DOCUMENTO TÉCNICO

La IA en la analítica de vídeo

Consideraciones para una analítica basada en aprendizaje automático y aprendizaje profundo

Marzo 2021

AXIS COMMUNICATIONS

Índice

1 **Resumen** 3

2 **Introducción** 4

3 **IA, aprendizaje automático y aprendizaje profundo** 4

- 3.1 Aprendizaje automático 5
- 3.2 Aprendizaje profundo 6
- 3.3 ¿Aprendizaje automático clásico o aprendizaje profundo? 7

4 **Fases del aprendizaje automático** 7

- 4.1 Recopilación y etiquetado de datos 7
- 4.2 Entrenamiento 8
- 4.3 Pruebas 9
- 4.4 Implementación 9

5 **Analítica en local** 9

6 **Aceleración de hardware** 10

7 **La IA todavía está en su desarrollo inicial** 10

8 **Consideraciones para un rendimiento óptimo de la analítica** 11

- 8.1 Posibilidades de uso de la imagen 11
- 8.2 Distancia de detección 12
- 8.3 Configuración de alarmas y grabaciones 12
- 8.4 Mantenimiento 13

9 **Privacidad e integridad personal** 14

10 **Apéndice** 15

- 10.1 Redes neuronales 15
- 10.2 Redes neuronales convolucionales (CNN) 16

1 Resumen

La analítica de vídeo basada en IA es uno de los temas más controvertidos en el sector de la videovigilancia. Algunas de las aplicaciones pueden acelerar considerablemente el análisis de datos y automatizar tareas repetitivas. Sin embargo, las soluciones de IA actuales no pueden sustituir la experiencia ni la capacidad de decisión de un operador humano. El auténtico potencial se encuentra en la combinación: aprovechar las ventajas de las soluciones de IA para ayudar al operador humano a trabajar mejor.

El concepto de IA engloba los algoritmos de aprendizaje automático y de aprendizaje profundo. Los dos tipos construyen automáticamente un modelo matemático utilizando grandes cantidades de datos de muestra, o *datos de entrenamiento*, para aprender a calcular los resultados sin necesidad de una programación específica. Un algoritmo de IA se desarrolla a partir de un proceso iterativo, en el que se repite un ciclo de recopilación de datos de entrenamiento, etiquetado de datos de entrenamiento, aplicación de datos etiquetados para entrenar el algoritmo y prueba del algoritmo entrenado hasta alcanzar el nivel de calidad deseado. Después de esta fase, el algoritmo ya está listo para usarse en una aplicación de analítica, que puede adquirirse e instalarse en un entorno de vigilancia. En este punto, el entrenamiento ha finalizado y la aplicación ya no aprende nada más.

Una tarea habitual de la analítica de vídeo basada en IA es la detección visual de humanos y vehículos en una transmisión de vídeo y la diferenciación entre unos y otros. Un algoritmo de *aprendizaje automático* ha aprendido la combinación de características visuales que define estos objetos. Un algoritmo de *aprendizaje profundo* es más sofisticado y, con el entrenamiento adecuado, puede detectar objetos mucho más complejos. Sin embargo, también requiere un esfuerzo de desarrollo considerablemente superior y más recursos de cálculo cuando se utiliza la aplicación terminada. Por tanto, en una situación de vigilancia bien definida, merece la pena valorar si una aplicación de aprendizaje automático optimizado y especializada puede ser suficiente.

Los avances en los algoritmos y el aumento de la potencia de procesamiento de las cámaras han permitido ejecutar analítica de vídeo basada en IA directamente en la cámara (local) sin tener que recurrir a un servidor para realizar los cálculos. De este modo, la funcionalidad en tiempo real es mejor, ya que las aplicaciones tienen acceso inmediato a material de vídeo sin comprimir. Si las cámaras cuentan con aceleradores de hardware dedicados, como una MLPU (unidad de procesamiento de aprendizaje automático) o una DLPU (unidad de procesamiento de aprendizaje profundo), la analítica local puede funcionar con un consumo inferior al que requeriría una CPU o una GPU (unidad de procesamiento gráfico).

Antes de poder instalar una aplicación de analítica de vídeo basada en IA, es importante analizar y seguir detenidamente las recomendaciones del fabricante, basadas en las condiciones previas y limitaciones conocidas. Cada instalación de vigilancia es única y el rendimiento de la aplicación debe evaluarse caso por caso. Si se observa que la calidad es inferior a lo previsto, deben analizarse las causas de forma global, sin poner el foco únicamente en la aplicación de analítica en sí. El rendimiento de la analítica de vídeo depende de muchos factores vinculados al hardware de la cámara, la configuración de la cámara, la calidad del vídeo, la dinámica de la escena y la iluminación. En muchos casos, si conocemos el impacto de estos factores y realizamos los ajustes correspondientes conseguiremos mejorar el rendimiento de la analítica de vídeo en una instalación concreta.

Ante la presencia cada vez mayor de la IA en la vigilancia, es obligado valorar si las ventajas en cuanto a la eficiencia operativa y nuevos escenarios de uso justifican la aplicación de la tecnología y en qué condiciones.

2 Introducción

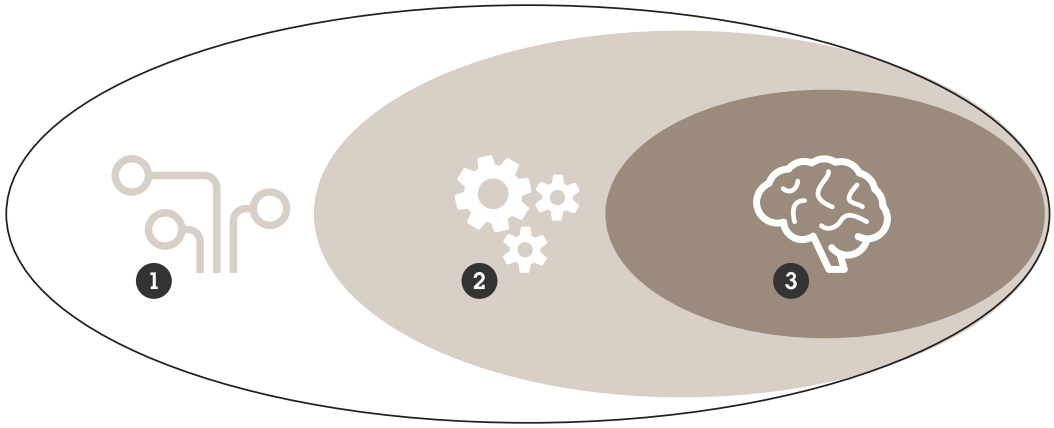
La IA, o inteligencia artificial, ha sido objeto de numerosos avances y encendidos debates desde la aparición de los primeros ordenadores. Aunque sus plasmaciones más revolucionarias todavía están por llegar, las tecnologías basadas en IA se utilizan actualmente para desempeñar tareas claramente definidas en aplicaciones como el reconocimiento de voz, los motores de búsqueda o los asistentes virtuales. La IA se utiliza también y cada vez más en el sector sanitario, por ejemplo en diagnósticos de radiografías o en exámenes oculares.

