



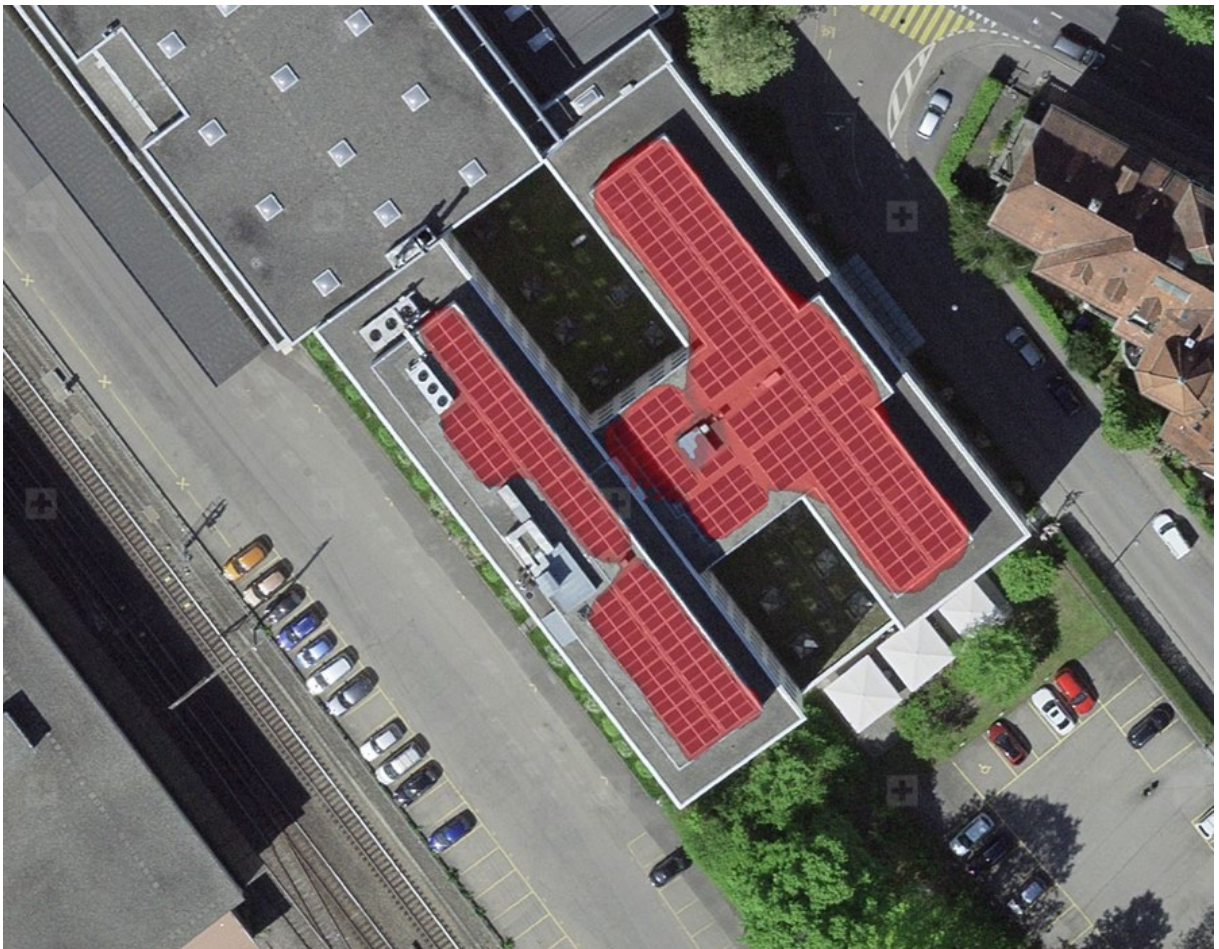
## Schlussbericht

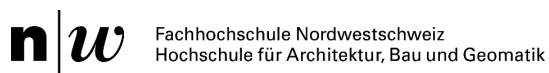
---

# SoIAI

## Automatische Erkennung von Solarenergieanlagen mit Deep Convolutional Neural Networks

---





**Datum:** 23.06.2021

**Ort:** Muttenz

**Auftraggeberin:**

Bundesamt für Energie BFE  
Forschungsprogramm Photovoltaik  
CH-3003 Bern  
[www.bfe.admin.ch](http://www.bfe.admin.ch)  
[energieforschung@bfe.admin.ch](mailto:energieforschung@bfe.admin.ch)

**Autoren:**

Adrian Meyer, Institut Geomatik, [adrian.meyer@fhnw.ch](mailto:adrian.meyer@fhnw.ch) (Kontakt)  
Denis Jordan, Institut Geomatik, [denis.jordan@fhnw.ch](mailto:denis.jordan@fhnw.ch)  
Martin Christen, Institut Geomatik, [martin.christen@fhnw.ch](mailto:martin.christen@fhnw.ch)

**BFE-Bereichsleitung:** Stefan Oberholzer, [stefan.oberholzer@bfe.admin.ch](mailto:stefan.oberholzer@bfe.admin.ch)  
**BFE-Programmleitung:** Stefan Oberholzer, [stefan.oberholzer@bfe.admin.ch](mailto:stefan.oberholzer@bfe.admin.ch)  
**BFE-Vertragsnummer:** SI/501806-01

**Für den Inhalt und die Schlussfolgerungen sind ausschliesslich die Autoren dieses Berichts verantwortlich.**



## Zusammenfassung

Die Schweizerische Statistik der erneuerbaren Energien schätzt die Nutzung der Sonnenenergie anhand von Markterhebungen ab. Um eine umfassendere Statistik zu erreichen, wurden bestehende Solaranlagen aus hochauflösenden Luftbildern der swisstopo mithilfe künstlicher Intelligenz auf der Basis von Deep Learning (DL) Algorithmen automatisch identifiziert und quantifiziert. Mit einer Gesamt-Modellgenauigkeit von ca. 90 % erreicht der Objektdetektor eine so hohe Präzision sowie Trefferquote, dass er sich für die Erhebung einer landesweiten Statistik sowohl von Photovoltaik, als auch von Solarthermie-Anlagen eignet.

## Résumé

La statistique suisse des énergies renouvelables estime l'utilisation de l'énergie solaire sur la base d'études de marché. Afin d'obtenir des statistiques plus complètes, les installations solaires existantes ont été automatiquement identifiées et quantifiées à partir d'images aériennes de haute résolution (swisstopo) à l'aide d'une intelligence artificielle basée sur des algorithmes de Deep Learning (DL). Avec une précision globale du modèle d'environ 90 %, le détecteur d'objets atteint un niveau de précision et un taux de réussite suffisamment haut pour que les résultats obtenus puissent être intégrés dans le cadre d'une statistique de l'utilisation des énergies solaires thermiques et photovoltaïques en Suisse.

## Summary

The Swiss renewable energy statistics estimate the use of solar energy on the basis of market surveys. To extend these statistics, existing solar installations were identified and quantified fully automatically using deep learning (DL) algorithms based on high resolution aerial photographs provided by Swisstopo. Boasting an overall accuracy of approximately 90 % the object detection displays such high precision and recall metrics that the acquisition on country-level statistics is feasible for both, photovoltaics as well as solar thermal systems.





# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung .....</b>	<b>7</b>
<b>2</b>	<b>Ausgangslage .....</b>	<b>9</b>
<b>3</b>	<b>Vorgehen und Methode.....</b>	<b>11</b>
3.1	Datenquellen und alternative Methoden .....	11
3.2	Machine Learning mittels DCNNs .....	12
3.3	Vortrainierte neuronale Netze .....	15
3.4	Testen des Stanford DeepSolar-Algorithmus .....	16
3.5	DCNN Transfer Learning für SolAI.....	17
<b>4</b>	<b>Resultate.....</b>	<b>18</b>
4.1	Generierung Basisdatensatz (WP 1).....	18
4.2	Installation von Hard- und Software (WP 2).....	19
4.3	Crowd Contribution Webservice (WP 3) .....	20
4.3.1	Webinterface zum Digitalisieren von Solaranlagen.....	20
4.3.2	Generierung des Trainingsdatensatzes .....	22
4.4	Erstellung der Deep Learning Networks (WP 4) .....	24
4.4.1	Vordetektionen mittels Faster-RCNN .....	24
4.4.2	Entwicklung des pixelgenauen Mask-RCNN.....	26
4.5	Anwendung des Netzwerks auf die Schweiz (WP 5).....	29
4.6	Anwendung auf ältere Luftbilddatensätze (WP 6).....	30
4.7	Evaluation (WP 7) .....	31
4.7.1	Vorgehen .....	31
4.7.2	Konfusionszählung: Nur Photovoltaik.....	33
4.7.3	Metriken: Nur Photovoltaik .....	33
4.7.4	Konfusionszählung: Nur Solarthermie.....	35
4.7.5	Metriken: Nur Solarthermie .....	36
4.7.6	Konfusionszählung: Kombiniertes Modell .....	37
4.7.7	Metriken: Kombiniertes Modell .....	38
4.7.8	Vergleich und Limitationen der Modelle .....	39
4.8	Veröffentlichung und Open Source Zugang (WP 8) .....	39
<b>5</b>	<b>Publikationen und Vorträge.....</b>	<b>40</b>
<b>6</b>	<b>Referenzen .....</b>	<b>42</b>



## Abkürzungsverzeichnis

AI	Artificial Intelligence (Künstliche Intelligenz)
AV	Amtliche Vermessung
BFE	Bundesamt für Energie
DCNN	Deep Convolutional Neural Network
DL	Deep Learning (Kurzform für AI-Lernprozesse mittels DCNNs)
Faster RCNN	Bekannte Deep Learning Netzwerkarchitektur (DCNN) zur Objektdetektion mittels Bounding Boxes (Rahmen), findet z.B. Einsatz im autonomen Fahren.
FCIR	False Colour Infrared (Falschfarben-Infrarotbilder in der Fernerkundung)
F1-Score	Gebräuchliche Metrik für die Gesamtgenauigkeit von Klassifikationen, basiert auf Konfusionsmatrix, harmonisches Mittel aus Precision (Nutzergenauigkeit) und Recall (Produzentengenauigkeit)
FHNW	Fachhochschule Nordwestschweiz
FN	False Negative – Falsch-Negative Detektion, Element der Konfusionsmatrix
FP	False Positive – Falsch-Positive Detektion, Element der Konfusionsmatrix
GPU	Graphics Processing Unit (Prozessor der Grafikkarte)
GWR	Gebäude und Wohnungsregister des BFS (Bundesamt für Statistik)
HTTPS	Internet-Protokoll zur sicheren Netzwerkkommunikation
IGEO	Institut Geomatik der Fachhochschule Nordwestschweiz
Mask RCNN	Bekannte Deep Learning Netzwerkarchitektur (DCNN) zur semantischen Objektdetektion mittels Masken (pixelgenauer Segmentation), basiert im wesentlichen auf Faster RCNN mit Erweiterung
NIR	Nahes Infrarot (oft bedeutungsäquivalent verwendet mit FCIR)
RCNN	«Regional CNN», Subtyp der DCNNs, programmiert für den Zweck der Lokalisation von Bildbereichen
RGB	Rot, Grün, Blau (Kanalzusammensetzung «normaler» Farbbilder)
PV	Photovoltaik
SSH	Secure Shell (kryptographisches Netzwerkprotokoll für IT-Fernsteuerung)
THM	Solarthermie-Anlagen (Warmwasserkollektoren)
THR	Threshold (Schwellwert), hier als Filteroperation für den sog. Confidence Score, ein Sicherheitswert in % welche ein Detektor für jede einzelne Detektion ausgibt
TIF	Tagged Image File (ein kompressionsfreies Bilddatenformat mit hohem Speicherbedarf)
TP	True Positive – Richtig-Positive Detektion, Element der Konfusionsmatrix
VRT	Virtual Raster Tile (ein Format zur Lokalisierung grosser Luftbilddatensätze)
WP	Work Package



# 1 Einleitung

Die Nutzung erneuerbarer Energiequellen befindet sich in einer starken Entwicklung, insbesondere steigt der gesellschaftliche und politische Druck nach Alternativen zu nicht erneuerbaren Energiequellen und damit verbunden der Ausbau von Photovoltaik- und Solarthermie-Anlagen kontinuierlich und weltweit. Die erschwinglichen Preise für zuverlässige und effiziente Technologien gestalten die Nutzung von Solarenergie, selbst im Vergleich zu weniger nachhaltigen Formen der Energiegewinnung (insb. Fossile Energieträger), als ressourcenschonende und zunehmend wirtschaftlich interessante Möglichkeit der Energiegewinnung in der Schweiz (Quelle BFE).

Für die Abschätzung des Ausbaus von Solaranlagen wird in der Schweiz bereits ein Solarkataster geführt (<https://sonnendach.ch>). Dieses zeigt die für Solarenergie geeignete Potentialfläche mittels Neigung und Orientierung von Dächern und Fassaden bereits für viele Kantone, bzw. grösstenteils schweizweit an.

Bisher fehlen in der Schweiz jedoch noch zuverlässige Lage- und Flächendaten der bereits installierten Photovoltaik- und Solarthermie-Anlagen, um eine vollständige Auswertung im Zusammenspiel mit den Solarpotential-Kataster zu ermöglichen.

Eine manuelle Erfassung der Anlagepositionen sowie ihrer physikalischen Dimensionen wäre angesichts des hohen Umfangs bereits installierter Flächen sowohl in zeitlicher wie auch in personeller Hinsicht zu aufwändig. Hierzu bieten sich vermehrt innovative datengetriebene Lösungen mit Hilfe von Computeralgorithmen an, welche durch die Automatisierung des Objekterkennungs-Prozesses Ressourcen einsparen und neue Möglichkeiten in der Erfassung vieler Objekte eröffnen. Insbesondere die enorm hochauflösenden, landesweit verfügbaren und kalibrierten Luftbilder der swisstopo (Swissimage 10cm) stellen für dieses Unterfangen erstmals eine reichhaltige Datenbasis mit hoher Informationsdichte und einheitlicher Qualität dar, die sich optimal im Zusammenspiel mit Algorithmen zur automatisierten Objektdetektion einsetzen lassen. Dabei konnten in den letzten Jahren Deep Learning Modelle als algorithmische Basis für die Objektdetektion eindrucksvolle Ergebnisse in unterschiedlichen Anwendungsfeldern, insbesondere für Aufgabenstellungen der Fernerkundung, aufzeigen (Yu et al. 2018).

Dieser Schlussbericht bezieht sich auf das Projekt SolAI, welches am Institut Geomatik der Fachhochschule Nordwestschweiz (FHNW) in Zusammenarbeit mit dem Bundesamt für Energie (BFE) und der swisstopo November 2018 gestartet wurde. Das Projekt zielte darauf ab, bestehende Solaranlagen aus den Luftbildern der swisstopo mithilfe künstlicher Intelligenz auf der Basis von Deep Learning Algorithmen, automatisch zu identifizieren und hinsichtlich ihrer Fläche zu quantifizieren.

Ein solcher Ansatz soll dann als Basis für die Umsetzung der Energiestrategie und statistischer Schätzmodelle, dem Solarmarkt dienen (Meyer et al. 2019).

Das Projekt bestand dabei aus mehreren Phasen. Zunächst wurde ein Basisdatensatz aus den vorhandenen bekannten Koordinaten von Solaranlagen mit den zugehörigen Luftbildern generiert, um diese in einem für das Deep Learning geeignete Format (kleine Kacheln) vorliegen zu haben. Anschliessend wurde das eigens für das Projekt ausgebaute High Performance Computing Cluster mit neuer Hard- und Software bestückt. Der anspruchsvolle Installationsprozess konnte bedarfsgerecht abgeschlossen werden.

Mehrere Projektmitarbeiter beschäftigten sich anschliessend mit der Entwicklung eines teilautomatisierten Crowd Contribution Clients – einer Online-Software, die der Erhebung von Digitalisierungsdaten diene. Diese Datensätze (im Folgenden auch als «Labels» oder «Annotations» bezeichnet) bildeten die Basis für das Anlernen der Modelle (sog. «Training») dar und wurden im Wesentlichen aus manuell gezeichneten polygonalen Umrissen um vorhandene Solarinstallationen dargestellt.



Das eigentliche Digitalisieren («Labelling») wurde im Rahmen von geleiteten Workshops durch ein Team von eigens hierfür geschulten Experten durchgeführt. So liessen sich 30'000 Solaranlagen auf knapp 8'000 Luftbildkacheln mit einer Grösse von je ca. 80 x 80 m digitalisiert annotieren. Anlagen für Photovoltaik (ca. 70 %), Solarthermie (ca. 15 %) sowie unklare Objekte (ca. 15 %) wurden dabei getrennt klassifiziert.

Für die Erstellung des Detektionsmodells wurde auf Algorithmen der Klasse der Deep Convolutional Neural Network (DCNN) zurückgegriffen, verschachtelte Netzwerkgraphen welche die Strukturen in neuronalen Pyramidenzellen mathematisch nachbilden und sich daher für Bildanalysen eignen. Ausgewiesene Experten für diese Algorithmen führten eine Vielzahl von Trainingsdurchläufen durch, um einen möglichst präzisen Detektor zu generieren. Die Wahl fiel dabei auf ein DCNN des Typus «COCO-pretrained Mask RCNN», ein effizientes Kerngerüst für universelle Objektdetektion (Matterport), welches mit dem Framework Pytorch/Torchvision (Facebook AI Research) unter Verwendung der Programmiersprache Python angepasst wurde. Eine grosse Herausforderung stellte die Anwendung des modellierten DCNN auf den grossen Luftbilddatensätzen der ganzen Schweiz dar. Dieses Big-Data-Problem bestand aus der Speicherung und Verarbeitung vieler Terabyte an Bilddaten. Hierfür wurden eigens hoch-effiziente, prozessnahe Parallelisierungs-Algorithmen geschrieben, um die benötigte Rechenzeit von mehreren Monaten auf wenige Tage herunterzuberechnen.

Den Abschluss des Projekts stellten die statistische Analyse und Evaluation der durch die Detektionsalgorithmen gewonnenen Ergebnisse dar, für welche Metriken wie Precision, Recall und abgeleitete Kombinationsmetriken wie dem F1-Score verwendet wurden.

