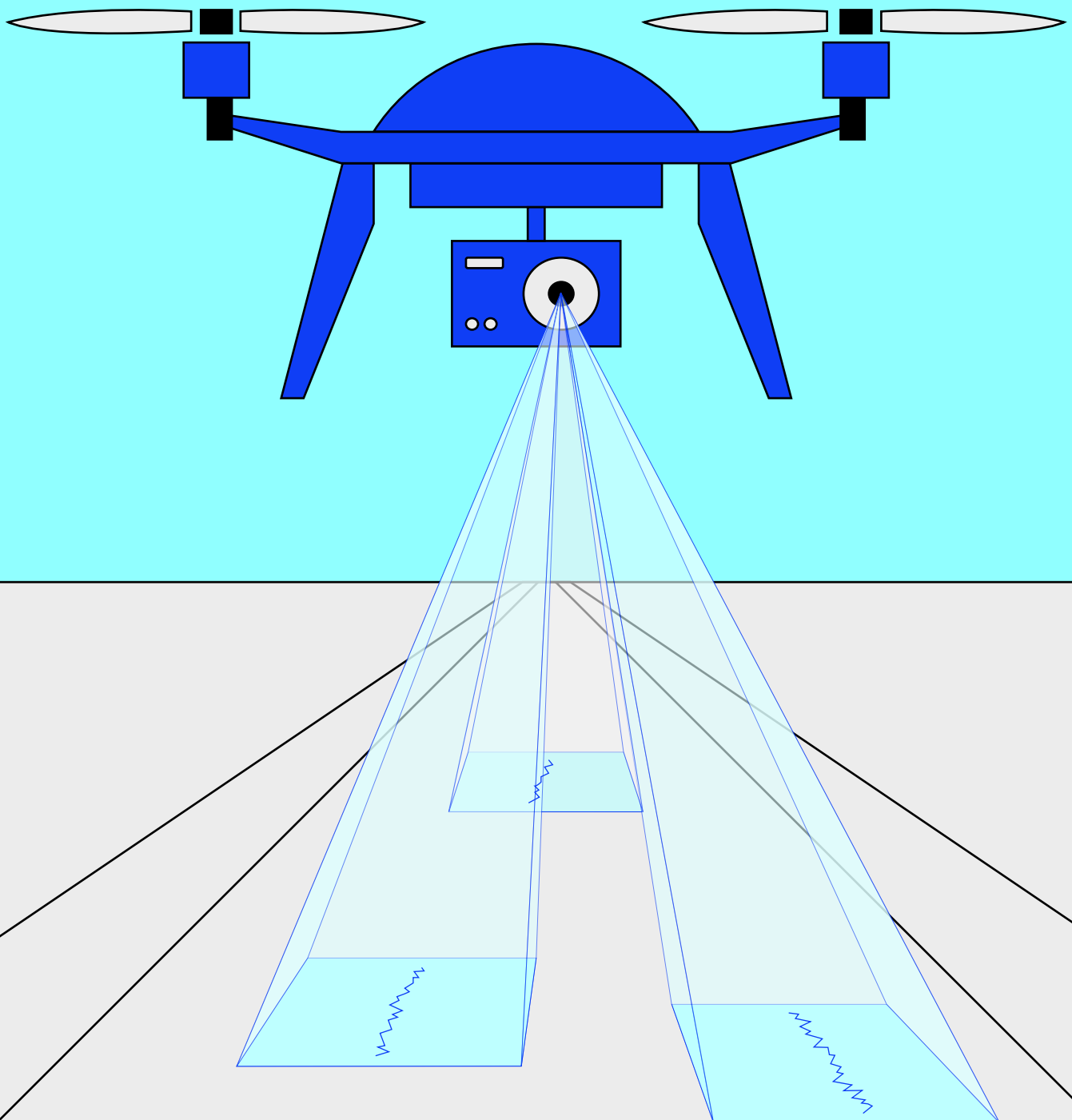


# Automatisierte Infrastrukturwartung

Drohneninspektionen mit Bilderkennung



Die Infrastrukturwartung von Strassen, Brücken und Staumauern bietet grosses Potenzial für den Einsatz von künstlicher Intelligenz (KI). KI-basierte Bildererkennung kann kleinste Risse oder Beschädigungen systematisch und automatisiert erfassen. Infrastrukturbetreiber führen aktuell die meisten Inspektionen noch manuell durch. Im Rahmen der Innovation-Sandbox für KI haben IBM Research und pixmap gmbh auf dem Militärflugplatz in Dübendorf ein Pilotprojekt umgesetzt, um das Potenzial von KI-basierten Inspektionen zu beurteilen. Eine Drohne hat qualitativ hochwertiges Bildmaterial der Landebahn erstellt, um mithilfe von KI-Modellen automatisiert Beschädigungen zu erkennen. Die Erkenntnisse werden dazu verwendet, den Einsatz von KI bei der Inspektion und Wartung von weiteren Infrastrukturelementen voranzutreiben. Das Bildmaterial wird anderen Innovationsakteuren zur Verfügung gestellt. Die Kooperation zwischen Verwaltung, Militär, Forschung und Privatwirtschaft leistet mit diesem Projekt einen Beitrag zur Weiterentwicklung des internationalen KI-Standorts im Metropolitanraum Zürich.

#### **Innovation-Sandbox für Künstliche Intelligenz (KI)**

Das vorliegende Dokument wurde im Rahmen der Innovation-Sandbox für KI erarbeitet. Die Sandbox ist eine Testumgebung für die Umsetzung von KI-Projekten aus verschiedenen Sektoren. Die breit abgestützte Initiative aus Verwaltung, Wirtschaft und Forschung soll verantwortungsvolle Innovation fördern, indem das Projektteam und teilnehmende Organisationen eng mit regulatorischen Fragestellungen arbeiten und die Nutzung von neuartigen Datenquellen ermöglichen.

[Mehr Informationen](#)



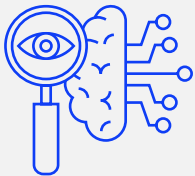
|    |   |          |
|----|---|----------|
| I. | <b>Herausforderungen<br/>manueller Inspektionen</b> | <b>4</b> |
|----|---|----------|

---



|     |  |          |
|-----|--|----------|
| II. | <b>Drohneinsatz zur Erstellung<br/>des Bildmaterials</b> | <b>6</b> |
|-----|--|----------|

---



|      |  |          |
|------|--|----------|
| III. | <b>Automatisierte Inspektion<br/>durch Bilderkennung</b> | <b>9</b> |
|------|--|----------|

---



|     |   |           |
|-----|---|-----------|
| IV. | <b>Potenzial von KI in der<br/>Infrastrukturwartung</b> | <b>14</b> |
|-----|---|-----------|

---



|    |  |           |
|----|--|-----------|
| V. | <b>Anhang: technischer IBM<br/>Research Report</b> | <b>17</b> |
|----|--|-----------|

---

I.

# Herausforderungen manueller Inspektionen



Die Wartung und Inspektion von Infrastrukturelementen wie Strassen, Brücken und Staumauern ist von immenser Bedeutung für die Sicherheit der Bevölkerung und die Aufrechterhaltung von wirtschaftlichen Aktivitäten. Infrastrukturbetreiber stehen allerdings vor einer enorm grossen Aufgabe mit hoher Komplexität: Die Länge des Strassennetzes in der Schweiz beträgt über 84 000 km.<sup>1</sup> Mehr als 40 000 Brücken überspannen Schluchten, Täler und Flüsse.<sup>2</sup> Über 200 Talsperren und Tausende kleinere Stauanlagen machen die Schweiz zum Land mit der höchsten Staudammdichte.<sup>3</sup> Wie kann diese riesige Menge an Infrastrukturelementen effizient überwacht werden? Viele Inspektionsverfahren basieren auf der manuellen Erhebung von Rissen, Beschädigungen und anderen Unregelmässigkeiten des Materials. Die Folge ist eine Reihe von Herausforderungen, die von mangelnder Effizienz bis hin zu menschlichen Fehlern reichen. Das Potenzial der Automatisierung durch KI blieb bisher ungenutzt. Kapitel 1 geht zuerst auf die vier Kernherausforderungen manueller Inspektionen ein und erläutert danach das Potenzial der automatisierten Inspektion durch Bilderkennung.

## Mangelnde Effizienz

Bei manuellen Inspektionen muss ein Mensch physisch die gesamte Fläche einer Infrastruktur, bspw. von Landebahnen oder Strassenabschnitten, überprüfen. Das bedeutet nicht nur, dass manuelle Inspektionen viel Zeit und personelle Ressourcen in Anspruch nehmen, sondern auch, dass die Instandhaltungszyklen oft länger sind, als sie es sein sollten. Das manuelle Abfahren oder Ablaufen von

Infrastrukturelementen führt zu Verzögerungen, die mithilfe von automatisierten Systemen vermeidbar wären. In vielen Fällen ist die manuelle Inspektion auch sehr kostenintensiv.

## Schlechte Dokumentation

Während manuelle Inspektionen zwar Daten über Schäden und Probleme liefern, bleiben sie oft hinter dem zurück, was heute technisch möglich ist. Anstatt eines umfassenden digitalen Abbilds (**Digital Twin\***), das den Zustand einer Infrastruktur über einen längeren Zeitraum kontinuierlich dokumentiert, wird oft eine Liste von problematischen Bereichen oder Schäden erstellt. Die Position der Schäden wird oft nur grob angegeben, was die Behebung erschwert. Ein fehlendes digitales Modell mit automatisierter Erkennung bedeutet auch, dass die Daten oftmals nicht in Betriebssysteme integriert sind und dementsprechend nicht als Entscheidungsgrundlage genutzt werden können.

## Menschliche Fehlerquellen

Menschen machen Fehler. Sie können aufgrund von Konzentrationsproblemen auftreten, insbesondere bei repetitiven und langwierigen Aufgaben wie der Inspektion langer Strassenabschnitte. Darüber hinaus gibt es das Problem der Konsistenz: Was jemand als Schaden ansieht, erachtet jemand anderes als unbedeutend. Nicht erfasste Schäden oder Ver-

\* Die im Text hervorgehobenen Begriffe werden im Glossar erklärt.

wechslungen können zu teuren Reparaturen oder gar zu Infrastrukturrisiken führen.

## Gefährdung der Sicherheit

Manuelle Inspektionen bringen oft Gefährdungen für das Inspektionspersonal mit sich, insbesondere wenn sie an exponierten Lagen, etwa in der Nähe von Hochspannungsleitungen oder in der Höhe, durchgeführt werden müssen. Jedes Jahr gibt es Unfälle, die durch manuelle Inspektionen verursacht werden. Während Sicherheitsausrüstung und -schulungen dazu beitragen können, die Risiken zu reduzieren, bleibt die Frage, ob es sinnvoll ist, Menschen solchen Gefahren auszusetzen, wenn technologische Alternativen verfügbar sind.

## Potenzial für automatisierte Inspektionen

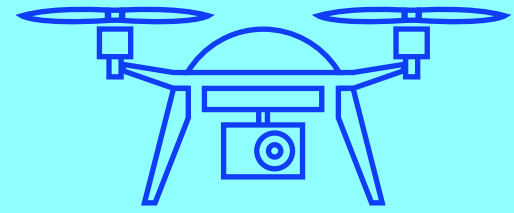
Die genannten Herausforderungen zeigen das grosse Potenzial von automatisierten Inspektionen. Die Fortschritte in der Bilderkennungstechnologie eröffnen neue Möglichkeiten, um einige der Herausforderungen zu überwinden und die Effizienz und die Genauigkeit von Inspektionsverfahren signifikant zu steigern. Im vorliegenden Projekt haben sich verschiedene Akteure aus Verwaltung, Forschung und Wirtschaft zusammengeschlossen, um einen Beitrag zur automatisierten Infrastrukturwartung zu leisten. IBM Research hat im Frühling 2022 einen Projektvorschlag in die Innovation-Sandbox für KI eingereicht. Die Idee war es, Erfahrungen aus vergangenen Projekten mit KI-basierten Brückeninspektionen zu nutzen, um diese auf die Landebahnen von Flughäfen auszuweiten. Die Innovation-Sandbox für KI unterstützte das Innovationsvorhaben, indem qualitativ hochwertige Bilddaten bereitgestellt wurden und indem mit dem Militärflugplatz Dübendorf ein Projektpartner aus dem Innovationsökosystem des Metropolitanraums Zürich gewonnen werden konnte. Die Zusammenarbeit für die Erstellung des Bildmaterials erfolgte mit der pixmap gmbh, die auf die Inspektion und Vermessung durch Drohnen und Flugroboter spezialisiert ist.

Der Report ist nachfolgend in die folgenden Teile gegliedert: Kapitel 2 beschreibt das Vorgehen und die Erkenntnisse aus den drei Drohnenmissionen von pixmap gmbh. Anschliessend gibt Kapitel 3 einen Überblick über die automatisierte Auswertung des Bildmaterials durch die Bilderkennungstechnologie von IBM Research. Kapitel 4 zieht ein Fazit und zeigt mögliche Handlungsfelder, um die automatisierte Inspektion von Infrastrukturelementen im Metropolitanraum Zürich voranzutreiben. Der Anhang beschreibt die technischen Details der Bilderkennung durch IBM Research.

