

SynthEO

Unlocking the Power of AI in Earth Observation with Synthetic Data to Tackle Pressing Environmental Challenges

Programm / Ausschreibung	Expedition Zukunft, Expedition Zukunft 2022, Expedition Zukunft Start 2022	Status	abgeschlossen
Projektstart	01.04.2024	Projektende	31.05.2025
Zeitraum	2024 - 2025	Projektlaufzeit	14 Monate
Keywords	Generative AI, synthetic data, privacy, satellite imagery, environmental monitoring		

Projektbeschreibung

Unsere Gesellschaft steht vor gewaltigen Umweltproblemen – diese reichen von rapide fortschreitender Abholzung, zu den Auswirkungen des Klimawandels, begleitet von Landdegradation, Verschmutzung von Land und Wasser, ineffizienter Abfallwirtschaft, der wachsenden Gefahr von Waldbränden und rasanter Urbanisierung.

Um diese Herausforderungen zu bewältigen, spielt künstliche Intelligenz (KI) in der Erdbeobachtung eine immer wichtigere Rolle. Um das volle Potenzial von KI zu entfalten, ist es wichtig, eine zentrale Hürde zu überwinden: die Komplexität der Erstellung von Trainingsdaten für KI-Modelle.

Tiefe neuronale Netzwerke, ein Grundbaustein von Künstlicher Intelligenz, benötigen große Mengen sorgfältig vorbereiteter Trainingsdaten. Die Erstellung solcher Datensätze ist jedoch ressourcenintensiv und zeitaufwendig. Der Mangel an hochwertigen Trainingsdaten behindert die breitere Einführung von KI im Bereich der Erdbeobachtung.

Synthetische Trainingsdaten, insbesondere synthetische Satellitenbilder, bieten eine bahnbrechende Lösung für diese Herausforderungen. Sie nutzen generative KI, um große Mengen synthetischer Trainingsdaten zu erstellen und die zentrale Engstelle im KI-Modelltraining zu überwinden. Synthetische Daten erfüllen den Bedarf an großen und spezialisierten Datensätzen und sind konform mit Datenschutz- und Sicherheitsanforderungen. Sie beschleunigen die Einführung von KI in der Erdbeobachtung und fördern die Entstehung und Umsetzung neuer Anwendungsfälle und Dienstleistungen, die bis 2031 voraussichtlich mehr als 5,5 Milliarden Euro Umsatz generieren werden. Angesichts sich entwickelnder Vorschriften, insbesondere des EU AI Acts und der DSGVO, gewinnt die Verwendung von synthetischen Daten an Schwung, da sie den Weg für konforme und skalierbare KI-Lösungen ebnet.

Das Hauptziel von SynthEO besteht darin, einen Prototypen zu entwickeln, der in der Lage ist, synthetische Satellitenbilder und entsprechende Segmentierungsmasken aus minimalen bis nicht vorhandenen Daten zu generieren. .

SynthEO wird dazu beitragen, die Vision von Another Earth zu validieren: die Entwicklung der führenden synthetischen Datenmaschine, die unbegrenzte Mengen hochwertiger visueller Trainingsdaten für das Training von KI-Modellen erstellen kann, um die breitere Anwendung von KI in der Erdbeobachtung zu unterstützen.

Abstract

As a society, we are facing massive environmental challenges that range from rapid deforestation to the impact of climate change, with issues such as land degradation, pollution of both land and water, inefficient waste management, the escalating threat of wildfires, and the rapid pace of urbanisation. Amidst these challenges, Artificial Intelligence (AI) in Earth observation is emerging with the potential to address and mitigate some of the most pressing environmental concerns our society faces today.

However, to unlock the real potential of Artificial Intelligence to tackle pressing environmental challenges, we need to overcome the key barrier that is the challenge of training data for AI models.

Deep learning, a fundamental component of AI, requires large amounts of carefully prepared training data to function and to learn the complex features from large amounts of data. It's the training data from which the AI learns to understand the data it's processing. Creating training data involves labour-intensive tasks such as data sourcing, manual labelling, and segmentation - resulting in a resource-intensive, time consuming and expensive process hindering the adoption of AI.

Furthermore the requirement of diversity in datasets to prevent AI bias and ensure compliance with regulations like GDPR and the EU AI Act adds further complexity. The lack of high quality training data is hindering the broader adoption of AI in Earth observation, preventing the disruption needed to help tackle the environmental challenges of our time.

Synthetic Training data, specifically Synthetic Satellite imagery, is emerging as a groundbreaking solution to the challenges in training data. Synthetic Satellite imagery uses generative AI to create large amounts of synthetic training data, that addresses this critical bottleneck in AI model training. Synthetic data not only meets the need for large and specialised datasets but is also compliant with privacy and security requirements. Synthetic training data has the potential to fast-track the growth of Earth observation value-added services, projected to exceed €5.5 billion in revenues by 2031. In the face of evolving regulations, including the EU AI Act and GDPR, the adoption of Synthetic Data is gaining momentum as it paves the way for compliant and scalable AI solutions.

The main objective of SynthEO is to develop a prototype that is able to generate synthetic satellite images and corresponding segmentation masks from little to no input. SynthEO will help validate the larger vision of Another Earth to build the leading Synthetic Data Engine that is able to create unlimited amounts of high quality visual training data for AI model training to support the adoption of AI in Earth Observation.

Endberichtkurzfassung

This project successfully developed SynthEO, a prototype capable of generating synthetic landcover and optical satellite data pairs with 1m Ground Sample Distance, exceeding initial resolution targets. This synthetic data, derived from the OpenEarthMap dataset and validated through downstream segmentation models, can be used to train AI models for real-world applications. Key achievements include:

SynthEO Prototype Development: Successfully generated high-resolution synthetic landcover and optical satellite data. While initial plans for synthetic landcover generation via GANs were adjusted to a Diffusion-based approach, and the direct generation of landcover data proved challenging due to dataset size, the team adapted by leveraging Geospatial Foundation

Models for landcover data creation. Synthetic RGB data generation was successfully implemented.

Performance Validation: The best model trained on synthetic data achieved results nearly on par with models trained solely on real data (within 0.05 mIoU difference), demonstrating the effectiveness of the generated synthetic data.

Market Studies: Conducted extensive market research, including interviews with stakeholders across five industry verticals (Carbon Markets, Vegetation Management, Mining and Raw Materials, Energy, Environmental Monitoring). This led to the identification of priority customer profiles (Heads of R&D, ML/AI, Innovation in geospatial companies), validation of demand for synthetic EO data, and the establishment of a V1 pricing model.

Projektpartner

- Another Earth EOD FlexCo
- FMPG Synth OG