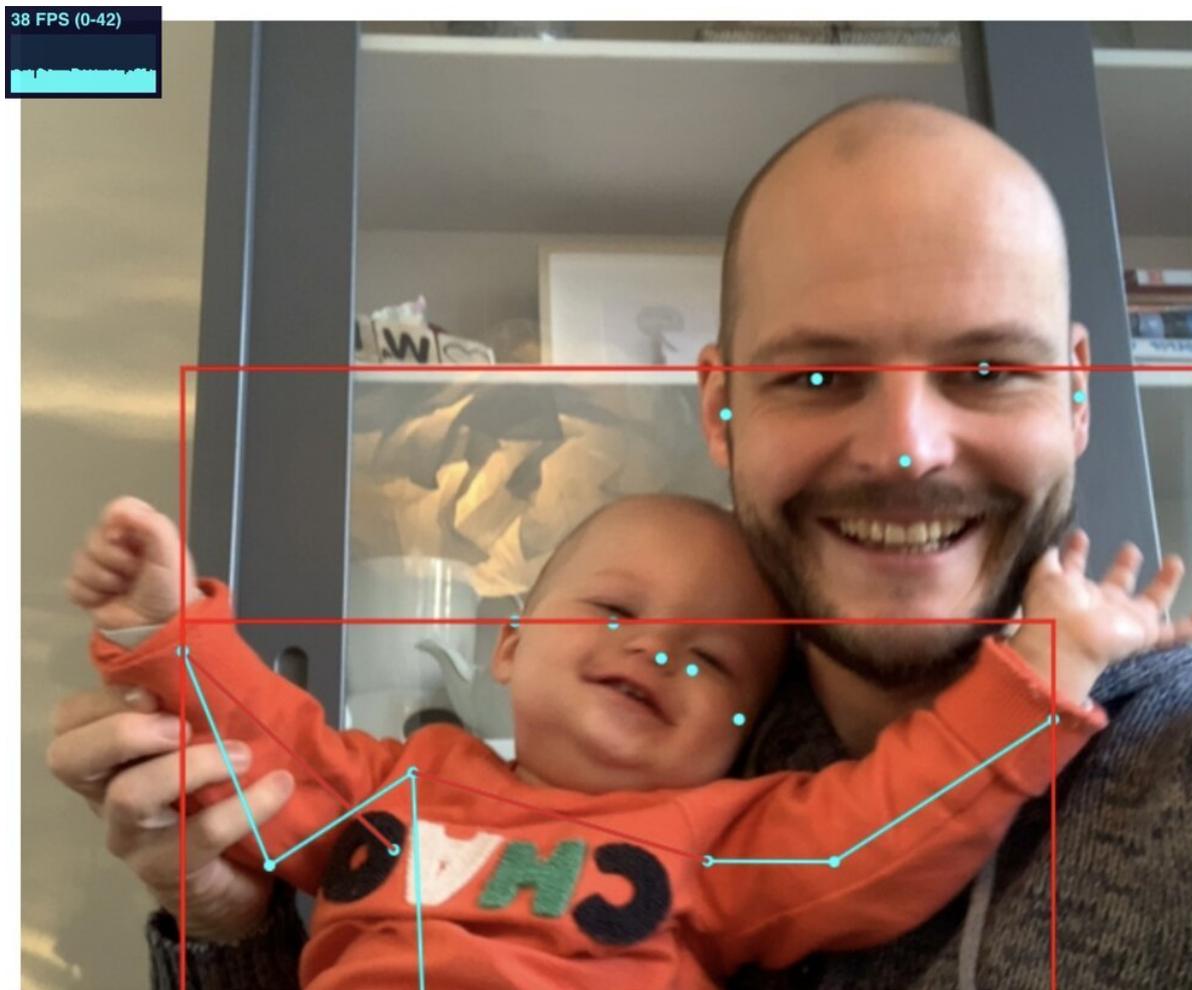


Reconocimiento de postura humana en tiempo real a través de visión por computadora

Uso de TensorFlow y PoseNet en una transmisión de video

Willem L. Middelkoop

Dec. 1, 2019



Para un nuevo proyecto emocionante, he estado experimentando con la visión por computadora usando TensorFlow. Quería lograr la detección de poses humanas en tiempo real para impulsar proyecciones de video interactivas y juegos. ¡Hora de sumergirse en el mundo del aprendizaje automático, los tensores y la visión por computadora!