- 1 Infrastruktur und Streckenlänge | Bundesamt für Statistik (admin.ch)
- 2 Über Flüsse und Schluchten – Explora (ethz.ch)
- 3 Einzigartige Schweizer Stauseen | House of Switzerland

## II.

# Drohneneinsatz zur Erstellung des Bildmaterials



Die Grundlage für eine automatisierte Detektion von Rissen und Beschädigungen von Infrastrukturelementen ist umfassendes Bildmaterial. Es gibt verschiedene Methoden und Quellen, um zu solchen Daten zu gelangen. In manchen Fällen reicht es, frei zugängliche Satellitenaufnahmen zu verwenden, bspw. wenn die Infrastrukturschäden auch bei tiefer Auflösung klar ersichtlich sind. Falls jedoch für eine detaillierte Inspektion hochauflösendes Bildmaterial benötigt wird, kommt man um eine spezifische Aufnahme nicht herum. Die Erfassung kann mit einem Fahrzeug erfolgen oder aus der Luft, z.B. mit einer Drohne. Im vorliegenden Fall hat die Firma pixmap gmbh Drohnen eingesetzt. Drohnen haben den grossen Vorteil, systematische Aufnahmen zentimetergenau, wiederholbar und in hoher Auflösung machen zu können. Zudem sind sie leicht transportierbar und relativ kostengünstig zu betreiben.

Für das Pilotprojekt stellte der Militärflugplatz Dübendorf seine Landebahn von über 2,8 km Länge zur Verfügung. Durch den Flugplatzbetreiber wurde ein repräsentativer Pistenabschnitt von 200 x 40 m definiert. Pixmap gmbh nahm diesen Bereich in der geforderten Maximalqualität mittels Drohne auf und stellte IBM Research die Daten zur Auswertung zur Verfügung. Die sehr hohe Auflösung erlaubte es dem KI-Team von IBM, im Nachgang zu evaluieren, ob auch weniger hochauflösendes Bildmaterial für eine automatisierte Auswertung der Risse und Beschädigungen ausreichen würde. Dies ist insbesondere im Hinblick auf den operativen Einsatz von automatisierten Inspektionen relevant, bspw. falls die Höchstqualität zu teuer wäre oder die Datenerhebung für regelmässige Wartungsarbeiten zu lange dauern würde.



Abbildung 1: Die 2,8 km lange Piste des Militärflugplatzes in Dübendorf, Kanton Zürich

Für die Planung des Drohneneinsatzes im vorliegenden Kontext mussten einige Herausforderungen beachtet werden.

**Regulierung:** Je nach Mission und Einsatzort gibt es regulatorische Anforderungen für den Einsatz von Drohnen. Mit der Übernahme der Drohnenregulierung der EU/EASA per 1. Januar 2023 wurden die gesetzlichen Vorgaben nochmals strenger. Mittels Betriebsbewilligung muss der Betreiber der Drohne nachweisen, dass sowohl Risiken für Personen am Boden (Ground Risks) als auch Kollisionen mit anderen Luftfahrzeugen (Air Risks) mit grösster Wahrscheinlichkeit vermieden werden können. Im vorliegenden Fall wurde die Piste gesperrt und die Flughöhe der Drohne auf maximal 10 m limitiert, womit der übrige Flugplatzbetrieb mit Helikoptern nicht tangiert wurde.

**Anforderungen an die Drohne/Kamera:** Für eine Aufnahme in Submillimeterauflösung muss eine

qualitativ hochwertige Kamera auf einer Drohne mit Präzisions-**GPS** eingesetzt werden. Moderne Drohnen lassen sich mit Vollformatkameras (Sensorgröße von 36 × 24 mm) und Auflösungen von 40 bis 100 MP bestücken. Wichtig ist auch eine Kamera mit kurzen Verschlusszeiten, um Bewegungsunschärfe zu vermeiden, und einem kurzen Auslöseintervall, im vorliegenden Fall 0,7 s. Um die Drohne zentimetergenau über die Piste lenken zu können, muss sie mit einem speziellen GPS-System (**RTK-GNSS**) ausgerüstet sein und autonom nach vordefinierten Wegpunkten fliegen können.

**Wetterabhängigkeit:** Es ist zu beachten, dass solche Missionen wetterabhängig sind. Die Piste muss trocken sein, die Drohne darf keinen Schatten auf die aufgenommene Fläche werfen, und starke Windböen sind zu vermeiden. Bei der Planung sind Reservetermine für die Drohnenaufnahmen deshalb wichtig.

Im vorliegenden Projekt konnte die Firma pixmap gmbh die Drohnenmissionen wie geplant am 13. Mai 2023 durchführen.



Abbildung 2: **Quadkopter** beim Überfliegen der Piste auf ca. 5 m Höhe

Um über eine möglichst breite Datenbasis zu verfügen, hat man den Hauptauftrag leicht erweitert, und es wurden drei unterschiedliche Missionen mit verschiedenen Aufnahmeparametern geflogen.

### **Mission 1: Ein neues Niveau der Auflösung**

Die typische Auflösung für Drohnenaufnahmen im Vermessungsbereich liegt zwischen 1 cm und 3 cm. Die vorliegende Mission ging jedoch weit darüber hinaus. Die geforderte Maximalauflösung von 0,25 mm gegenüber 1 cm entspricht einem Faktor von 40- bzw. 1600-mal mehr Punkten pro Flächeneinheit. Dies führte zu extremen Flugparametern: Mit einer Hochleistungskamera ausgerüstet musste die Drohne auf einer Flughöhe von nur 3 m operieren. Eine sehr langsame Fluggeschwindigkeit von 0,7 m/s und ein Bild alle 0,7 s waren nötig, um Bilder mit leichter Überlappung zu erfassen und Bewegungsunschärfe zu vermeiden. Dies führte zu sehr langen Flugzeiten von total 2 h bei etwa 11 500 Bildern.

### **Mission 2: Tiefere Auflösung als Alternative**

Während Mission 1 auf maximale Auflösung fokussierte, zielte die zweite Mission auf Skalierbarkeit ab, die Möglichkeit also, später eine ganze Landebahn aufnehmen zu können. Eine etwas geringere Auflösung von 0,75 mm erlaubte eine 10-mal schnellere Erfassung mit entsprechend kleinerem Datenvolumen.

### **Mission 3: Fokus auf Kartierung**

Diese Mission zielte darauf ab, mittels Fotogrammetrie ein noch umfassenderes Gesamtbild der Landebahn zu erstellen. Alle Einzelbilder wurden dabei durch Spezialsoftware (Pix4Dmapper) zu einem einzigen, georeferenzierten **Orthofoto** und einem Höhenmodell verrechnet. Anstelle von Tausenden einzelnen Aufnahmen entstand so ein einziges, unverzerrtes Gesamtbild, das dann weiter analysiert werden konnte. Eine fotogrammetrische Erfassung ist aber nur mit einer wesentlich höheren Überlap-

pung der einzelnen Aufnahmen möglich, was wiederum die erzielbare Auflösung schmälert. Möchte man die gesamte Piste fotogrammetrisch aufnehmen, kann mit heutiger Technik eine Auflösung von rund 1,5 bis 2 mm erzielt werden. Für den im vorliegenden Projekt festgelegten Pistenabschnitt konnte eine Auflösung von 0,6 mm erreicht werden.

## Schlussfolgerungen der Drohnenmissionen

Die Bilanz ist positiv: Alle Missionen konnten beim ersten Versuch erfolgreich durchgeführt werden. Auf Anrieb liessen sich hochwertige Daten ohne Erfassungslücken generieren. Die umfassende Auswertung durch IBM Research (siehe Kapitel 3) zeigt, dass das Ziel, Pistenschäden automatisch zu erkennen, selbst mit der leicht verminderten Auflösung von 0,75 mm (M2) zu erreichen ist. Das bedeutet, dass sich das hier angewendete Verfahren bereits heute auf ganze Landebahnen oder längere Strassenabschnitte applizieren lässt. Entsprechende Missionen müssen mit dem notwendigen Know-how sorgfältig geplant und mit hochwertiger Ausrüstung durchgeführt werden mit Fokus auf Zielsetzungen, die der Betreiber präzise festlegt. Mit fortschreitender Entwicklung der Drohrentechnik werden auch erweiterte Anforderungen wie eine fotogrammetrische Erfassung im Submillimeterbereich bald möglich sein, und der Aufwand wird tendenziell abnehmen. Der Einsatz von Drohnen zur Inspektion von Infrastrukturelementen ist definitiv von grosser Bedeutung.

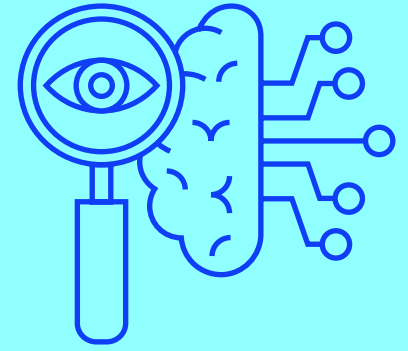
|                     | <b>Mission 1 (M1)</b> | <b>Mission 2 (M2)</b>            | <b>Mission 3 (M3)</b>  |
|---------------------|-----------------------|----------------------------------|------------------------|
| Zielsetzung         | Beste Auflösung       | Skalierbarkeit auf gesamte Piste | Mapping/Fotogrammetrie |
| Auflösung           | 0,25 mm               | 0,75 mm                          | 0,60 mm                |
| Fluggeschwindigkeit | 0,7 m/s               | 4,2 m/s                          | 1,1 m/s                |
| Flugzeit            | 120 min               | 10 min                           | 60 min für ca. 1/3     |
| Anzahl Aufnahmen    | ~ 11 5000             | ~ 1200                           | <b>1 Orthofoto</b>     |

Tabelle 1: Vergleich der drei Drohnenmissionen



### III.

# Automatisierte Inspektion durch Bilderkennung



IBM Research nutzte das hochaufgelöste Bildmaterial aus den drei Drohnenmissionen von pixmap gmbh als Grundlage für die Entwicklung von automatisierten KI-basierten Inspektionsmethoden.

Durch Fortschritte im Bereich **Deep Learning** sind immer mehr Anwendungen in der Bilderkennung realisierbar, die früher als unmöglich galten. Die rasanten technologischen Entwicklungen werden durch die Verfügbarkeit von **annotierten Daten** und spezialisierter Hardware, bspw. **GPU** (Graphics Processing Unit), vorangetrieben, die das Training von KI-Modellen erleichtern. Die Methode gehört zum Bereich **Machine Learning**, wobei das Modell aus grossen Mengen von Daten lernt. Der aktuelle Stand der Technologie ermöglicht im vorliegenden Projekt die Erkennung von kleinen Rissen in der Infrastruktur.

Im vorliegenden Projekt bestand die Herausforderung darin, dass sehr selten **annotierte Daten** im Bereich der Bauinfrastruktur öffentlich verfügbar sind und die meisten Bilddaten keine sichtbaren Defekte zeigen. Lösungen wie **Few-Shot Learning**, **Transfer Learning** und **Self-Supervised Learning** werden erforscht, um diese Herausforderung zu überwinden.

IBM Research Zürich hat im Rahmen dieses Projekts spezielle Expertise im Bereich der automatisierten visuellen Inspektion von Bauinfrastrukturen eingebracht, insbesondere bei Betonpfeilern von Brücken. Ziel der eingesetzten Technologien ist es, durch automatisierte Inspektionen die Kosten der Infrastrukturwartung zu senken, eine systematische Dokumentation der Bauwerke zu ermöglichen und Risikobewertungen vorzunehmen. Dabei wurden auch aktuelle Einschränkungen diskutiert und Empfehlungen für die Zukunft gegeben. Der Anhang dieses Reports beschreibt die Methoden und Tech-

nologien hinter der automatisierten visuellen Defekterkennung von IBM Research im Detail.

## Datenorganisation

Pixmap gmbh führte drei Erfassungsmissionen mit unterschiedlichen Anforderungen durch, was vor allem im Hinblick auf die Datenvolumina relevant war. IBM Research konnte so die Effizienz und die Genauigkeit der gesammelten Bilddaten optimieren. M1 zielte auf höchste Detailgenauigkeit, was jedoch direkte Auswirkungen auf die Gesamtflugzeit, die Auswertungszeit und den Umfang der Datenverarbeitung und der Kosten hatte. Um diese Probleme zu lösen und Kosten zu senken, war die Datenorganisation bei M2 weniger anspruchsvoll. M2 erfasste weniger detaillierte Bilder mit einer **GSD** (Ground Sampling Distance) von 0,75 mm pro Pixel, also einem 3-mal kleineren Detailreichtum als bei M1. Dadurch verkleinerte sich das Datenvolumen um das 9-fache, was bei der Datenverarbeitung Zeit und Speicherplatz sparte. IBM Research hat zudem festgestellt, dass weniger überlappende Bilder und etwas mehr Unschärfe zusätzlich Verarbeitungszeit einsparen und die Datenmenge reduzieren. Diese Änderungen sorgten für eine effiziente Datenorganisation, während die Qualität der Daten für Analysen ausreichend hoch blieb. M3 diente vorrangig als Vergleichsmission mit spezifischen Überlappungsanforderungen und deckte weniger Gesamtfläche ab. IBM Research analysierte von jeder Mission die Menge an Bilddaten separat, um die optimale Methode für die Organisation und Verarbeitung von Daten in Bezug auf Detailgenauigkeit und Effizienz zu finden.