Das akkurate Datenmaterial, die innovative DL Technologie und die sehr hohen erreichten Genauigkeiten von ca. 90 %, die im Rahmen des SolAI Projektes vorgestellt werden, sollen wesentlich zu einer Veröffentlichung in der internationalen Fachliteratur beitragen. Nach erfolgter Publikation sollen Modelle durch Open Source Lizenzen (vorbehaltlich etwaiger Datenschutzprüfungen) zugänglich gemacht werden. Ziel ist dabei, einen wissenschaftliche Diskurs herbeizuführen und durch Benchmarking verbesserte Algorithmen bereitzustellen und nutzbar zu machen.





## 2 Ausgangslage

Die Statistik der erneuerbaren Energien des BFE<sup>1</sup> schätzt die Produktion von Solarstrom und Solarwärme in der Schweiz mehrheitlich auf der Basis von Markterhebungen ab. Die dadurch gewonnenen Angaben über den Energieertrag können auf der Grundlage der verwendeten Methodik nur ansatzweise validiert werden. Um eine verlässliche Quantifizierung der installierten Solaranlagen und deren Erträge im Rahmen der Energiestrategie 2050 zu erreichen, bedarf es einer flächendeckenden Analyse mit einer homogenen Datenerhebung und Analysemethodik.

Das vorliegende Projekt adressiert diese ambitionierte Zielsetzung mit Hilfe von «State of the Art» Methoden der automatisierten Mustererkennung auf der Basis von DL-Algorithmen. Im Projekt SolAI wird ein tiefes neuronales Netzwerk (DCNN) erzeugt, welches der Detektion (Lokalisierung und Grössenabschätzung) von Solaranlagen aus Luftbildern dient.

Der vorgestellte Zugang ist äusserst vielversprechend, weil moderne DL-Verfahren in der Lage sind, selbstständig spezifische und kontextrelevante Merkmale in Bildern zu erlernen, die hier genutzt werden sollen, um Solaranlagen vollautomatisiert zu erkennen.

Verschiedene Arbeiten haben gezeigt, dass die automatisierte Detektion von Solaranlagen mit Hilfe von Fernerkundungsdaten grundsätzlich möglich ist. So zeigt die Arbeit von Lee et. al (2018) eine Detektion von fehlerhaften Solarpanels mit RGB- und Thermalbildern, allerdings einschränkend für Photovoltaik-Freiflächenanlagen und anhand von Drohnenbildern mit entsprechend guten Sichtbarkeitsverhältnissen und hoher Auflösung. Weitere Arbeiten verwenden fernerkundliche Infrarot-Nahaufnahmen von einzelnen Solarpanels, wie beispielsweise Feng et. al (2013) zur Detektion von Oberflächenrissen.

Mittels Machine Learning wurden auch Modelle erzeugt, welche mit grosser Genauigkeit Voraussagen zur Solarstromerzeugung machen können (Shahriar 2015 & Iyengar 2017). Eine genaue Vorhersage der Solarstromerzeugung ist entscheidend, um lokale erzeugte Solarenergie voll auszuschöpfen. Für die Abschätzung der aktuellen Solarstromproduktion könnten die Ergebnisse in einem Nachfolgeprojekt mit den Daten aus dem nationalen Solarkataster ([www.sonnendach.ch](http://www.sonnendach.ch)) des Bundesamtes für Energie sowie aktuellen Strahlungsdaten von MeteoSchweiz kombiniert werden.

Ein prominentes Beispiel für die bereits gelungene Detektion von Solarpanels aus Satellitenbildern (Google Maps) stellt das Projekt *DeepSolar* der Stanford University dar (Yu et al. 2018). Die grundsätzliche Anwendbarkeit der Objektdetektion mittels DCNNs wurde mit diesem Ansatz bereits bestätigt. Es zeigt sich jedoch ein grosser Interpretationsspielraum in den erreichten Genauigkeiten. Mittels moderner Verfahren und genauerer Eingangsdaten knüpft das Institut Geomatik der FHNW an diese Ergebnisse an, mit dem Ziel reproduzier- und nutzbare Klassifikationen rapportieren zu können.

Das IGEO verfügt hinsichtlich der automatischen Klassifizierung von Bilddaten anhand datengetriebener Methoden über eine langjährige Expertise. Hervorzuheben sind Projekte mit industriellen Partnern und Ämtern auf Kantonal- und Bundesebene, wie zum Beispiel für die Automatisierung der Arealstatistik für das Bundesamt für Statistik (BFS) (Jordan et al. 2019). Im Rahmen des Projektes hat das Institut Geomatik in einer Kooperation mit ExoLabs GmbH einen Softwareprototypen entwickelt, der mittels modernster Algorithmen des Maschinellen Lernens und einer Multisensorfusion aus Luft- (RGB, Infrarot, Höhe), Zeitreihen aus multispektralen Satellitenbildern sowie weiteren Datenquellen zukünftig zu einer operativen Effizienzsteigerung zur Erstellung der Arealstatistik führen soll. Lösungen zur Anonymisierung von Mobile Mapping Daten oder etwa die Klassifizierung von Landwirtschaftsflächen auf der Basis von Deep Learning Algorithmen sind weitere Beispiele für Projekte, die am Institut Geomatik im Bereich künstlicher Intelligenz für die Geodatenanalyse durchgeführt wurden.

<sup>1</sup> <https://www.bfe.admin.ch/bfe/de/home/versorgung/statistik-und-geodaten/energiestatistiken/teilstatistiken.html>



Für die Erkennung von Solarpanels sind umfangreiche Kenntnisse in der automatisierten Objektdetektion mittels DCNNs hilfreich. Die Masterthesis von Adrian Meyer am Institut Geomatik beschäftigte sich bereits 2018 mit der Objektdetektion aus Luftbildern am Beispiel von Wildtieren. Die Arbeit erhielt von der Deutschen Gesellschaft für Photogrammetrie und Fernerkundung den 2. Karl Kraus Nachwuchsförderpreis (Meyer et. al. 2019).

In dem nun für das BFS aktuell durchgeführten Projekt wurde schliesslich an der FHNW AI (inkl. DL) zur Entwicklung eines produktiv einsetzbaren auf Multisensordaten basierenden Softwareprototyps eingesetzt. Ziel war es, das Automatisierungspotential für die Arealstatistik zu untersuchen und eine «State of the Art»-Lösung zu implementieren, mit der höhere Auflösungen und Genauigkeiten erreicht werden sollen.

Im Rahmen der Fragestellung für das SolAI Projektes wurde die Expertise des Instituts Geomatik abgeholt und spezifisch zur Erkennung von Solaranlagen angepasst (siehe Abschnitt 3).



## 3 Vorgehen und Methode

### 3.1 Datenquellen und alternative Methoden

Grundsätzlich eignen sich verschiedene Datensätze und algorithmische Herangehensweisen für die vorliegende Klassifikationsaufgabe. Neben der swisstopo Luftbilder mit einer Auflösung von 10cm (seit 2018) können Infrarotbilder (NIR, FCIR oder Thermal), aber auch amtliche Informationen (z.B. Gebäudergrundrisse oder -volumen) für die bevorstehende Klassifikationsaufgabe liefern. Die Daten sollen nach ihrer Evaluierung schrittweise und gemäss dem erwarteten Potenzial bereitgestellt werden, um diese in den Klassifikationsalgorithmus einfließen zu lassen.

Photovoltaik- und Solarthermie-Anlagen gibt es von zahlreichen unterschiedlichen Herstellern, in zahlreichen unterschiedlichen Grössen, Varianten, Materialien und Farben. Die Objekterkennung mittels *Scale-Invariant Feature Transform* (SIFT) nach Lowe (1999), damit verwandten Methoden (z.B. *SURF*, *KAZE*, *ORB*; Tareen 2018) oder klassischem *Template Matching* hat sich für viele Anwendungen wie beispielsweise zur Gebäudedetektion (Tang & Li 2010) bewährt, da die Objekte so unabhängig von Skalierung, Rotation und Translation erkannt werden können. Bei diesen Ansätzen müsste jedoch eine enorme Vielzahl von unterschiedlichen Varianten bei Photovoltaik- und Solarthermie-Anlagen modelliert oder niedrige Genauigkeitsraten in Kauf genommen werden.

Hier zeigt sich der besondere Vorteil moderner DCNN-Verfahren. Diese sind durch interne höherdimensionale Bildmanipulationen in der Lage, skalen- und kongruenzinvariante Merkmale aus Bilddaten zu extrahieren. Mittels dieser Merkmale kann letztlich unempfindlich gegenüber den oben genannten Limitationen eine genauere Bildklassifikation erreicht werden (Josifovski 2015).



## 3.2 Machine Learning mittels DCNNs

Der grundsätzliche Aufbau von DCNNs basiert auf dem Prinzip eines mehrschichtigen Perzeptrons. Ein Perzeptron kennt Eingabe- und Ausgabevariablen: Eine Verschaltung von Knotenpunkten in einem Graphennetzwerk (siehe Abb. 1) analog den Zellkörpern von Neuronen, die über gewichtete Kanten miteinander verbunden sind (analog den Synapsen) und bestimmte Schwellwerte zur Aktivierung ihrer Nachbarn erreichen müssen (analog zur Auslösung eines Aktionspotentials) (McCulloch & Pitts, 1943; Rosenblatt, 1958). In der Struktur der Verschaltung ähneln DCNNs dem Aufbau des Sehprozesses bei lebenden Organismen (sog. Pyramidenzellen zwischen kortikalen Schichten). Das Verständnis über diesen Prozess geht zurück auf Forschung aus den 50er- und 60er-Jahren an Säugetiergehirnen, wobei komplette „mentale Karten“ des visuellen Raumes in den sog. Rezeptiven Feldern des visuellen Kortex nachgewiesen werden konnten, welche in aufeinander folgenden hierarchischen Schichten organisiert sind (Hubel & Wiesel, 1968). In „niedrigeren“ Schichten werden Wahrnehmungssignale aufgenommen und prozessiert, in „höheren“ Schichten hingegen auf abstrakterer Ebene analysiert und verglichen.

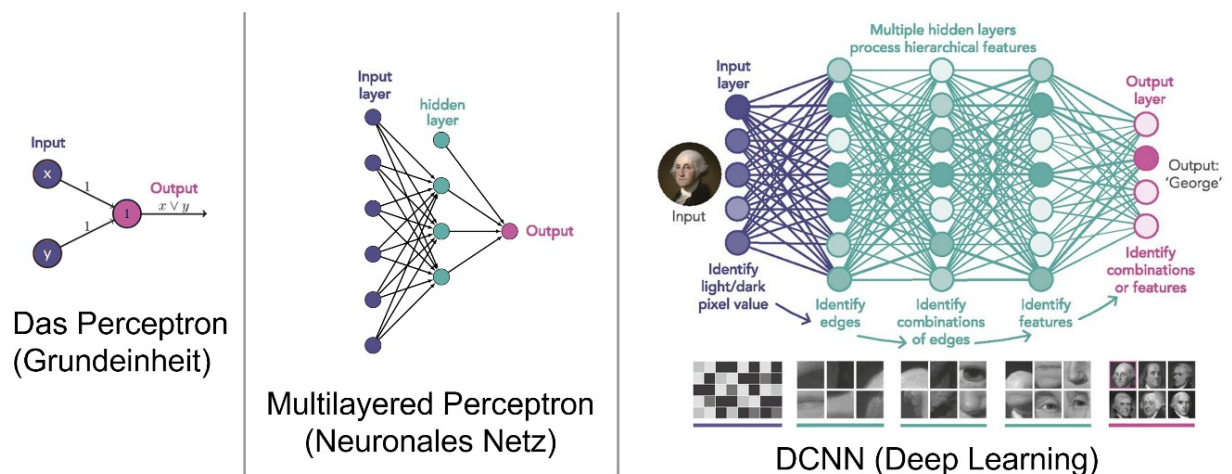


Abbildung 1: Grundsätzlicher Aufbau neuronaler Netze als Graphenmodelle (teilweise übernommen von Waldrop, 2019).

Ein zentrales Element von DCNNs ist ihre Möglichkeit zur adaptiven Informationsrückführung unter Anpassung der Kantengewichte vieler Neuronen gleichzeitig. Dieses Verhalten wird „Backpropagation“ genannt und geht auf Forschungen zum Modell des Neocognitrons in den 80er-Jahren zurück (Fukushima & Miyake, 1982). In den folgenden Jahren wurde das Konzept mehrfach angewandt, verbessert, generalisiert und vereinfacht (LeCun et al., 2015).

Während die mathematischen Konzepte bereits länger bekannt waren, stand erst in den späten 90er-Jahren die notwendige Rechenleistung zur Verfügung, um erste praxisrelevante Anwendungen in der Bilderkennung umzusetzen (LeCun et al., 1998). Es wurden damals 32x32 Pixel grosse Bilddateien von Zahlen auf handgeschriebenen Schecks klassifiziert. Der Anwendungsfall ist heute ein Standardbeispiel für die Beschreibung der Funktionsweise der Netzwerke und resultierte z.B. in der Schaffung der bekannten MNIST-Datenbank, welche für das Benchmarking verschiedener Algorithmen eingesetzt wird (Deng et al., 2009).

Erst in jüngster Zeit gelangen jedoch Erfolge vorwärts gerichteter, tiefer neuronaler Netzwerke, die bei herausfordernden Anwendungen auch systematisch bessere Ergebnisse als bisherige Lernverfahren erreichten (Graves & Schmidhuber, 2009) und zahlreiche Programmierwettbewerbe gewannen.



Mit dem Aufkommen massiv parallelisierbarer Implementierungen durch Nutzung schneller Grafikprozessoren konnte der Aufbau der Netzwerke in den Folgejahren so verfeinert werden, dass zum ersten Mal übermenschliche Performanz in der Mustererkennung erreicht wurde. Ein häufig zitiertes Beispiel ist die Verkehrszeichenerkennung im Rahmen des *IJCNN*-Wettbewerbs 2011, wo ein DCNN-Algorithmus im Vergleich zur menschlichen Performanz (Fehlerrate 1,16 %) nur halb so viele Fehlklassifikationen vornahm (Fehlerrate 0,54 %; Ciregan et al., 2012; Stallkamp et al., 2011). Ein weiterer viel beachteter Fortschritt auf dem Gebiet der Bildklassifizierung wurde 2012 mit dem *AlexNet-CNN* in der *ImageNet*-Challenge erzielt (Krizhevsky et al., 2012). Erfolge wie diese leiteten einen Paradigmenwechsel im Umgang mit der Technologie ein, indem eine Vielzahl neuer Anwendungsformen erschlossen werden konnten.

In den heutigen Implementierungen werden die Inputvariablen (häufig als multidimensionale Vektormatrizen, sog. „Tensoren“) zunächst über eine neuronale Konvolutionsschicht (*Convolutional Layer*) verarbeitet. Die Konvolution, in der Mathematik eine Faltung durch eine Matrixmultiplikation, beschreibt eine grosse Anzahl möglicher Varianten in der funktionalen Analysis, aus zwei unterschiedlichen Funktionen durch Überlagerung eine dritte, neue, bisher unbekannte Funktion zu erzeugen, welche aber gewisse Eigenschaften mit ihren Ursprungsfunktionen teilt. Mehrere der aus der Konvolution resultierenden neuronalen Aktivierungssignale werden jeweils anschliessend in einer Schicht zur Dimensionsreduktion zusammengeführt (*Pooling Layer*). In neueren Algorithmen zur Dimensionsreduktion wird nur das jeweils stärkste Signal bevorzugt (*Max-Pooling Layer*).

In früheren und einfacheren DCNNs wechselten sich Konvolutions- und Pooling-Schicht nur zwei Mal ab (Sundaresan & Lin, 1998), später 16-19 Mal (Simonyan & Zisserman, 2014), in jüngsten Modellen können auch 152 oder mehr Schichtwechsel hintereinander verschaltet werden (Szegedy et al., 2017).

Den Abschluss eines DCNNs stellt im Regelfall eine vollständig verbundene Neuronenschicht (*Fully Connected*) dar. Die Anzahl Neuronen entspricht in dieser Schicht im Regelfall der Anzahl Klassen, in welche die Input-Tensoren klassifiziert werden sollten. Die korrekte Klassifizierung wird in dieser Schicht mit einem hohen Grad an Abstraktion ausgewertet. Dies ermöglicht die effektive Fehlerrückführung über die Verbindung zu allen anderen Schichten des Netzwerks (Long et al., 2015).

Heute finden sich DCNNs in den verschiedensten Formen in einer Vielzahl von Anwendungen. Abhängig vom Verwendungszweck kommen bei der Bildverarbeitung mittels DCNNs verschiedene Typologien zum Einsatz, die sich in Komplexität und Aufbau unterscheiden (Abb. 2). Allen Typologien gemeinsam ist dabei die automatische Extraktion von Bildmerkmalen (*Feature Engineering*).

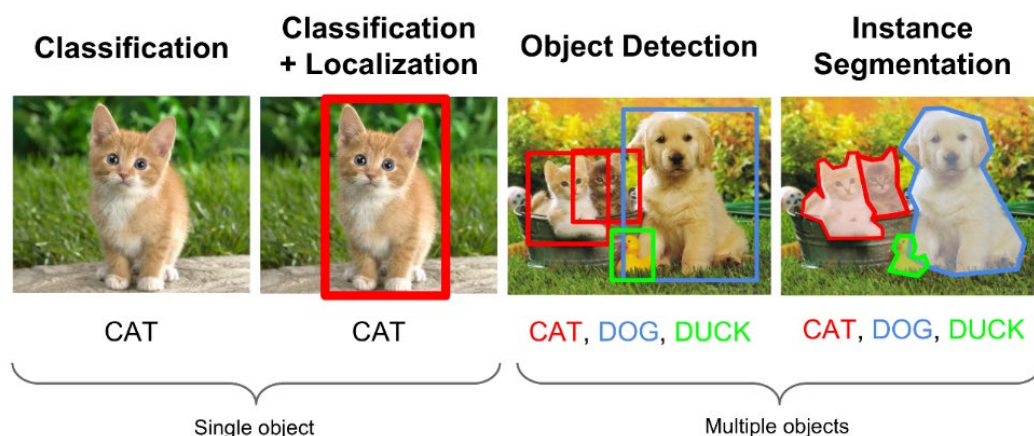


Abbildung 2: DCNN-Typologien (übernommen von Ouaknine, 2018).



