

Sistemas Inteligentes

Sesion 2/12

December 2, 2021

1 Introducción

En esta sesión hemos continuado el temario de la anterior sesión, nos hemos centrado en el reconocimiento de imágenes.

Antes de entrar en detalle hay unos términos importantes, **clasificación**, **localización** y **detección**.

- **Clasificación**

La clasificación se entiende como etiquetar la imagen con la clase a la que pertenece (gato, persona, casa, etc).

- **Localización**

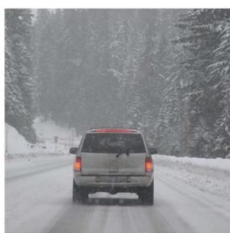
La localización por su parte se entiende por marcar los objetos dentro de la imagen, concretamente usando *Bounding Box*.

- **Detección**

La detección se entiende como la combinación de los dos términos anteriores, es decir, dentro de una imagen la clasificaremos en una clase y la marcaremos usando *Bounding Box*.

What are localization and detection?

Image classification



Classification with localization



Detection

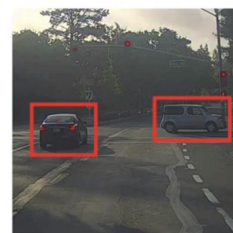


Figure 1: Diferencias entre clasificación, localización y detección

2 Reconocimiento de objetos

Para el tema del reconocimiento de objetos tenemos distintas técnicas, cada técnica tiene el mismo fin aunque difiera la forma en la que se haga, esto dependerá del problema al que nos enfrentemos, por lo que tendremos que saber cuando aplicar algoritmos más o menos complejos a nuestros problemas. Hay dos técnicas que se usan mucho, estas técnicas son **Reconocimiento de características** y **Machine Learning / Deep Learning**, aunque tenemos unas técnicas que son más sencillas de usar:

- **Coincidencia de plantilla**

Esta técnica se usa en el procesamiento digital de imágenes, esta técnica es interesante usarla cuando los objetos que queremos detectar/clasificar tienen unas características comunes. Por ejemplo, gatos.

- **Segmentación de imágenes y análisis de blobs**

Esta técnica nos permite saber las características de los objetos de interés, y eliminar los objetos no deseados. Un blob se puede definir como una región conectada de píxeles (un objeto).

3 Reconocimiento de características

Una técnica que usa mucho es el reconocimiento de objetos sabiendo sus características. Esto se basa en que tenemos una imagen modelo de la cual sacamos sus características (un vector), luego hacemos lo mismo para la imagen de prueba, por lo que si las características de las dos imágenes coinciden hemos acertado.

En la práctica, lo que haremos es usar la distancia euclídea para verificar si se trata de la misma característica, se usarán los vectores de características de cada imagen, si la distancia que hay entre ambos vectores es menor que umbral que hemos establecido con anterioridad entonces coinciden.

Tras esto lo que hay que hacer es hacer una verificación geométrica, es decir, hay que ver si hay una transformación geométrica desde la imagen de prueba a la imagen modelo o viceversa. Se aplica una rotación, escalado y transición para ver si son iguales.

4 Machine Learning y Deep Learning

La gran diferencia entre **Machine Learning** y **Deep Learning** es su funcionamiento por dentro, por su parte **ML** se basa en un aprendizaje supervisado (p.e Adaboost), esto significa que tenemos que hacer es recopilar las imágenes que van a formar nuestro Dataset de entrenamiento, después tenemos que indicarle las características más relevantes de cada imagen del Dataset, esto lo podríamos hacer mediante algoritmos como **SIFT**, **Adaboost**, entre otros, una vez extraído las características se lo tendremos que pasar al modelo para que aprenda. Para finalizar habrá que clasificar las imágenes en las clases que se correspondan.

El uso de **ML** está acotado al reconocimiento de objetos sencillos, debido a que tenemos que indicar que características tenemos que reconocer con anterioridad.

Por otro lado, el funcionamiento de **Deep Learning** se basa en que aprenda por su cuenta, es decir, no tenemos que especificarle que características tiene que usar, estos algoritmos usan Redes Neuronales, concretamente Redes Neuronales Convolucionales (CNN), estas redes ya las hemos visto con anterioridad en la asignatura, básicamente tienen unas capas por donde la información pasa y

va aprendiendo a clasificar las imágenes. Algo interesante de esto es que podemos usar redes ya entrenadas y solo modificar las ultimas capas para que clasifique nuestro los objetos de nuestro problema.

Una pregunta que nos podemos hacer es **¿Cuándo debemos usar Machine Learning o Deep Learning?**, la respuesta es “sencilla“, esto dependerá de si disponemos de un Dataset de imágenes muy grande, debido a que **Deep Learning** necesita un Dataset enorme, otra variable a tener en cuenta es si es tenemos un hardware potente, debido a que **DL** necesita mucha potencia de cálculo.

