y [clasificadores](https://es.wikipedia.org/wiki/Clasificador_(matem%C3%A1ticas)) que sean invariantes a los cambios que pueda tener un objeto, como su posición o iluminación.[[3]](https://es.wikipedia.org/wiki/Visi%C3%B3n_artificial#cite_note-3)​[[4]](https://es.wikipedia.org/wiki/Visi%C3%B3n_artificial#cite_note-4)​

**Análisis de video**

[[editar](https://es.wikipedia.org/w/index.php?title=Visi%C3%B3n_artificial&action=edit&section=15)]

[](https://commons.wikimedia.org/wiki/File:Computer_vision_sample_in_Sim%C3%B3n_Bolivar_Avenue,_Quito.jpg)Detección de vehículos en movimiento en una Avenida de Quito, Ecuador

El término análisis de vídeo describe un amplio número de nuevas tecnologías y evoluciones en el campo de la vigilancia con vídeo y la seguridad. Estos cambios están produciendo sistemas de seguridad más efectivos y eficientes. El análisis de vídeo es fundamental en el sector de vídeo vigilancia y seguridad. En un sistema de CCTV([circuito cerrado de televisión](https://es.wikipedia.org/wiki/Circuito_cerrado_de_televisi%C3%B3n" \o "Circuito cerrado de televisión)) tradicional es habitual visualizar el contenido de hasta 16 cámaras simultáneamente. Esta tarea resulta complicada para un vigilante de seguridad pues hay estudios que aseguran que después de 22 minutos de supervisión este pierde hasta el 95 por ciento de la actividad de la escena. Con el análisis de vídeo se alerta al vigilante cuando hay movimiento o señala en qué cámara hay mayor probabilidad de actividad sospechosa o peligrosa.

Una de las capacidades básicas del análisis de vídeo es la detección de movimiento: tecnología que identifica y alerta cuando ocurre el movimiento. Sofisticadas adaptaciones de la detección de movimiento incluyen sensores que detectan el movimiento en direcciones no autorizadas. Véase [Detector de movimiento](https://es.wikipedia.org/wiki/Detector_de_movimiento).

Algunas aplicaciones del análisis de vídeo son la detección de objetos abandonados en lugares llenos de gente, controlar obras de arte en los museos y detectar vehículos no autorizados que ingresen a determinadas áreas. La detección de matrículas de vehículos y congestión de tránsito son características que se están desenvolviendo.

A pesar de la constante evolución del análisis de vídeo, algunas de sus aplicaciones no son muy rigurosas. Tecnologías como el reconocimiento facial y la detección de movimientos sospechosos todavía no son fiables y a menudo producen falsas alarmas. Véase [Sistema de reconocimiento facial](https://es.wikipedia.org/wiki/Sistema_de_reconocimiento_facial).

**Visión 3D**

[[editar](https://es.wikipedia.org/w/index.php?title=Visi%C3%B3n_artificial&action=edit&section=16)]

La visión 3D informática se encarga de proporcionar la capacidad de emular la visión humana a un ordenador. Con dicha capacidad el ordenador podrá generar un modelo tridimensional de un objeto o escena, generalmente a partir de una imagen en 2D. Existen técnicas o sistemas que permiten captar la profundidad de los objetos o en una escena, como por ejemplo: [sistemas estereoscópicos](https://es.wikipedia.org/wiki/Estereoscop%C3%ADa), mediante múltiples cámaras, sistemas basados en el tiempo de vuelo o los sistemas basados en el escáner de luz estructurada.

Las aplicaciones son numerosas, en el campo de la automoción por ejemplo, [Google self-Driving car](https://www.google.com/selfdrivingcar) utiliza la detección de objetos, junto con radares y sensores para conducir por la vía pública de una forma autónoma.

**La reconstrucción en 3D a partir de imagen 2D**

[[editar](https://es.wikipedia.org/w/index.php?title=Visi%C3%B3n_artificial&action=edit&section=17)]

A partir de una imagen 2D obtener una reconstrucción en 3D plantea una serie de problemas ya que existen muchos objetos con formas similares o estos mismos objetos pueden aparecer en la imagen de forma diferente. Hay que limitar el proceso de reconstrucción asumiendo la similitud de las formas entre los objetos. Para comparar la imagen de consulta se deben crear una serie de modelos 3D para cada clase. Se usan estos modelos para generar una base de datos de objetos de la clase con sus respectivos mapas de profundidad, a esto lo llamaremos un aprendizaje supervisado porque se conocerán todos los objetos a tratar. Esto nos dará ejemplos de mapas 3D factibles asociados a formas que se utilizarán para estimar la imagen.

Un ejemplo de los bloques que debe contener un sistema de reconocimiento 3D a partir de una imagen 2D:

* **Adquisición de la imagen:** En esta etapa se captura una proyección en dos dimensiones de la luz reflejada por los objetos de la escena.
* [**Segmentación**](https://es.wikipedia.org/wiki/Segmentaci%C3%B3n_(procesamiento_de_im%C3%A1genes))**:** detección de bordes y regiones: Permite separar los diferentes objetos de la escena.
* **Extracción de características**: Se obtiene una representación numérica en forma de vector por cada imagen.
* **Reconocimiento y localización:** Mediante técnicas, como pueda ser la triangulación, se localiza al objeto en el espacio 3D.
* **Interpretación o estimación:** A partir de la información obtenida se estima la escena.