Die bereits erwähnte Imagenet-Challenge war dabei ein reines Klassifikationsproblem – d.h. für ein ganzes Bild wurde entschieden, welcher Klasse es zugehörig ist. Erst später wurden sog. R-CNNs (Regional CNNs mit Region Proposal Network, Ren et.al. 2016) eingeführt, die zusätzlich die Lokalisierung eines Objektes in einem Bild ermöglichten. Mit weiterführenden Ansätzen wie Faster R-CNN gelang dann die effiziente Differenzierung mehrerer unterschiedlich klassierter Objekte innerhalb eines Bildes durch iteratives Positionieren einer grossen Zahl möglicher Bounding Boxes und Kombination mit der automatischen Feature-Analyse innerhalb sog. Fully Connected Layers (FCN). Diese Art der Objektdektion wird heute noch häufig eingesetzt, weist hohe Genauigkeiten und einen vertretbaren Aufwand beim Generieren von Trainingsdaten auf.

Um jedoch eine Instanzsegmentierung zu erreichen, wie sie für die präzise Detektion von Solarpanels letztlich notwendig ist, muss der Bounding-Box-basierte Algorithmus des Faster R-CNN um eine Binärmaske erweitert werden, die die genauen Umrisse eines Objektes im Bild beschreibt (siehe Abb. 2). Zusätzlich zu den Bounding Boxes müssen bei der Erhebung von Trainingsdaten also Binärmasken erfasst werden, was den Markierungsaufwand sehr stark erhöht.



Abbildung 3: Binärmasken in der automatischen Instanzsegmentierung mit Mask R-CNN (übernommen von Farooq, 2018).



### 3.3 Vortrainierte neuronale Netze

Für die grossen Datensätze im Rahmen internationaler Challenges wie *Imagenet* oder *COCO* wird ein neuronales Netzwerk trainiert, indem Kantengewichte zwischen den Neuronenschichten nach dem Zufallsprinzip initialisiert werden. Sobald mit dem Training begonnen wird, werden die Gewichte so verändert, dass die Klassifikationsaufgabe auf unbekanntem Testdaten mit weniger Fehlern ausgeführt wird (Optimierung). Ist ein bestimmter festgelegter Schwellwert bei der Klassifikationsgenauigkeit überschritten, so können die Kantengewichte gespeichert („eingefroren“) werden, um das Netzwerk auf neue, unbekannte Daten anzuwenden (sog. *Inferencing*).

Um den enormen Rechenaufwand zu reduzieren, der bei der Verwendung zufälliger initialer Kantengewichte anfällt, werden jedoch bei neuen Klassifikationsaufgaben in der Regel sog. *Pre-Trained Models*, also vortrainierte Netzwerke dieser internationalen Challenges eingesetzt. Diese verfügen so bereits über die Fähigkeit, Features aus RGB-Bildern zu extrahieren und befinden sich daher näher an der optimalen Lösung des Klassifikationsproblems als komplett zufällig initialisierte Netzwerke (siehe Abb. 4). Dadurch sinkt nicht nur der Rechenaufwand sondern auch die absolut erreichbare Genauigkeit steigt in vielen Fällen.

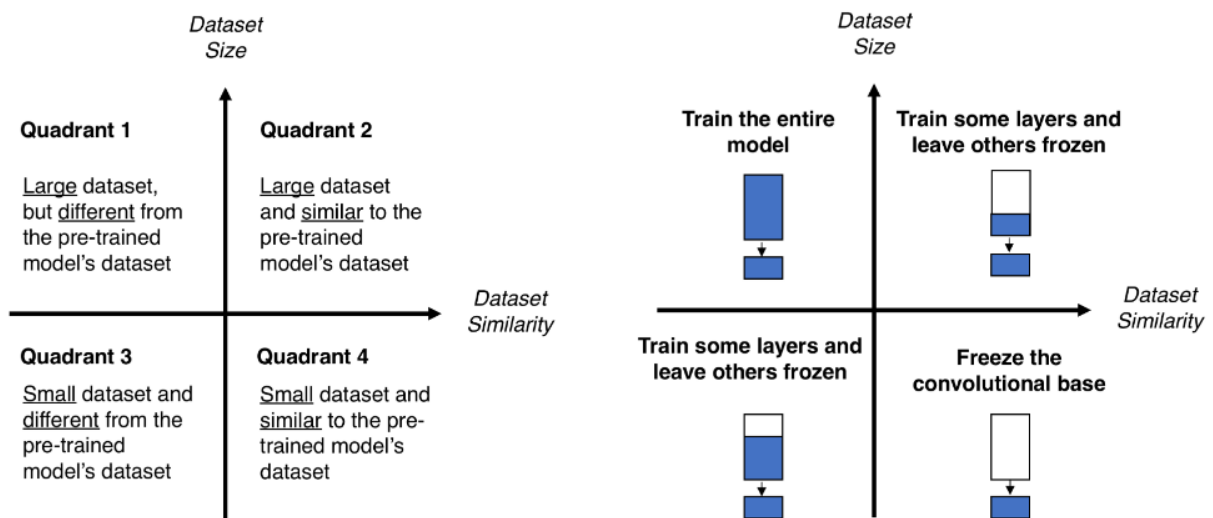


Abbildung 4: Verschiedene Strategien bei der Verwendung vortrainierter Modelle in Abhängigkeit des vorhandenen Datensatzes (übernommen von Marcelino, 2018).

In Abb. 4 werden verschiedene Ansätze beschrieben, um festzulegen welcher Anteil der Kantengewichte eines vortrainierten Netzwerks neu trainiert werden sollte. Im Falle von Solaranlagen bedeutet dies, dass sich diese zwar qualitativ etwas vom vortrainierten Datensatz unterscheiden (beide RGB, aber Orthophotos statt Zentralperspektive), aber mit wachsender Datenbasis auf das Einfrieren niedrigerer Ebenen des Ursprungnetzwerks verzichtet werden kann.



### 3.4 Testen des Stanford DeepSolar-Algorithmus

DeepSolar (Abb. 5) ist ein dem vorliegenden Projekt verwandter Ansatz, der auf die Verwendung einer pretrained Faster-RCNN-Architektur setzt, um Solarpanels zu detektieren. Anschliessend kommt ein auf *Class Activation Maps* basierender Ansatz zum Einsatz, um eine Instanzsegmentierung von Solarpanels zu erreichen.

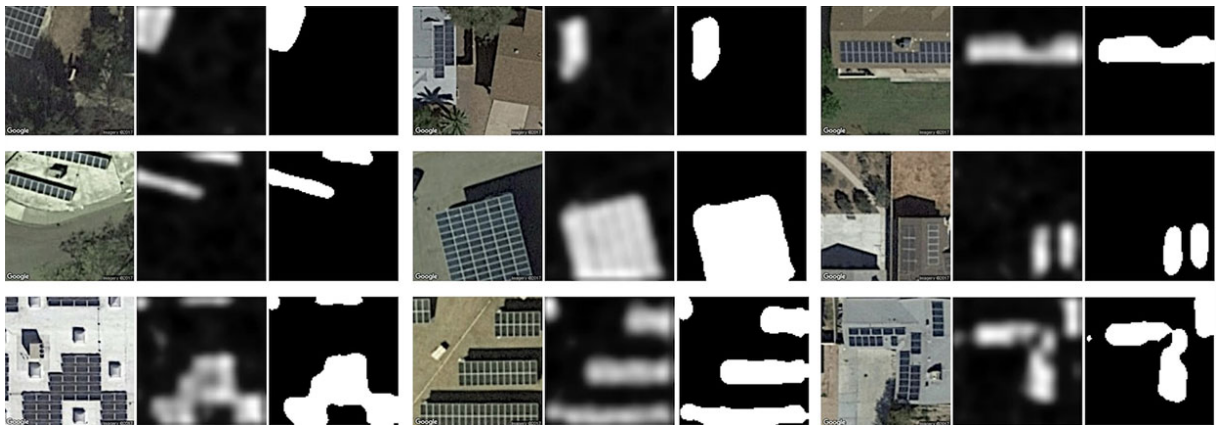


Abbildung 5: Binärmasken aus dem DeepSolar-Projekt (übernommen aus Yu et. al. 2018).

Die Verwendung dieses bereits verfügbaren Algorithmus wurde ursprünglich erwogen, musste aber bei nähergehender Betrachtung aus mehreren Gründen verworfen werden:

- Es werden wesentlich niedriger aufgelöste Trainingsdaten verwendet (Google Maps) als beim vorliegenden Projekt (Swissimage).
- Die Instanzsegmentierung wird **nicht** mit der modernsten Methode erreicht (Mask-RCNN). Es kommt für diesen Aufgabenteil eine teilüberwachte Klassifikation zum Einsatz, wogegen bei SolAI eine **vollständig unüberwachte** Automatisierung angestrebt wird.
- Der Algorithmus ist zwar per Github als Open Source verfügbar, jedoch sind die Ergebnisse trotzdem nicht reproduzierbar: Das initiale Training kann nicht durchgeführt werden, da der verwendete Originaldatensatz nicht öffentlich zugänglich ist.
- Die erreichten Genauigkeiten liegen mit 27,3 % in Wohngebieten und 18,8 % in Gewerbegebieten weit unter den von SolAI angestrebten Genauigkeiten.





### 3.5 DCNN Transfer Learning für SolAI

Für das SolAI-Projekt wird auf ein vortrainiertes Mask-RCNN-Netzwerk mit Resnet50-Backbone zurückgegriffen. Die Implementierung erfolgte in Pytorch, das Pretraining auf dem instanz-segmentierten COCO Dataset. Das Neutrainig der Kantengewichte erfolgt mittels GPU-Parallelisierung auf selbst digitalisierten Datensätzen (Abb. 6, genauer Prozess siehe 4.4).

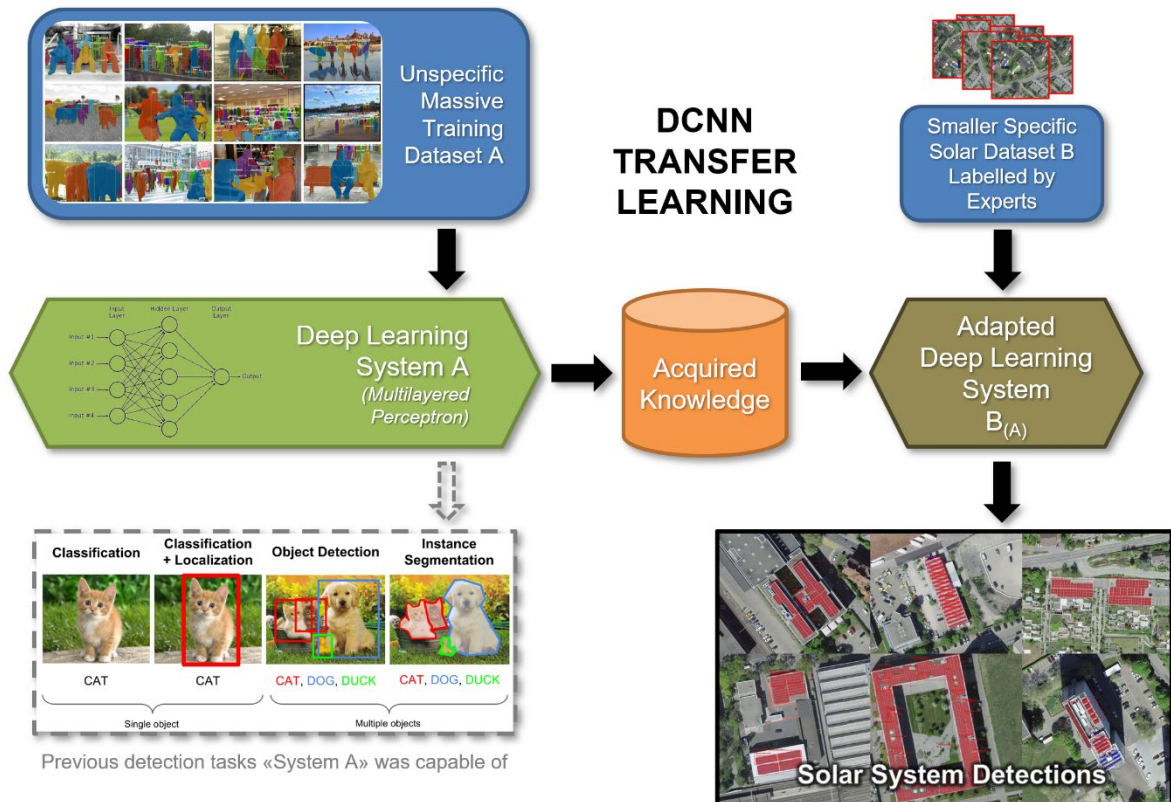


Abbildung 6: Schematische Darstellung des zur Anwendung kommenden Transfer Learning Prozesses mittels DCNNs.

Schliesslich lassen sich die vom Modell errechneten Positionen, Flächen und Typen der detektierten Anlagen als GIS-Datensatz zur Verfügung stellen, sodass dieser auch auf kommunaler und kantonaler Ebene, sowie z.B. für Feuerwehr, Netzplaner usw. gewinnbringend eingesetzt werden kann. Durch die Möglichkeit, dem entwickelten Lernverfahren Daten zu verschiedenen verfügbaren Zeitständen präsentieren zu können, lassen sich zusätzlich zur aktuellen Ertragsstatistik Prognosen errechnen. Damit kann das aktuelle und zukünftige Nutzungspotenzial in der Schweiz und in einzelnen Gebieten genauer bestimmt werden. Diese Erkenntnisse dienen einer Standortbestimmung für die Umsetzung der Energiestrategie 2050. Die Ergebnisse sollen in der Bundesgeodateninfrastruktur visualisiert und präsentiert werden.



## 4 Resultate

### 4.1 Generierung Basisdatensatz (WP 1)

Der im Rahmen des Projektes erstellte Basisdatensatz basiert auf dem Gebäude- und Wohnungsregister (*GWR*) des Bundesamts für Statistik in Kombination mit den Dachneigungen und -orientierungen aus dem *swissBuildings3D*-Datensatz, sowie dem *SwissImage*-Luftbilddatensatz der Swisstopo in 25cm- und 10cm-Auflösung (wo verfügbar, siehe Abb. 7).

Um den Umriss jedes Gebäudes wird später vollautomatisch mit einem Pufferabstand eine Luftbildkachel erzeugt, auf welcher die Binärmasken für die Solarpanels digitalisiert werden können.

Um dieses Ziel zu erreichen, stehen die Basisdaten in einem standardisierten Format (VRT/TIF) auf einem abgesicherten Server zur Verfügung. Diese Formatkombination bietet schnelle Verarbeitungsraten, stellt jedoch hohe Anforderungen an die Speicherkapazität.

Die Luftbilder der swisstopo (2016) mit einer 25cm pro Pixel Auflösung bestehen aus 3433 Teilkacheln (TIF) mit einer jeweiligen Bildgrösse von 17500x12000 Pixeln und haben eine insgesamt Grösse von ca. 2 TB. Insgesamt hat dieser Datensatz also über 720 Milliarden Pixel. Ab dem Jahr 2020 werden diese Daten sogar mit einer Auflösung von 10cm pro Pixel flächendeckend über die ganze Schweiz erhältlich sein. Dies zieht bereits einen etwa 6x höheren Speicherbedarf nach sich. Werden noch Bilddaten älterer Zeitstände hinzugenommen, so steigt die Datenmenge weiter an. Auch der Einschluss weiterer spektraler Kanäle wie bei FCIR oder NIR-TIF (z.B. *SwissImage RS*) benötigt zusätzliche grosse Speicherkapazitäten.

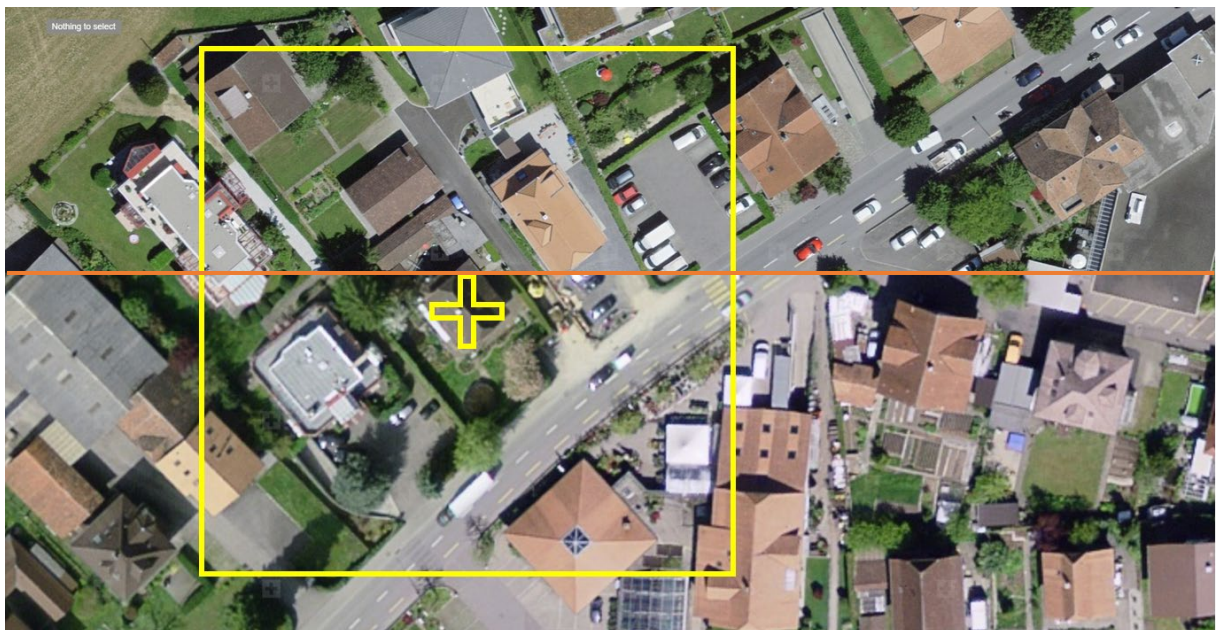


Abbildung 7: Zentroid (Kreuz) einer Beispielkachel (gelber Rahmen), im Vergleich SwissImage 10cm (oben) und 25cm (unten).