Abbildung 3: Originalbild von M1. Alle Bilder von M1 haben eine Auflösung von 0,25 mm/Pixel, führen jedoch zu einem sehr hohen Datenvolumen.

## Algorithmische Bildzusammensetzung

Das Hauptziel für alle drei Missionen war es, viele einzelne Bilder zu einem grossen, detaillierten und genauen Übersichtsbild zusammenzufügen, das die gesamte Landebahn zeigt. Die Herausforderung lag in der extremen Genauigkeit, die notwendig war, um Risse und Schäden detailliert zu erkennen.

IBM Research hat einen speziellen Algorithmus entwickelt, der die Bilder in mehreren Phasen effizient und genau zusammenfügt. Die Methode konzentrierte sich darauf, die Positionierungs- und Ausrichtungsfehler zu minimieren, um ein kohärentes und präzises Gesamtbild zu erstellen. Damit sich das immense Datenvolumen effektiv verarbeiten liess, wurde der Datensatz in kleinere Segmente unterteilt, die unabhängig voneinander bearbeitet und später zusammengeführt wurden.

**Anspruchsvolle Bildzusammensetzung:** Das GPS-System, das von pixmap gmbh genutzt wurde, konnte die Positionierung der einzelnen Bilder bis auf eine Abweichung von 1 cm optimieren. Das entspricht einem viel höheren Detailgrad als bei herkömmlichen Aufnahmen und macht die Bildzusammensetzung sehr anspruchsvoll.

**Komplexe Algorithmen:** Der entwickelte Algorithmus arbeitet in mehreren Stufen, um Daten zu optimieren, auszurichten und schliesslich zu einem grossen Bild zusammenzufügen.

**Datenmanagement:** Um die riesige Datenmenge zu bewältigen, wurde sie in kleinere, handhabbare Teile zerlegt, die später wieder zusammengeführt wurden.

Für eine effiziente Bearbeitung wurden Methoden implementiert, um bereits bearbeitete Daten zu speichern und wiederzuverwenden. Dadurch liess sich wertvolle Rechenzeit einsparen.

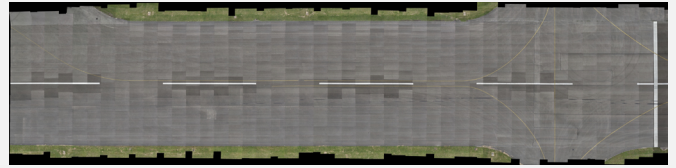


Abbildung 4: Übersichtsbild des Landebahnabschnitts von M2, basierend auf 816 Originalbildern. Auf dem Übersichtsbild kann mit einem Zoom jede beliebige Stelle im Detail digital inspiziert werden.

## KI-Technologie für zuverlässige Risserkennung

Das **Foundation Model** von IBM Research, das für die Risserkennung entwickelt wurde, basiert auf einem automatisierten visuellen Inspektionsmodell, das mithilfe von Vision-Transformer-Architektur und **Self-Supervised Learning** ausgebildet wurde. Es durchläuft mehrere Stufen: Am Anfang wird es mit allgemeinen Bildern trainiert, dann mit bereits verfügbaren Bildern ziviler Infrastrukturen wie Brücken, um anschliessend für die spezifische Aufgabe – das Erkennen von Rissen – verfeinert zu werden. Dieses Modell wurde auf die Bilder der Missionen M1 und M2 angewendet und lieferte zuverlässige Ergebnisse, auch wenn veräusserte Detektionen an den Rändern zwischen Bordstein und Landebahnoberfläche auftraten. Die Erkennung von Rissen, die auf Bildern sichtbar sind, ist mit diesem Modell zuverlässig möglich. Die Daten von M2 sind ausreichend für die Analyse und Erkennung von relevanten Rissen und bieten ein gutes Verhältnis von Qualität und Aufwand, sodass die Erfassung der gesamten Landebahn in einem halben Tag machbar ist.



Abbildung 5: Die automatisch detektierten Risse werden rot eingefärbt und mit Informationen zur Länge und Konfidenz des Defekts ergänzt.

| Mission             | M1   | M2   | M2 vs. M1     |
|---------------------|------|------|---------------|
| Rissinstanzen total | 3920 | 2629 | 32,9% weniger |
| Konfidenz (>0,5)    | 691  | 586  | 15,2% weniger |

Tabelle 2: Übersicht über die von M1 und M2 erkannten Risse. Das KI-Modell arbeitete in einem sensitiven Modus, sodass viele Risse erkannt wurden. Bei der Konfidenz sind die Zahlen niedriger; hier wurden nur Rissinstanzen mit einem Vertrauenswert von mindestens 0,5 einbezogen. Der Unterschied von M2 zu M1 ist in diesem Fall anteilmässig deutlich kleiner, was bedeutet, dass das Modell auch unter erschwerten Bedingungen gute Ergebnisse liefert.

## OCL-Tool zur Präsentation von Ergebnissen

Das One-Click-Learning (OCL)-Tool von IBM Research ist ein Forschungsinstrument, das KI-Ergebnisse auf umfassende Weise präsentiert. Es kann KI-Ergebnisse visuell darstellen, generieren, demonstrieren und kontinuierlich weiterentwickeln. Beim hier vorgestellten Projekt wurden die Bild- und Vorhersageansichten zur Darstellung von Rissen und Defekten auf der Landebahn in Dübendorf verwendet.

Funktionen und Ansichten des OCL-Tools:

**Bildbetrachtung:** Erlaubt den Zugriff auf alle bereitgestellten Bilder (>13 000) in strukturierten Ordnern mit Fokus auf die drei unterschiedlichen Missionen und Bildqualitäten sowie die Navigation innerhalb dieser Struktur.

**Vorhersageansicht:** Zeigt Ergebnisse der KI-Modelle und extrahiert zugehörige Attribute wie Risslänge automatisch, um pixelgenaue Segmentierungsmasken von Defekten in der Landebahn darzustellen.

**Übersichtsansicht:** Erlaubt dem Benutzer das Betrachten grosser Landebahnabschnitte und eine Echtzeitnavigation. Es ermöglicht auch das Verständnis des Kontexts, in dem ein Defekt detektiert wurde.

**Zusammengeführte Vorhersagen und Überblick:** Ermöglicht es dem Benutzer, die Zusammenhänge und Positionen aller Defekte zu verstehen, und führt mehrfache Detektionen eines Defekts in verschiedenen Bildern zu einer einzigen Vorhersage zusammen.

**Statistische Zusammenfassung:** Stellt aggregierte Statistiken über alle erkannten Defekte bereit, sowohl für einzelne Bilder als auch für die gesamte zusammengeführte Übersicht.

**Reportingfunktionalität:** Erstellt Reports für alle im Tool erfassten Daten mit Detailansichten jedes Defekts, wichtigen Attributen und Direktverlinkungen zum OCL-Tool für eine vergrösserte Ansicht des visualisierten Defekts. Eine Georeferenzierung erlaubt es dem Pistenwartungspersonal, die Defekte auf der Landebahn manuell zu überprüfen und Schäden mit einem Verguss instand zu setzen.

Im Projekt wurde das OCL-Tool primär für die Demonstration und Überprüfung der Ergebnisse eingesetzt, insbesondere im Kontext von M1 und M2, um die Analyse und Interpretation der grossen Datenmengen zu erleichtern. IBM Research hat einen vollständigen Bericht und spezifische Rissdetektionen für die M2-Daten bereitgestellt. Dies ist insbesondere für die Verwendung der Projektergebnisse durch die zuständigen Infrastrukturbetreiber relevant.

### III. Automatisierte Inspektion durch Bilderkennung

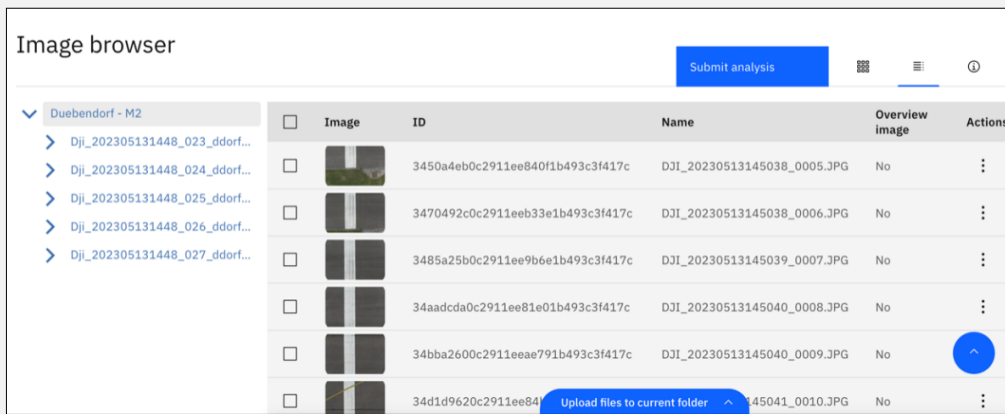


Abbildung 6: OCL-Bildbetrachtung, die den Zugriff auf alle Originaldaten des Projekts ermöglicht

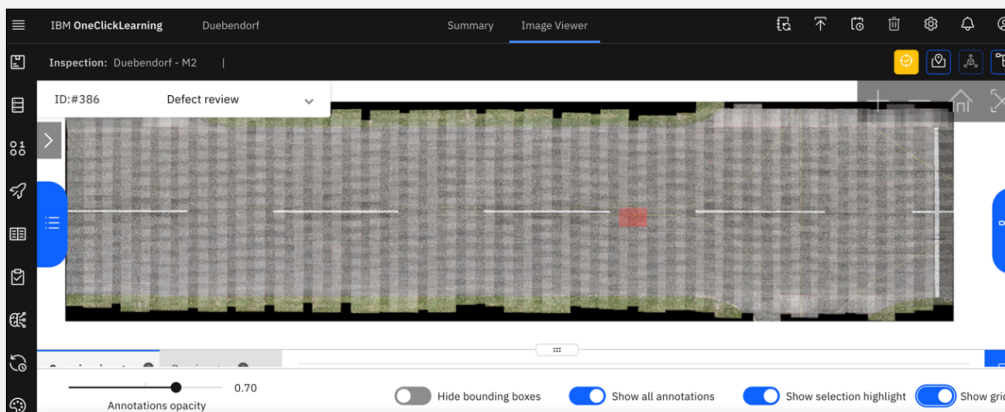


Abbildung 7: OCL-Übersichtsansicht, die den Zugriff auf die gesamte Szene in voller Auflösung ermöglicht

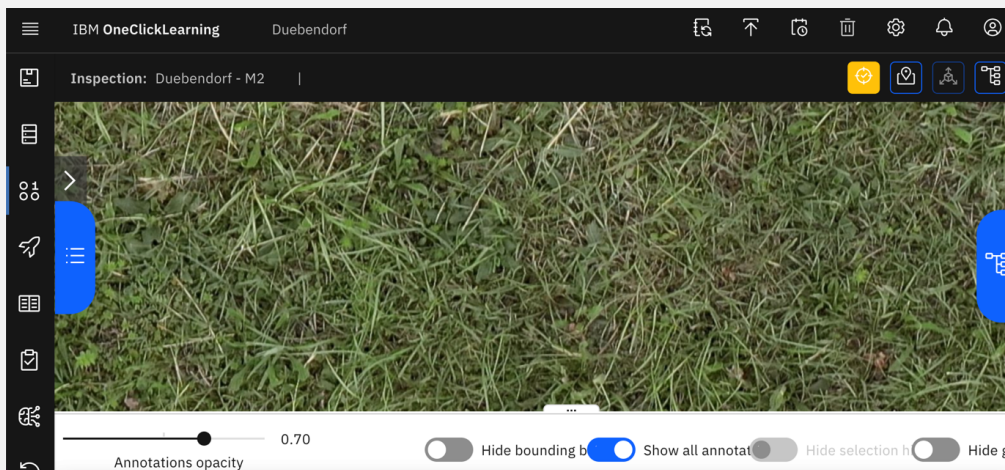


Abbildung 8: OCL-Übersichtsansicht, die Details in voller Auflösung vollständig bewahrt

### III. Automatisierte Inspektion durch Bilderkennung



Abbildung 9: OCL-Bildbetrachtung, auf der Rissdetektionen auf dem Originalbild in Rot markiert sind