La analítica de vídeo basada en IA es uno de los temas más controvertidos en el sector de la videovigilancia y también una de las tecnologías en la que hay más expectativas depositadas. En el mercado hay aplicaciones que utilizan algoritmos de IA para acelerar el análisis de datos y automatizar tareas repetitivas. Sin embargo, en el contexto más amplio de la vigilancia, la IA debe entenderse hoy y en el futuro más inmediato como un recurso más en el proceso de diseño de unas soluciones eficaces.

Este documento técnico describe el contexto tecnológico de los algoritmos de aprendizaje automático y aprendizaje profundo y de sus posibilidades de desarrollo y aplicación en el terreno de la analítica de vídeo. Esta descripción incluye un breve repaso del hardware de aceleración basado en IA y también las ventajas e inconvenientes de la analítica basada en IA local en comparación con las técnicas basadas en servidores. El documento analiza también cómo pueden optimizarse los requisitos de rendimiento de la analítica basada en IA, teniendo en cuenta un amplio abanico de factores.

3 IA, aprendizaje automático y aprendizaje profundo

La inteligencia artificial (IA) es un concepto muy amplio asociado a máquinas capaces de resolver tareas complejas gracias a unas características que les dotan de una cierta inteligencia. El aprendizaje profundo y el aprendizaje automático son disciplinas vinculadas a la IA.



1 *Inteligencia artificial*
2 *Aprendizaje automático*
3 *Aprendizaje profundo*

3.1 Aprendizaje automático

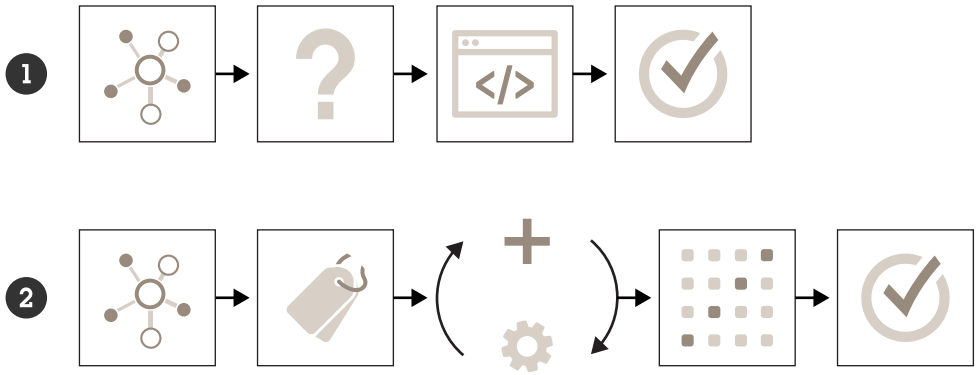
El aprendizaje automático es una disciplina de la IA que utiliza algoritmos de aprendizaje estadísticos para construir sistemas capaces de aprender automáticamente y mejorar gracias al entrenamiento sin necesidad de una programación específica.

En este apartado, hacemos una distinción entre la programación tradicional y el aprendizaje automático en el contexto de la *visión artificial*, esto es, la disciplina mediante la cual los ordenadores pueden llegar a entender qué sucede en una escena a partir del análisis de imágenes o vídeos.

La visión artificial programada tradicional se basa en métodos que calculan las *características* de una imagen, por ejemplo programas informáticos que buscan bordes pronunciados o esquinas. En este caso, un desarrollador de algoritmos que sepa cuáles son los elementos más importantes de los datos de imagen deberá definir manualmente estas características. A continuación, el desarrollador combina estas características para que el algoritmo pueda determinar qué hay en la escena.

Los algoritmos de aprendizaje automático construyen automáticamente un modelo matemático utilizando grandes cantidades de datos de muestra, o *datos de entrenamiento*, para aprender a tomar decisiones calculando resultados sin necesidad de una programación específica. Las características todavía deben definirse manualmente, pero el algoritmo aprende a combinarlas gracias a la exposición a un gran volumen de datos de entrenamiento *etiquetados*. En este documento, para hablar de esta técnica basada en el uso de características definidas manualmente en combinaciones aprendidas utilizamos el término *aprendizaje automático clásico*.

En otras palabras: en una aplicación de aprendizaje automático tenemos que entrenar al ordenador para conseguir el programa que queremos. Un humano recopila y etiqueta los datos, en ocasiones con la ayuda de ordenadores servidores, que se ocupan del preetiquetado. El resultado se introduce en el sistema y el proceso continúa hasta que la aplicación ha aprendido lo suficiente para detectar lo que queremos, por ejemplo un tipo de vehículo concreto. Entonces, el modelo entrenado se convierte en el programa. Sin embargo, una vez terminado el programa el sistema ya no aprende nada más.



1 *Programación tradicional:*
Los datos se recopilan. Se definen los criterios del programa. El programa se codifica (por un humano). Listo.
2 *Aprendizaje automático:*
Los datos se recopilan. Los datos se etiquetan. El modelo se somete a un proceso de entrenamiento iterativo. Entonces, el modelo entrenado finalizado se convierte en el programa. Listo.