## 4.2 Installation von Hard- und Software (WP 2)

Das bestehende GPU-Cluster des Instituts Geomatik wurde im Rahmen des Projekts ausgebaut, um die erforderlichen Kapazitäten für ein komplexes neuronales Netz effizient bereitstellen zu können. Hierfür wurde ein *HP Apollo 6500 Compute Cluster* in das Rechenzentrum im Campus Muttenz der FHNW installiert welches auf vier aktuelle Hochleistungsgrafikkarten des Typs *Nvidia Tesla V100* zugreift.

Softwareseitig wurde auf dem Cluster ein neues Unix-basiertes Betriebssystem installiert, auf welches per SSH ferngesteuert zugegriffen werden kann.

Als Entwicklungsumgebung wird das kollaborative browserbasierte System *JupyterHub* eingesetzt (Abb. 8). *JupyterHub* bietet Benutzern geschützten Zugriff auf Rechenumgebungen und Ressourcen, ohne die Benutzer mit Installations- und Wartungsaufgaben zu belasten. Benutzer wie Studenten, Forscher und Datenwissenschaftler können ihre Arbeit in ihren eigenen Arbeitsbereichen an gemeinsamen Ressourcen erledigen, die von Systemadministratoren effizient verwaltet werden können. So wird eine vorkonfigurierte datenwissenschaftliche Umgebung für jeden Projekt-Benutzer von SoIAI bereitgestellt.

Ein grosser Vorteil von *Jupyter*-Notebooks ist deren flexibler Einsatz. Programmcode kann mit wenig Aufwand auf andere Umgebungen portiert werden. Anpassungen und Weiterentwicklungen erfolgen durch die niedrigschwellige Browserumgebung. Durch Erweiterungen und Ressourcenmanagement können die Umgebungen schnell an sich verändernde Anforderungen angepasst werden.

Die Versionierung erfolgt mittels *Github*, einer Cloud-Contribution-Plattformlösung für Codeentwicklungen, mit welcher auch ein späterer Open-Access-Zugang bereitgestellt werden kann.

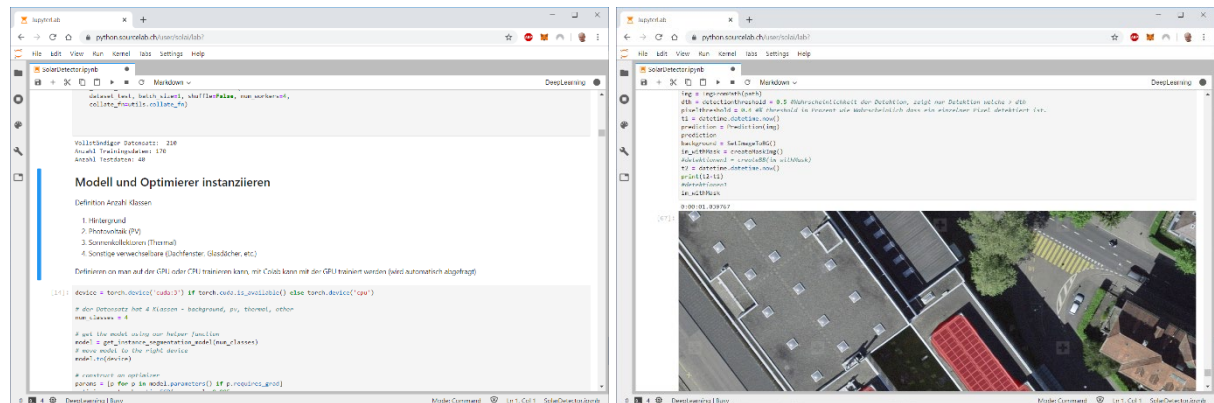


Abbildung 8: Software-Entwicklung direkt auf dem Hochleistungsrechner im Webinterface JupyterLab



## 4.3 Crowd Contribution Webservice (WP 3)

### 4.3.1 Webinterface zum Digitalisieren von Solaranlagen

Zur Erhebung von Trainingsdaten wurde ein Crowd Contribution Webservice entwickelt, der den beschriebenen Basisdatensatz aus WP1 verarbeitet und eine effiziente Erfassung von digitalisierten, polygonzug-basierten Binärmasken ermöglicht.

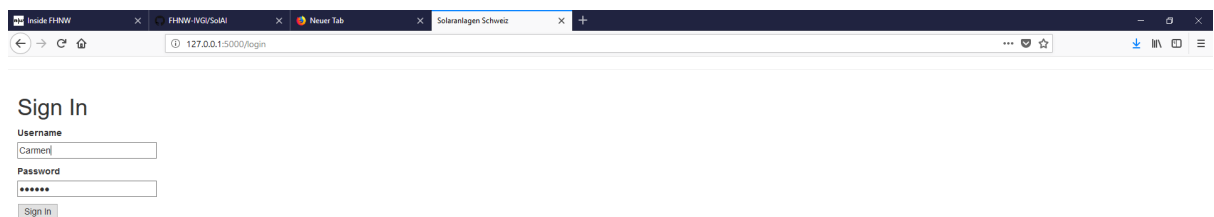


Abbildung 9: Login-Interface.

Aus Sicherheits-, sowie Datenschutz- und Lizenzgründen (insbesondere hinsichtlich des SwissImage-Datensatzes) ist der Zugang zum Crowd-Contribution-Interface mittels einer Login-Maske geschützt (Abb. 9). Die Übertragung von Daten wird verschlüsselt (HTTPS-Protokoll).

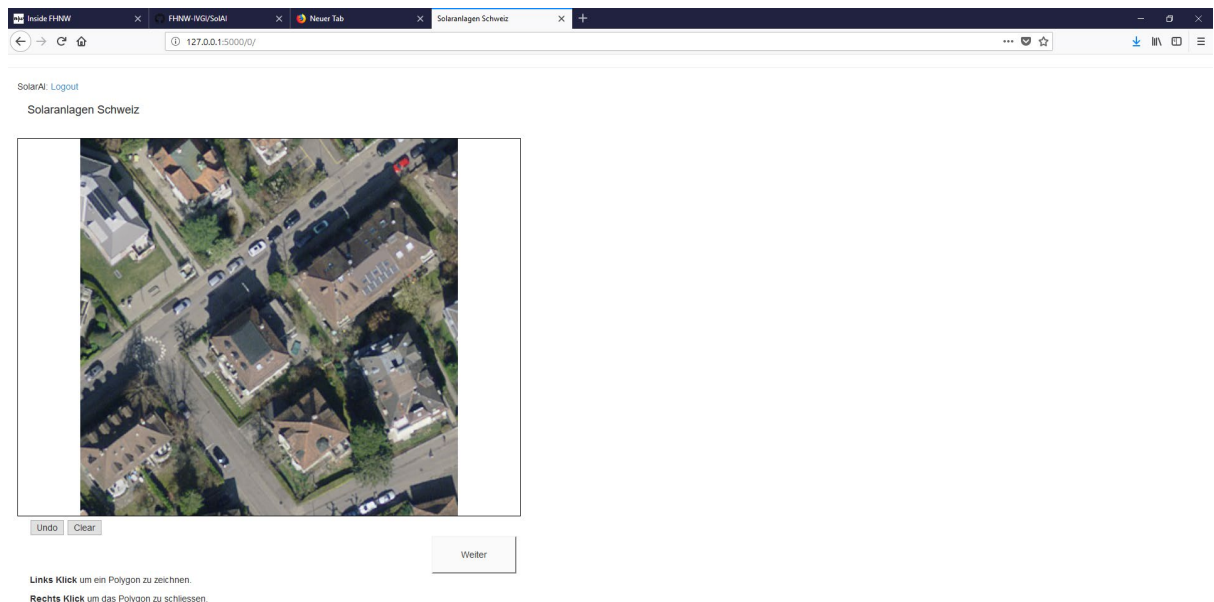


Abbildung 10: Beispiel einer randomisiert ausgewählten Luftbildkachel.

In einem Viewport werden randomisiert Luftbildkacheln um Schweizer Gebäude angezeigt (Abb. 10). Diese Kacheln werden beim Aufrufen des Viewports automatisch im Hintergrund generiert und gespeichert. Ist auf dem Luftbild bereits eine Solaranlage durch einen anderen Bearbeiter eingezeichnet, so kann diese verifiziert werden. Über Klicks in den Viewport können neue Solaranlagen erfasst werden. Mittels eines Dialogfensters kann anschliessend auch die Klassifizierung der Solaranlage gewechselt werden (Abb. 11).

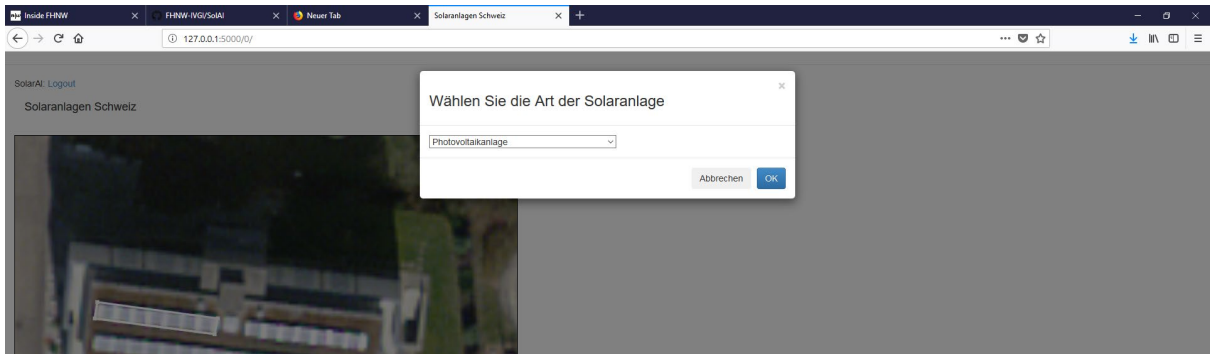


Abbildung 11: Erfassung einer neuen Solaranlage.

Zusätzlich zur Auswahl von Photovoltaik-Anlagen ist die Erfassung von Solarthermie-Systemen, sowie die Erfassung sonstiger Strukturen (nicht-Solar, z.B. Dachfenster, Vordächer, Gewächshäuser, o.ä. mit Verwechslungspotential) möglich. Im Hintergrund werden anschliessend Shapefiles und GeoJSON-Datienobjekte für jede Kachel angelegt.

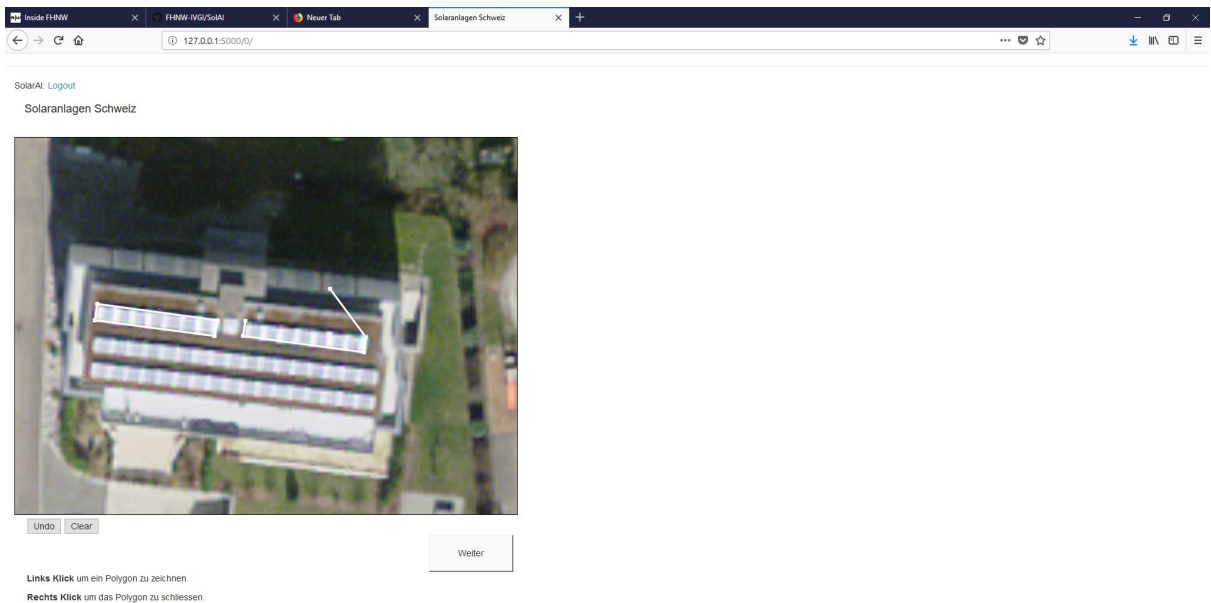


Abbildung 12: Auch Funktionen zum Rückgängig-Machen fehlerhafter Digitalisierungen, sowie nachträgliche Bearbeitungen sind möglich.

Zusammenhängende Zeilen von Solarpanels werden gemeinsam erfasst. Sobald der Abstand eines Moduls oder einer Zeile grösser als die Modulbreite ist, wird eine getrennte Erfassung angestrebt. Auch können Fehler in der Digitalisierung durch eine «Rückgängig»-Funktion korrigiert werden (Abb. 12).



#### 4.3.2 Generierung des Trainingsdatensatzes

Der grösste Teil der Kacheln wurde in einem koordinierten einwöchigen Workshop von geschulten Annotations-ExpertInnen durch manuelle Digitalisierung erfasst (Abb. 13).

Als Datenbasis für die Generierung der zu markierenden Kacheln standen punktuelle Solarstandortdaten aus der Förderdatenbank des Kantons Aargau sowie Faster-RCNN-basierte Vordetektionen (siehe 4.4) zur Verfügung.

Gelabelt wurde ausschliesslich auf Daten der Flugjahre 2017 und 2018, da 2019 zum Workshop noch nicht zur Verfügung stand (Abb. 14).

Der Trainingsdatensatz umfasst Vektoren von ca. 31'400 Solaranlagen (davon ca. 22'000 PV) auf 7'839 Luftbildkacheln mit der Ausdehnung 80x80m, d.h. bei 10cm GSD 800x800 Px (Abb. 15).



Abbildung 13: Labeling Workshop mit 10 eigens für diese Aufgabe geschulten Annotations-ExpertInnen

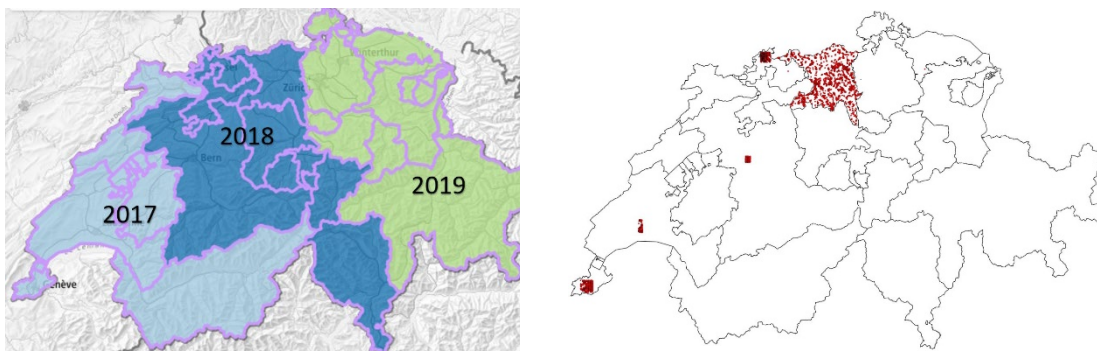


Abbildung 14: Links - Flugjahre Swissimage 10cm. Rechts - Positionen der Labels für Solaranlagen.

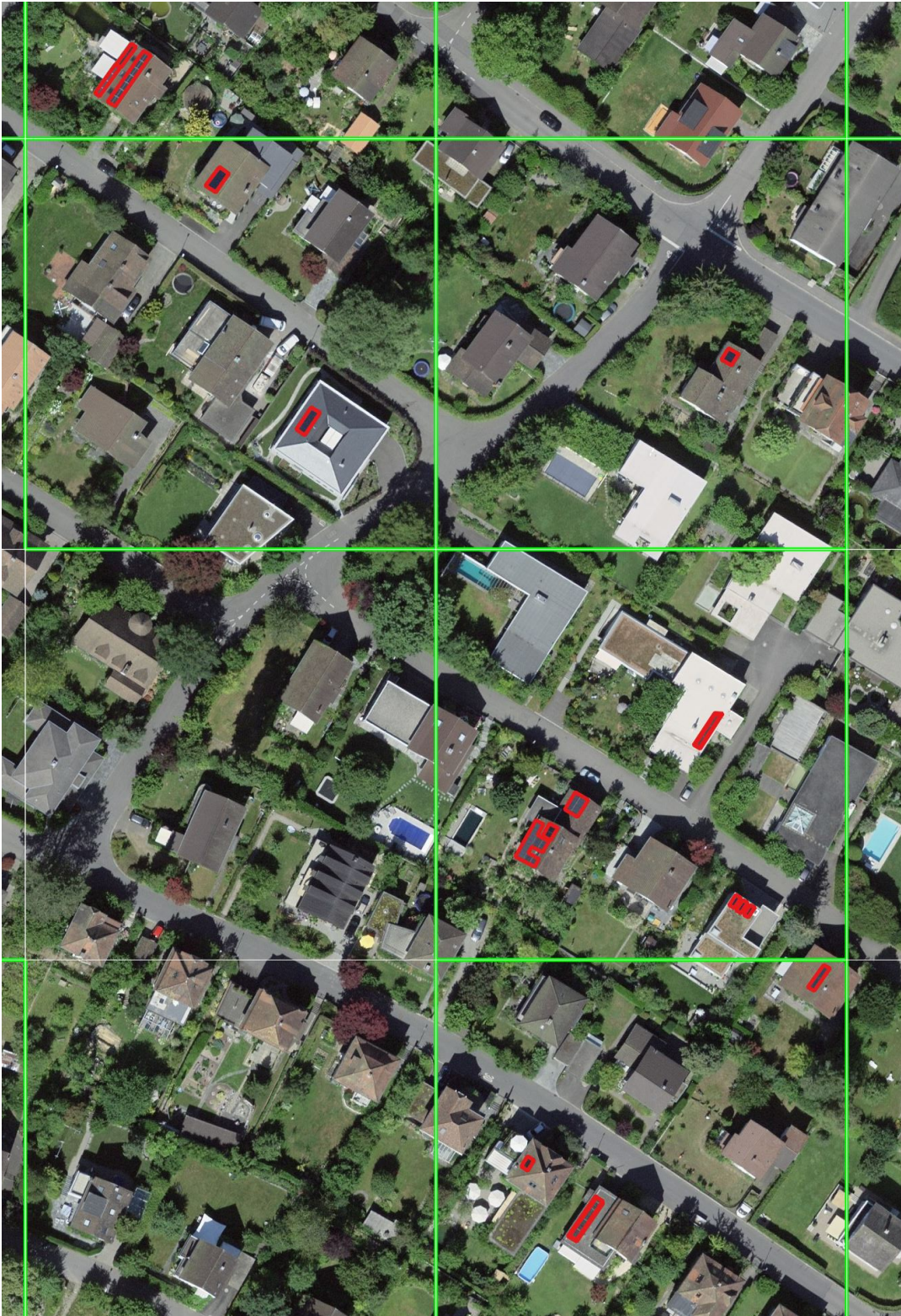


Abbildung 15: Beispiele für trainierbare Labels (rot) auf 80m Kacheln (grün) in einem Aargauer Wohngebiet.