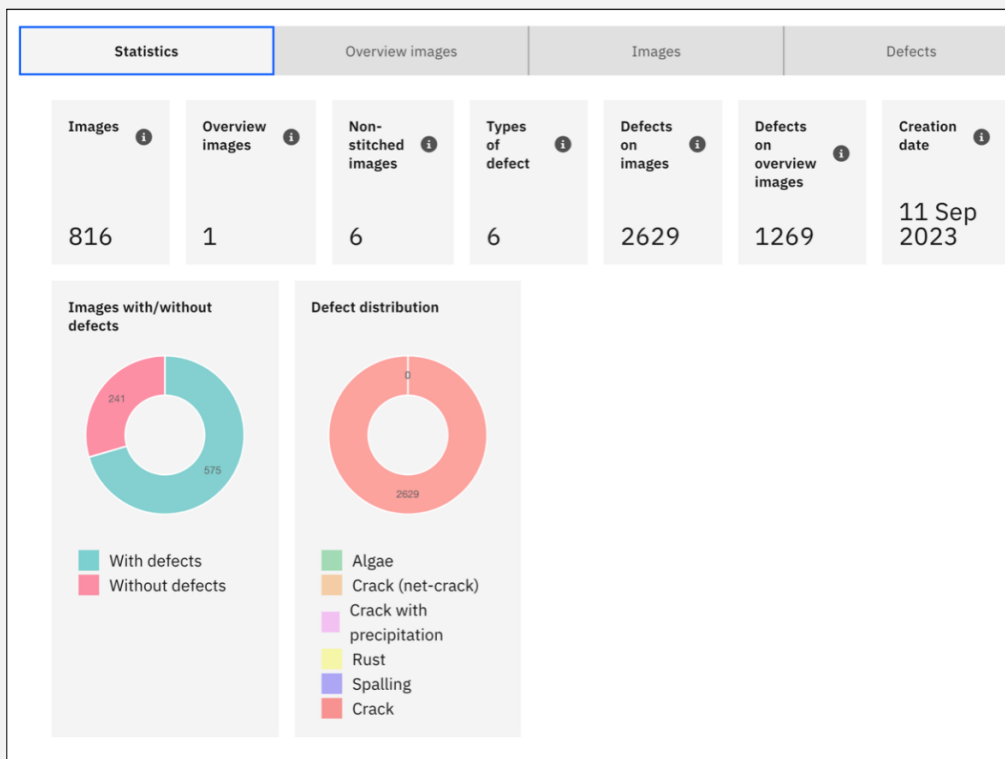


Abbildung 10: Statistische Zusammenfassung der M2-Daten auf OCL

## IV.

# Potenzial von KI in der Infrastruktur- wartung



Das Innovation-Sandbox-Projekt **«Automatisierte Infrastrukturwartung»** hat erfolgreich demonstriert, wie grosse Datenmengen systematisch erfasst und von KI-Modellen bewertet werden können. Risse konnten in allen drei Missionen (M1, M2 und M3) korrekt identifiziert werden. Um in der Praxis grosse Flächen in angemessener Zeit zu scannen, ist es unerlässlich, dass die Erfassung effizient durchgeführt wird. IBM Research hat gezeigt, dass M1 und M2 gegenüber M3 überlegen sind. Zusätzlich besagt eines der Hauptresultate, dass die Auflösung von M2 ausreicht, um die wichtigen Risse zu erfassen, vollständig zu dokumentieren und solide Entscheidungen bezüglich des Gesamtzustands der Infrastruktur zu treffen. Die gesamte Piste kann in einem halben Tag durch eine Drohne erfasst werden, was eine kontinuierliche Überprüfung in der Praxis ermöglicht (bspw. halbjährlich im Frühling und im Herbst, um saisonale Unterschiede fortlaufend zu erfassen).

Hochauflösende Bilddaten von Pisten wie der Landebahn in Dübendorf sind im Normalfall schwer zugänglich. Im Rahmen der Innovation-Sandbox für KI reale Daten nutzen zu können, ist deshalb für den Forschungs- und Innovationsstandort des Metropolitanraums Zürich sehr wichtig. Institutionen wie IBM Research erhalten so die Möglichkeit, die neuesten KI-Algorithmen und -Strategien in einem relevanten Kontext zu bewerten und zu verbessern. Daher wird diese Art von Daten auch in den nächsten Jahren dazu beitragen, die künftigen Entwicklungen der KI-Technologie im Bereich der automatisierten Bilderkennung voranzutreiben.

Zusätzlich wird mit jedem Projekt, in dem KI-Anwendungen erfolgreich eingesetzt werden, immer sicherer, dass die entwickelten **Foundation Models** – in diesem Fall von IBM Research – in einem breiten

Kontext zuverlässig funktionieren. Das bedeutet, dass diese Art der Bilderkennung auch für die Inspektion von Fassaden von Grossgebäuden, Brücken, Staudämmen, Tunneln oder Strassenoberflächen angewendet werden kann.

## Bestätigung der Mehrwerte einer automatisierten Landebahninspektion

Das vorliegende Projekt hat bestätigt, dass die Bilderkennung sehr hohes Potenzial bei der automatisierten Inspektion von Infrastrukturelementen bietet. Auch wenn die KI-Anwendung noch nicht im operativen Betrieb eingesetzt wird, kann man davon ausgehen, dass die in Kapitel 1 genannten vier Herausforderungen der manuellen Inspektion adressiert werden können.

## Mehr Effizienz

Das Projekt hat gezeigt, dass der Einsatz von Drohnen zur Erstellung von Bildmaterial zeitsparend ist. Durch die Verwendung von Bilderkennung kann zudem eine höhere Effizienz erzielt werden, verglichen mit traditionellen Inspektionsmethoden, die ohne KI-basierte Berichterstattung von Bodenpersonal durchgeführt werden. Selbst wenn KI-Technologien zum Einsatz kommen, muss in den meisten Fällen eine Fachperson vor Ort die erkannten Schäden final validieren und beurteilen. Das Bodenpersonal kann die Inspektion allerdings basierend auf einer bestehenden Entscheidungsgrundlage vornehmen. Idealerweise könnten die Reports mit den grössten Schäden direkt an die Wartungsunternehmen wei-

tergeleitet werden, die für die Behebung der Schäden zuständig sind. Dadurch würde der gesamte Prozess optimiert und beschleunigt.

## Bessere Dokumentation

Die Möglichkeit, digitale Abbilder von Infrastrukturelementen zu erstellen und diese kontinuierlich zu überprüfen, stellt einen erheblichen Mehrwert dar. Aktuell fehlt in der Praxis oftmals eine vollständige, personenunabhängige und einheitliche Dokumentation. Die digitale Abbildung ist besonders nützlich für die detaillierte und systematische Überprüfung. Dies fördert die Qualitätssicherung, da bestehende Risse und bereits behobene Schäden präzise und über einen längeren Zeitraum hinweg überwacht werden können.

## Weniger menschliche Fehlerquellen

Durch den Einsatz von automatisierten Systemen können menschliche Fehler minimiert werden, die oft wegen unterschiedlicher und inkonsistenter Einschätzungen von Fachleuten auftreten. Insbesondere bei Personalwechseln oder -ausfällen können traditionelle Inspektionsmethoden zu unterschiedlichen Beurteilungen führen. Die Automatisierung der Inspektionen durch KI ermöglicht eine konsistente und objektive Analyse und Bewertung der Infrastrukturelemente, wodurch die Ergebnisse zuverlässiger und nachvollziehbarer werden.

## Höhere Sicherheit

In gefährlichen Umgebungen wie Staudämmen oder Brücken, können Drohnen zur Durchführung von Inspektionen eingesetzt werden, um die Risiken für die Menschen zu minimieren. Obwohl dies im aktuellen Projekt nicht im Vordergrund stand, ist es ein wichtiger Aspekt, vor allem wenn man bedenkt, dass Inspektoren sich in solchen Umgebungen potenziellen Gefahren aussetzen müssen. Durch den Einsatz von Bilderkennungstechnologie in solchen Umgebungen kann die Sicherheit des Personals erhöht und gleichzeitig eine detaillierte und präzise Analyse der Strukturen gewährleistet werden.

## Ausblick

Die Innovation-Sandbox soll das Innovationsökosystem im Metropolitanraum Zürich stärken. Das vorliegende Projekt ist ein Beitrag dazu. Der flächendeckende Einsatz von Bilderkennung in der Infrastrukturwartung befindet sich allerdings noch in weiter Ferne. Deshalb schlägt das Projektteam folgende Handlungsfelder vor, um das Potenzial am Innovationsstandort Zürich künftig noch besser zu nutzen:

### 1. Integration der automatisierten Inspektion in bestehende Prozesse

Um den Mehrwert von automatisierten Inspektionen zu maximieren, ist es wichtig, diese Technologien nahtlos in die bestehenden Prozesse der Infrastrukturbetreiber zu integrieren. Das beinhaltet unter anderem die Entwicklung von Schnittstellen, um verschiedene Anwendungen miteinander zu verknüpfen und die Resultate in einem Format bereitzustellen, das die Weiterverarbeitung durch Infrastrukturbetreiber ermöglicht. Insbesondere bei Begehungen braucht das Wartungspersonal Zugriff auf ein digitales Modell der Landebahn mit GPS-Funktion, um die automatisiert erkannten Risse vor Ort zu finden und zu verifizieren. Zudem sollten Best Practices über Organisationsgrenzen hinweg verbreitet werden, um Wissen und Erfahrungen effektiv zu teilen und somit die Einführung und Nutzung von automatisierten Inspektionstechnologien zu erleichtern.

### 2. Neue Open-Data-Ansätze im Innovationsökosystem

Für die Stärkung des Innovationsökosystems im Metropolitanraum Zürich ist es essenziell, neue Open-Data-Ansätze zu entwickeln. Dabei sollten basierend auf konkreten Anwendungsfällen aus Wirtschaft, Forschung und Verwaltung mehr Daten verfügbar gemacht werden. Das ermöglicht es anderen Akteuren des Innovationsökosystems, ähnliche Projekte umzusetzen. Durch die Bereitstellung und Nutzung grosser Datenmengen von Landebahnen, Brücken und Staudämmen können innovative Lösungen entwickelt und die Potenziale von KI-basierten Inspektionen besser ausgeschöpft werden.

### **3. Übertragbarkeit auf andere Infrastrukturelemente**

Die Übertragbarkeit der getesteten Technologien und Methoden auf andere Infrastrukturelemente wie Brücken, Strassen und Staudämme muss geprüft werden. Jeder Anwendungsfall bringt spezifische Chancen und Herausforderungen mit sich, weshalb ein interdisziplinärer Austausch über verschiedene Infrastrukturkategorien hinweg wichtig ist. Ein solcher Austausch fördert die Entwicklung angepasster Lösungen für diverse Infrastrukturelemente und ermöglicht so die breite Anwendung innovativer Bilderkennungstechnologien in der Infrastrukturwartung.

### **4. Proaktiver Umgang mit regulatorischen Fragestellungen**

Die Überwindung regulatorischer Hürden ist essenziell für die Implementierung KI-basierter Inspektionen durch Drohnen. Wichtige Massnahmen hierbei sind die Einrichtung von Testumgebungen für experimentelle Anwendungsfällen. Ein intensiver Austausch zwischen Regulatoren wie dem Bundesamt für Zivilluftfahrt (BAZL) und Industrieverbänden (bspw. Drone Industry Association Switzerland) ist ebenfalls wichtig, um Anforderungen frühzeitig zu definieren, Gesetze zukunftsgerichtet anzupassen und Zertifizierungsprozesse zu beschleunigen. Diese Strategien können Innovationsbarrieren abbauen und die Einführung neuer Inspektionstechnologien im Metropolitanraum Zürich und in der gesamten Schweiz fördern.

### **5. Stärkung der gesellschaftlichen Akzeptanz von Drohnen**

Um die Implementierung von KI-basierter Drohnentechnologie für die Infrastrukturwartung erfolgreich zu realisieren, ist die Stärkung der gesellschaftlichen Akzeptanz von Drohnen von entscheidender Bedeutung. Es ist wichtig, einen offenen und inklusiven Dialog mit der Bevölkerung zu führen, um Chancen und Risiken transparent zu kommunizieren und Missverständnisse zu klären. Das Wissen über die Drohnentechnologie in der breiten Gesellschaft liesse sich durch Informationsveranstaltungen, Workshops und Bildungsinitiativen vergrössern. Zusätzlich könnten Begegnungszonen und Erlebnisparks eingerichtet werden, wo Menschen direkt mit Drohnen in Kontakt kommen, ihre Funktionen und Anwendun-

gen erleben und somit ein besseres Verständnis für die Technologie entwickeln können. Diese Massnahmen können dazu beitragen, Vorurteile abzubauen und das Vertrauen in die Technologie zu fördern, was wiederum die Implementierung von innovativen Inspektionstechnologien unterstützt.



v.

# Anhang: technischer IBM Research Report



Das von pixmap gmbh bereitgestellte Bildmaterial lieferte IBM Research die Grundlage für eine automatisierte Inspektion durch KI-basierte Bildererkennung. Die jüngsten Forschungstrends im Bereich **Deep Learning** ermöglichen immer spezifischere Anwendungen, die vor einem Jahrzehnt noch undenkbar waren. Zwei Hauptfaktoren treiben diese Entwicklung voran: Erstens erlaubt die weite Verfügbarkeit annotierter Daten die Entwicklung, das Testen und die Verbesserung von Algorithmen. Zweitens beschleunigt die Verfügbarkeit von spezialisierter Hardware wie **GPU** die Experimentierzeit und ermöglicht das Training, das Testen und die Implementierung grösserer KI-Modelle.