En comparación con la programación tradicional, la ventaja de la IA al diseñar un programa de visión artificial es la capacidad de procesar un gran volumen de datos. Un ordenador puede procesar miles de

imágenes sin perder la concentración, mientras que un programador humano terminará por cansarse y desconcentrarse. Gracias a esta ventaja, la aplicación desarrollada utilizando la IA puede ser mucho más precisa. Sin embargo, cuanto más compleja es la aplicación, más difícil es para la máquina obtener el resultado deseado.

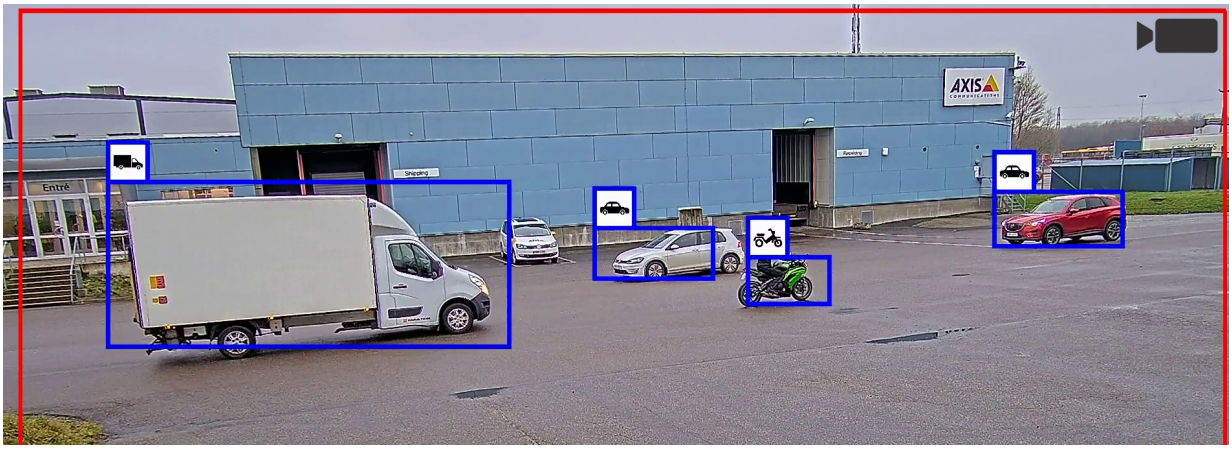
3.2 Aprendizaje profundo

El aprendizaje profundo es una versión perfeccionada del aprendizaje automático, en la que tanto la extracción de características como su combinación en estructuras de reglas complejas para obtener unos resultados se aprenden a partir de un modelo basado en datos. El algoritmo puede definir automáticamente las características que debe buscar en los datos de entrenamiento. También puede aprender estructuras muy profundas de combinaciones de características encadenadas.

La base del funcionamiento de los algoritmos empleados en el aprendizaje profundo es el funcionamiento de las neuronas y su utilización por parte del cerebro para obtener unos conocimientos de nivel superior, combinando la información generada por las neuronas en una jerarquía profunda, o una *red*, de reglas encadenadas. El cerebro es un sistema cuyas combinaciones son producto del trabajo de las neuronas, lo que elimina la distinción entre la extracción de características y su combinación, y las sitúa en un mismo plano. Los investigadores han simulado estas estructuras en las denominadas *redes neuronales artificiales*, que constituyen el tipo de algoritmo utilizado más a menudo en el aprendizaje profundo. Consulte el apéndice de este documento para ver un breve resumen del funcionamiento de las redes neuronales.

Utilizando algoritmos de aprendizaje profundo es posible construir detectores visuales sofisticados y entrenarlos automáticamente para detectar objetos muy complejos, independientemente de la escala, la rotación y otras variaciones.

El motivo de esta flexibilidad es que los sistemas de aprendizaje profundo pueden aprender a partir de un mayor volumen de datos más variados en comparación con los sistemas de aprendizaje automático clásicos. En la mayoría de los casos, superan considerablemente a los algoritmos de visión artificial diseñados por humanos. Por este motivo, el aprendizaje profundo resulta especialmente indicado para problemas complejos, en los que los humanos no pueden formar fácilmente combinaciones de características, como la clasificación de imágenes, el procesamiento del lenguaje o la detección de objetos.



La detección de objetos basada en el aprendizaje profundo permite clasificar objetos complejos. En este ejemplo, la aplicación de analítica no solo puede detectar vehículos, sino también clasificar el tipo de vehículo.

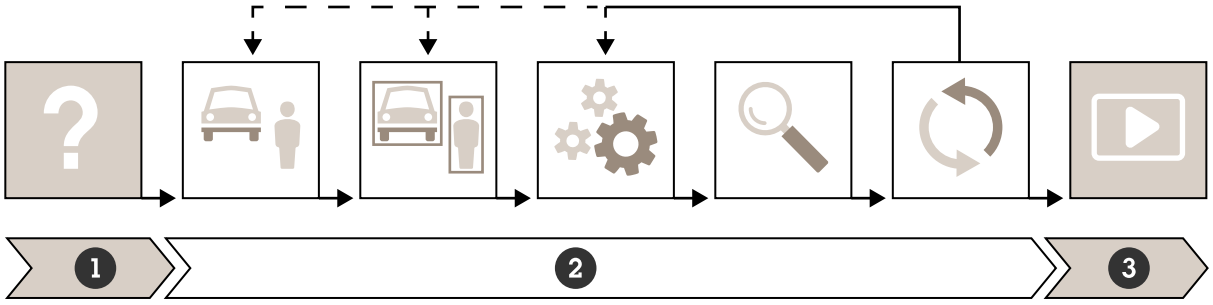
3.3 ¿Aprendizaje automático clásico o aprendizaje profundo?

Aunque los algoritmos son similares, un algoritmo de aprendizaje profundo normalmente utiliza un conjunto mayor de combinaciones de características aprendidas que un algoritmo de aprendizaje automático. Por este motivo, la analítica basada en aprendizaje profundo puede ser más flexible y, con el entrenamiento adecuado, puede aprender a realizar tareas mucho más complejas.

