Richtlinie zum „Software-Sprint“

VCAT – Visual Collections and Training Data

Schlussbericht

Zuwendungsempfänger: Adam Harvey / Jules LaPlace

Name des Zuwendungsempfängers: Adam Harvey / Jules LaPlace

Das Projekt welches diesem Bericht zugrunde liegt wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IS18S58 ermöglicht.Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

Kurze Darstellung der Aufgabenstellung und Motivation

Was war Deine Motivation? Welches Problem wolltest Du mit Deinem Projekt lösen? Wie war die geplante Vorgehensweise zur Problemlösung (auch Angabe der wichtigsten Meilensteine)?

Die Grundmotivation die zur Entwicklung des VCAT-Projekt geführt hat, ist ein Mangel an Trainingsdaten die für Computer-Vision genutzt werden können. Trainingsdaten sind ein wesentlicher Bestandteil von Neuronalen Netzen, die im Bereich des Computer Vision eingesetzt werden. Die öffentlich zugänglichen Datensätze sind jedoch generisch und nicht auf die speziellen Forschungsziele von Menschenrechtsgruppen und investigativen Journalisten anwendbar. Das VCAT-Projekt wurde entwickelt, um eine Open-Source-Softwarelösung für dieses Problem bereitzustellen, indem 3D-modellierte synthetische Bildtrainingsdaten entwickelt wurden. Das Projekt hatte zum Ziel, zwei Anwendungen zu entwickeln: eine zur Generierung synthetischer Daten und die andere zum Aufbau einer visuellen Suchmaschine zur Bereitstellung der Bildverarbeitungsmodelle.

Beitrag des Projektes zu den Zielen der Förderinitative „Software-Sprint“

Wer ist die Zielgruppe für Deine Lösung? Wie profitiert sie von den Ergebnissen Deines Projekts? Welche Bezüge gibt es zu den Themenfeldern und Zielen des Software Sprints?

Zielgruppe des Projekts sind Forschungsgruppen, die mit visuellen Medien aus Konfliktzonen arbeiten. Diese Gruppen können von VCAT profitieren, indem sie das synthetische Datensystem verwenden, um sehr spezifische neuronale Objekterkennungsnetze aufzubauen.

Des Weiteren profitieren die Zielgruppen, indem sie die browserbasierte visuelle Suchmaschine verwenden, um große Videoarchive zu verarbeiten.

Ausführliche Darstellung der Ergebnisse

Welche konkreten Ergebnisse hast Du erzielt? Konnten alle Meilensteine erreicht werden? Welche zusätzlichen Erkenntnisse hast Du aus der Projektarbeit gewonnen, auch im Hinblick auf die Begleitung durch die Open Knowledge Foundation?

VCAT erreichte sein Hauptziel, die Entwicklung einer Open-Source-Prototyp-Softwareanwendung und -Methodik zur Generierung synthetischer Bildtrainingsdaten für Objekte in Konfliktzonen von Syrien und Jemen. Die beiden Hauptanwendungen sind: das synthetische Datengenerierungssystem und die visuelle Suchmaschine.

Das synthetische Datensystem umfasst eine Sammlung von Python-Skripten und -Anwendungen, die mit einer Blender 3D-Modellierungssoftware 3D-Szenen erzeugen und automatisch in annotierte Bilder umwandeln, die für die Erkennung von Trainingsobjekten und Algorithmen zur Bildklassifizierung verwendet werden. Das synthetische Datengenerierungssystem kann eine Vielzahl von Variablen **randomisieren**, einschließlich Hintergrund, Himmel, Boden, Objektposition, Beleuchtung, Drehung, Größe des Kamerasensors, Zoom der Kamera und Farbtemperatur.

Nach einer Auswertung anderer 3D-Modellierungssoftware wurde das Programm Blender wegen seiner hochwertigen Renderings, der Open-Source-Lizenz und der Möglichkeit, aus der Ferne zu rendern, ausgewählt. Dadurch kann das VCAT-System auf entfernten Cloud-Servern betrieben werden, um die Vorteile einer höheren Rechenleistung beim Erzeugen großer Datensätze zu nutzen.

Die finale Ausgabe des Systems, wie auf der GitHub-Seite mit synthetischen Daten (https://github.com/vframeio/vframe_synthetic) dokumentiert, umfasst die fotorealistischen Renderings, farbcodierte Annotationsbilder, CSV-formatierte Annotationsdateien und animierte Demos, die das Beheben von Fehlern und Anzeigen der Ergebnisse erleichtern. Das System ermöglicht die Generierung synthetischer Daten mit 2 verschiedenen Render-Engines: Cycles und Eeevee. Ersteres wird für hochwertiges Rendering und letzteres für Prototyping verwendet. Die Renderzeiten pro Bild für hochwertige Daten reichen von 15-200 Sekunden, je nach Bildgröße und Hardware. Bei Daten geringerer Qualität betragen die Renderzeiten mit der Eeevee Render-Engine typischerweise 1-10 Sekunden pro Bild. Während der Entwicklung verwendete VCAT zwei NVIDIA GTX1080-Ti GPUs mit einem Intel i7-7700 für das Rendering.