Visión por computadora

Realmente emocionante es el campo científico interdisciplinario de la visión por computadora, que se ocupa de cómo se puede hacer que las computadoras obtengan una comprensión de alto nivel a partir de imágenes digitales y transmisiones de video. Incluye desafíos para adquirir, procesar y analizar imágenes digitales. En última instancia, querrás que la computadora tome decisiones basadas en su comprensión del (complejo) mundo que la rodea.

TensorFlow

TensorFlow es una plataforma de código abierto de extremo a extremo para el aprendizaje automático, originalmente desarrollada por Google. Se basa en el trabajo científico de varios académicos y desde entonces se ha convertido en un ecosistema estable y robusto de herramientas, bibliotecas y recursos comunitarios.

El nombre "TensorFlow" se deriva del concepto de un "flujo de tensores". Un tensor se describe mejor como una "cosa", como algo que la computadora reconoció o intenta analizar. Técnicamente, un tensor es una matriz multidimensional con valores numéricos. Esto permite a la computadora comparar diferentes "cosas", ver cómo sus características son similares (o diferentes).

El principal desafío en el procesamiento de imágenes en tiempo real es reducir la complejidad de la imagen en fragmentos procesables que se puedan expresar como tensores. Por lo general, esto implica muchos pasos que comprenden el algoritmo de visión por computadora. Pasos como reconocer una forma, recortar esa forma en particular, cambiar su tamaño, eliminar su color, comparar los contornos, etc. Los tensores fluyen a través de estos pasos, de 'sin analizar' a 'reconocido', de ahí el nombre TensorFlow.

Potencia de procesamiento

Lo mejor del aprendizaje automático moderno que utiliza TensorFlow es que, en general, no se necesita hardware (muy) caro para lograr resultados. Esto se debe a que el "trabajo pesado" se realiza al entrenar modelos de aprendizaje automático, que se pueden imaginar como "tablas de referencia". Estos modelos se crean analizando conjuntos de datos conocidos, donde se le da a la computadora la 'respuesta' y determina las características del sujeto. El entrenamiento de modelos requiere tiempo, potencia de procesamiento y datos de entrenamiento.