Sin embargo, en la analítica especializada en vigilancia, un algoritmo de aprendizaje automático clásico optimizado puede ser suficiente. Si el alcance está bien definido, los resultados pueden ser similares a los obtenidos por un algoritmo de aprendizaje profundo, pero con un menor volumen de operaciones matemáticas y, por tanto, con un coste y un consumo de energía inferiores. Además, requiere muchos menos datos de entrenamiento, lo que se traduce en un esfuerzo de desarrollo inferior.

4 Fases del aprendizaje automático

El desarrollo de un algoritmo de aprendizaje automático sigue una secuencia de pasos e iteraciones, resumidos sucintamente a continuación, antes de poder implantar una aplicación de analítica terminada. La base de una aplicación de analítica es un algoritmo o varios, por ejemplo un detector de objetos. En el caso de aplicaciones basadas en aprendizaje profundo, la base del algoritmo es el modelo de aprendizaje profundo.



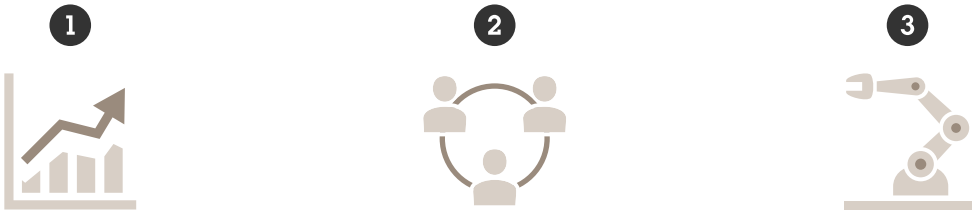
1. *Preparación: definición de la finalidad de la aplicación.*
2. *Entrenamiento: obtención de datos de entrenamiento. Etiquetado de los datos. Entrenamiento del modelo. Prueba del modelo. Si la calidad no es la esperada, los pasos anteriores se repiten en un ciclo de mejora iterativo.*
3. *Implementación: instalación y utilización de la aplicación terminada.*

4.1 Recopilación y etiquetado de datos

Para desarrollar una aplicación de analítica basada en IA es necesario recopilar un gran volumen de datos. En videovigilancia, normalmente estos datos son imágenes y fragmentos de vídeo de personas y vehículos u otros objetos de interés. Para que los datos sean reconocibles por una máquina o un ordenador hace falta etiquetarlos, para clasificar los objetos relevantes en categorías. El etiquetado de datos es una tarea eminentemente manual y extremadamente laboriosa. Los datos preparados deben cubrir una variedad de muestras lo suficientemente amplia para que puedan ser aplicables al contexto en el que se utilizará la aplicación de analítica.

4.2 Entrenamiento

El entrenamiento, o aprendizaje, es cuando se introducen datos etiquetados en el modelo y se aplica un entorno de entrenamiento para modificar y mejorar el modelo en un ciclo iterativo hasta alcanzar la calidad deseada. En otras palabras: el modelo se optimiza para resolver la tarea definida. El entrenamiento puede realizarse principalmente de tres formas.



1 *Aprendizaje supervisado: el modelo aprende a realizar predicciones precisas*
2 *Aprendizaje no supervisado: el modelo aprende a identificar clusters*
3 *Aprendizaje de refuerzo: el modelo aprende de los errores*

4.2.1 Aprendizaje supervisado

El aprendizaje supervisado es el método más empleado en el aprendizaje automático actual. Puede describirse como un aprendizaje basado en ejemplos. Los datos de entrenamiento están claramente etiquetados, lo que implica que los datos de entrada ya están asociados a un resultado de salida deseado.

El aprendizaje supervisado por lo general requiere un gran volumen de datos etiquetados y el rendimiento del algoritmo entrenado depende directamente de la calidad de los datos de entrenamiento. En el terreno cualitativo, lo más importante es utilizar un conjunto de datos que represente todos los datos de entrada potenciales que pueden producirse en una situación de uso real. En el caso de detectores de objetos, el desarrollador debe asegurarse de entrenar el algoritmo con una amplia variedad de imágenes, con diferentes instancias de objetos, orientaciones, escalas, situaciones de iluminación, fondos y distracciones. Solo si los datos de entrenamiento son representativos de la situación de uso prevista la aplicación de analítica final podrá realizar unas predicciones precisas al procesar nuevos datos no manejados durante la fase de entrenamiento.

4.2.2 Aprendizaje no supervisado

El aprendizaje no supervisado utiliza algoritmos para analizar y agrupar conjuntos de datos no etiquetados. No es un método de entrenamiento habitual en el sector de la vigilancia, porque es un modelo que requiere un gran esfuerzo de calibración y prueba, sin garantías de obtención de unos resultados predecibles.

Los conjuntos de datos deben ser relevantes para la aplicación de analítica, pero no hace falta que estén etiquetados o marcados claramente. El trabajo de etiquetado manual desaparece de la ecuación, pero el número de imágenes o vídeos necesarios para el entrenamiento aumenta considerablemente. Durante la fase de entrenamiento, el modelo identifica, con la ayuda del entorno de entrenamiento, características comunes en los conjuntos de datos. De este modo, durante la fase de desarrollo puede agrupar los datos a partir de patrones y, al mismo tiempo, detectar anomalías que no pueden asignarse a ninguno de los grupos aprendidos.

4.2.3 Aprendizaje de refuerzo

El aprendizaje de refuerzo se utiliza en ámbitos como la robótica, la automatización industrial o la planificación de estrategias de negocio, pero a causa de la gran cantidad de feedback necesario, es un

método poco utilizado actualmente en contextos de vigilancia. El aprendizaje de refuerzo se basa en adoptar medidas adecuadas para maximizar la *recompensa* potencial en una situación específica, una recompensa que es mayor cuando el modelo adopta las decisiones correctas. El algoritmo no utiliza pares de datos/etiquetas para el entrenamiento, sino que se optimiza probando sus decisiones a través de la interacción con el entorno y midiendo la recompensa. El objetivo del algoritmo es aprender una política de acciones que ayuden a maximizar la recompensa.

4.3 Pruebas

Una vez entrenado el modelo, debe someterse a unas pruebas exhaustivas. Este paso implica normalmente una parte automatizada, complementada con unas exhaustivas pruebas en situaciones de uso reales.

