

Sesión 2/12:

Daniel Asensi Roch

Resumen Algoritmo Face Recognition ViolaJones expuesto en clase:

Desarrollado por Paul Viola y Michael Jones en 2001, el marco de detección de objetos Viola-Jones puede detectar objetos en imágenes de manera rápida y precisa y funciona particularmente bien con el rostro humano. A pesar de su antigüedad, el marco sigue siendo un actor líder en la detección de rostros junto con muchas de sus contrapartes de CNN. El marco de detección de objetos Viola-Jones combina los conceptos de características similares a Haar, imágenes integrales, el algoritmo AdaBoost y el clasificador en cascada para crear un sistema de detección de objetos que es rápido y preciso. Por lo tanto, para comprender el marco, primero debemos comprender cada uno de estos conceptos individualmente y luego averiguar cómo se conectan para formar el marco.

A menudo, en Computer Vision, las características se extraen de las imágenes de entrada en lugar de utilizar sus intensidades (valores RGB, etc.) directamente. Las características similares a las de Haar son un ejemplo. Otros ejemplos incluyen Histograma de gradientes orientados (HOG), Patrones binarios locales (LBP), etc. Una característica similar a Haar consta de regiones oscuras y regiones claras. Produce un valor único tomando la suma de las intensidades de las regiones claras y restando eso por la suma de las intensidades de las regiones oscuras. Hay muchos tipos diferentes de funciones similares a Haar, excepto el Marco de detección de objetos Viola-Jones. Los diferentes tipos de características similares a Haar nos permiten extraer información útil de una imagen, como bordes, líneas rectas y líneas diagonales que podemos usar para identificar un objeto (es decir, el rostro humano).

Reconocimiento de Objetos y Personas Mediante DeepLearning:

<https://machinelearningmastery.com/object-recognition-with-deep-learning/>

La detección de objetos es una tarea importante de visión por computadora que se utiliza para detectar instancias de objetos visuales de ciertas clases (por ejemplo, humanos, animales, automóviles o edificios) en imágenes digitales como fotos o fotogramas de video. El objetivo de la detección de objetos es desarrollar modelos computacionales que proporcionen la información más fundamental que necesita la aplicación de visión por computadora.

La detección de personas es una variante de la detección de objetos que se utiliza para detectar una "persona" de clase primaria en imágenes o fotogramas de vídeo. La detección de personas en secuencias de video es una tarea importante en los sistemas de videovigilancia modernos. Los algoritmos de aprendizaje profundo recientes proporcionan resultados sólidos de detección de personas. La mayoría de las técnicas modernas de detección de personas se entrenan en vistas frontales y asimétricas.

En los últimos años, los rápidos avances de las técnicas de aprendizaje profundo han acelerado enormemente el impulso de la detección de objetos. Con las redes de aprendizaje profundo y la potencia informática de las GPU, el rendimiento de los detectores y rastreadores de objetos ha mejorado enormemente, logrando avances significativos en la detección de objetos.

El aprendizaje automático (ML) es una rama de la inteligencia artificial (IA), y esencialmente implica aprender patrones a partir de ejemplos o datos de muestra a medida que la máquina

accede a los datos y tiene la capacidad de aprender de ellos (aprendizaje supervisado en imágenes anotadas). El aprendizaje profundo es una forma especializada de aprendizaje automático que implica el aprendizaje en diferentes etapas. Para aprender más sobre el trasfondo tecnológico

Los métodos de Deep Learning generalmente dependen de una capacitación supervisada. El rendimiento está limitado por la potencia de cálculo de las GPU que aumenta rápidamente año tras año.

Ventajas: la detección de objetos de aprendizaje profundo es significativamente más resistente a la oclusión, las escenas complejas y la iluminación desafiante.

Contras: se requiere una gran cantidad de datos de entrenamiento; el proceso de anotación de imágenes es laborioso y costoso. Por ejemplo, etiquetar 500 000 imágenes para entrenar un algoritmo de detección de objetos DL personalizado se considera un pequeño conjunto de datos. Sin embargo, muchos conjuntos de datos de referencia