## 4.4 Erstellung der Deep Learning Networks (WP 4)

### 4.4.1 Vordetektionen mittels Faster-RCNN

Zunächst wurde die Entwicklung eines Faster-RCNN-Modells zur Objektdetektion angestrebt (Abb. 16). Dieses ist lediglich in der Lage, eine Lokalisierung der Solaranlagen vorzunehmen, indem um eine Detektion eine sog. „Bounding Box“ gezeichnet wird. Das Annotieren der Trainingsgebiete gestaltete sich hiermit wesentlich leichter, da lediglich grobe Rahmen eingezeichnet werden müssen und dementsprechend noch keine vollständigen Annotationsdaten aus dem Crowd Contribution Webservice vorliegen mussten. Erst in einem weiteren Schritt wird dieses Netzwerk anschliessend um die Funktionen zur Instanz-Segmentierung (Mask-RCNN) erweitert.

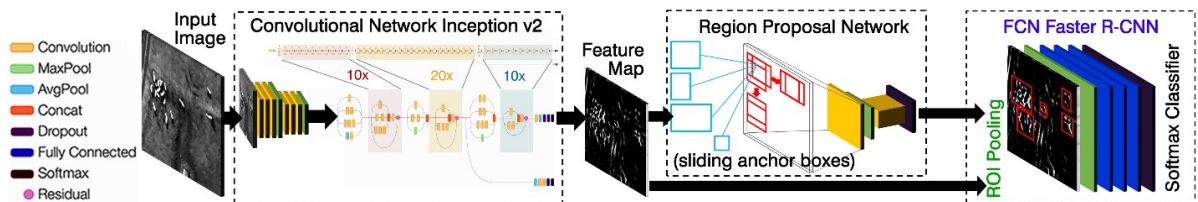


Abbildung 16: Aufbau der verwendeten Faster-RCNN-Architektur (übernommen aus Meyer et. al. 2019).

Das Netzwerk wurde in dieser ersten Entwicklungsstufe auf Basis eines vom Data Scientist eigenständig markierten Datensatzes von 600 Solaranlagen trainiert und dient der technischen Evaluation des eingesetzten Netzwerktyps und somit dem Proof of Concept. Der Trainingsvorgang umfasste 200'000 Iterationen. Ein kleiner Teil der markierten Daten (15 % , ca. 90 Solaranlagen) wurde dabei für eine Vorevaluation zurückbehalten.

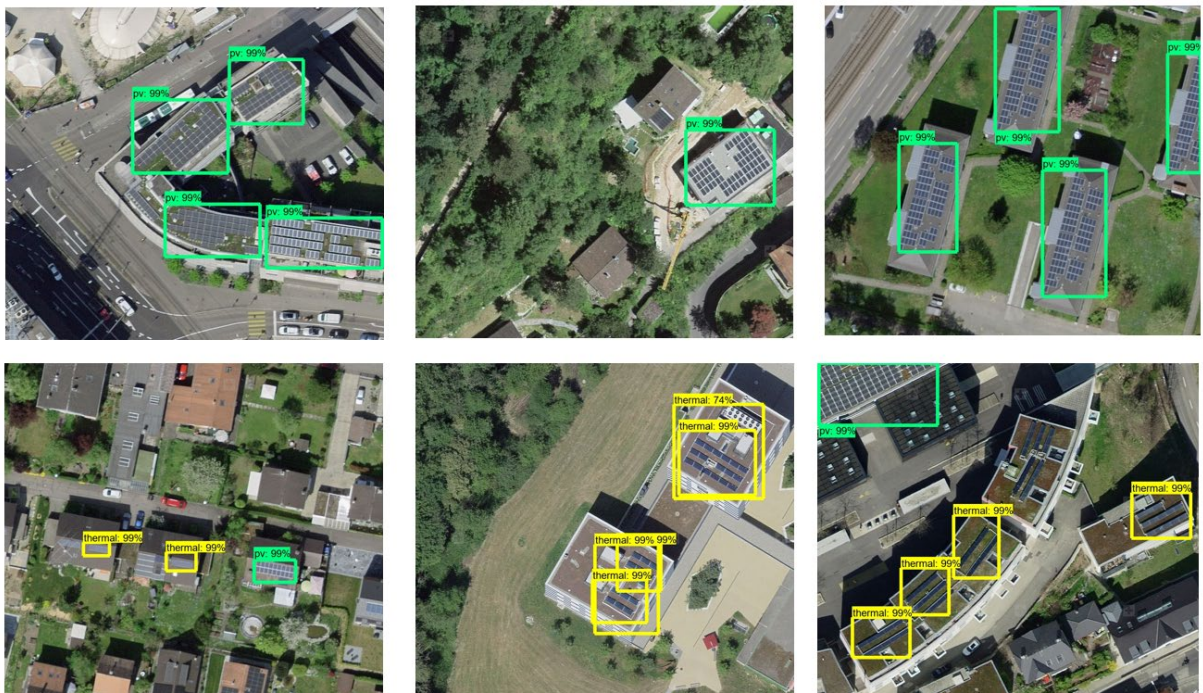


Abbildung 17: Beispielergebnisse beim Inferencing des trainierten Faster-RCNN-Modells auf unbekannten Testdaten.





Grössere Solaranlagen wurden in diesem Test auf dem unabhängigen Testdatensatz vom Netzwerk bereits sehr zuverlässig erkannt (Abb. 17). Auch die Unterscheidung zwischen Solarthermie und Photovoltaik funktioniert in vielen Fällen bereits trotz der kleinen Datenbasis gut.

Average Precision (AP) @[ IoU=0.50:0.95   area= all   maxDets=100 ]	= 0.472
Average Precision (AP) @[ IoU=0.50   area= all   maxDets=100 ]	= 0.782
Average Precision (AP) @[ IoU=0.75   area= all   maxDets=100 ]	= 0.480
Average Precision (AP) @[ IoU=0.50:0.95   area= small   maxDets=100 ]	= -1.000
Average Precision (AP) @[ IoU=0.50:0.95   area=medium   maxDets=100 ]	= 0.247
Average Precision (AP) @[ IoU=0.50:0.95   area= large   maxDets=100 ]	= 0.619
Average Recall (AR) @[ IoU=0.50:0.95   area= all   maxDets= 1 ]	= 0.349
Average Recall (AR) @[ IoU=0.50:0.95   area= all   maxDets= 10 ]	= 0.588
Average Recall (AR) @[ IoU=0.50:0.95   area= all   maxDets=100 ]	= 0.606
Average Recall (AR) @[ IoU=0.50:0.95   area= small   maxDets=100 ]	= -1.000
Average Recall (AR) @[ IoU=0.50:0.95   area=medium   maxDets=100 ]	= 0.428
Average Recall (AR) @[ IoU=0.50:0.95   area= large   maxDets=100 ]	= 0.723

Abbildung 18: Evaluationsmetriken für die Genauigkeit der Detektion.

Trotz des vergleichsweise kleinen Trainingsdatensatzes übertrifft die durchschnittliche Nutzergenauigkeit (*Average Precision*, AP) des Netzwerks mit insgesamt 47,2 % die vom DeepSolar-Projekt erreichten Genauigkeiten bereits deutlich (Intersect over Union zwischen 50 % und 95 %, vgl. 3.4, Abb. 18). Für grössere Anlagen (*area=large*) liegt die durchschnittliche Präzision derzeit sogar bei 61,9 %. Kleinere und mittlere Anlagen werden nur in Einzelfällen erkannt. (Details zu Metriken wie Precision und Recall siehe 4.7)

Diese ersten Evaluationen zeigten, dass für sehr kleine Anlagen, sowie Anlagen am Rand der Kacheln noch zusätzliche Trainingsdaten benötigt wurden. Diese Fälle mussten im folgenden insbesondere durch die Erweiterung des Algorithmus zum Mask-RCNN adressiert werden. Trainingsdaten für Instanzsegmentierung aus dem Crowd-Contribution-Tool lassen sich auch problemlos wieder in Bounding Boxes übersetzen sodass mit grösseren Eingangsdatensätzen auch für den Fall der reinen Lokalisation eine deutliche Steigerung der Genauigkeit (auch bei kleineren Anlagen) erwartet werden kann.



#### 4.4.2 Entwicklung des pixelgenauen Mask-RCNN

Die verwendete Mask-RCNN-Architektur wurde mit dem Deep-Learning-Framework Pytorch (Paszke et al., 2019) implementiert. Die Parallelisierung erfolgt mit CUDA-fähigen GPUs auf dem High-Performance-Computing-Cluster der FHNW-Serveranlage in Muttenz. Das Mask-RCNN-Backbone wird durch eine ResNet-50-Implementierung (Waleed, 2017) gebildet und von einem Feature Pyramid Network (FPN) begleitet. Diese Kombination von Code-Elementen resultiert in einem neuronalen Netz, das mehr als 40 Mio. Parameter nutzt. Der Datensatz besteht einerseits aus ca. 8'000 RGB-Bildern (80 x 80m, abgeleitet aus Swissimage) und andererseits aus den dazugehörigen ca. 30'000 Merkmalsregionen (überlagerte Annotationen), die durch Pixelmasken repräsentiert werden (siehe 4.3.2).

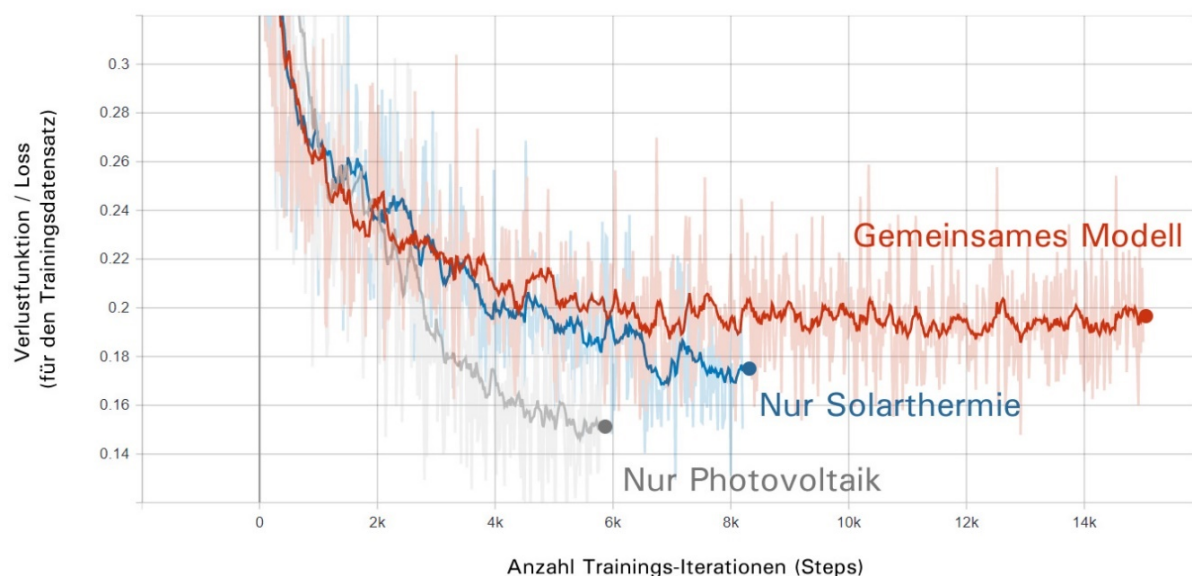


Abbildung 19 - Die Trainings-Verlustfunktion sinkt während des Trainingsvorgangs.

Das Training wird iterativ durchgeführt, indem Teilmengen des gekachelten Datensatzes dem Netzwerk präsentiert werden, um die Kantengewichte im Netzgraphen zu verändern (Optimierung der Klassifikationen). Der Fortschritt wird Schritt für Schritt durch statistische Minimierung der Verlustfunktionen (Training Loss) gemessen.

In Abb. 19 zeigt sich, dass differentielle Einzel-Modelle (also solche welche nur auf Subsets des Datensatzes mit PV oder Thermal trainiert wurden) schneller tiefere Verlustwerte erreichen als ein kombiniertes Modell. Es können dabei nur Kacheln trainiert werden, die Masken (Labels) enthalten. Zwei kleinere Teilmengen aller labelhaltigen Kacheln wurden aus dem Training zurückbehalten, so dass insgesamt nur 80 % der trainierbaren Kacheln dem Netz zur Verlustminimierung präsentiert werden. Der Validierungsdatensatz (10 %) und der Testdatensatz (10 %) enthalten ebenfalls Labels, sind aber statistisch unabhängig vom Trainingsdatensatz. Der Testdatensatz dient als unvoreingenommene Reserve, um später die Detektorleistung auf zuvor "ungesehenen", aber gelabelten Daten auszuwerten (Evaluation siehe 4.7).

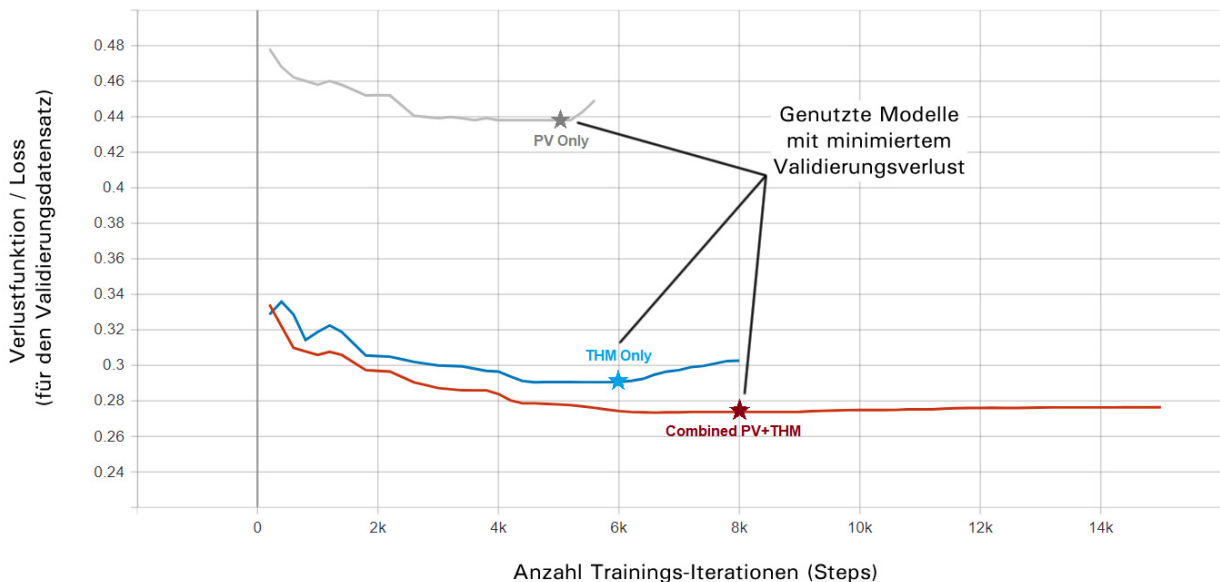


Abbildung 20: Die Validierungs-Verlustfunktion steigt nach einem Scheitelpunkt wieder an. Am Minimum ist das Modell auf Generalisierungsfähigkeit optimiert, das Training kann nach Bemerkten des Anstiegs gestoppt werden.

Der Validierungsdatensatz wird verwendet, um nach jeder 100sten Iteration eine wiederkehrende Auswertung während des Trainings durchzuführen, welche dem Netzwerk zur Fehlerrückführung dient. Darüber hinaus bietet der Validierungsdatensatz die Möglichkeit zu erkennen, wann das Training gestoppt werden kann: Wenn die Verlustfunktion auf dem Validierungsset ein Minimum erreicht hat, würde nach diesem Punkt ein weiteres Training das Modell in ein sog. «Overfitting»-Szenario bringen. Bildlich gesprochen würde nach dem minimalen Scheitelpunkt der Validierungs-Verlustfunktion das Modell nur noch Trainingsdaten «auswendig lernen», dabei jedoch die Generalisierungsfähigkeit für die Anwendbarkeit auf neue Daten verlieren.

In Abb. 20 zeigt sich, dass ein kombiniertes Modell aus PV und Thermal bereits bei dieser Art von Zwischenevaluation den niedrigsten Gesamtverlustwert aufweist und demnach die höchste Generalisierungsfähigkeit gewonnen hat.

Jeder Detektion (auch auf dem Validierungsdatensatz) wird vom Detektor ein Konfidenz-Score zugewiesen (Abb. 21) - eine Art Sicherheitsmetrik aus dem Lernprozess). Legt man nun sichere und unsichere Detektionen zusammen, kann ausserdem die Performanz des Modells unter Verwendung von Bounding Boxes (siehe 4.4.1) oder Maskensegmenten (ebf. Abb. 21) getrennt als Sondertest untersucht werden. Während beim kombinierten Modell die durchschnittliche Nutzergenauigkeit (AP, average precision) für Segmente bei 83 % und für Bounding Boxes bei 82 % lag, waren die Werte für die Einzelmodelle etwas niedriger: 81 % für Segmente und Bounding Boxes im «Nur PV»-Fall und 77 % für Segmente, bzw. 79 % für Bounding Boxes im «Nur Thermal»-Fall. Diese Zwischentests zeigen, dass ein Rückgriff auf Bounding Boxes als vermeintlich sicherere Detektionsvariante in den neuen Mask-RCNN-Modellen nicht mehr greift und daher direkt eine Optimierung des Modells hinsichtlich der Segmentausgabe angestrebt werden kann. Die hier erwähnten Precision-Werte liegen niedriger als tatsächlich erreichbare Metriken (z.B. F1-Score, siehe 4.7), da auch Detektionen mit niedriger Konfidenz in die Auswertung miteinbezogen werden.