In diesem Kontext wird die automatisierte visuelle Defekterkennung bei der Bauinfrastruktur (bspw. bei Brücken, Gebäuden oder Landebahnen) erst realistisch. Das Ziel ist es, mit automatisierten Inspektionen Wartungskosten zu senken, eine systematische Dokumentation der Infrastruktur zu ermöglichen und eine angemessene Risikobewertung vorzunehmen. Machine-Learning-Techniken eröffnen dabei ein attraktives Geschäftsfeld. Dabei lernt das System aus einer grossen Menge annotierter Daten und wird durch **Self-Supervised Learning** trainiert, um auch Vorhersagen zu neuen, bisher ungesehenen Bilddaten treffen zu können. Deep-Learning-Techniken sind eine Untergruppe von Machine-Learning-Algorithmen, die in vielen Anwendungen der Bildererkennung zuletzt Durchbrüche erzielt haben.

Die typischen Datensätze zur Bildererkennung in der Forschung sind extrem gross; z.B. besteht ImageNet aus über 1 Million annotierten Bildern. In diesem Fall sind für jedes Bild Labels zu verschiedenen Klassen verfügbar. Im Gegensatz dazu stellen

sich bei der Defekterkennung in der Bauinfrastruktur mehrere Herausforderungen: Erstens ist die Menge der öffentlich verfügbaren annotierten Daten in diesem Bereich sehr klein; zweitens sind systematisch gesammelte sichtbare Defekte in bestehenden Datensätzen selten; drittens existieren entweder keine detaillierten Annotationen für die relevanten Segmente (bspw. Risse und Beschädigungen) oder ihre Erstellung ist sehr zeitaufwendig.

Im Bereich **Deep Learning** gibt es neue Ansätze, um diese grundlegenden Probleme zu lösen. **Few-Shot Learning** wird eingesetzt, wenn nur wenige Annotationen verfügbar sind. **Transfer Learning** trainiert KI-Modelle in einem Bereich und wendet sie auf einen leicht anderen Bereich an. Die neuesten **Foundation Models** werden zunächst mit immer grösseren Mengen verfügbarer Daten trainiert. **Self-Supervised Learning** wiederum wird eingesetzt, weil damit Bilder erkannt und rekonstruiert werden können, ohne dass sämtliche Details in den Bildern annotiert sein müssen. Um diese Methode erfolgreich zu nutzen, ist allerdings ein Feintuning mit einigen annotierten Bilddaten nötig.

IBM Research Zürich hat Expertise im Umgang mit Bilddaten im Bereich der automatisierten visuellen Inspektion von Bauinfrastrukturen, insbesondere bei der Inspektion von Betonpfeilern von Brücken. Das Projektziel im Rahmen der Innovation-Sandbox ist es, den ganzen Prozess einer Bildererkennung durchzuspielen und den Mehrwert der automatisierten visuellen Inspektion zu demonstrieren. Zusätzlich wird auf aktuelle Einschränkungen und zukünftige Empfehlungen eingegangen.

Die Beiträge von IBM Research in diesem Sandbox-Projekt sind:

- Vergleich von drei Erfassungsmethoden (Missionen 1 bis 3)
- Bildzusammensetzungstechnologie für geringe Überlappungsbereiche
- Demonstration von KI-Modellen zur Rissdetektion
- Präsentation aller Ergebnisse über das OCL-Tool
- kurzes Video zur Zusammenfassung der Ergebnisse und zur Verwendung des OCL-Tools
- Berichterstattung über alle Erkenntnisse in einem PDF-Dokument

## Organisation grosser Datenmengen

Drei Erfassungsmethoden wurden von pixmap gmbh durchgeführt (siehe Kapitel 2), um Informationen zu den Auswirkungen auf die Flugzeit zu erhalten und die Ergebnisse der Bilderkennung zu vergleichen. IBM Research hat die Bilddaten für alle drei Missionen unabhängig voneinander verarbeitet, um abschliessende Erkenntnisse für die Varianten zu gewinnen und einen direkten Vergleich der Erfassungsmethoden zu ermöglichen.

M1 hatte mit 0,25 mm/Pixel eine konservative Anforderung an die Auflösung. Das anspruchsvolle Ziel wurde durch pixmap gmbh in allen erfassten

Bilddaten von M1 erreicht. Die sehr hohen Anforderungen hatten direkte Auswirkungen auf die Gesamtflugzeit, die Auswertungszeit und das Datenvolumen. Insbesondere das Datenvolumen beeinflusst, wie effizient Daten gespeichert, verarbeitet und analysiert werden können.

Um den drei herausfordernden und kostenintensiven Faktoren entgegenzuwirken, hat IBM Research in M2 die Anforderungen an die **GSD** gelockert. **GSD** beschreibt die Distanz zwischen zwei aufeinanderfolgenden Pixeln, die auf dem Boden gemessen werden. Die Anforderungen waren bei M2 mit einer Ziel-**GSD** von 0,75 mm/Pixel um den Faktor 3 kleiner als bei M1. Die Auswirkungen waren vielfältig: Die tieferen Anforderungen hatten einen exponentiellen Effekt auf das erwartete Datenvolumen pro Flächeneinheit und reduzierten dieses um das 9-fache. Des Weiteren gab es weniger Überlappung und eine etwas höhere Toleranz für Bewegungsunschärfen, was die Erfassungszeit und das Datenvolumen reduzierte. Doch jede Abweichung von den standardmässig empfohlenen Einstellungen birgt das Risiko einer höheren Fehlerquote beim Erstellen der Übersichtsbilder und/oder unterschiedlicher Ergebnisse bei der KI-Analyse. Insbesondere das Risiko, winzige Defekte zu übersehen, war bei M2 höher, da der theoretisch kleinste erkennbare Defekt sich bei M2 im Vergleich zu M1 um den Faktor 3 vergrössert. Nichtsdestotrotz hat sich die Methode

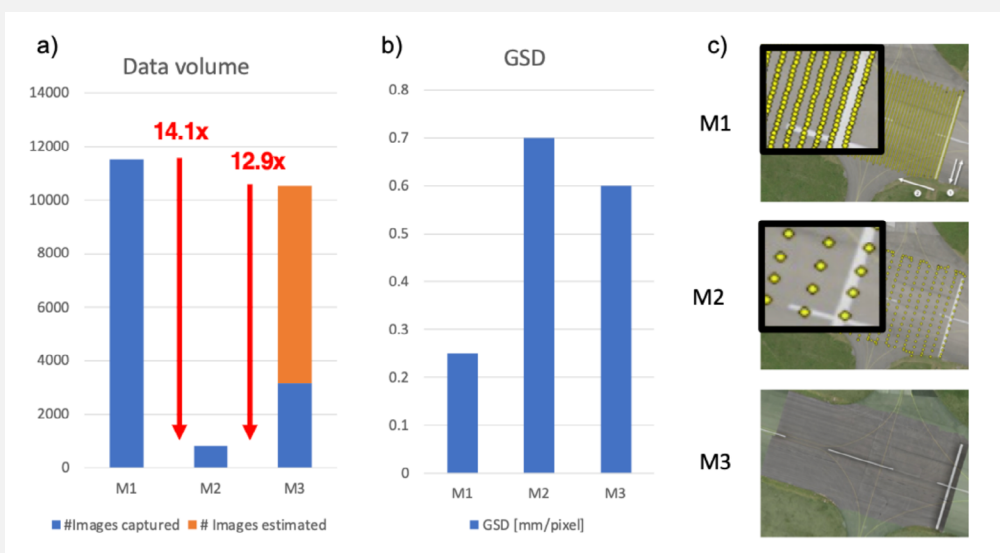


Abbildung 1: Übersicht über M1, M2 und M3; a) Anzahl der erfassten Bilder; b) **GSD** pro Mission; c) Vorschau darauf, wie die Daten erfasst wurden. M3 wurde nur auf 30% der Gesamtfläche durchgeführt, daher wurde die geschätzte Anzahl zusätzlicher Bilder hinzugefügt, die erforderlich wären, um dieselbe Fläche von 200 x 40 m wie in M1 und M2 abzudecken.

von M2 als praktikable Lösung erwiesen, die 10-mal schneller ist und denselben Bereich abdecken kann wie die Methode von M1. M3 wurde mit hohen Überlappungsanforderungen durchgeführt, die denjenigen von externen Tools entsprechen. pixmap gmbh lieferte zusätzlich das georeferenzierte M3-**Orthofoto** als Vergleich. M3 deckte nur etwa 30% des gesamten Testbereichs ab, der bei M1 und der schnellen M2 vollständig abgedeckt wurde.



Abbildung 2: Repräsentatives Beispiel der ursprünglich erfassten Bilder von M1. Die Bilder haben eine Auflösung von 0,25 mm/Pixel.



Abbildung 3: Repräsentatives Beispiel der ursprünglich erfassten Bilder von M2. Die Bilder haben eine Auflösung von 0,75 mm/Pixel. Im Vergleich zu M1 hat M2 eine 3-mal grössere GSD. Das wird deutlich an der ungefähr 9-mal grösseren Fläche, die das Referenzbild von M2 abdeckt.

## Algorithmus zur Bildzusammensetzung

Ein wichtiges Ziel des Projekts war es, aus all den Einzelaufnahmen ein einziges Übersichtsbild zu erstellen, das als digitales Modell den ganzen Abschnitt der Landebahn zeigt und diesen für Ana-

lysen durchsuchbar macht. Typischerweise erfordern Bildzusammensetzungsalgorithmen hohe Überlappungen von etwa 70%, um zuverlässig zu funktionieren. Zusätzlich müssen Drohnenmissionen für kartografische Anwendungen normalerweise schwächere Anforderungen mit Genauigkeit von einigen Zentimetern erfüllen. Diese Umstände waren ausschlaggebend dafür, dass die Bilddaten des Projekts an der Grenze des technologisch machbaren erstellt wurden. Die hohen Anforderungen führen zudem dazu, dass Ungenauigkeiten des GPS-Standorts, Abweichungen von der geplanten Route aufgrund externer Bedingungen (bspw. Wind) und kamerabezogene Fehler (bspw. Abweichungen von der erwarteten Ausrichtung) einen stärkeren Einfluss haben. Doch für die konsequente Identifizierung von einzelnen Rissen ist eine pixel- und submillimetergenaue Standortausrichtung erforderlich. Reguläre GPS-Standortempfänger erreichen unter moderaten Annahmen im Gegensatz dazu nur eine metergenaue Positionierung. In diesem Projekt verwendete pixmap gmbh ein **RTK**-fähiges GPS, das eine nominale Genauigkeit von 1 cm liefert. Üblicherweise entspricht die Positionsgenauigkeit der Aufnahmen ungefähr 2 bis 3 **GSD**, also rund 2- bis 3-mal der Auflösung der Aufnahme. Im vorliegenden Fall jedoch liegt dieser Faktor bei 40 oder mehr. Dies macht die Bildrekonstruktion für M1 und M2 herausfordernd.

Für dieses Projekt hat IBM Research die üblicherweise verwendete Bildzusammensetzungstechnologie weiterentwickelt, um die beschriebenen Probleme zu lösen. Die verwendete Variante des Algorithmus arbeitet in drei Hauptstufen. Erstens werden Informationen extrahiert, um ein globales Optimierungsproblem zu definieren. Zweitens wird das Optimierungsproblem gelöst. Und drittens werden alle Daten nachverarbeitet, um alle Kacheln zu erstellen, welche die vom Betrachter angezeigten Daten füllen. Der erste Schritt beinhaltet eine Nachbarschaftssuche von eng beieinanderliegenden Bildern basierend auf den GPS-Metadaten, um die Verarbeitungszeit zu verbessern. Für jedes etablierte Paar wird ein zweistufiges Bildzusammensetzungsproblem gelöst, um die Bildinstanzen visuell auszurichten. Entsprechende Punktpaare aus beiden Bildern werden für die abgeglichenen Schlüsselpunkte, welche die Filterkriterien erfüllen, extrahiert. Die beiden Cluster von Schlüsselpunkten müssen im finalen Übersichtsbild übereinstimmen, was Fehler-

gleichungen für ein einzelnes Paar definiert. Paargleichungen werden für die Sammlung von etablierten Paaren extrahiert. Aufgrund der planaren Natur der vorliegenden Daten wird das endgültige Optimierungsproblem direkt im 2-D-Bildraum des Übersichtsbilds formuliert.