**La reconstrucción en 3D a partir de múltiples imágenes 2D**

[[editar](https://es.wikipedia.org/w/index.php?title=Visi%C3%B3n_artificial&action=edit&section=18)]

El objetivo de la reconstrucción 3D con múltiples imágenes es averiguar la geométrica de una escena capturada por una colección de imágenes. Por lo general, la posición de la cámara y los parámetros internos se supone que se conocen o se pueden estimar a partir del conjunto de imágenes. La automática correspondencia de los objetos de la imagen suele ser ambigua e incompleta, se recomienda tener un conocimiento previo sobre los objetos.

[Diagrama

El contenido generado por IA puede ser incorrecto.](https://commons.wikimedia.org/wiki/File:Multicamera.PNG)Modo multicámara

**Técnicas de obtención de imágenes 3D**

[[editar](https://es.wikipedia.org/w/index.php?title=Visi%C3%B3n_artificial&action=edit&section=19)]

[](https://commons.wikimedia.org/wiki/File:FedStereo_stereoscopic_camera_anaglyph.jpg)Cámara estereoscópica FedStereo hecha en Ucrania.

* [**Multicámara**](https://es.wikipedia.org/wiki/Multic%C3%A1mara)**:** El objetivo de esta técnica es captar el objeto o la escena con varias cámaras calibradas para obtener diferentes puntos de vista y así generar datos de profundidad.
* **Cámaras esteroscópicas:** Permiten emular la visión humana creando dos imágenes como si fueran los ojos. Véase [Cámaras éstereoscópicas.](https://es.wikipedia.org/wiki/C%C3%A1mara_estereosc%C3%B3pica)
* [**Escáner de luz estructurada**](https://es.wikipedia.org/wiki/Esc%C3%A1ner_de_luz_estructurada)**:** Mediante una proyección de un patrón de luz, el dispositivo es capaz de capturar la forma y las características del objeto. Véase [Escáner de luz estructurada](https://es.wikipedia.org/wiki/Esc%C3%A1ner_de_luz_estructurada).
* [**Time of Flight (TOF)**](https://es.wikipedia.org/wiki/Televisi%C3%B3n_3D#Time-of-Flight_.28TOF.29) : Mediante un sensor que emite una señal de infrarrojo, esta señal incide sobre la escena o objeto y vuelve rebotada sobre la cámara. La cámara genera una imagen en escala de grises que nos da la información de profundidad. Véase [Televisión 3D: TOF](https://es.wikipedia.org/wiki/Televisi%C3%B3n_3D#Time-of-Flight_.28TOF.29).

[](https://commons.wikimedia.org/wiki/File:Kinect_Sensor_at_E3_2010_(front).jpg)Kinect Sensor en la exposición E3 de 2010 (frente de la cámara).

**Herramientas de desarrollo**

[[editar](https://es.wikipedia.org/w/index.php?title=Visi%C3%B3n_artificial&action=edit&section=20)]

[](https://commons.wikimedia.org/wiki/File:TOF_Kamera.jpg)TOF Cámara

En la actualidad, una gran parte de los algoritmos más conocidos de [visión informática](https://aprendiendoinformatica.net/vision-artificial/) ya han sido implementados en librerías específicas para tal fin, es decir, que generalmente no es necesario desarrollarlos. De este modo, un desarrollador de sistemas de visión informática se puede centrar en explotar las funcionalidades que le ofrecen esas librerías para resolver sus propios problemas de la forma más idónea y que me mejor se adapte a cada caso.

Una de las librerías más conocidas es [OpenCV](https://opencv.org/), gratuita y libre, disponible para lenguajes tan populares como Python o C++. Esta librería proporciona las [herramientas fundamentales para la lectura y el guardado de imágenes](https://aprendiendoinformatica.net/opencv/primeros-pasos/) por parte de un código de programación, funciones básicas para mejorar la calidad de las imágenes, algunos métodos de segmentación como [la umbralización de imágenes](https://aprendiendoinformatica.net/opencv/umbralizacion/) e incluso una pequeña gama de funciones de aprendizaje máquina, aunque esto se escapa un poco más de la idea principal de la librería.

Cabe destacar que también existe una implementación destinada a los dispositivos [Android](https://es.wikipedia.org/wiki/Android), cuyas funcionalidades se pueden explotar desde cualquier código implementado en [Android Studio](https://es.wikipedia.org/wiki/Android_Studio), haciendo las configuraciones oportunas.

También existen otras librerías de Python como [Scikit-Image](https://scikit-image.org/) o [Scikit-Learn](https://scikit-learn.org/stable/), que son de gran importancia en el ámbito del procesamiento de imagen y el aprendizaje automático. En general, se observa que Python es un lenguaje muy utilizado en este ámbito de la visión por ordenador.

Por último, cabe destacar la existencia de otras librerías especializadas en Aprendizaje Profundo (conocido en inglés por el término [Deep Learning](https://es.wikipedia.org/wiki/Aprendizaje_profundo)[[5]](https://es.wikipedia.org/wiki/Visi%C3%B3n_artificial#cite_note-5)​) como es el caso de [TensorFlow](https://www.tensorflow.org/?hl=es-419) o Torch (siendo [PyTorch](https://pytorch.org/) su implementación específica en Python). Así mismo, estas últimas librerías soportan la aceleración por hardware proporcionada por las tarjetas gráficas [NVIDIA](https://www.nvidia.com/es-es/) por medio de la librería CUDA.

Dado que los procesadores gráficos de las tarjetas gráficas se adaptan muy bien al cálculo de operaciones matriciales (aspecto característico del aprendizaje máquina, siendo todavía más crítico en el aprendizaje profundo), en la actualidad esta se ha convertido en la forma más común de trabajar con modelos profundos, de ahí el gran éxito de la librería [CUDA](https://es.wikipedia.org/wiki/CUDA).