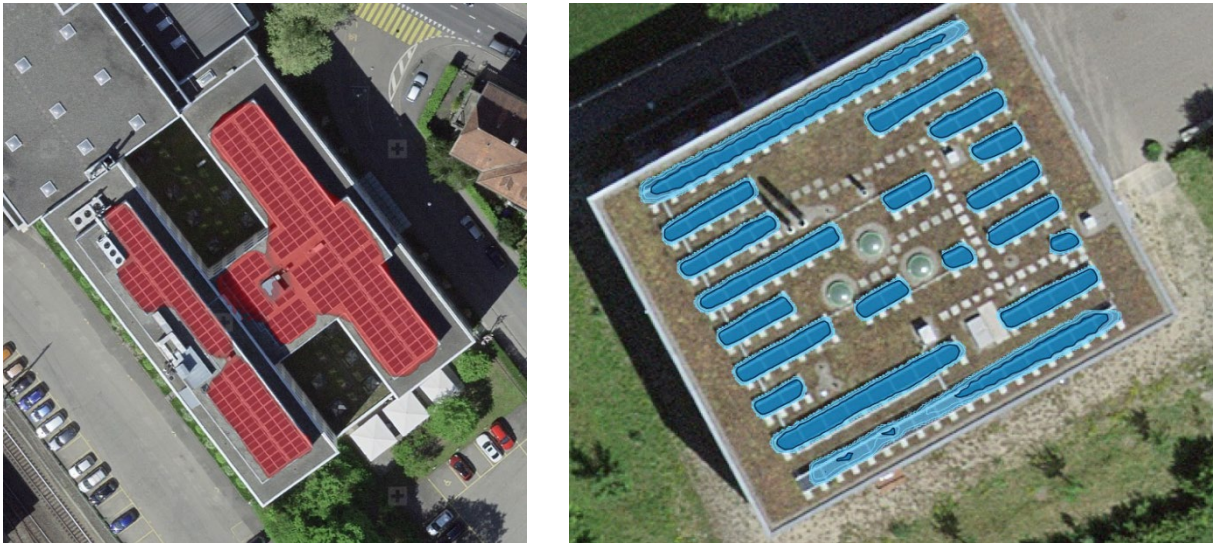


Abbildung 21: Beispielhafte Anwendung des trainierten Mark-RCNN auf Validierungsdaten. Links: Ausgabe der Segmente als Binärmaske (rot) mit einem universellen niedrigen Konfidenz-Schwellwert; grössere Teile des Dachs sind in der Detektion enthalten. Rechts: Ausgabe der Segmente mit hohem Konfidenzschwellwert (dunkelblau), hellblaue Bereiche würden in diesem Fall verworfen.

Die Passgenauigkeit der Masken (Mask Accuracy) wurde an den Trainingsdatensatz angepasst. Die Modelle erreichen hierbei sehr hohe Genauigkeiten von über 94 % (Abb. 22). Diese hohen Genauigkeitswerte sind jedoch statistisch nicht unabhängig, da sie auf den Trainingsdaten erzeugt werden und repräsentieren daher eher den Trainingsfortschritt als eine erreichte Generalisierungsfähigkeit.

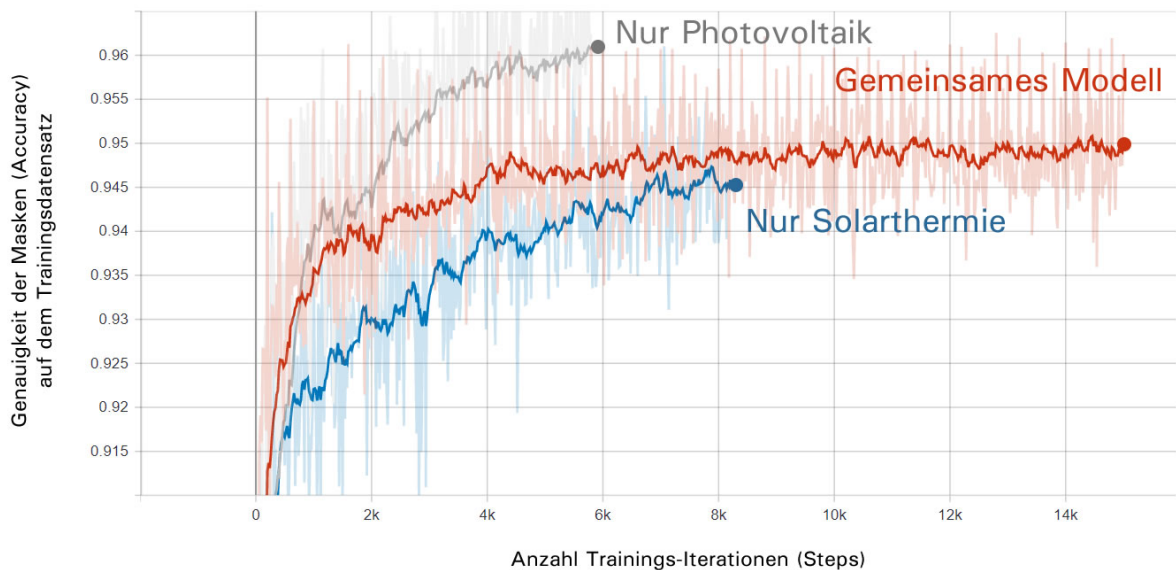


Abbildung 22: Die Passgenauigkeit der Masken im Trainingsdatensatz ist grundsätzlich sehr hoch.



## 4.5 Anwendung des Netzwerks auf die Schweiz (WP 5)

Die Anwendung des Netzwerks vom Dezember 2020 auf die Luftbilddatensätze der ganzen Schweiz stellte sich als herausforderndes Unterfangen heraus, da die hohe Datenmenge und Prozessierungskapazität eine einfache serielle Abarbeitung aller Luftbildkacheln nicht innerhalb einer realistischen Zeitspanne erlaubten. Ein einfacher Durchlauf wurde mit einer seriellen ununterbrochenen Prozessierungsdauer von ca. 100 Tagen geschätzt.

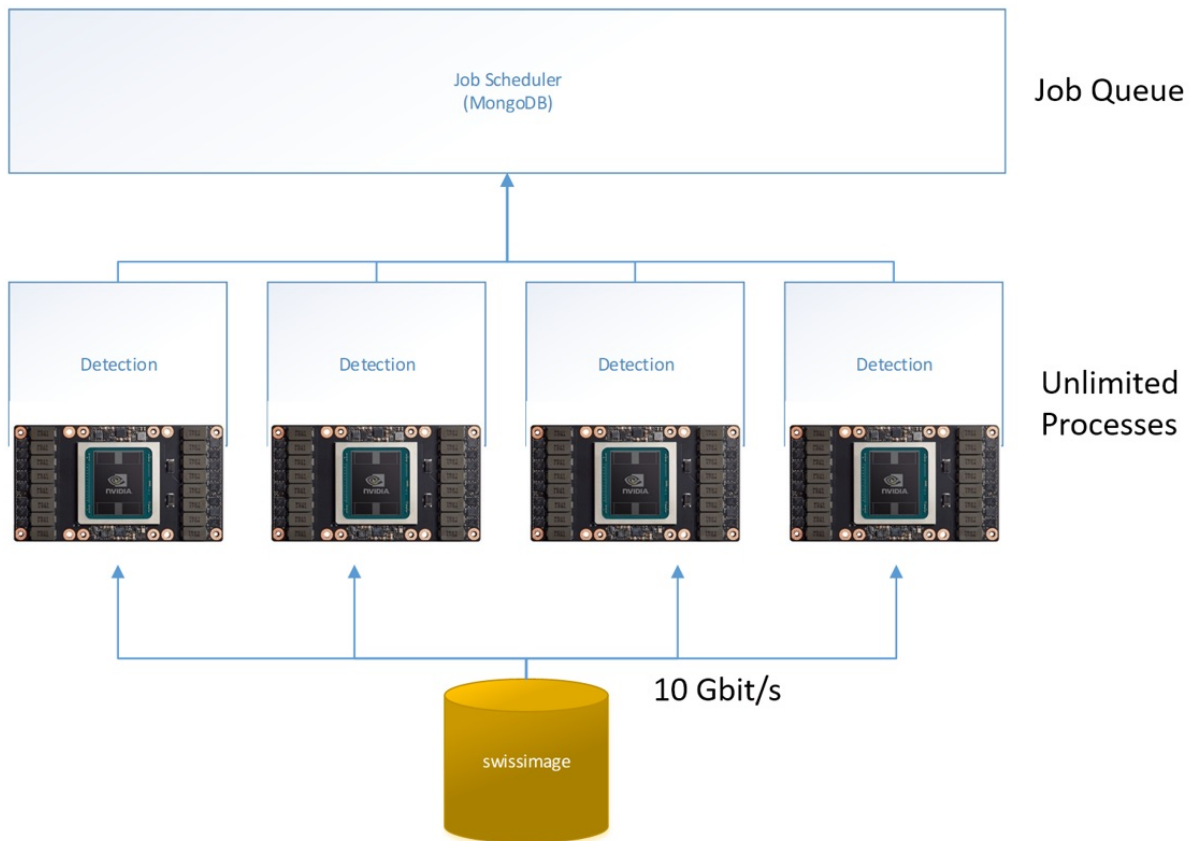


Abbildung 23: Schematische Darstellung der Inference-Parallelisierung

Durch die speziell für diese Aufgabe umgesetzte Implementierung einer Parallelisierungsmechanik mittels Shell Scripts konnte die Dauer eines Inference-Komplettdurchlaufs (Prädiktionen über das gesamte Staatsgebiet der Schweiz) auf ca 4 bis 10 Tage (je nach Rechenauslastung) gesenkt werden. Dies wurde ermöglicht durch einen Job Scheduler in Verbindung mit einer NoSQL-Datenbank (MongoDB), welcher die vier GPUs gleichzeitig ansteuert und kontinuierlich unter Volllast setzt (Abb. 23).



## 4.6 Anwendung auf ältere Luftbilddatensätze (WP 6)

Das neuronale Netzwerk funktioniert am besten mit 10 cm-Auflösung. Eine Anwendung auf ältere Luftbilddatensätze ist nicht direkt möglich. Für jede weitere Auflösungsstufe muss ein eigenes neuronales Netzwerk erstellt werden. Dies ist nicht praktikabel, da die erfassten Annotationen zwischen den beiden Zeitständen nicht kompatibel sind. Ausserdem ist die Auflösung auf älteren Orthophotos häufig schlicht zu schlecht, um Solaranlagen zuverlässig zu detektieren.

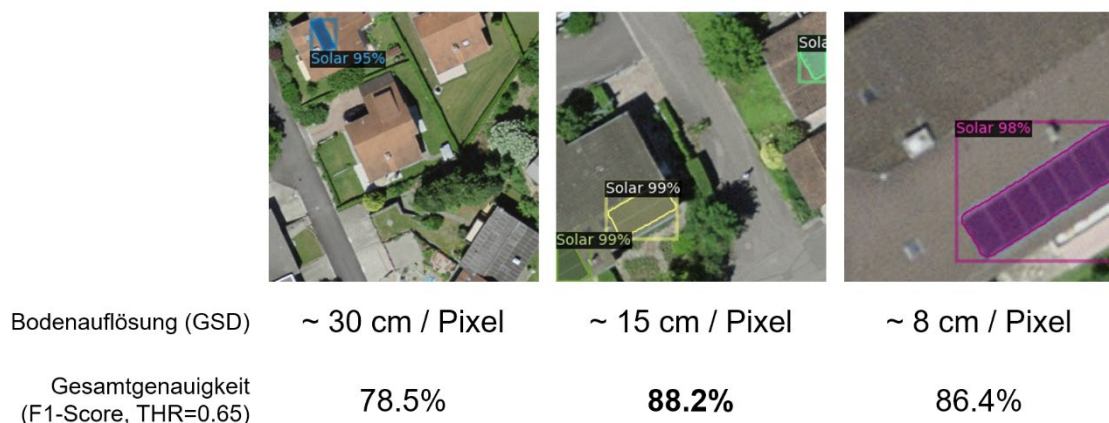


Abbildung 24: Tests auf Resampling-Daten mit verschiedenen Zoom-Levels.

Um den Einfluss der Bodenauflösung zumindest abschätzen zu können, wurden verschiedene Resampling-Bildgrößen aus dem Swissimage-Datensatz extrahiert, trainiert, optimiert und ausgewertet (Abb. 24). In der Gesamtperformanz schnitt das kombinierte 15cm-Modell noch gleich gut wie das finale 10cm-Modell ab (siehe 4.7, 88,2 % F1-Score), eine Vergrößerung auf 8cm/Px reduzierte die Genauigkeit bereits leicht, eine Reduktion auf 30cm/Px zeigte schon wesentlich deutlichere Einbussen. Sollte ein Modell für die älteren 25cm-Swissimage-Daten daher in Zukunft trainiert werden, könnten lediglich Genauigkeiten etwa in diesem Bereich erwartet werden. Die Herausforderung der fehlenden Annotationen, d.h. des ressourcen-intensiven Relabelings bleibt jedoch bestehen, weshalb dieser Ansatz nicht weiter verfolgt wurde.



## 4.7 Evaluation (WP 7)

### 4.7.1 Vorgehen

Die Evaluation des Modells wurde auf dem unabhängigen Testdatensatz ausgeführt – also jenen 15 % der gelabelten Daten, die weder im Training verwendet noch im Validierungsdatensatz zur Fehlerrückführung verwendet wurden. Verwendet wurde das Assessment Script des Object Detection Frameworks des Swiss Territorial Datalab (STDL).

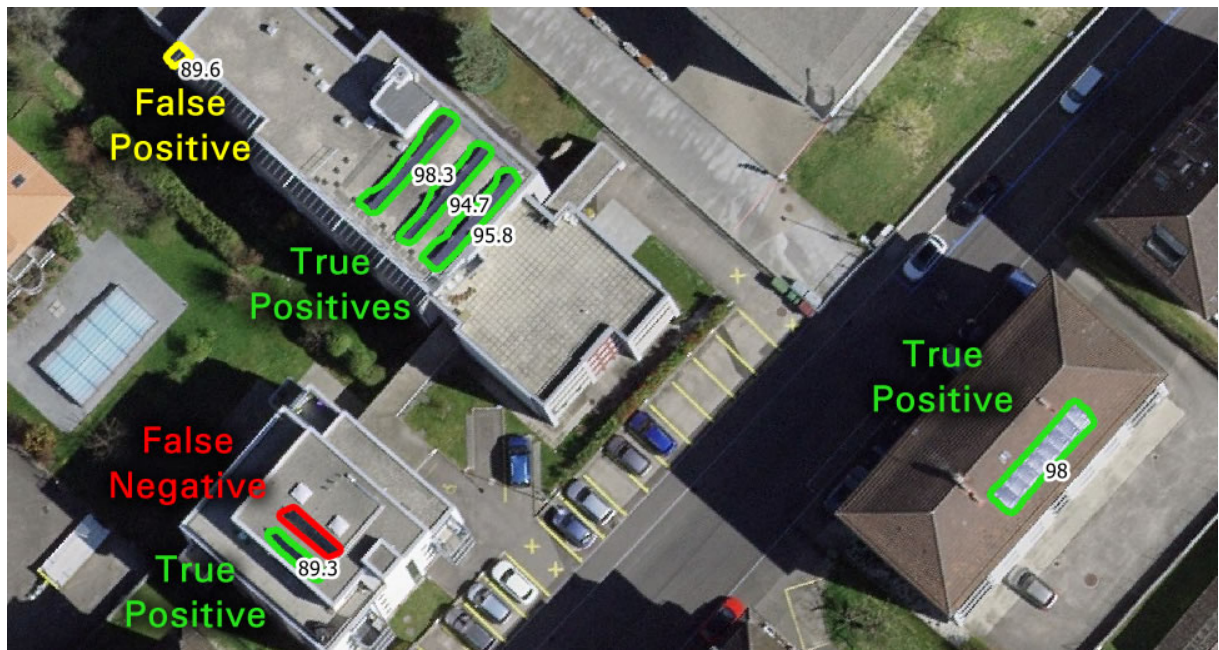


Abbildung 25: Beispiele für die Klassifikation von Prädiktionen auf dem Testdatensatz.

Mit dem Testdatensatz kann die Überlappung zwischen den Vorhersagen (Prädiktionen aus dem Inferencing) und den manuell erhobenen Labels überprüft werden (Abb. 25). Wird eine hinreichend grosse Überlappung zwischen einem Label und einer Vorhersage gefunden (i.d.R. 50 %), so wird diese Erkennung als "True Positive" (TP) betrachtet. Hat der Detektor ein Label komplett übersehen oder nur unzureichend detektiert, so wird diese Detektion als "Falsch Negativ" (FN) gewertet. Hat der Detektor eine Solaranlage vorhergesagt, welche in den gelabelten Daten nicht vorhanden war, wird dies als "Falsch positiv" (FP) betrachtet.

Alle Detektionen auf der Fläche des Testdatensatzes können nun mit diesem verglichen, d.h. diese drei Fälle (TP, FP und FN) können gezählt werden. Diese Auswertung bezeichnet man auch als Konfusionsmatrix oder Konfusionszählung. Richtig negative (TN) werden bei der Objektdetektion nicht ausgewertet, da es sich dabei um den (unzählbaren) Hintergrund handelt.



Das Zählen von TPs, FPs und FNs auf dem Testdatensatz ermöglicht darüber hinaus die Berechnung von Standardmetriken wie Nutzergenauigkeit (Precision), Produzentengenauigkeit/Trefferquote (Recall) und Gesamtgenauigkeit (F1-Score). Precision ist dabei der Anteil TPs an der Gesamtheit aus TP+FP, während Recall als der Anteil TPs an der Gesamtheit aus TP+FN definiert ist (Abb. 26).