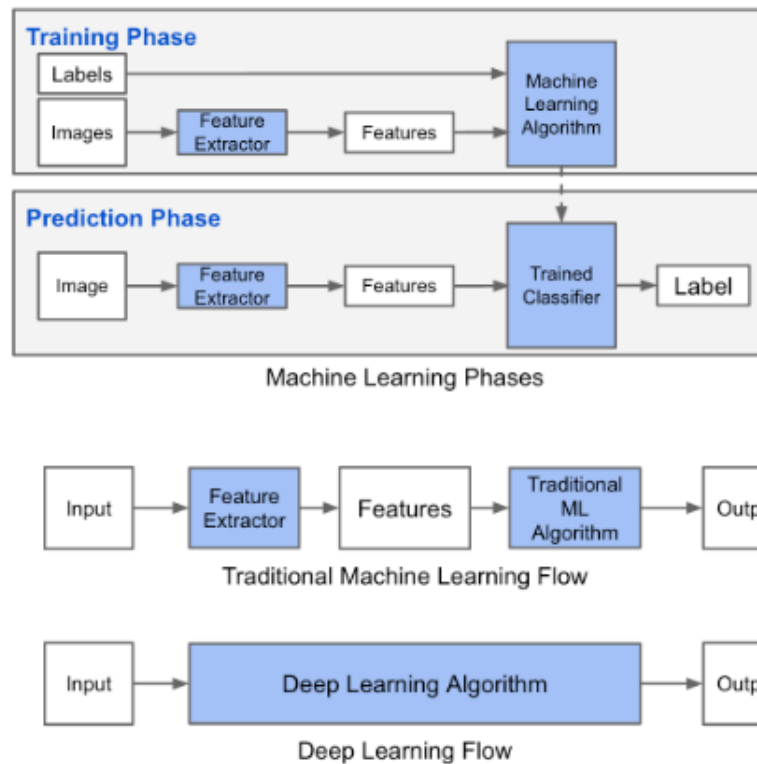


Figure 2: Deep Learning vs Machine Learning

5 Reconocimiento de caras

Para el tema del reconocimiento facial se usa el algoritmo Viola & Jones, ya que es un algoritmo que funciona muy bien, ya que es rápido (tarda cuestión de milisegundos), es fiable y tiene un número bajo de falsos positivos. Un problema de este algoritmo es que su uso se limita a caras frontales o con un poco de giro, ya que basa su funcionamiento en detectar los ojos, también la iluminación es factor importante a la hora de usar este algoritmo, porque su base está en trabajar con las diferencias luminosas.

La estructura que emplea es Adaboost de una forma similar a la que hemos trabajado en la parte práctica. Para el tema de extracción de características, se basa como hemos comentado anteriormente en regiones luminosas de la cara, como puede ser la nariz, los pómulos, etc. Sacaremos regiones

rectangulares (unas 4) vecinas. Además, lo aplicará a imágenes integrales, cuya ventaja es que hace todo los cálculos en una sola iteración, ya que el valor actual de cada píxel viene dado por la suma de los valores del cuadrado de arriba, es decir, de los números de la izquierda y de arriba, esto hace que sea muy rápido de calcular.

6 Reconocimiento de movimiento

Para detectar movimiento existen varias tecnologías, algunas de las más famosas son las siguientes:

- **Vídeo**

Se usan para imágenes 2D que se encuentran en movimiento, la gran limitación es que no podemos analizar en 3 dimensiones, esto puede suponer ciertos problemas en ámbitos como la ciberseguridad.

- **Sistema de captura de movimiento**

El uso de este tipo de sistemas está extendido en industrias como el cine o los videojuegos, su funcionamiento es sencillo, una persona se pone un traje con puntos reflectantes, y se usan unas cámaras que están en distintos ángulos en la sala, también hay unos puntos reflectantes, las cámaras detectan estos puntos al reflejar la luz y capturan el movimiento en tiempo real. Este tipo de técnicas dan muy buenos resultados pero su inconveniente es que los equipos son muy caros, por lo que su uso queda relegado a grandes producciones de cine por ejemplo.

- **Nubes de puntos / Mapas de profundidad**

La ventaja de usar estas tecnologías es que se tiene en cuenta la tercera dimensión (coordenada z). El funcionamiento de esto es “sencillo”, se usa escala de grises y se mide la profundidad de la imagen respecto a la cámara (máscara de profundidad). Una tecnología que usaba esto era Microsoft Kinect, aunque su uso en un principio estaba destinado al mundo de los videojuegos también se empleó en el ámbito científico debido a su reducido costo.

7 Artículo: La inquietante apuesta china por el reconocimiento facial

El reconocimiento facial en el día a día de China se ha vuelto un problema, ya que supone una problemática moral al tener una vigilancia tan intrusiva en el día a día de los ciudadanos, esta vigilancia se hace a escala estatal del país, algunos de sus usos es que permite la detección de criminales, detección del estado de ánimo, también su uso se hace para acceso al metro sin tener que pasar ninguna tarjeta, estos casos que se plantean no parecen muy descabellados pero hay otros como un informe en unas pantallas de los conductores que se salten las normas de tráfico. Esto si lo planteamos de una manera general no supone un gran problema, el problema surge cuando la vigilancia es extrema, el año pasado contaban con 170 millones de cámaras distribuidas por todo el país, se esperaba que se implantasen hasta 400 millones, esto supone una situación asfixiante donde el ciudadano tiene que actuar de una manera perfecta, una situación parecida a las de obras literarias como **1984**, **George Orwell** por poner un ejemplo, donde un **Gran Hermano** dicta como tiene que vivir la vida el ciudadano, que es lo correcto y que no lo es, en conclusión una dictadura del pensamiento.

Reconocimiento facial en China, ¿es ético?