Die Verlustfunktion berechnet, wie gut das Modell basierend auf den gegebenen Werten Vorhersagen trifft. Im vorliegenden Fall besteht der Verlust (Loss) aus drei Komponenten: visuelle Ausrichtung, Grössenregularisierung und Positionsregularisierung. Der Verlust der visuellen Ausrichtung wird über die Paargleichungen berechnet, welche die Schlüsselpunkte für beide Bilder ausrichten. Die Grössenregularisierung wird über die Länge der projizierten Bildkanten eingeführt, wodurch sichergestellt wird, dass sie einer Referenzlänge der ursprünglichen Bildgrösse entsprechen. Ohne diesen Verlust neigt das Optimierungsproblem dazu, Bilder zu verkleinern, da dies zwar nicht die Ausrichtung, aber dafür alle Fehlergleichungen konsequent verbessert. Die Positionsregularisierung wird eingeführt, damit alle projizierten Mittelpunkte der optimierten Bildpositionen dasselbe Muster befolgen wie in der GPS-Position jedes Bildes. Wie oben dargestellt, ist die nominale Genauigkeit des GPS begrenzt, daher wurde ein relativ schwacher Regularisierungsfaktor verwendet, um mögliche in diesen Informationen vorhandene Fehler zu berücksichtigen. Dennoch hilft der Positionsregularisierungsverlust, Bilder global genau auf die Positionen zu setzen, auf die sie gehören. Ohne diesen Schritt neigt der Algorithmus dazu, die Ausrichtung nahtlos zu lösen, damit paarweise Überlappungen zwischen den Bildern gut aussehen. Gleichzeitig besteht so aber die Tendenz, dass sich Fehler ansammeln und ein Driftmuster sichtbar wird, wenn die Bilder nicht global korrekt platziert sind. Unter Berücksichtigung aller drei Verlustfaktoren konnte IBM Research die Daten von M1 und M2 zuverlässig verarbeiten, um Übersichtsergebnisse für die gesamte Szene zu produzieren.

Die Rekonstruktion für die M2-Daten erfolgt direkt für die gesamte Szene, die aus 816 Bildern besteht. Für  $n$  Bilder existieren  $n \times (n-1) / 2$  Paare, was insgesamt 332 520 Bildpaare ergibt. Durch die Anwendung des Clustering auf die GPS-Daten und die Berücksichtigung von Bildpaaren, bei denen die Drohne in einer 5-m-Nähe positioniert wurde, wird die Anzahl der betrachteten Paare auf 19 023 Paare

reduziert. Dadurch wird die Anzahl der Gleichungen um 94,3% reduziert.

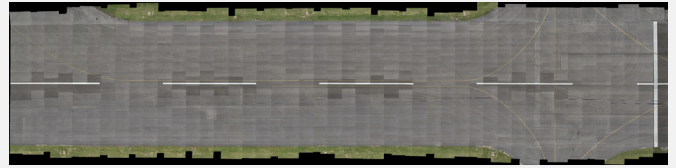


Abbildung 4: Vorschau auf die Übersicht von M2. Selbst in der herausfordernden Konfiguration mit geringer Überlappung war der Zusammensetzungsalgorithmus in der Lage, den gesamten Landebahnabschnitt bestehend aus 816 Originalbildern vollständig zu rekonstruieren. Das Testgebiet umfasst 200 m Landebahn, die eine Breite von 40 m hat. Das Übersichtsbild ist mit einer **GSD** von 0,75 mm/Pixel verfügbar.

M1 bestand aus einer grossen Datenmenge von über 10 000 Dateien, was es schwierig machte, alles auf einmal zu verarbeiten. Um solche Daten effizienter zu behandeln, die Arbeit pro Aufgabe zu reduzieren, potenziell stark von paralleler Aufgabebearbeitung zu profitieren, eine teilweise Wiederherstellung zu ermöglichen und Iterationszyklen beim Debugging zu verkürzen, teilte IBM Research die Daten in Gruppen auf, die unabhängig voneinander behandelt wurden. Dieses Teile-und-herrsche-Verfahren zeigte, dass es sinnvoll war, ein Limit von maximal 200 Bildern pro Optimierungslauf einzuhalten. Um die Struktur des Projekts beizubehalten, die eine Lieferung der M1-Daten in 14 Blöcken vorsah, wurde jeder Block in 5 Datenpakete unterteilt. Daher wurden 70 unabhängige Probleme gelöst, die im Durchschnitt jeweils aus 164,7 Bildern bestanden, wobei das oben beschriebene Vorverarbeiten, die globalen Optimierungsverläufe und das Nachverarbeiten befolgt wurden.

Um Daten weiterhin effizient zu verarbeiten, hat IBM Research Zusammenführungsfunktionen festgelegt, die die Ergebnisse mehrerer gelöster Teilprobleme nehmen und das daraus resultierende grössere Problem lösen. Dieser Prozess wurde effizient implementiert, sodass bereits in einem Teilproblem geleistete Arbeit gespeichert und wiederverwendet werden konnte. Zum Beispiel wurde die Extraktion aller Merkmale und ihrer Feature-Vektoren nicht neu berechnet, sondern neu geladen, wodurch eine erhebliche Menge an Rechenzeit eingespart wurde. Zusätzlich wurden Fehlergleichungen für die Zusammenführungsprobleme nur für diejenigen Bildpaare aktualisiert, die neue Gleichungen

für Bildgruppenpaare verbunden. Diejenigen, die bereits in vorhandenen Gruppen bestanden, wurden neu geladen. Der Algorithmus konnte alle M1-Daten in jedem der 70 Blöcke problemlos verarbeiten. IBM Research führte die Zusammenführungsfunktion 14-mal aus, um alle 5 Teilblöcke in jeden Block zusammenzuführen; so, wie die Daten bereitgestellt wurden.

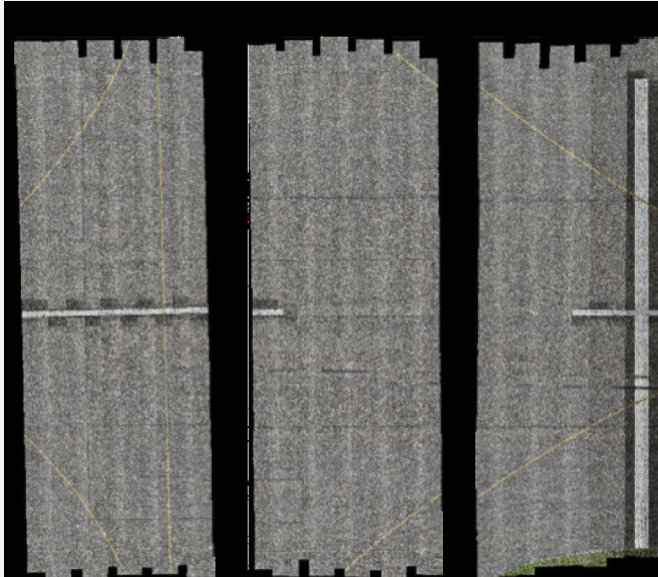


Abbildung 5: Vorschau der zusammengeführten Stitching-Ergebnisse pro Block von M1. a), b) und c) zeigen jeweils den rekonstruierten Überblick entsprechend einem Block, so, wie die Daten bereitgestellt wurden. Die Ergebnisbilder sind in voller M1-Auflösung von 0,25 mm/Pixel verfügbar. Im Durchschnitt besteht jeder der 14 zusammengeführten Blöcke aus 823,6 originalen M1-Bildern.

## KI-Technologie für zuverlässige Risserkennung

Das für die Risserkennung verwendete KI-Modell ist ein Ergebnis des Grundlagenmodells für automatisierte visuelle Inspektion, das IBM entwickelt hat, um Erkennungsaufgaben an allgemeiner ziviler Infrastruktur anzugehen. Um Informationen aus Bildern zu extrahieren, verwendet das Modell die aktuelle Vision-Transformer-Architektur, die auf Self-Attention-Blöcken basiert anstatt auf den traditionellen konvolutionellen Blöcken. Dieses Grundlagenmodell für automatisierte visuelle Inspektion wird hierarchisch mithilfe der Vorlauffrennungspipeline von IBM ausgebildet. Die verschiedenen Stufen dieser Pipeline passen das Modell schrittweise an eine

bestimmte Aufgabe an, wobei hauptsächlich nicht **annotierte Daten** verwendet werden.

In der ersten Stufe der Pipeline wird das Modell mit gross angelegten Alltagsobjekt-Datensätzen unter Verwendung des **Self-Supervised Learning** trainiert. Das beinhaltet das Lösen einer Pre-text-Aufgabe, z.B. das Rekonstruieren eines Bildes, nachdem Teile davon verdeckt wurden. So lernt das Modell, Bilddaten zu verarbeiten und sich auf bestimmte Teile davon zu konzentrieren. Das Ergebnis ist ein Modell, das nützliche Informationen aus Bildern extrahieren kann. Der obige Schritt wird noch einmal wiederholt, allerdings wird in der zweiten Stufe der Pipeline ein technischerer Datensatz verwendet. In diesem Fall verwendete das Modell einen zivilen Infrastrukturdatensatz, der aus Bildern von Brücken, Strassen usw. besteht. Das Ergebnis dieser Stufe ist ein auf zivile Infrastruktur zugeschnittenes Modell.

Der letzte Schritt besteht darin, das Modell mit überwachtem Lernen zu verfeinern, um die endgültige beabsichtigte Aufgabe auszuführen. Das Modell wird mit beschrifteten Bildern trainiert und angewiesen, Pixel zu lokalisieren und zu erkennen, die zu bestimmten Fehlern gehören, in diesem Fall Rissen. Das Ergebnis ist ein Grundlagenmodell für visuelle Inspektion, das speziell auf die Erkennung von Rissen in Bildern der zivilen Infrastruktur abgestimmt ist.

## Zusammenfassung der projektspezifischen Ergebnisse

IBM Research hat das oben beschriebene KI-Modell auf alle Originalbilder von M1 und M2 angewendet. Für M3 wurde der Fall untersucht, in dem ein extern zusammengesetztes Bild eingefügt wurde. Da die gesamte Szene in Kacheln der Grösse  $1024 \times 1024$  Pixel unterteilt war, wurde die Hauptanalyse dieser Daten auf diesen Kacheln durchgeführt. IBM Research betrachtete die M3-Daten als zusätzliche Daten, die eine externe Montage demonstrieren, aber nur 30% der gesamten Testfläche abdecken. Daher konzentrierte sich das KI-Team auf den Vergleich der M1- und M2-Daten. Es gilt, zu beachten, dass sowohl M1 als auch M2 in Bezug auf das Erfassungsvolumen und die Qualität gegenüber M3 überlegen sind. M1 und M2 erreichen eine viel bessere

Qualität in etwa der gleichen Zeit oder liefern die gleiche Qualität mit weniger Aufwand. Demzufolge sind die M2-Ergebnisse ausreichend gut, um eine gute Bewertung des Abschnitts der untersuchten Landebahn vorzunehmen. Die Erfassung der gesamten Landebahn (ca. 2,8 km) in weniger als einem Tag ist realisierbar, wenn die Anforderungen erfüllt sind.

| Mission             | M1   | M2   | M2 vs. M1     |
|---------------------|------|------|---------------|
| Rissinstanzen total | 3920 | 2629 | 32,9% weniger |
| Konfidenz (>0,5)    | 691  | 586  | 15,2% weniger |

Tabelle 2: Übersicht über die Anzahl der erkannten Rissinstanzen für M1 und M2. Die Gesamtzahl besteht aus allen erkannten Instanzen. Sie ist relativ gross, da das KI-Modell in einem sensiblen Modus betrieben wurde, bei dem alle Vorhersagen mit einem Vertrauenswert über 0,2 gemeldet wurden. Zu beachten ist, dass die zweite Zeile niedrigere Zahlen meldet, bei denen Instanzen mit einem Vertrauenswert von mindestens 0,5 berücksichtigt wurden. Obwohl M2 eine 3-mal schlechtere **GSD** als M1 hat, werden nur etwa 33% weniger Defekte (bzw. 15%) gemeldet.

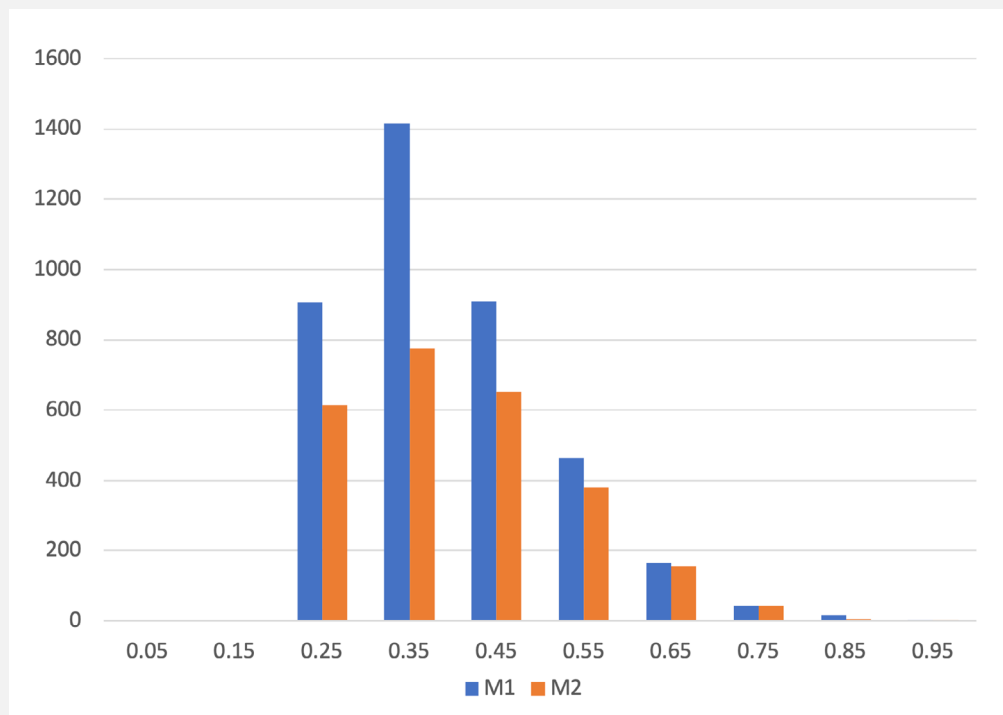


Abbildung 6: Häufigkeit der durch KI-Modelle vorhergesagten Confidence Scores (Vertrauenswerte), berechnet aus allen Ergebnissen von M1 und M2. Ergebnisse unter einem Wert von 0,2 wurden nicht berücksichtigt. Beide Ergebnisse zeigen eine ähnliche Verteilung, jedoch neigt das Modell dazu, mehr Instanzen zu detektieren als auf dem Bildmaterial von M2 vorhergesagt. Dieser Unterschied ist in Richtung des Maximums der Verteilungen grösser. Jedoch werden zum vertrauenswürdigen Ende der Verteilung ähnliche Ergebnisse erzielt. Das deutet darauf hin, dass die M2-Konfiguration gut geeignet ist, um ähnliche Ergebnisse wie M1 zu liefern, insbesondere wenn die Hauptergebnisse aus einer gefilterten Ansicht von Rissen und Beschädigungen mit hohen Confidence Scores gezogen werden.

