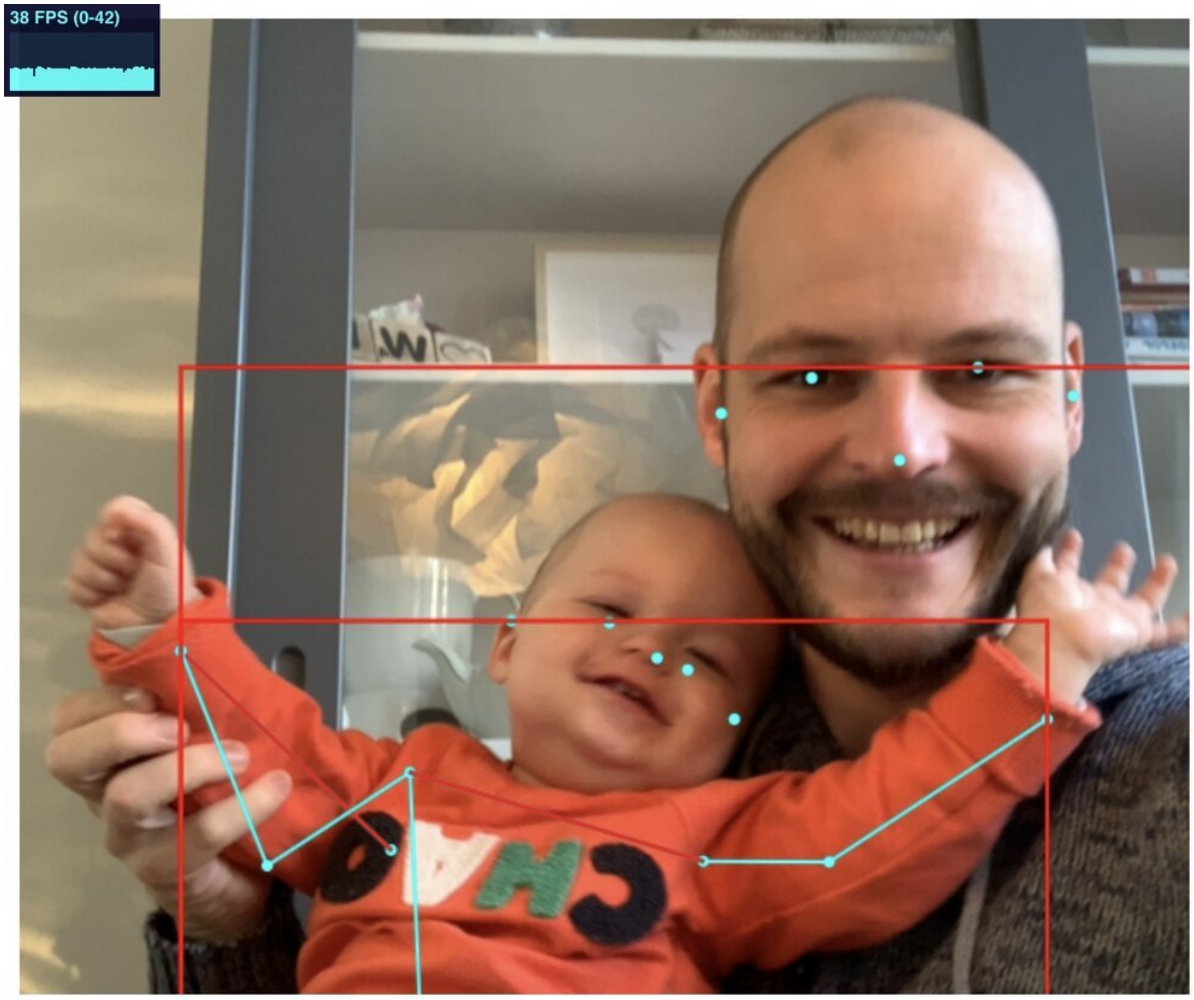
Echtzeit-Menschliche Posenerkennung durch Computer Vision

Verwendung von TensorFlow und PoseNet für einen Video-Feed

Willem L. Middelkoop
Dec. 1, 2019



Für ein spannendes neues Projekt habe ich mit Computer Vision unter Verwendung von TensorFlow experimentiert. Ich wollte Echtzeit-Menschenerkennung erreichen, um interaktive Videoprojektionen und Spiele zu steuern. Zeit, in die Welt des maschinellen Lernens, der Tensoren und der Computer Vision einzutauchen!

Computer Vision

Besonders spannend ist das interdisziplinäre wissenschaftliche Feld der Computer Vision, das sich damit beschäftigt, wie Computer dazu gebracht werden können, ein hochgradiges Verständnis aus digitalen Bildern und Videostreams zu gewinnen. Es beinhaltet Herausforderungen bei der Erfassung, Verarbeitung und Analyse digitaler Bilder. Letztendlich möchte man, dass der Computer Entscheidungen basierend auf seinem Verständnis der (komplexen) Welt um ihn herum trifft.