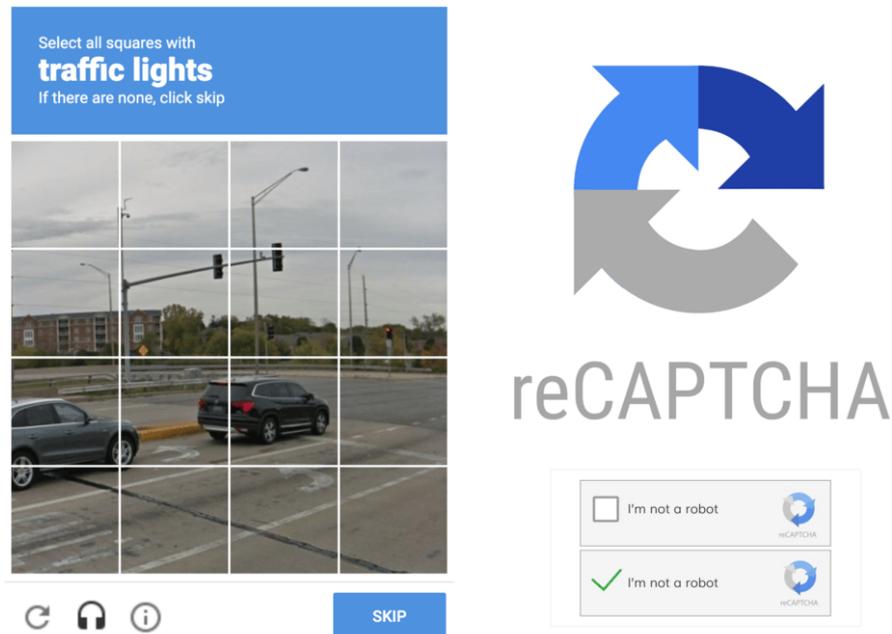
Datos de entrenamiento

Google hizo algo realmente inteligente, distribuyó el entrenamiento de sus modelos de aprendizaje automático a todos nosotros. ¡Sí, eso te incluye a ti!

Recaptcha

Es probable que hayas encontrado "reCAPTCHA" o "Recaptcha". Un desafío visual que debes completar cuando pides algo en línea o intentas registrarte en algún lugar. Este desafío proporciona una protección contra (spam) bots para los propietarios de tiendas web y sitios web. Su objetivo es distinguir a los humanos de los bots. Lo hace proporcionando imágenes (complejas) a los humanos, pidiéndoles que identifiquen objetos

en estas imágenes. Como señales de tráfico, cruces, coches, etc. [Google usa este feedback humano](#) para entrenar sus modelos.

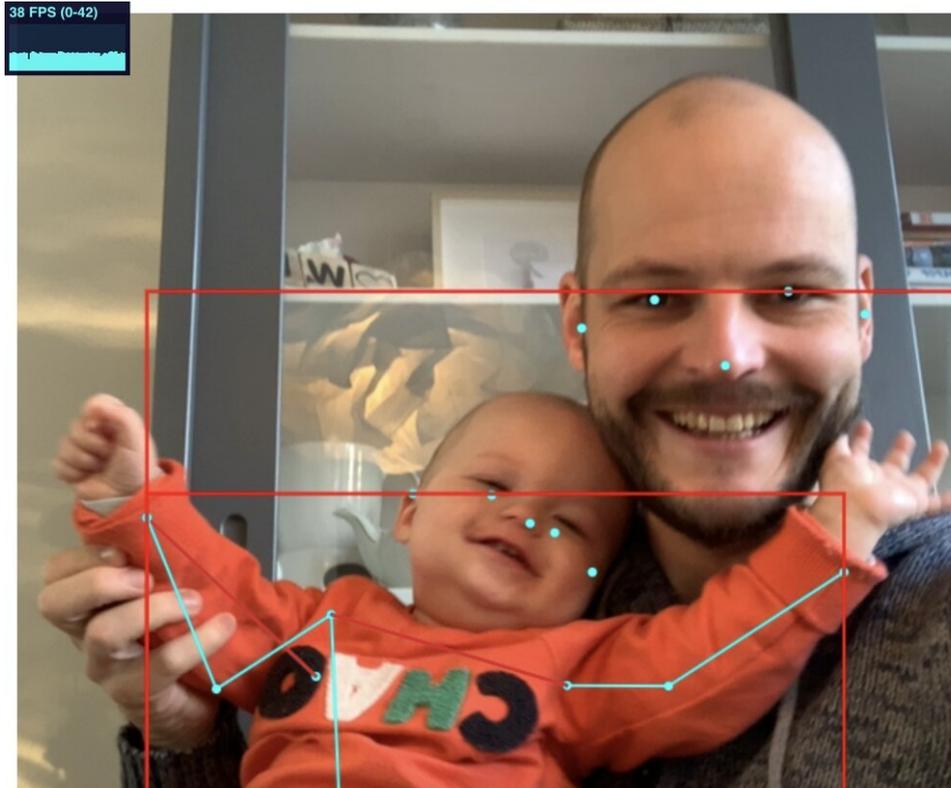


ReCaptcha: ¡llevas años entrenando la IA de Google!

Lo mejor es que muchos de los hallazgos de Google están disponibles públicamente, hay muchos proyectos interesantes de código abierto que nos permiten a nosotros, simples desarrolladores, hacer cosas increíbles con el aprendizaje automático.