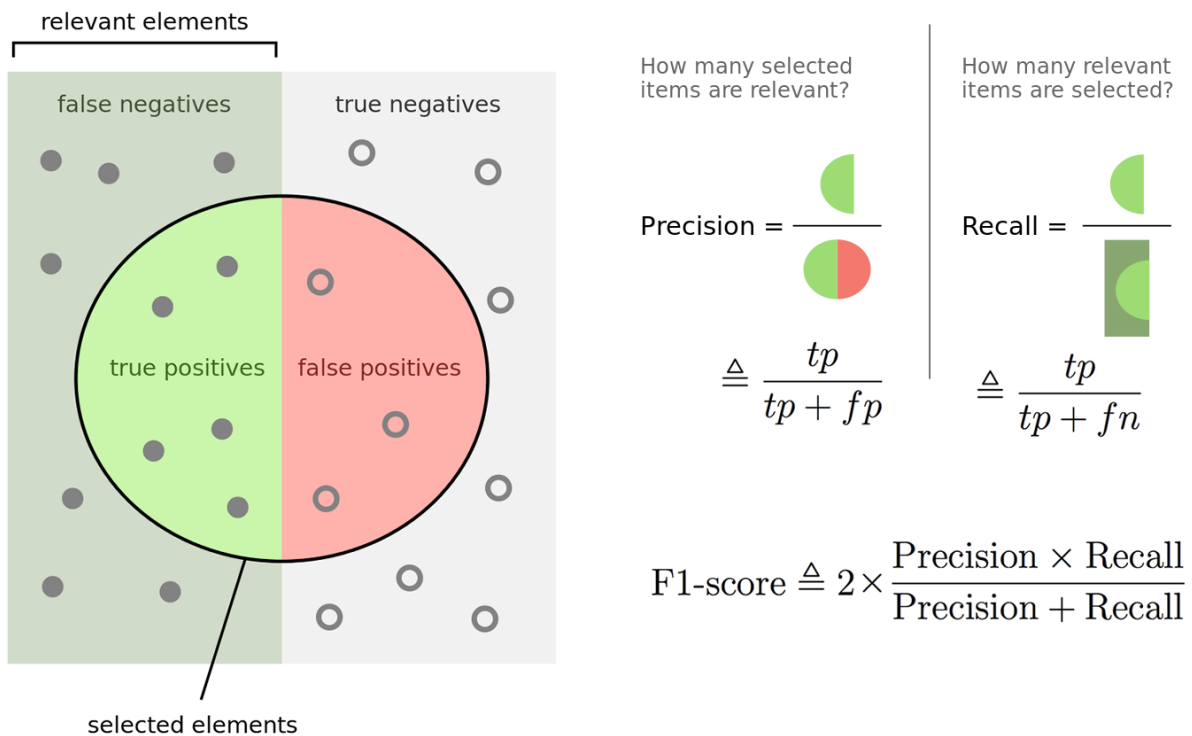


Abbildung 26: Veranschaulichung der Berechnung der Standardmetriken (teilweise übernommen von Wikimedia Commons, 2021).

Die Zählungen, sowie die Metriken können als Funktion des minimalen Konfidenzschwellenwerts (Threshold) aufgezeichnet werden, der auf einen akzeptablen Prozentsatz für eine bestimmte Erkennungsaufgabe eingestellt werden kann. Ein niedriger Schwellwert sollte im Allgemeinen weniger FN-Fehler ergeben (Recall optimiert), während ein hoher Schwellwert weniger FP-Erkennungen ergeben sollte (Precision optimiert).

Der F1-Score ist mathematisch gesehen das harmonische Mittel aus Precision und Recall und bietet somit eine konservative Einschätzung der Modellgenauigkeit unabhängig von der Betrachtungsrichtung. Das Modell mit der besten Leistung anhand des maximalen F1-Scores ist also dasjenige das gleichzeitig so wenige FP-Fehler wie möglich sowie so wenige FN-Fehler wie möglich macht. Dieses so eingestellte Modell wurde verwendet, um das Inferencing über die gesamte Schweiz durchzuführen.





#### 4.7.2 Konfusionszählung: Nur Photovoltaik

In der Evaluation für das reine Photovoltaik-Modell befinden sich  $n = 310$  Testsamples. Das Verhalten des Modells bewegt sich in einem Normalbereich. Bei niedrigen Schwellwerten treten nicht übermäßig viele FP auf, bei hohen Schwellwerten ist der FN-Anteil immer noch vergleichsweise niedrig. Über alle Bereiche des Konfidenz-Schwellwerts ist die Anzahl richtiger Detektionen (TP) der bei weitem höchste Wert.

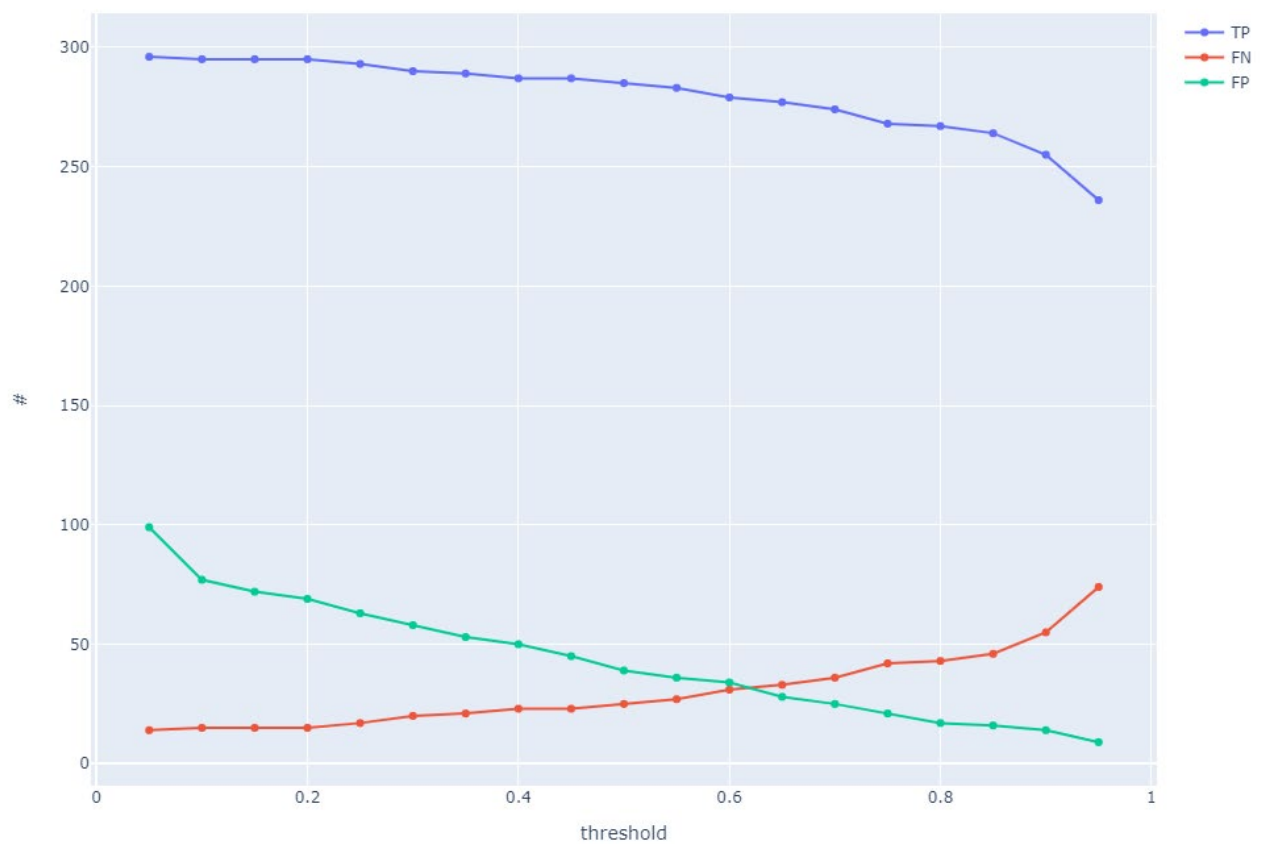


Abbildung 27: Konfusionszählung für das reine Photovoltaik-Modell



### 4.7.3 Metriken: Nur Photovoltaik

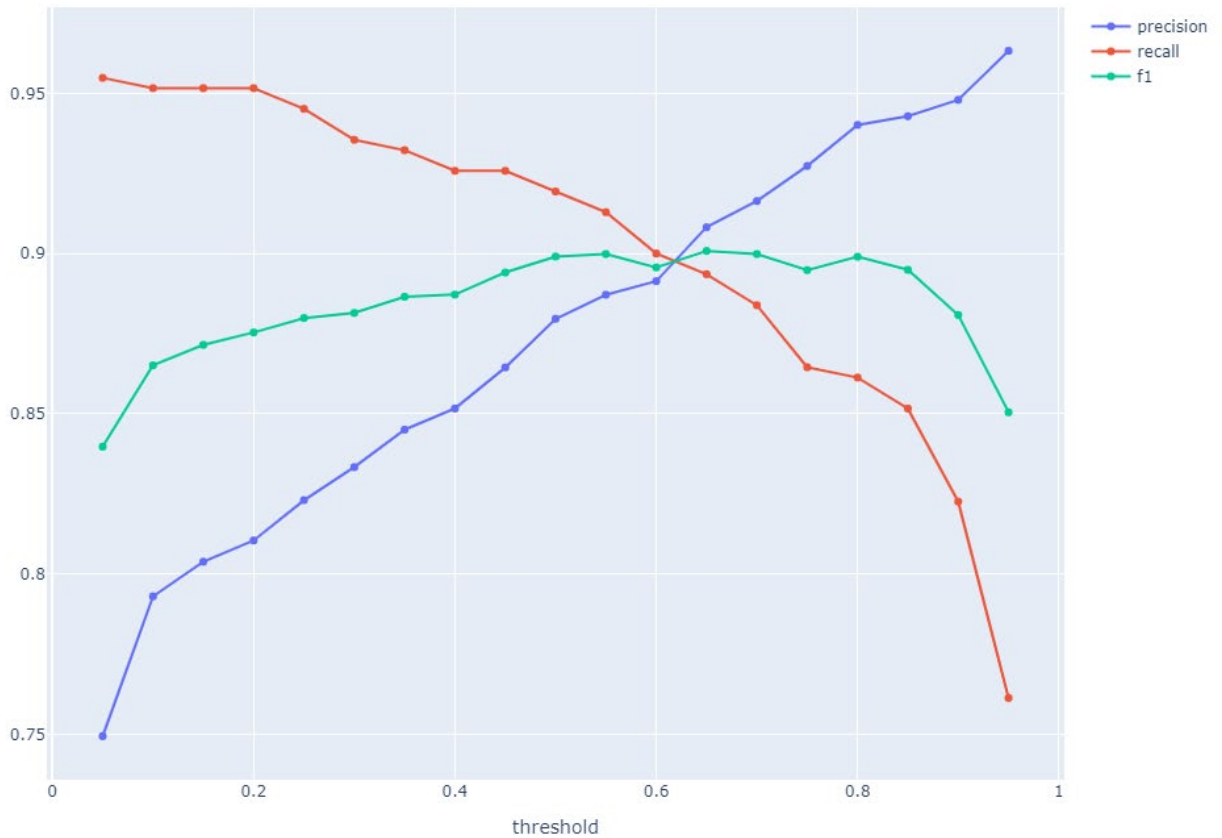


Abbildung 28: Metriken für das reine Photovoltaik-Modell

Das Modell zeigt sich in der Metrik-Evaluation sehr robust. Bei niedrigen Schwellwerten wird eine Trefferquote (Recall) von über 95 % erreicht, bei hohen Schwellwerten sinkt der Recall, dafür steigt die Präzision auf einen Wert von 96 % an. Die maximale Gesamtgenauigkeit erreicht das Modell bei einem Schwellwert von 65 % mit einem F1-Score von 90 %.



#### 4.7.4 Konfusionszählung: Nur Solarthermie

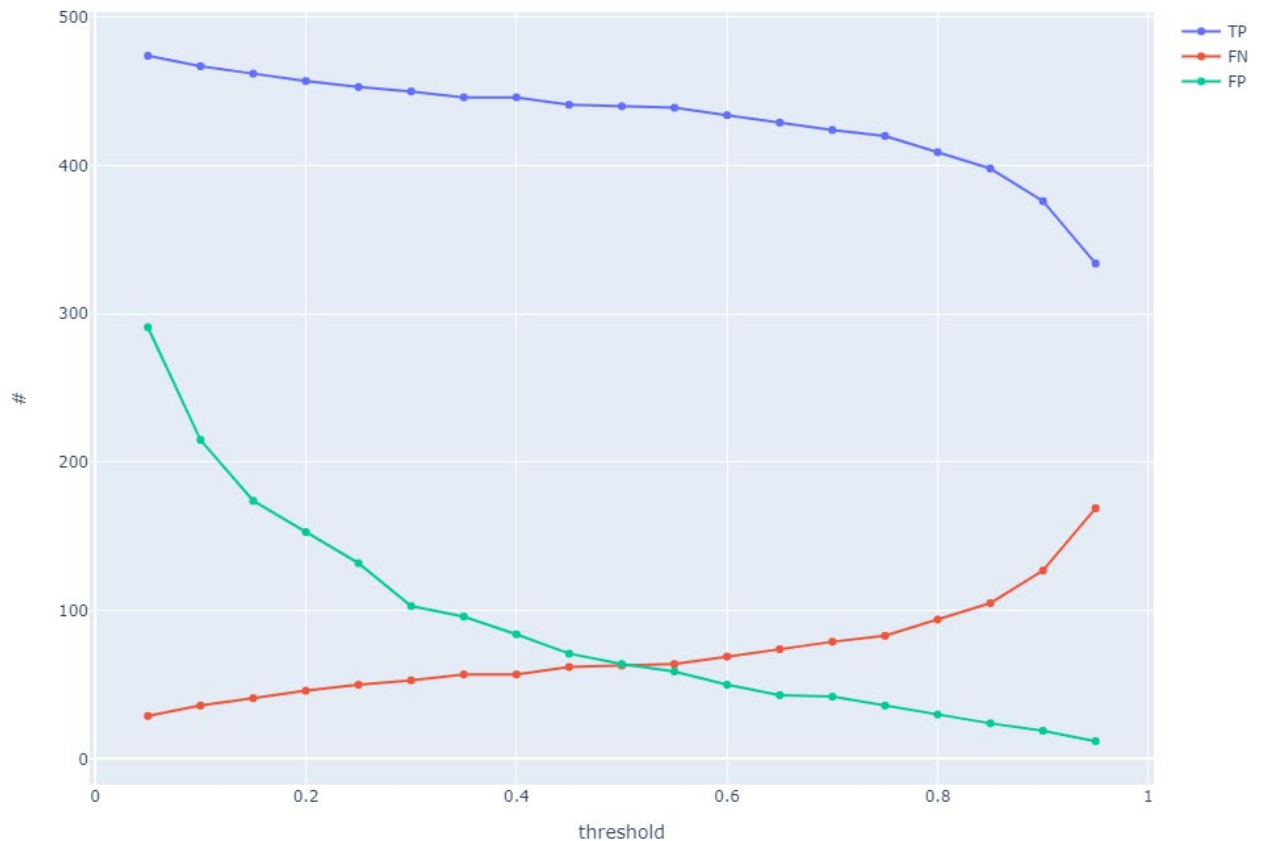


Abbildung 29: Konfusionszählung für das reine Solarthermie-Modell

In der Evaluation für das reine Solarthermie-Modell befinden sich  $n = 503$  Testsamples. Das Verhalten des Modells bewegt sich im Normalbereich ist jedoch nicht als so präzise einzustufen wie das reine PV-Modell. Am unteren Ende des Schwellwertbereichs werden fast 300 FP detektiert. Häufig sind dies Dachfenster, Pergolas, Swimming Pools oder gelegentlich auch ein Fahrzeug. Die wesentlich kleinere Grösse von Solarthermie-Anlagen macht deren Detektion schwieriger. Das zeigt sich auch beim oberen Ende des Schwellwertbereichs, wo der Anteil FN stark ansteigt. Der Detektor «verpasst» hier also einige Solarthermie-Anlagen. Dennoch zeigt sich hier ein robustes Modell, das ebenfalls über den gesamten Schwellwertbereich mit der höchsten Anzahl richtige Detektionen (TP) hervorbringt.



#### 4.7.5 Metriken: Nur Solarthermie

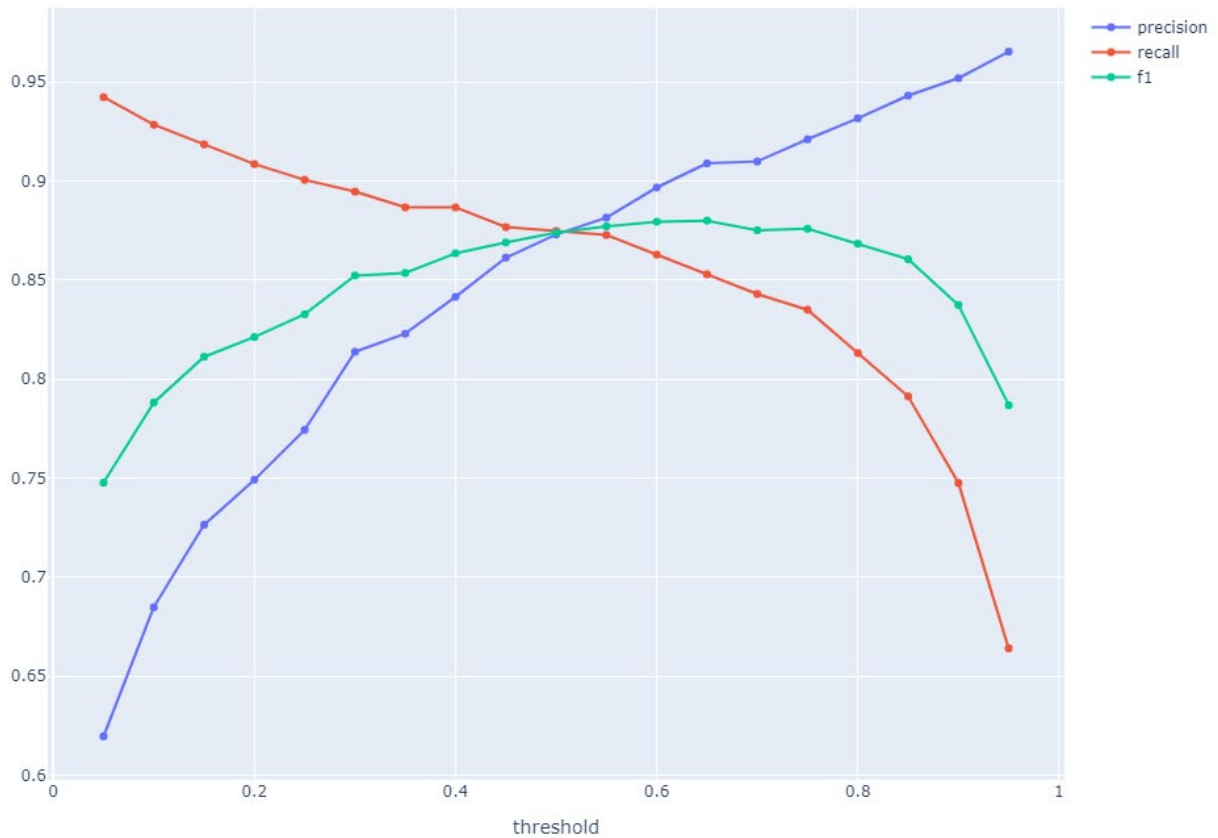


Abbildung 30: Metriken für das reine Solarthermie-Modell

Bei den Metriken zeigt sich das Thermal-Modell dennoch fast so robust wie das PV-Mobell. Bei niedrigen Schwellwerten erreicht der Recall erwartungsgemäss sehr hohe Werte, bei hohen Schwellwerten steigt die Präzision stark an. Die anwendbare Bandbreite ist bei dem Modell jedoch etwas geringer. In einem Schwellwertbereich von 30 % bis 85 % bewegt sich der F1-Score mit über 85 % Gesamtgenauigkeit daher immer noch in einem sehr starken Rahmen. Die maximale Gesamtgenauigkeit wird ebenfalls bei einem Schwellwert von 65 % erreicht, der max. F1 Score steigt hier auf 88 % an.