En la parte automatizada, se evalúa la aplicación utilizando nuevos conjuntos de datos, no procesados por el modelo durante su entrenamiento. Si la evaluación no es satisfactoria, el proceso vuelve a empezar: se recopilan nuevos datos de entrenamiento, se añaden etiquetas o se mejoran y se reentrena el modelo.

Una vez alcanzado el nivel de calidad deseado, empieza una prueba sobre el terreno. En esta prueba, la aplicación se expone a escenarios reales. La magnitud de la prueba y su nivel de variación dependen del alcance de la aplicación. Cuanto más reducido sea el alcance, menos serán las variaciones que habrá que probar. Cuanto mayor sea el alcance, más pruebas serán necesarias.

Los resultados se comparan y evalúan de nuevo. Este paso puede provocar que el proceso vuelva a empezar de nuevo. Otro posible resultado sería la definición de precondiciones, para explicar un escenario conocido en el que no se recomienda utilizar la aplicación o se recomienda solo en algunos casos.

4.4 Implementación

La fase de implementación se conoce también como fase de inferencia o predicción. La *inferencia* o *predicción* es el proceso de ejecución de un modelo de aprendizaje automático entrenado. El algoritmo utiliza lo aprendido durante la fase de entrenamiento para producir el resultado deseado. En el contexto de la analítica para vigilancia, la fase de inferencia es la aplicación ejecutada en un sistema de vigilancia para supervisar escenas reales.

Para garantizar la funcionalidad en tiempo real al ejecutar un algoritmo basado en aprendizaje automático en datos de entrada de audio o vídeo, por lo general hace falta una aceleración de hardware específica.

5 Analítica en local

Antiguamente la analítica de vídeo de alto rendimiento estaba basada siempre en servidores, ya que requería más energía y refrigeración de lo que podía ofrecer una cámara. Sin embargo, los avances en los algoritmos y el aumento de la potencia de procesamiento de los dispositivos locales en los últimos años han abierto la puerta a la posibilidad de ejecutar analítica de vídeo avanzada basada en IA de forma local.

Las ventajas de las aplicaciones de analítica locales son evidentes, ya que tienen acceso a material de vídeo sin comprimir con una latencia mínima, lo que permite aplicaciones en tiempo real sin el coste adicional ni la complejidad asociados a la transferencia de datos a la nube para realizar los cálculos. Además, la analítica local también tiene unos costes de hardware e implantación inferiores, ya que el sistema de vigilancia no necesita tantos recursos de servidores.

En algunas aplicaciones, la combinación de un procesamiento local y basado en servidor puede resultar interesante, ya que una parte del procesamiento se realiza en la cámara y otra en el servidor. Un

sistema híbrido de este tipo facilita la ampliación de aplicaciones de analítica, ya que puede utilizarse en transmisiones de diferentes cámaras.

6 Aceleración de hardware

Aunque a menudo es posible ejecutar una aplicación de analítica concreta en varios tipos de plataformas, la aceleración de hardware dedicada permite obtener un rendimiento muy superior cuando la potencia es limitada. Con los aceleradores de hardware es posible implementar aplicaciones de analítica con un consumo inferior. En algunos casos, pueden complementarse con recursos de cálculo en servidores y en la nube.

- **GPU (unidad de procesamiento gráfico)**. Las GPU se desarrollaron principalmente para aplicaciones de procesamiento gráfico, pero también se utilizan para acelerar la IA en plataformas en la nube y servidores. Aunque a veces se utilizan también en sistemas integrados (local), las GPU no son un recurso óptimo, desde el punto de vista de la eficiencia energética, para tareas de inferencia de aprendizaje automático.
- **MLPU (unidad de procesamiento de aprendizaje automático)**. Una MLPU puede acelerar la inferencia en algoritmos de aprendizaje automático clásicos, para realizar tareas de visión artificial con un consumo muy reducido. Está diseñada para la detección en tiempo real de un número limitado de tipos de objetos simultáneos, por ejemplo humanos y vehículos.
- **DLPU (unidad de procesamiento de aprendizaje profundo)**. Las cámaras con una DLPU integrada pueden acelerar la inferencia en algoritmos generales de aprendizaje profundo con un consumo reducido, lo que permite una clasificación más selectiva de los objetos.

7 La IA todavía está en su desarrollo inicial

A menudo podemos tener la tentación de comparar el potencial de una solución de IA con lo que un humano es capaz de hacer. Mientras los operadores de videovigilancia humanos solo pueden mantener la atención durante un breve período de tiempo, un ordenador puede seguir procesando grandes cantidades de datos extremadamente rápido y sin cansarse. Sin embargo, pensar que las soluciones de IA podrán sustituir a los operadores humanos sería caer en un importante error conceptual. El auténtico potencial se encuentra en una combinación pragmática: aprovechar las ventajas de las soluciones de IA para ayudar al operador humano a trabajar mejor.

Las soluciones de aprendizaje automático o aprendizaje profundo a menudo se describen como soluciones capaces de aprender automáticamente o mejorar con la experiencia. Sin embargo, los sistemas de IA disponibles en la actualidad *no* aprenden nada nuevo después de su implantación y *no* recuerdan eventos específicos que se hayan producido. Para mejorar el rendimiento del sistema, tiene que volver a entrenarse con datos más precisos y relevantes durante sesiones de aprendizaje supervisado. Un aprendizaje sin supervisión normalmente requiere muchos datos para generar clusters, por lo que normalmente no se utiliza en aplicaciones de videovigilancia. En la actualidad, se utiliza principalmente para analizar grandes conjuntos de datos y detectar anomalías, por ejemplo en transacciones financieras. La mayoría de los sistemas descritos como modelos de "autoaprendizaje" en el terreno de la videovigilancia se basan en análisis de datos estadísticos y no en el reentrenamiento de modelos de aprendizaje profundo.

La experiencia humana sigue superando a la mayoría de las aplicaciones de analítica basada en IA con fines de vigilancia. Especialmente las aplicaciones pensadas para tareas muy generales, en las que la información contextual resulta fundamental. Una aplicación basada en aprendizaje automático puede detectar correctamente una persona a la fuga si ha recibido un entrenamiento específico, pero a diferencia

de un humano, capaz de situar los datos en contexto, la aplicación no sabe por qué la persona está corriendo, si para llegar al autobús o porque efectivamente está huyendo de un agente de policía. A pesar de las promesas de muchas empresas que aplican la IA en sus aplicaciones de analítica para vigilancia, la aplicación todavía no puede compararse a un humano a la hora de entender lo que ve en el vídeo.