Die zweite Hauptkomponente des Projekts ist die visuelle Suchmaschine. Die Suchmaschine ist einzigartig, weil sie einen ModelZoo oder eine Sammlung von Modellen beinhaltet, die erweitert werden können. Der Administrator der visuellen Suchmaschine kann dann aus verschiedenen Modellen auswählen, um Merkmalsvektoren zu berechnen, die für die Suche nach einer Sammlung von Videos und Bildern verwendet werden. Für spezifische Untersuchungen können verschiedene Merkmalsvektoren verwendet werden. So könnte beispielsweise ein Tankdetektor mit 3D-Modellen von Tanks trainiert und dann verwendet werden, um einen Suchindex zu erstellen, der in der Lage ist, zwischen verschiedenen Tanktypen zu unterscheiden und dem Benutzer genauere interaktive Suchergebnisse zu liefern.

Durch die Entwicklung des Projekts haben wir Einblicke in Bereiche gewonnen, in denen die Computer Vision Industrie die Zugänglichkeit fehlt, nämlich durch die Bereitstellung von Open Source Datensätzen und einer Schnittstelle zur Nutzung von Open Source Model-

len mit handelsüblicher Desktop-Hardware. Es ist in diesem Bereich in dem wir der Meinung sind, dass VCAT einen wesentlichen Beitrag zur Unterstützung von Menschenrechtsforschern und investigativen Journalisten leisten kann.

Zielgruppe, Nutzen und mögliche Weiterentwicklungen

Welcher Nutzen ergibt sich für die Zielgruppe aus den Ergebnissen Deines Projekts? Welche weiter-gehenden Effekte ergeben sich aus der Open-Source-Stellung der Ergebnisse? Gibt es Ideen für die Weiterentwicklung Deiner Lösung und Pläne zu deren Umsetzung?

Hat die Arbeit in dem Projekt Dich in Deiner persönlichen, fachlichen Weiterentwicklung unterstützt?

VCAT erreichte sein Hauptziel, in dem es eine Softwareanwendung, 3D-Modelle sowie eine Methodik zur Generierung synthetischer Bildtrainingsdaten für Objekte in Konfliktzonen von Syrien und Jemen entwickelte. Durch die Entwicklung haben wir Erkenntnisse darüber gewonnen, wo die Bildverarbeitungsindustrie nicht zugänglich ist, nämlich bei der Bereitstellung von Open-Source-Datensätzen. Wir glauben, dass VCAT hier einen wesentlichen Beitrag leisten kann.

Kurze Darstellung der Arbeiten, die zu keiner Lösung geführt haben

Gab es Arbeiten bzw. Lösungsansätze, die nicht weiter verfolgt wurden? Was waren die Hinter-gründe, und wie bist Du alternativ vorgegangen?

Wir fanden heraus, dass die Entwicklung neuer manueller Annotationswerkzeuge nicht notwendig war. Es gibt bereits viele bestehende Tools in diesem Bereich. Die Annektierung durch externe Arbeitskräfte ist teuer. Und die Weitergabe von Daten an Fremdfirmen ist rechtlich nicht möglich. Stattdessen werden die synthetischen Datenobjektdetektoren verwendet, um neue Datenquellen semi-automatisch zu annotieren.

Kurze Angabe von Präsentationsmöglichkeiten für mögliche Nutzer

Wo können sich Interessenten detailliert über Deine Projektergebnisse informieren (z.B. Webseite, GHitHub, Veröffentlichungen)?

Unter https://vframe.io/research/synthetic-datasets/ finden Sie Informationen über die Forschung und unter Sie https://github.com/vframeio/ finden Sie den Code.

Kurze Erläuterung zur Einhaltung der Arbeits- und Kostenplanung

Gab es im Projektverlauf Ereignisse, die eine Anpassung der Planung erforderlich machten – z.B. Mehr- oder Minderaufwand bei der Bearbeitung von Teilaufgaben?

Wir haben den Zeitplan geändert, um mehr Zeit für die Entwicklung der 3D-Modelle zu haben. Damit wurde fast ein Drittel des ursprünglich für die Entwicklung des Codes vorgesehenen Budgets verbraucht und es war schwieriger als ursprünglich geplant. Des Weiteren haben wir von der Verwendung von Unity und Unreal Engine auf Blender für die 3D-Entwicklung umgestellt, weil es Open-Source-freundlicher ist. Dies dauerte länger, da sich die

3D-Software während unserer Entwicklung noch in der undokumentierten Beta-Phase befand.

Kurze Darstellung von etwaigen Ergebnissen bei anderen Stellen

Gab es Entwicklungen anderer Personen oder Institutionen, die Einfluss auf Deine Arbeiten und die Zielsetzung hatten? Wenn ja, worin bestand dieser und wie bist Du damit umgegangen?

Während der Entwicklung beobachteten wir aufkommende Trends in den Bereichen KI/ML und Computer Vision und stellten fest, dass es ein stetiges Wachstum bei den Verbesserungen der Techniken zur Objekterkennung und Bildklassifizierung gab, aber wenig Wachstum bei den öffentlich zugänglichen Trainingsdatensätzen. Den größten Wachstums in Datensätze konnten im Bereich **verwandte** autonome Fahrzeuge (Mapillary-Datensatz), Roboternavigation und Antenne (xView-Datensatz) beobachtet werden. Wir haben auch eine zunehmende Anzahl von Start-ups für die Computer-Vision bemerkt, die Remote Annotation Services anbieten. Es gab auch eine gewisse Zunahme der Aktivitäten in der Unternehmensentwicklung von synthetischen Datensätzen, aber dies betraf auch hauptsächlich Industrieroboter oder autonome Fahrzeuge.