TensorFlow

TensorFlow ist eine End-to-End-Open-Source-Plattform für maschinelles Lernen, die ursprünglich von Google entwickelt wurde. Sie basiert auf wissenschaftlichen Arbeiten verschiedener Wissenschaftler und hat sich seitdem zu einem stabilen und robusten Ökosystem aus Tools, Bibliotheken und Community-Ressourcen entwickelt.

Der Name "TensorFlow" leitet sich vom Konzept eines "Flusses von Tensoren" ab. Ein Tensor lässt sich am besten als ein "Ding" beschreiben, wie etwas, das der Computer erkannt hat - oder zu analysieren versucht. Technisch gesehen ist ein Tensor ein mehrdimensionales Array mit numerischen Werten. Dies ermöglicht es dem Computer, verschiedene "Dinge" zu vergleichen und zu sehen, wie ähnlich (oder unterschiedlich) ihre Eigenschaften sind.

Die größte Herausforderung bei der Echtzeit-Bildverarbeitung besteht darin, die Komplexität des Bildes in verarbeitbare Teile zu reduzieren, die als Tensoren ausgedrückt werden können. Normalerweise umfasst dies viele Schritte, die den Algorithmus der Computer Vision bilden. Schritte wie das Erkennen einer Form, das Zuschneiden auf diese bestimmte Form, das Ändern der Größe, das Entfernen der Farbe, das Vergleichen der Konturen usw. Tensoren fließen durch diese Schritte, von "unanalysiert" zu "erkannt", daher der Name TensorFlow.

Rechenleistung

Das Tolle am modernen maschinellen Lernen mit TensorFlow ist, dass man im Allgemeinen keine (sehr) teure Hardware benötigt, um Ergebnisse zu erzielen. Dies liegt daran, dass die "schwere Arbeit" beim Trainieren von Machine-Learning-Modellen erledigt wird, die man sich als eine Art "Referenztabellen" vorstellen kann. Diese Modelle werden durch die Analyse bekannter Datensätze erstellt, wobei dem Computer die "Antwort" gegeben wird und er die Eigenschaften des Subjekts bestimmt. Das Trainieren von Modellen erfordert Zeit, Rechenleistung und Trainingsdaten.

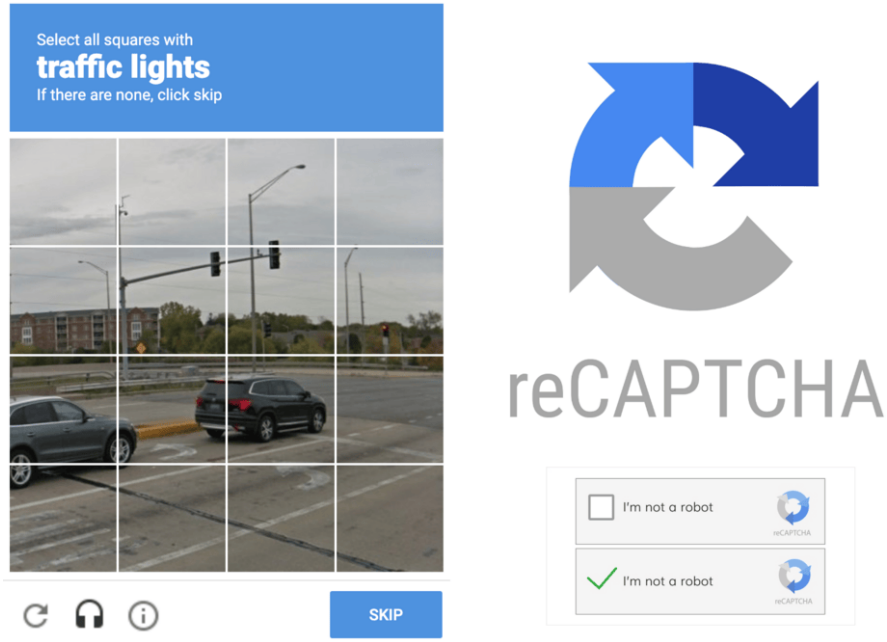
Trainingsdaten

Google hat etwas wirklich Cleveres getan: Es hat das Training seiner Machine-Learning-Modelle auf uns alle verteilt. Ja, das schließt Sie ein!