Reconocimiento de poses humanas usando: PoseNet

Uno de estos modelos de aprendizaje automático disponibles es [PoseNet](#). Es un modelo de visión que se puede utilizar para estimar la postura de una persona en una imagen o video. Esto se hace determinando dónde están las articulaciones clave del cuerpo, como los codos, las manos, las caderas, los tobillos de las rodillas, etc.

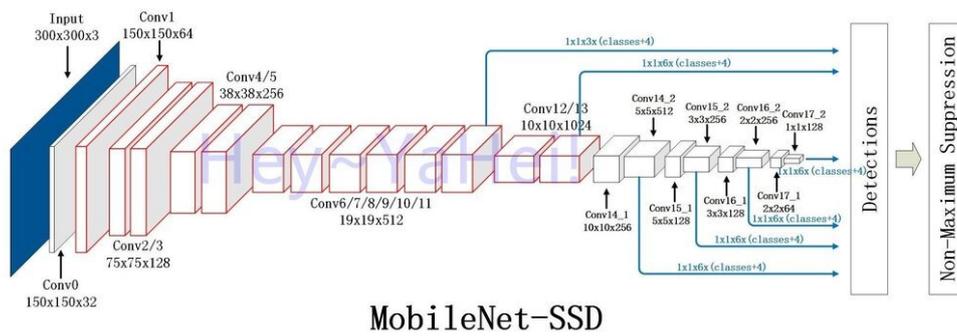


Yo y mi mini-yo siendo reconocidos por el modelo PoseNet

Algoritmo Single Shot MultiBox Detector (SSD)

El Single Shot MultiBox Detector es un algoritmo popular para detectar objetos en imágenes y transmisiones de video. Es popular debido a su velocidad. El principio es relativamente simple de entender: el algoritmo intenta reducir la complejidad de la imagen lo más rápido posible, realizando solo análisis profundos en mapas de características reducidos y simplificados.

Este enfoque funciona mejor si desea detectar objetos grandes, ya que se destacan rápidamente: generan suficientes características de alto nivel para hacer predicciones al principio del algoritmo. Esto es ideal para la detección del cuerpo, que es uno de los primeros pasos cuando se desea determinar las poses humanas (primero mirar *dónde* está una persona, luego *cuál* es su pose).



MobileNet-SSD

Algoritmo MobileNet-SSD (imagen: hey-yahei.cn)

El reconocimiento de imágenes y personas es algo en lo que la gente de [China](#) es bastante buena, por varias razones. Sus algoritmos funcionan bien en hardware relativamente débil. ¡Esto permite que la visión por computadora en todo tipo de dispositivos funcione bien, en tiempo real!

Casos de uso geniales

Poder detectar poses en tiempo real permite todo tipo de cosas geniales. ¡Literalmente, puedes convertirte en el controlador del juego o la entrada para una instalación de arte interactiva!

Espejo PomPom

Con 928 bolas esféricas de piel sintética, esta [instalación de arte de Daniel Rozin](#) es increíble. La escultura está controlada por cientos de motores que construyen siluetas de espectadores utilizando visión por computadora.



Espejo PomPom de falsos abullonados de Daniel Rozin

Convirtiéndote en un pájaro: La traición del santuario

Permitir que las personas se pierdan por un momento solo fue posible al reunir diferentes disciplinas. Esta asombrosa [instalación de arte interactiva dirigida por Chris Milk](#) es un tríptico gigante que lleva a los espectadores a través de varias etapas de vuelo utilizando visión por computadora.



Convirtiendo personas en pájaros usando visión por computador

Conclusión

Trabajar en proyectos creativos y experimentales es un privilegio increíble, es lo que hace que mi trabajo sea tan emocionante.

Con montones de datos cada vez mayores, la creación de sentido digital y autónoma es un desafío importante para el futuro: ¡uno en el que estoy ansioso por trabajar!