Por el mismo motivo, las aplicaciones de analítica basadas en IA también pueden activar falsas alarmas o incluso pasarlas por alto. Es algo que puede suceder por ejemplo en entornos complejos con mucho movimiento. Sin embargo, también podría tener su origen en una persona que transporta un objeto muy grande, lo que impide a la aplicación reconocer sus características humanas y dificulta por tanto su correcta clasificación.

La analítica basada en IA debe utilizarse actualmente a modo de apoyo, por ejemplo para ayudar a determinar la gravedad de un incidente antes de alertar a un operador humano, que será quien determine la respuesta necesaria. Con este modelo, la IA se utiliza para ayudar a escalar las situaciones y es el operador humano el que finalmente evalúa los incidentes potenciales.

8 Consideraciones para un rendimiento óptimo de la analítica

Para que una aplicación de analítica basada en IA cumpla con las expectativas de calidad necesarias, se recomienda evaluar y entender las condiciones previas y limitaciones conocidas, que normalmente figuran en la documentación de la aplicación.

Cada instalación de vigilancia es única y el rendimiento de la aplicación debe evaluarse caso por caso. Si se observa que la calidad es inferior a lo previsto, es importante analizar las causas sin poner el foco únicamente en la aplicación de analítica en sí. Todo análisis debe partir de un enfoque integral, porque el rendimiento de la aplicación de analítica depende de numerosos factores, la mayoría de los cuales podemos optimizar si somos conscientes de su impacto. Estos factores pueden ser, por ejemplo, el hardware de la cámara, la calidad del vídeo, la dinámica de la escena, el nivel de iluminación o bien la configuración, la posición y la dirección de la cámara.

8.1 Posibilidades de uso de la imagen

La calidad de imagen suele depender de una alta resolución y una alta sensibilidad a la luz de la cámara. Aunque la importancia de estos factores no puede ponerse en duda, también hay otros que son igual de importantes de cara a la *usabilidad* real de una imagen o de un vídeo. Por ejemplo, la mejor transmisión de vídeo obtenida con la más cara de las cámaras de vigilancia no servirá de nada si falta iluminación en la escena por la noche, si se ha cambiado la orientación de la cámara o si falla la conexión con el sistema.

La colocación de una cámara debe analizarse detenidamente antes de su instalación. Para que la analítica de vídeo ofrezca los resultados esperados, la cámara debe estar colocada de modo que permita una visión despejada y sin obstáculos de la escena en cuestión.

La usabilidad de la imagen también puede depender de cada situación de uso concreta. Un vídeo puede parecer correcto a simple vista, pero tal vez no tenga la calidad óptima para que la aplicación de analítica de vídeo funcione bien. De hecho, muchos métodos de procesamiento de imagen utilizados habitualmente para mejorar el aspecto del vídeo a ojos de los humanos son poco recomendables cuando se utiliza la analítica de vídeo. Por ejemplo, los métodos de reducción del ruido, las técnicas de amplio rango dinámico o los algoritmos de exposición automática.

Las cámaras de vídeo actuales a menudo incorporan iluminación IR integrada, para poder funcionar en la más absoluta oscuridad. Se trata de una ventaja innegable, ya que permite instalar las cámaras en

lugares con problemas de iluminación sin necesidad de instalar soluciones de iluminación adicionales. Sin embargo, si existe el riesgo de precipitaciones intensas de lluvia o nieve en un lugar, es mejor no depender de la luz procedente de la cámara o de un punto muy cercano a la cámara. Un exceso de luz puede rebotar contra gotas de lluvia o copos de nieve y volver a la cámara, lo que impedirá utilizar la analítica. En cambio, con la luz ambiental es más probable que la analítica funcione correctamente, incluso con una climatología adversa.

8.2 Distancia de detección

Resulta complicado determinar una distancia de detección máxima en una aplicación de analítica basada en IA, ya que el valor exacto en metros de la ficha técnica no aporta toda la información necesaria. La calidad de imagen, las características de la escena, las condiciones meteorológicas o las propiedades de los objetos, como su color o brillo, influyen enormemente en la distancia de detección. Es evidente que, por ejemplo, un objeto muy brillante contra un fondo oscuro en un día soleado puede detectarse visualmente desde distancias muy superiores en comparación con un objeto oscuro en un día lluvioso.

La distancia de detección también depende de la velocidad de los objetos detectados. Para obtener unos resultados precisos, una aplicación de analítica de vídeo tiene que poder "ver" el objeto durante un período de tiempo suficientemente largo. La duración de este periodo depende de la capacidad de procesamiento (velocidad de fotogramas) de la plataforma: cuanto menor sea la capacidad de procesamiento, más tiempo tendrá que ser visible el objeto para poder detectarse. Si el tiempo de obturación de la cámara no se corresponde con la velocidad del objeto, la distorsión por movimiento visible en la imagen también puede reducir la precisión de detección.

Los objetos rápidos pueden pasar inadvertidos más fácilmente si pasan más cerca de la cámara. Una persona a la fuga situada a mucha distancia de la cámara, por ejemplo, tal vez se detecte correctamente, mientras que una persona situada muy cerca de la cámara a la misma velocidad puede entrar y salir tan deprisa del campo de visión que tal vez no genere ninguna alarma.

En una aplicación de analítica basada en movimiento, los objetos que se mueven en dirección a la cámara o en la dirección opuesta plantean otro desafío. La detección resultará especialmente difícil en el caso de objetos que se mueven despacio, que solo provocarán pequeños cambios en la imagen en comparación con el movimiento visible en la escena.

Una cámara con una resolución superior normalmente no ofrece una distancia de detección más grande. La potencia de procesamiento necesaria para ejecutar un algoritmo de aprendizaje automático es proporcional al tamaño de los datos de entrada. Por tanto, la potencia de procesamiento necesaria para analizar la resolución completa de una cámara 4K es por lo menos cuatro veces superior que la necesaria para una cámara a 1080p. Es muy habitual ejecutar aplicaciones basadas en IA a una resolución inferior a la que puede ofrecer la cámara o la transmisión, a causa de las limitaciones en la capacidad de procesamiento de la cámara.