Recaptcha

Wahrscheinlich sind Sie schon einmal auf "reCAPTCHA" oder "Recaptcha" gestoßen. Eine visuelle Herausforderung, die Sie abschließen müssen, wenn Sie etwas online bestellen oder versuchen, sich irgendwo zu registrieren. Diese Herausforderung bietet einen Schutz

vor (Spam-)Bots für Webshop- und Website-Betreiber. Ihr Ziel ist es, Menschen von Bots zu unterscheiden. Dies geschieht, indem den Menschen (komplexe) Bilder vorgelegt werden und sie gebeten werden, Objekte in diesen Bildern zu identifizieren. Wie Verkehrszeichen, Kreuzungen, Autos usw. Google verwendet dieses menschliche Feedback, um seine Modelle zu trainieren.

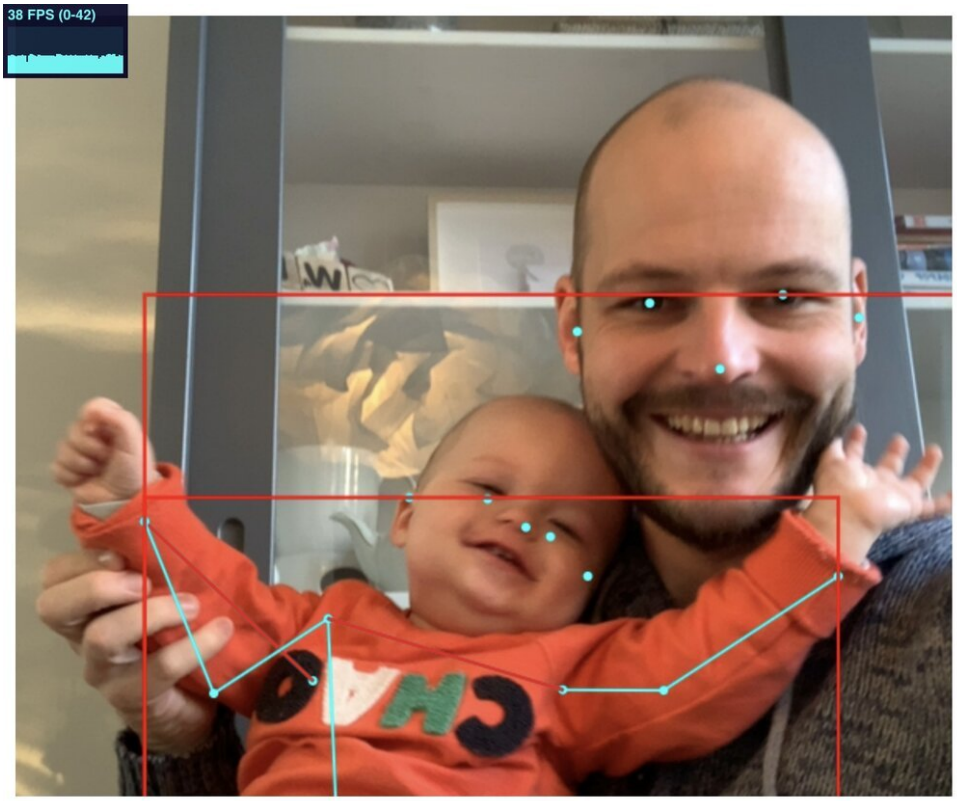


ReCaptcha: Du hast jahrelang Googles KI trainiert!

Das Tolle ist, dass viele der Erkenntnisse von Google öffentlich zugänglich gemacht werden. Es gibt viele interessante Open-Source-Projekte, die es uns - einfachen Entwicklern - ermöglichen, erstaunliche Dinge mit maschinellem Lernen zu tun.

Menschliche Posenerkennung mit: PoseNet

Eines dieser verfügbaren Machine-Learning-Modelle ist PoseNet. Es ist ein Visionsmodell, das verwendet werden kann, um die Pose einer Person in einem Bild oder Video zu schätzen. Dies geschieht durch die Bestimmung, wo sich die wichtigsten Gelenke des Körpers befinden, wie z. B. Ellbogen, Hände, Hüften, Knie, Knöchel usw.

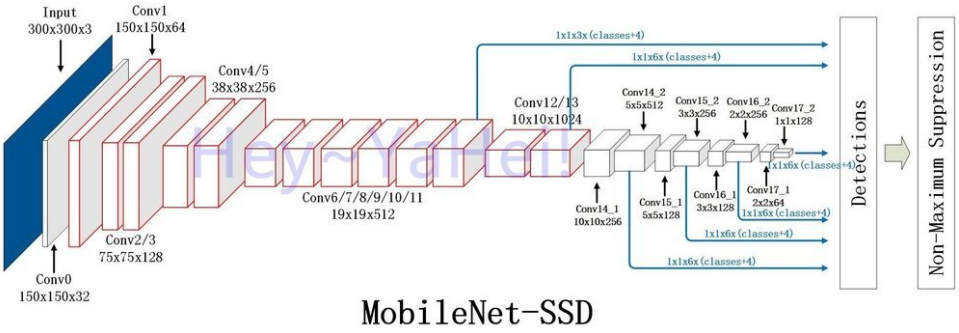


Ich und Mini-Ich werden vom PoseNet-Modell erkannt

Single Shot MultiBox Detector (SSD) Algorithmus

Der Single Shot MultiBox Detector ist ein beliebter Algorithmus zur Erkennung von Objekten in Bildern und Videostreams. Er ist wegen seiner Geschwindigkeit beliebt. Das Prinzip ist relativ einfach zu verstehen: Der Algorithmus versucht, die Bildkomplexität so schnell wie möglich zu reduzieren, indem er nur tiefe Analysen auf verkleinerten, vereinfachten Feature-Maps durchführt.