Als Nächstes führte IBM Research einen qualitativen Test durch, bei dem dieselbe Rissinstanz durch den OCL-Annotationsviewer ausgewählt wurde, um zu demonstrieren, dass alle drei Ansichten, basierend auf M1, M2 und M3, ähnliche Ergebnisse lieferten. Alle Modelle tendierten dazu, an der Grenze zwischen Bordstein und der eigentlichen Oberfläche der Landebahn falsche Rissvorhersagen zu treffen. Diese verrauschten Detektionen erklären die in den KI-Detektionen vorhandenen Störungen. Der tatsächliche Riss, der von Bedeutung ist, da er sich in der eigentlichen Oberfläche der Landebahn erstreckt, wurde jedoch in allen drei Fällen zuverlässig erkannt. Risse, die gross genug sind, sodass sie auf dem Bild sichtbar sind (selbst wenn sie mit der geringeren Auflösung von M2 aufgenommen wurden), können die KI-Modelle erkennen. In diesem Sinne ist die M2-Konfiguration ausreichend, um den Bedarf für die Gesamtanalyse zu decken und Risse in dieser Grösse zu erkennen.

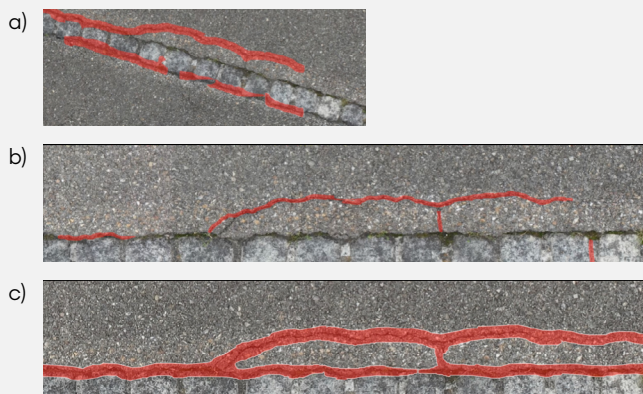


Abbildung 7: Vorhergesagter Defekt betrachtet durch OCL. Vergleich von herangezogenen Ansichten, die sich auf die gleiche Position der Landebahn konzentrieren. a) Ergebnisse von M3, die so ausgerichtet sind, dass Norden oben ist, daher ist das Bild gedreht. M1 und M2 sind so ausgerichtet, dass die Hauptachse der Landebahn mit der Bildausrichtung übereinstimmt; b) Ergebnisse für M1 (feine Konfiguration); c) M2 (schnelle Konfiguration). Der echte Riss, der sich in der Oberfläche der Landebahn erstreckt, wird in allen drei Konfigurationen zuverlässig erkannt.

## OCL-Tool zur Präsentation von Ergebnissen

Das OCL-Werkzeug ist ein Forschungsinstrument von IBM, das kontinuierlich weiterentwickelt wird, um KI-Ergebnisse zu demonstrieren, zu generieren, zu iterieren und anzuzeigen. Im Rahmen dieses Projekts wurden nur die Share- und Showcase-Funktionalitäten des Werkzeugs verwendet.

Hauptmerkmale von OCL:

- **Bildbetrachtung** für Zugriff auf Bilder
- **Vorhersageansicht** zum Zugriff auf alle KI-Modell-Ergebnisse für jedes Bild
- **Übersichtsansicht** zum Zugriff auf den gesamten Abschnitt der Landebahn
- **Übersicht der zusammengeführten Vorhersagen**
- **Statistische Zusammenfassung** der vorhergesagten Defekte
- **Reportingfunktionalität** zum Exportieren von Dokumentationen von Defekten

## Bildbetrachtung

Diese Ansicht ermöglicht den Zugriff auf alle bereitgestellten Bilder. Das Tool kann mit verschachtelter Ordnerstruktur arbeiten, um das grosse Volumen der bereitgestellten Daten zu navigieren. M1 ist in 14 Blöcke strukturiert, welche die Landebahn von Ost nach West abdecken, M2 ist in 5 Blöcke strukturiert, welche die Landebahn in dieselbe Richtung abdecken. Für M3 wurden keine Rohbilder geladen, da das Werkzeug dazu dient, den vollständigen Ablauf zu demonstrieren, wenn ein extern zusammengesetztes **Orthofoto** eingespeist wird.

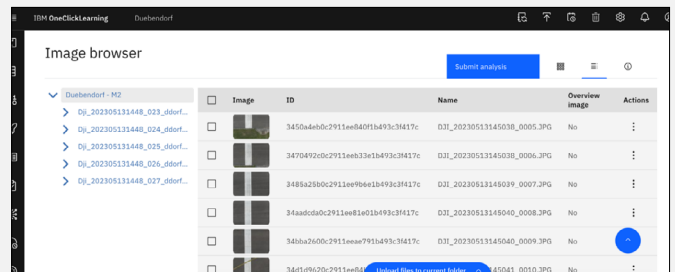


Abbildung 8: Screenshot der OCL-Bildbetrachtung. Sie ermöglicht den Zugriff auf alle Originaldaten des Projekts.

## Übersichtsansicht

Diese Funktionalität ermöglicht es, beliebig grosse Szenen mit voller Präzision anzusehen. Der Nutzer navigiert nahezu in Echtzeit. Um einen solchen Service bei benutzerdefinierten Daten bereitzustellen, hat IBM Research Front- und Backend-Services entwickelt und betreibt einen Kachelserver, der den aktuell angezeigten Teil auf Anfrage hosten und dynamisch laden kann. Für jede Ansicht müssen Daten

in einem spezifischen Format injiziert werden, um die vollständige Ansicht so zu kacheln, dass sie effizient bereitgestellt werden kann. Im Rahmen dieses Projekts hat IBM Research Funktionen implementiert, um grosse TIFF-Dateien in das Tool zu integrieren und die von M3 bereitgestellten Orthofotodaten direkt zu unterstützen. Basierend auf dem Bildzusammensetzungsalgorithmus werden alle erzielten Ergebnisse so abgelegt, gespeichert und aufbereitet, dass der Nutzer sie effizient betrachten kann.

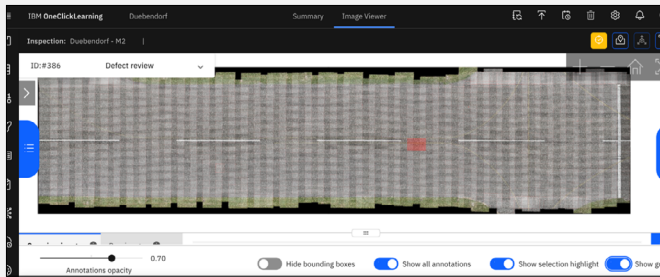


Abbildung 9: Screenshot der OCL-Übersichtsansicht. Sie ermöglicht den Zugriff auf die gesamte Szene in voller Auflösung. In dieser Ansicht zeigt M2 den gesamten Testbereich, wobei die Originalbilder hervorgehoben sind und die Grösse eines ausgewählten Originalbilds angezeigt wird. Diese Ansicht kann genutzt werden, um direkt zum Originalbild zu navigieren.

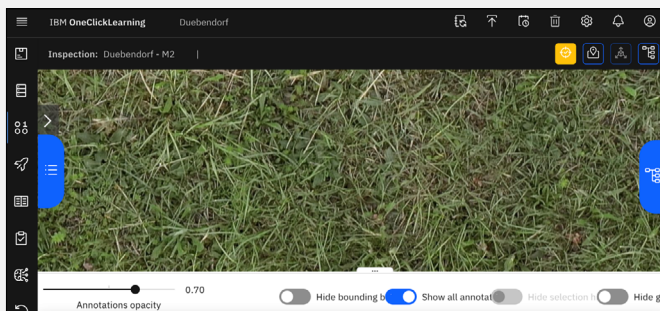


Abbildung 10: Screenshot der OCL-Übersichtsansicht. In voller Auflösung werden Details vollständig bewahrt. Um die erzielte Auflösung einordnen zu können, sind in dieser Abbildung einige Details zu Gräsern ersichtlich, die im Rahmen der M2 erstellt wurden.

## Vorhersageansicht

Sofern nicht anders angegeben, werden KI-Modelle immer teilweise auf den Originalbildern eingesetzt, um für einige Elemente Ergebnisse zu liefern. In diesem Innovation-Sandbox-Projekt hat IBM Research ein KI-Modell zur Instanzsegmentierung eingesetzt, um pixelgenaue Segmentierungsmasken von Rissen zu erhalten. Zugehörige Attribute wie die Länge des Risses werden automatisch aus der vorhergesagten

Geometrie extrahiert. Zusätzliche Attribute wie die vorhergesagte Bewertung werden ebenfalls aufgefüllt. Sie messen, wie sicher das KI-Modell in Bezug auf eine bestimmte Vorhersage von einem tatsächlichen Defekt ist.

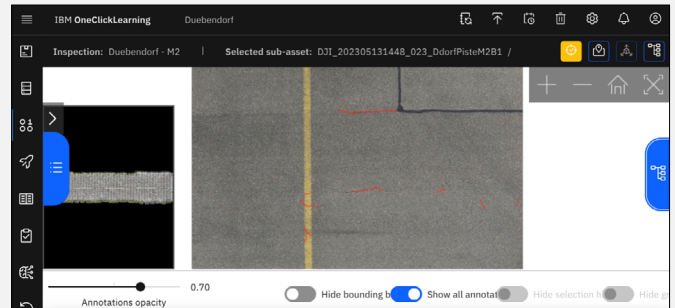


Abbildung 11: Screenshot der OCL-Bildbetrachtung. Rissdetektionen sind auf dem Originalbild in Rot markiert. Diese Ansicht zeigt aufgenommene Bilder und die dazugehörigen Ergebnisse des KI-Modells. Auf der linken Seite ist die Minikarte verfügbar, um den Teil hervorzuheben, in dem sich das Bild befindet.

## Vorhersagen auf Übersichtsbildern

Die Hauptanalyse wird auf den Originalbildern durchgeführt, wobei über 11 500 Bilder für M1 verarbeitet wurden. Das Inspizieren, Interpretieren und Schlussfolgern auf solchen elementweisen Ansichten können herausfordernd sein, da es für den Benutzer schwierig ist, den gesamten Kontext auf einmal zu erfassen. Die zusammengefügte Übersicht ermöglicht es, den direkten Kontext jedes Eintrags zu sehen und zu verstehen, wo der Defekt liegt und wie er möglicherweise mit anderen Defekten in der Nähe in Zusammenhang steht. Zusätzlich wurden alle Einträge vereinheitlicht und zu logischen Einheiten zusammengeführt. Beispielsweise wurde ein Riss im Überlappungsbereich mehrerer Bilder mehrmals erfasst und somit typischerweise mehrmals, in jedem Bild, in dem er vorkommt, detektiert. In der Übersicht wurden diese detektierten Vorkommen als eine einzige Vorhersage abgebildet. Zusätzlich verbesserte derselbe Prozess die Detektionsgenauigkeit, da Defekte mehrere Chancen zur Detektion hatten und, falls in einigen Bildern übersehen, in der Übersicht verblieben, wenn sie in mindestens einem Originalbild korrekt erkannt wurden. Für Tracking-Zwecke kann das Übersichtsbild den Standort jedes Originalbildes anzeigen und für jeden Defekt zu jeder Originaldetektion im Quellbildmaterial zurück-



verweisen. Daher konnte die Übersicht effizient genutzt werden, um direkt über die gesamten verfügbaren Daten zu navigieren.