8.3 Configuración de alarmas y grabaciones

Gracias a los diferentes niveles de filtros que aplica, la analítica de objetos genera muy pocas falsas alarmas. Sin embargo, la analítica de objetos solo funciona correctamente si se cumple la lista de condiciones previas definidas. De lo contrario, podría pasar por alto eventos importantes.

Si no tenemos claro que puedan cumplirse siempre todas las condiciones, se recomienda adoptar un planteamiento más conservador y configurar el sistema de modo que tenga algún otro activador de alarmas además de una clasificación de objetos concreta. De este modo se generarán más falsas alarmas, pero también se reducirá el riesgo de pasar por alto algo importante. Cuando las alarmas o los activadores están

vinculados directamente a un centro de control de alarmas, cada falsa alarma tiene un coste muy elevado. Por tanto, se hace evidente la necesidad de contar con una clasificación de objetos fiable que permita excluir las alarmas no deseadas. Sin embargo, la solución de grabación debe configurarse igualmente de modo que no dependa únicamente de la clasificación de objetos. Si se pasa por alto una alarma real, esta configuración permite evaluar a partir de la grabación el motivo de la omisión y mejorar a partir de aquí la instalación y la configuración.

Si la clasificación de objetos se realiza en el servidor durante la búsqueda de incidentes, se recomienda configurar el sistema para que grabe de forma continua y sin filtrar la grabación inicial. La grabación continua consume mucho almacenamiento, pero en cierto modo queda compensado con algoritmos de compresión modernos como Zipstream.

8.4 Mantenimiento

Una instalación de vigilancia debe ser objeto de un mantenimiento regular. Es recomendable realizar inspecciones físicas, y no únicamente visualizar el vídeo a través de la interfaz del VMS, para detectar y eliminar cualquier cosa que pueda obstruir o bloquear el campo de visión. También es importante en instalaciones estándar con solo grabación, pero lo es todavía más cuando se utiliza la analítica.

En el contexto de la detección de movimiento por vídeo básica, un obstáculo tan habitual como una telaraña movida por el viento puede aumentar el número de alarmas, lo que se traduciría en un espacio de almacenamiento mayor del necesario. Con la analítica de objetos, la telaraña crearía una zona de exclusión en la zona de detección. Sus hilos dificultarían la detección de los objetos y reducirían considerablemente las probabilidades de detección y clasificación.

Las telarañas pueden obstruir el campo de visión de una cámara de vigilancia.

La suciedad en el cristal frontal o la cúpula de una cámara difícilmente provocará problemas durante el día. Sin embargo, en condiciones de poca luz, cuando la luz impacta con una cúpula sucia por uno de sus laterales, por ejemplo la luz de los faros de un coche, pueden producirse reflejos inesperados que podrían reducir la precisión de la detección.

El mantenimiento asociado a la escena es tan importante como el mantenimiento de la cámara. Durante el ciclo de vida de una cámara, pueden pasar muchas cosas en la escena objeto de supervisión. Una simple comparación de las imágenes antes y después permitirá sacar a la luz posibles problemas. ¿Cómo era la

escena cuando se instaló la cámara y cómo es hoy? ¿Hace falta ajustar la zona de detección? ¿Debe ajustarse el campo de visión de la cámara o hace falta mover la cámara a otro lugar?

9 Privacidad e integridad personal

En el mundo de la seguridad y la vigilancia es preciso encontrar un equilibrio entre los derechos individuales a la privacidad y la integridad personal y la ambición de reforzar la seguridad evitando delitos o facilitando las investigaciones de incidentes. Cada instalación y caso de uso específico debe abordarse con la máxima consideración ética y también desde el conocimiento y la aplicación de la legislación local. También es necesario que la solución garantice la ciberseguridad e impida el acceso no intencionado a material de vídeo. Al mismo tiempo, la analítica en local y la generación de metadatos con fines estadísticos pueden reforzar la protección de la privacidad si únicamente se transmiten datos anonimizados para su procesamiento posterior.

Ante la aplicación cada vez más frecuente de la analítica automatizada en los sistemas de vigilancia, deben tenerse en cuenta algunos aspectos nuevos. Como las aplicaciones de analítica implican un riesgo de falsas detecciones, es importante que el proceso de decisión esté en manos de un operador o un usuario experimentado. Es lo que se conoce a menudo como mantener a un "humano en la ecuación". Además, es importante tener claro que la decisión humana puede verse afectada por cómo se genera y se presenta la alarma. Sin un entrenamiento adecuado y sin un conocimiento suficiente de la funcionalidad de la solución de analítica, es fácil llegar a conclusiones equivocadas.

Otra preocupación tiene relación con el proceso de desarrollo de los algoritmos de aprendizaje profundo, lo que en algunos casos obliga a aplicar la tecnología con muchas precauciones. La calidad de estos algoritmos depende fundamentalmente de los conjuntos de datos, esto es, los vídeos y las imágenes utilizados para entrenar el algoritmo. Diferentes pruebas han demostrado que si el material no se selecciona con cuidado, los sistemas de IA pueden presentar sesgos étnicos o de género en las detecciones. Esta situación ha dado pie a un encendido debate y también a limitaciones legales e iniciativas para garantizar que estos aspectos se tienen en cuenta durante el desarrollo de los sistemas.

Ante la presencia cada vez mayor de la IA en la vigilancia, es obligado valorar si las ventajas en cuanto a la eficiencia operativa y nuevos escenarios de uso justifican la aplicación de la tecnología.