Dieser Ansatz funktioniert am besten, wenn man große Objekte erkennen möchte, da diese schnell hervorstechen: Sie erzeugen frühzeitig im Algorithmus genügend High-Level-Features, um Vorhersagen zu treffen. Dies ist ideal für die Körpererkennung, die einer der ersten Schritte ist, wenn man menschliche Posen bestimmen möchte (zuerst schauen, *wo* sich eine Person befindet, dann *was* ihre Pose ist).



MobileNet-SSD-Algorithmus (image: hey-yahei.cn)

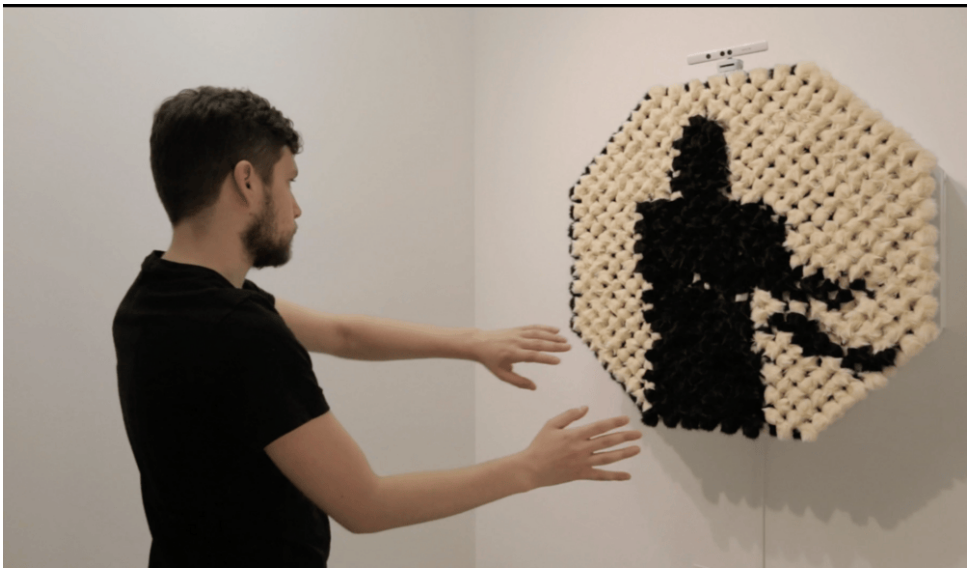
Bild- und Personenerkennung ist etwas, worin die Menschen aus China aus verschiedenen Gründen ziemlich gut sind. Ihre Algorithmen funktionieren gut auf relativ schwacher Hardware. Dies ermöglicht es, Computer Vision auf allen möglichen Geräten in Echtzeit gut funktionieren zu lassen!

Coole Anwendungsfälle

Die Möglichkeit, Posen in Echtzeit zu erkennen, ermöglicht alle möglichen coolen Dinge. Man kann - im wahrsten Sinne des Wortes - zum Gamecontroller oder zur Eingabe für eine interaktive Kunstinstallation werden!

PomPom Spiegel

Mit 928 kugelförmigen Kunstfell-Puffs ist diese Kunst Installation von Daniel Rozin erstaunlich. Die Skulptur wird von Hunderten von Motoren gesteuert, die mithilfe von Computer Vision Silhouetten von Betrachtern erstellen.



PomPom Faux-Puff-Spiegel von Daniel Rozin

Verwandlung in einen Vogel: The Treachery of Sanctuary

Menschen zu ermöglichen, sich für einen Moment zu verlieren, war nur durch die Zusammenführung verschiedener Disziplinen möglich. Diese erstaunliche interaktive Kunst Installation unter der Regie von Chris Milk ist ein riesiges Triptychon, das die Betrachter mithilfe von Computer Vision durch verschiedene Flugphasen führt.

Menschen mithilfe von Computer Vision in Vögel verwandeln

Fazit

An kreativen und experimentellen Projekten zu arbeiten ist ein unglaubliches Privileg, es ist das, was meine Arbeit so spannend macht.

Mit immer größer werdenden Datenmengen ist die digitale und autonome Sinngebung eine große Herausforderung für die Zukunft: eine, an der ich gerne arbeite!