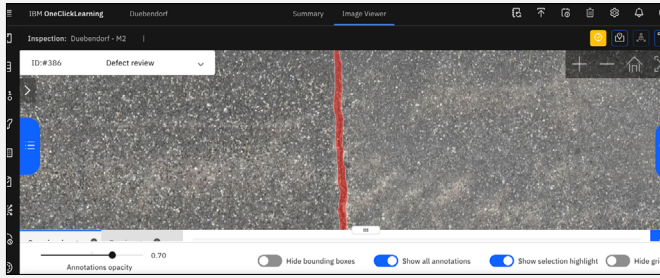


Abbildung 12: Screenshot des OCL-Übersichtsbildes. Ein im Rahmen von M2 detektierter Riss wird hervorgehoben und in der Übersichtsansicht dargestellt.

## Statistische Zusammenfassung

IBM Research hat aggregierte Zusammenfassungen bereitgestellt, die Statistiken über das Auftreten aller Defekte enthalten. Diese Statistiken können auch für Teilbereiche erstellt werden. Die grundlegende Statistik wird über die Originalbilder aggregiert und spiegelt eine elementweise Analyse wider. Zusätzlich wird diese Statistik aggregiert über das zusammengeführte Übersichtsbild erstellt, das weniger Annotationen enthält, da mehrere Detektionen zu einer einheitlichen Ansicht zusammengeführt werden.

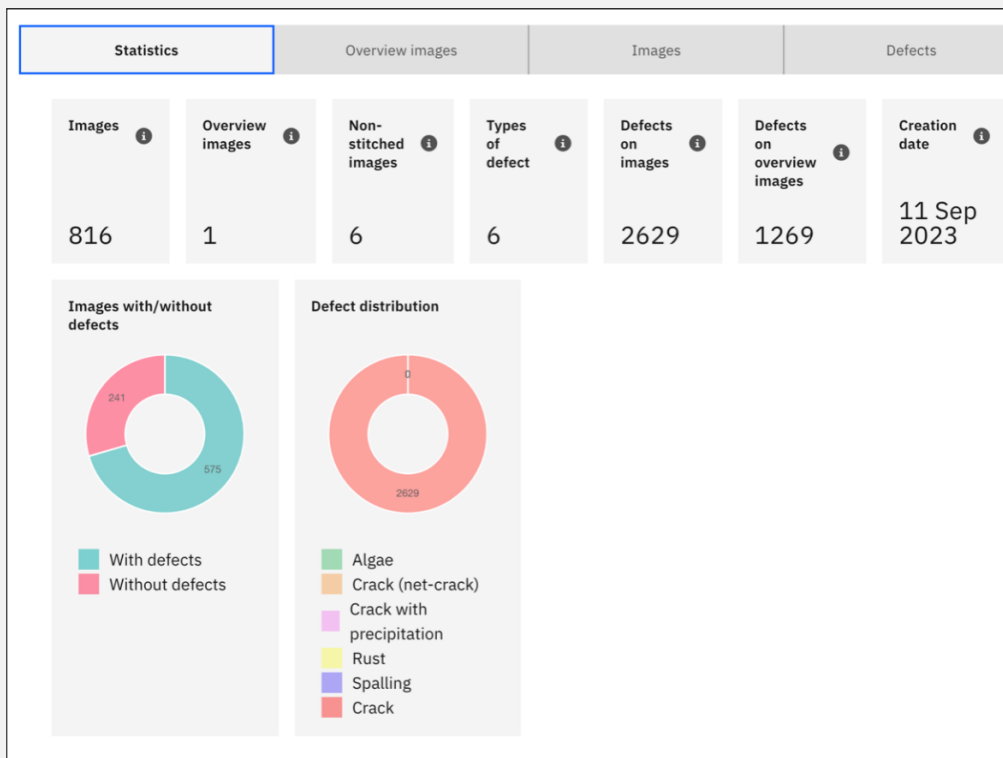


Abbildung 13: Screenshot der statistischen Zusammenfassung der M2-Daten auf OCL. Die M2-Daten bestehen aus 816 Bildern, die zu einem einzigen Übersichtsbild zusammengeführt wurden. Die Gesamtzahl der erkannten Defekte beträgt 2629 in den Originaldaten und beläuft sich nur auf 1269 Instanzen (48,2% der ursprünglichen Anzahl von Detektionen), wenn Vorhersagen auf dem Übersichtsbild zusammengeführt werden. Die Mehrheit von 70,5% (nämlich 575 Bilder) weist keine vorhergesagten Defekte auf.

## Reports

Für alle im OCL-Tool erfassten Daten können Reports erstellt werden, die alle Ergebnisse dokumentieren. Die Reports enthalten eine Übersichtstabelle sowie einen Detailbereich für jeden Defekt. Wenn das Tool zur Überprüfung der vorhergesagten Annotationen verwendet wurde, können zusätzliche Informationen wie Zustandsbewertung, Klassifizierung der Defektvarianten und Kommentare vom Fachexperten hinzugefügt werden, um den Bericht zu vervollständigen. Der Detailbereich des Berichts enthält eine detaillierte Ansicht des erkannten Defekts sowie eine Übersichtsansicht, die den Standort des Defekts im Feld zeigt, sodass er vor Ort leicht gefunden werden kann. Wichtige Attribute wie Längenmessungen, Vorhersagebewertungen und GPS-Standorte werden alle rapportiert. Alle Einträge sind direkt mit dem OCL-Tool verlinkt, sodass der Hyperlink direkt zur vergrößerten Ansicht des über unser Tool visualisierten Defekts führt.

In diesem Projekt hat IBM Research den vollständigen Bericht der M2-Daten sowie eine Ansicht ausgewählter Rissdetektionen bereitgestellt. Das vollständige Berichtsdokument wird als ergänzendes Material zur Verfügung gestellt.

## Schlussfolgerung und Ausblick

Das Projekt hat erfolgreich demonstriert, wie grosse Datenmengen von KI-Modellen systematisch erfasst und ausgewertet werden können. IBM Research konnte Risse in M1, M2 und M3 korrekt identifizieren. Um grosse Bereiche in angemessener Zeit erfassen zu können, ist es unerlässlich, dass die Erstellung des Bildmaterials effizient durchgeführt wird. IBM Research hat demonstriert, dass die M1- und M2-Konfigurationen der M3-Konfiguration überlegen sind. Zudem reicht die Auflösung von M2 aus, um die wichtigen Risse zu erfassen, die benötigt werden, um den Zustand der Landebahn vollständig zu dokumentieren. Als nächster Schritt kann die gesamte Landebahn von Dübendorf mit der M2-Konfiguration an einem halben Tag erfasst werden.

Die gesammelten Bilddaten sind der erste Datensatz dieser Art, der drei verschiedene Missionstypen für denselben Testbereich verfügbar macht. Des Weiteren wurde der inspizierte Bereich

durchgehend mit einer Auflösung von 0,25 mm/Pixel erfasst. Zugang zu realen Daten dieser Art ist sehr wichtig für die Forschung und Innovationsentwicklung, da die neuesten KI-Algorithmen und -Strategien damit in einem realen Kontext bewertet und verbessert werden können. Daher werden die Daten über das nächste Jahrzehnt hinweg einen Mehrwert bieten und helfen, die neuesten KI-Entwicklungen im Bereich Bilderkennung voranzutreiben.

Darüber hinaus bestätigt sich mit jedem Projekt, in dem KI erfolgreich eingesetzt wird, dass die entwickelten **Foundation Models** in einem breiteren Kontext zuverlässig funktionieren. Damit steigen auch die Erfolgsaussichten, dass dieselben KI-Modelle auf weitere Anwendungsfälle wie Fassaden grosser Gebäude, Brücken, Staudämme, Tunnel oder Strassenoberflächen ausgeweitet werden können.

# Glossar

**Annotierte Daten** Daten, die mit zusätzlichen Informationen, sogenannten Annotationen, versehen sind. Bei Bildern können dies bspw. Markierungen sein, die Objekte oder Bereiche von Interesse kennzeichnen. Annotierte Daten sind insbesondere beim Machine Learning und bei KI von grosser Bedeutung.

**Deep Learning** ein Unterbereich des Machine Learning, der auf künstlichen neuronalen Netzwerken basiert und besonders effektiv bei der Erkennung von Mustern und der Verarbeitung von unstrukturierten Daten ist.

**Digital Twin** eine digitale Repräsentation eines physischen Objekts, eines Prozesses oder eines Systems, die für Analysen, Simulationen und Steuerung genutzt werden kann.

**Few-Shot Learning** eine Methode des Machine Learning, bei der Modelle so entwickelt werden, dass sie mit einer sehr geringen Anzahl von Beispielen auskommen.

**Foundation Model** ein grosses vortrainiertes Modell im Bereich der KI, das als Grundlage für den Bau spezifischer Modelle dient. Foundation Models werden mit grossen Mengen an Daten trainiert, um eine breite Palette von Fähigkeiten zu entwickeln, und können dann für spezifische Aufgaben angepasst werden.

**GPU (Graphics Processing Unit)** eine Hardwarekomponente, die speziell für die Verarbeitung von Grafiken und Bildern entwickelt wurde. In der KI und beim Machine Learning wird sie zur schnellen Verarbeitung von Berechnungen genutzt.

**GPS (Global Positioning System)** ein satellitengestütztes System, das die Positionsbestimmung von Objekten auf der Erde ermöglicht.

**GSD (Ground Sampling Distance)** der Abstand zwischen zwei Pixeln auf der Erde, gemessen in Metern. Dieser Wert ist ein Mass für die Auflösung eines Bildes.

**Machine Learning (maschinelles Lernen)** ein Bereich der KI, der es Computern ermöglicht, aus Daten zu lernen und Entscheidungen zu treffen, ohne explizit programmiert zu werden.

**Orthofoto** ein Luftbild, das so bearbeitet wurde, dass es die geometrischen Eigenschaften des Geländes aufweist. Verzerrungen aufgrund der Perspektive und der Höhenunterschiede sind korrigiert.

**Quadkopter** ein Drohnentyp, der durch vier Rotoren angetrieben wird, die sich auf einer Ebene befinden. Quadkopter sind aufgrund ihrer Stabilität und Manövrierfähigkeit beliebt und werden oft für Fotografie, Videografie und Inspektionen eingesetzt, etwa in Projekten, bei denen hochwertiges Bildmaterial erstellt wird.

**RTK (Real Time Kinematic)** eine Technik zur Verbesserung der Genauigkeit von GPS-Systemen, die in Echtzeit arbeitet und besonders in der Vermessung eingesetzt wird.

**Self-Supervised Learning** ein unüberwachtes Lernverfahren, bei dem Modelle lernen, indem sie in einem bestimmten Kontext Vorhersagen treffen. Dabei ist das Vorhergesagte allerdings bereits bekannt.

**Transfer Learning** eine Technik im Machine Learning, bei der ein für eine Aufgabe entwickeltes Modell auf eine neue, aber ähnliche Aufgabe übertragen wird.

# Beteiligte Personen und Organisationen

## Autoren



**Raphael von Thiessen,**  
Leiter Innovation-Sandbox für KI,  
Standortförderung Kanton Zürich



**Florian Scheidegger,**  
Researcher, IBM Research Zürich



**Reto Weiss,**  
CEO, pixmap gmbh

### Fallbeispiele aus der Innovation-Sandbox für Künstliche Intelligenz (KI)

Als Fallbeispiel innerhalb der Innovation-Sandbox für KI diente ein Projekt von IBM Research Zürich. Die Organisation hat im Frühling 2022 einen Projektvorschlag eingereicht. IBM Research Zürich mit Sitz in Rüschlikon ist das europäische Forschungszentrum von IBM und ein führendes Institut in Bereichen wie Informationstechnologie, Cloud und KI. Die Inhalte des vorliegenden Reports wurden zwischen Juni und Oktober 2023 basierend auf dem konkreten Anwendungsfall **«Automatisierte Infrastrukturwartung»** verarbeitet.

# Impressum

## Herausgeber

Standortförderung AWA, Kanton Zürich  
Verein Metropolitanraum Zürich  
Innovation Zurich

## Projektkonzeption und -koordination

Raphael von Thiessen  
Standortförderung Kanton Zürich  
8090 Zürich  
raphael.vonthiessen@vd.zh.ch

## Konzeption in Zusammenarbeit mit:

Stephanie Volz  
Isabell Metzler  
Patrick Arnecke

## Autoren

Raphael von Thiessen  
Florian Scheidegger  
Reto Weiss

## Gestaltung

Sibylle Brodbeck, sibyllebrodbeck.ch

## Publikation

Dieser Report erscheint ausschliesslich digital und in den Sprachen Deutsch und Englisch

## Copyright

Alle Inhalte dieser Publikation, insbesondere Texte und Grafiken, sind urheberrechtlich geschützt. Das Urheberrecht liegt bei der Standortförderung Kanton Zürich. Die Publikation darf mit den Urheberangaben weitergegeben werden und es darf daraus mit vollständiger Quellenangabe zitiert werden.

© 2023 | Kanton Zürich

## Projekt-Steering

- Standortförderung im Amt für Wirtschaft und Arbeit, Kanton Zürich
- Statistisches Amt, Kanton Zürich
- Digitale Verwaltung und E-Government, Staatskanzlei Kanton Zürich
- Amt für Wirtschaft, Kanton Schwyz
- Metropolitanraum Zürich
- ETH AI Center
- UZH Center for Information Technology, Society, and Law (ITSL)
- swissICT
- ZHAW entrepreneurship