10 Apéndice

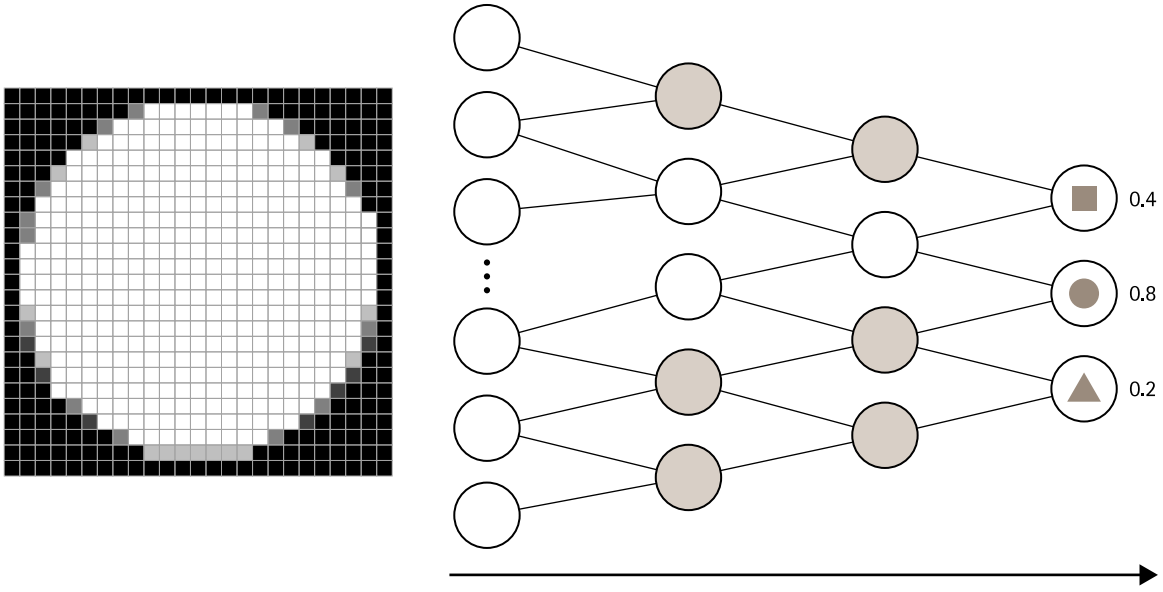
Este apéndice proporciona información adicional sobre las redes neuronales artificiales, que constituyen la base del aprendizaje profundo.

10.1 Redes neuronales

Las redes neuronales son una familia de algoritmos utilizados para reconocer relaciones en conjuntos de datos en un proceso similar en cierto modo al funcionamiento del cerebro humano. Una red neuronal consta de una jerarquía de diferentes capas de nodos (o neuronas) interconectados y la información se transmite a través de las conexiones, entre la capa de entrada y la capa de salida, pasando por la red.

La base del funcionamiento de las redes neuronales es que una muestra de datos de entrada puede reducirse a un conjunto finito de características que representen adecuadamente los datos de entrada. Entonces estas características pueden combinarse y nos ayudarán a clasificar los datos de entrada, por ejemplo describiendo el contenido de una imagen.

La siguiente ilustración presenta un ejemplo en el que una red neuronal se utiliza para identificar a qué clase pertenece la imagen de entrada. Cada píxel de la imagen está representado por un nodo de entrada. Todos los nodos de entrada están asociados a los nodos de la primera capa. Estos nodos generan valores de salida que se transfieren como valores de entrada a la segunda capa, y así sucesivamente. En cada capa, el proceso aplica también funciones de ponderación, valores de sesgo y funciones de activación.



Ejemplo de una imagen de entrada (izquierda) y una red neuronal (derecha). Cuando llega a la capa de salida, la red ha determinado las probabilidades de cada posible categoría (cuadrado, círculo o triángulo). La categoría con el valor de probabilidad más elevado será la forma más probable de la imagen de entrada.

Este proceso se conoce como *propagación hacia delante*. En caso de discordancia en el resultado de la propagación hacia delante se modifican ligeramente los parámetros de la red aplicando una *propagación hacia atrás*. Durante este proceso de entrenamiento iterativo se va mejorando gradualmente el rendimiento de la red.

Después de la implantación, por lo general una red neuronal no conserva la memoria de las secuencias hacia delante anteriores. Por tanto, no mejora con el tiempo y solo puede detectar los tipos de objetos o resolver los tipos de tareas para los que ha recibido el correspondiente entrenamiento.

10.2 Redes neuronales convolucionales (CNN)

Las *redes neuronales convolucionales* (CNN) son un subtipo de redes neuronales artificiales especialmente interesantes para tareas de visión artificial y tienen gran parte de responsabilidad en la rápida evolución del aprendizaje profundo. En el caso de la visión artificial, la red se entrena para detectar automáticamente características específicas de la imagen, como bordes, esquinas o diferencias de colores, lo que le permite identificar formas de objetos en una imagen.

Para conseguirlo, aplica una operación matemática conocida como *convolución*. Se trata de una operación altamente eficiente, ya que el resultado de cada nodo individual depende solo de un pequeño entorno de los datos de entrada, obtenido a partir de la capa anterior, sin necesidad de utilizar todo el volumen de datos de entrada. En otras palabras: en una CNN los nodos no están conectados a todos y cada uno de los nodos de la capa anterior, sino únicamente a un pequeño subconjunto. Las convoluciones se complementan con otras operaciones que reducen el volumen de los datos sin perder la información esencial. Al igual que en una red neuronal artificial estándar, cuanto más profundo viajan los datos por la red, más abstractos son.

Durante la fase de entrenamiento, la CNN aprende la mejor forma de aplicar las capas. Esto es, cómo las convoluciones deben combinar las características de la capa anterior para que el resultado de la red coincida al máximo con las etiquetas de los datos de entrenamiento. Durante la inferencia, la red neuronal convolucional entrenada aplica de forma secuencial las capas de convoluciones resultantes del entrenamiento.

T10163986/ES/M2.2/2103

Acerca de Axis Communications

Axis contribuye a crear un mundo más inteligente y seguro a través de soluciones para mejorar la seguridad y el rendimiento empresarial. Como empresa de tecnología de red y líder del sector, Axis ofrece soluciones de videovigilancia, control de acceso y sistemas de audio e intercomunicación. Se ven reforzadas por aplicaciones de análisis inteligentes y respaldadas por formación de alta calidad.

Axis tiene alrededor de 4000 empleados dedicados en más de 50 países y colabora con socios de integración de sistemas y tecnología en todo el mundo para ofrece soluciones personalizadas. Axis se fundó en 1984 y la sede está en Lund, Suecia