#### 4.7.6 Konfusionszählung: Kombiniertes Modell

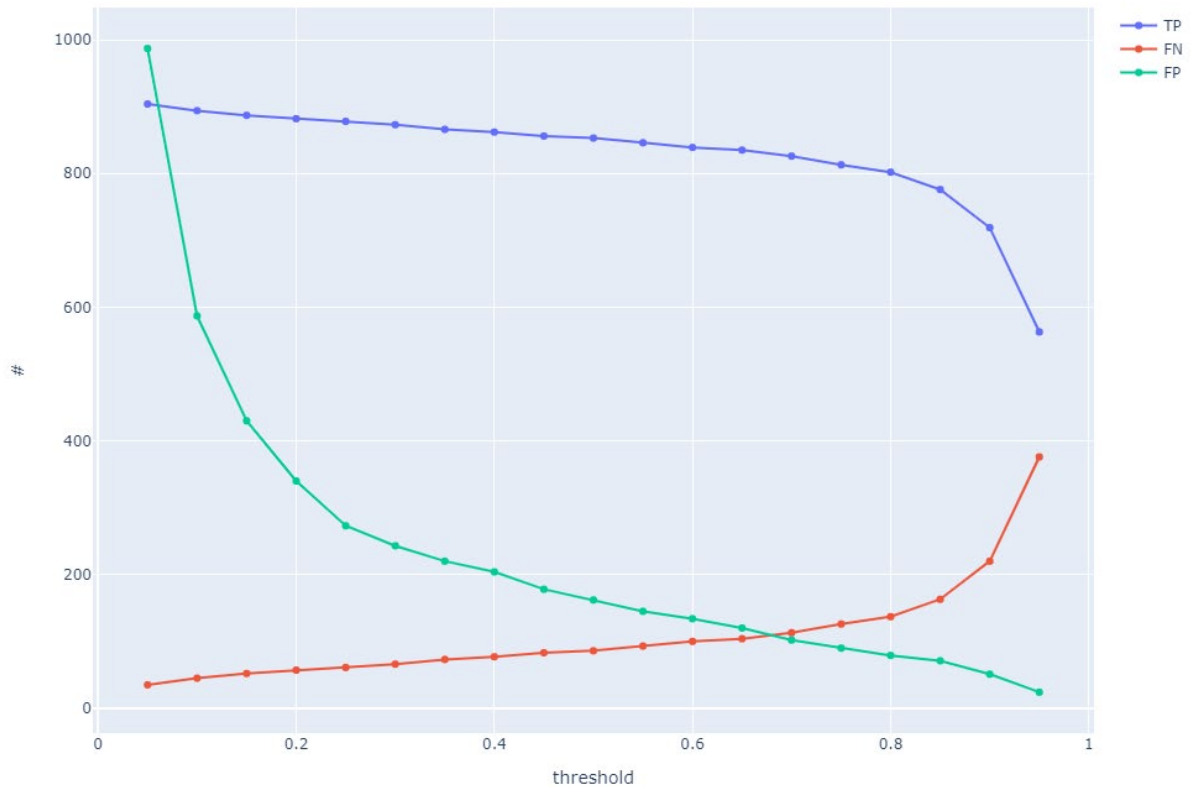


Abbildung 31: Konfusionszählung für das kombinierte Modell

In der Auswertung des kombinierten Modells beinhalten sich mit  $n = 939$  die grösste Anzahl Testsamples. Auch hier bewegt sich das Verhalten des Modells im Normalbereich. Bei sehr niedrigen Schwellwerten unter 20 % treten jedoch überproportional viele FP-Detektionen auf. Auch hier handelt es sich häufig um Dachfenster, Pergolas, Swimming Pools, Gewächshäuser, Glasdächer oder andere Strukturen. Auch am oberen Ende des Schwellwertbereichs über 90 % zeichnet sich eine stärkere Konzentration von «verpassten» Detektionen (FN) ab. Die Anzahl richtiger Detektionen macht jedoch ab 10 % Schwellwert bereits den grössten Anteil aus. Im Zentralbereich spricht dies für ein Modell mit hoher Generalisationsfähigkeit.



#### 4.7.7 Metriken: Kombiniertes Modell

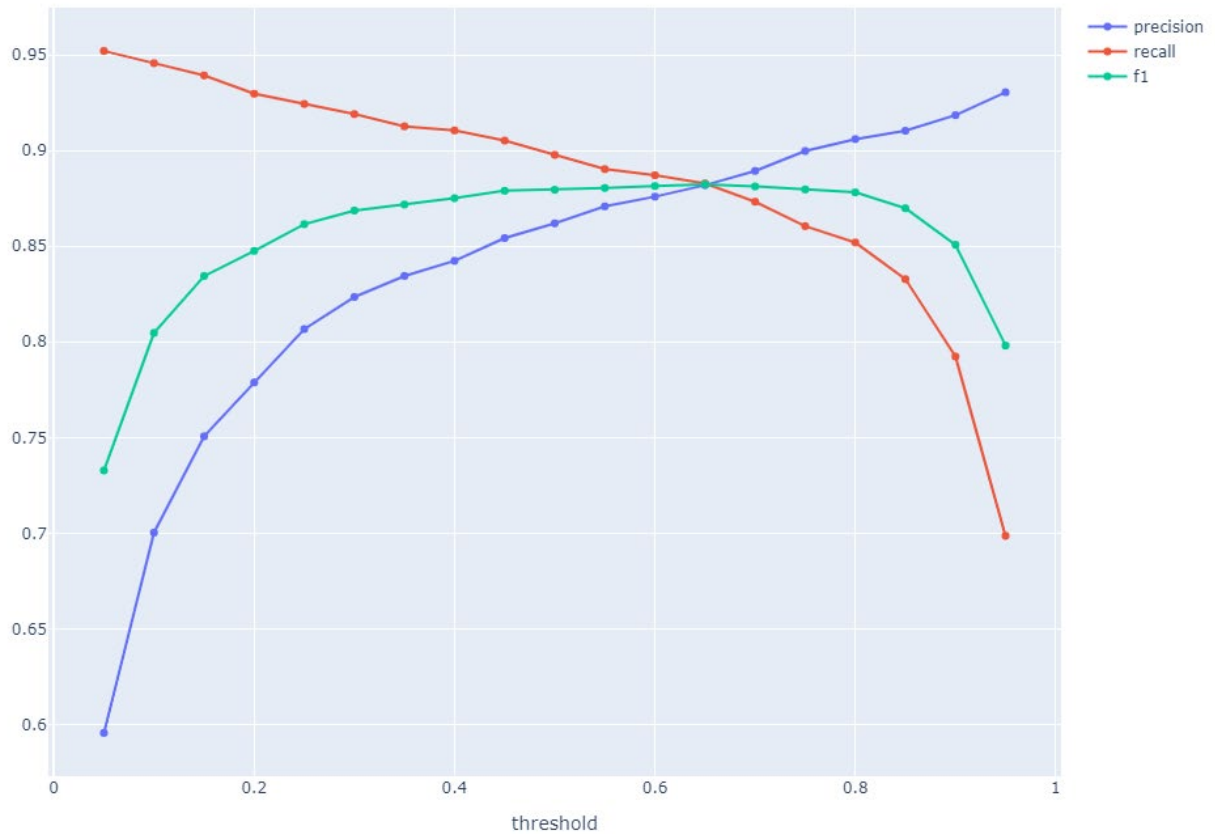


Abbildung 32: Metriken für das kombinierte Modell

In der Metrik-Auswertung zeigt sich das Modell sehr robust. Bei niedrigen Schwellwerten erreicht der Recall erwartungsgemäss sehr hohe Werte, bei hohen Schwellwerten steigt die Präzision stark an. Die anwendbare Bandbreite des Modells ist etwas geringer als beim reinen PV-Modell, aber deutlich besser als beim reinen Thermal-Modell: Zwischen 20 % und 90 % Konfidenzschwellwert liegt der F1-Score über 85 %. Das Maximum wird auch hier bei einem Schwellwert von 65 % erreicht und liegt bei 88 %.



#### Vergleich und Limitationen der Modelle

Sol-AI Detektoren	F1-Score Gesamtgenauigkeit (Konfidenzschwellwert THR $\geq$ 0.65)	Precision Nutzergenauigkeit (Konfidenzschwellwert THR $\geq$ 0.95)	Recall Produzentengenauigkeit/ Trefferquote (Konfidenzschwellwert THR $\geq$ 0.05)
Nur Photovoltaik (PV) N = 310 Testsamples	90.1 %	96.3 %	95.5 %
Nur Solarthermie (THM) N = 503 Testsamples	88.0 %	96.5 %	94.2 %
Kombiniertes Modell (PV + THM) N = 939 Testsamples	88.2 %	93.0 %	95.2 %

Tabelle 1: Metriken der Modelle (Maximalscores)

Die Detektion von Solarthermie stellt nach wie vor ein schwierigeres Unterfangen als die Detektion Photovoltaik dar. Dies begründet sich in der geringeren Grösse, sowie einer höheren Heterogenität der Thermalsysteme. Häufig ist es auch für menschliche Experten schwierig selbst auf den höchst aufgelösten Luftbildern der Swisstopo eine eindeutige Klassifikation vorzunehmen. Das Verwechslungspotential ist daher bei Solarthermie wesentlich höher als bei Photovoltaik.

Eine gewisse Limitation der Modelle besteht in der Anwendung auf den 2019 Swissimage Layer der Ostschweiz. Diese zusätzlichen 5 TB an Bilddaten standen erst in einer späteren Projektphase zur Verfügung, wodurch auf diesen Daten keine Labels erzeugt werden konnten. Da sich die 2019er Luftbilddaten aber vom Aufnahmezeitpunkt und der Saison zum Teil stärker insbesondere von den im Trainingsdatensatz stark vertretenen 2018er Daten des Aargau unterscheiden (siehe 4.3.2), muss für diese Zonen möglicherweise von einer veränderten Modellperformanz ausgegangen werden.

## 4.8 Veröffentlichung und Open Source Zugang (WP 8)

Das Projekt ist auf einem privaten github Repository. Aus Datenschutzgründen ist es momentan noch nicht klar, ob die Ergebnisse auch veröffentlicht werden können.



## 5 Publikationen und Vorträge

### Workshop “Object Detection with Deep Learning” Geopython 2019

Ein funktionierender Prototyp des Solardetektors wurde an der Geopython Konferenz am 24.06.2019 in Muttenz einem interessierem internationalen Fachpublikum (ca. 70 Teilnehmern) vorgestellt (Abb. 31). Das auf Google Colab veröffentlichte Jupyter Notebook ermöglichte den Teilnehmern, selbstständig eine einfache Version eines Solar-Panel-Detektors aufzubauen und so den grundlegenden technischen Prozess inklusive einiger Paketinstallationen anhand von nur wenigen hundert Zeilen Code zu verstehen. Das interaktive Notebook kann weiterhin über [diesen Link](#) aufgerufen und ausgeführt werden.

### Codesprint an der Europython 2019

An der Europython (13.07.2019) wurden in einem Codesprint/Hackaton Funktionen für das Cloud Collaboration Tool von motivierten Workshopteilnehmern diskutiert und weiterentwickelt.



Abbildung 33: Links - Adrian Meyer, Daniel Rettenmund und Denis Jordan (v.l.n.r.) während dem Remote Sensing / Deep Learning Workshop an der Geopython 2019. Rechts – Codesprint an der Europython 2019.

### Vortrag an der PyCon.DE 2019

Das Projekt wurde an der PyCon.DE / PyData in Berlin (9. bis 13. Oktober 2019) vorgestellt. Martin Christen hielt den Vortrag «Detecting and Analyzing Solar Panels in Switzerland using Aerial Imagery» und erhielt viel Interessantes Feedback, insbesondere im Bereich Optimierung der Loss-Funktion für Solarpanels.



Abbildung 34: PyCon.DE ist die grösste Python Konferenz im Deutschsprachigem Raum und hatte im 2019 über 1100 Teilnehmende.





### **Vortrag an der PV Tagung 2020**

Das Projekt wurde an der PV Tagung im März 2020 vorgestellt. Aufgrund der Covid-19 Situation wurde der Vortrag aufgezeichnet. Folien können unter diesem Link heruntergeladen werden:

[https://www.swissolar.ch/fileadmin/user\\_upload/Tagungen/PV-Tagung\\_2020/PVT20\\_5.4 Martin\\_Christen.pdf](https://www.swissolar.ch/fileadmin/user_upload/Tagungen/PV-Tagung_2020/PVT20_5.4_Martin_Christen.pdf)

### **Vortrag an der Python Machine Learning Konferenz 2020**

Das Projekt wurde an der GeoPython 2021 vorgestellt. Die Konferenz fand vom 21.-22. September 2020 aufgrund der Situation mit Covid-19 online statt. Adrian Meyer hat das Projekt vorgestellt. Es gab zahlreiche Interaktionen mit dem Publikum. <http://2020.pythonml.org/schedule.html>

### **Vortrag an der EuroPython 2020**

Adrian Meyer hielt am 23. Juli 2020 den Vortrag «Detecting and Analyzing Solar Panels in Switzerland using Aerial Imagery » an der international hochkarätig besetzten EuroPython 2020 Konferenz. Über 100 Teilnehmer hörten den online durchgeführten Vortrag. Link

### **Journal Publikation**

Eine Journal Publikation ist für 2021/22 geplant und wird im 2021 eingereicht.



## 6 Referenzen

- D. **Ciregan**, U. Meier, J. Schmidhuber. 2012. Multi-column deep neural networks for image classification. Presented at the Computer vision and pattern recognition (CVPR), 2012 IEEE conference on, IEEE, pp. 3642–3649.
- J. **Deng**, W. Dong, R. Socher, L.J. Li, K. Li, L. Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. Presented at the Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE, pp. 248–255.
- U. **Farooq**. 2018. From R-CNN to Mask R-CNN. [https://medium.com/@umerfarooq\\_26378/from-r-cnn-to-mask-r-cnn-d6367b196cfd](https://medium.com/@umerfarooq_26378/from-r-cnn-to-mask-r-cnn-d6367b196cfd) [05.09.2019].
- B. **Feng**, X. Shen, J. Long, H. Chen, "A Novel Crack Detection Algorithm for Solar Panel Surface Images," 2013 International Conference on Computer Sciences and Applications, Wuhan, 2013, pp. 650-654. doi: 10.1109/CSA.2013.158
- K. **Fukushima**, S. Miyake. 1982. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition, in: Competition and Cooperation in Neural Nets. Springer, pp. 267–285.
- A. **Graves**, J. Schmidhuber, 2009. Offline handwriting recognition with multidimensional recurrent neural networks. Presented at the Advances in neural information processing systems, pp. 545–552.
- D. H. **Hubel**, T. N. Wiesel. 1968. Receptive fields and functional architecture of monkey striate cortex. J Physiol 195, 215–243.
- S. **Iyengar**, N. Sharma, D. Irwin, P. Shenoy, K. Ramamritham. 2017. A Cloud-Based Black-Box Solar Predictor for Smart Homes. ACM Trans. Cyber-Phys. Syst. 1, 4, Article 21 (August 2017), 24 pages. DOI: <https://doi.org/10.1145/3004056>
- D. **Jordan**, N. Lack, S. Hochuli, A. F. Meyer, M. Schär. 2019. Automatisierte Klassifizierung der Landnutzung: Deep Learning basierter Ansatz für die Arealstatistik der Schweiz. Geomatik 9/2019, pp. 260-264.
- A. **Krizhevsky**, I. Sutskever, G. E. Hinton. 2012. Imagenet classification with deep convolutional neural networks. Presented at the Advances in neural information processing systems, pp. 1097–1105.
- S. **Lee**, K. E. An, B. D. Jeon, K. Y. Cho, S. J. Lee, D. Seo, "Detecting faulty solar panels based on thermal image processing," 2018 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, 2018, pp. 1-2. doi: 10.1109/ICCE.2018.8326228
- J. **Long**, E. Shelhamer, T. Darrell. 2015. Fully convolutional networks for semantic segmentation. Presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431–3440.
- J. **Josifovski**, Object Recognition: SIFT vs Convolutional Neural Networks. 2015. [https://tams.informatik.uni-hamburg.de/lehre/2015ws/seminar/ir/pdf/slides/JosipJosifovski-Object\\_Recognition\\_SIFT\\_vs\\_Convolutional\\_Neural\\_Networks.pdf](https://tams.informatik.uni-hamburg.de/lehre/2015ws/seminar/ir/pdf/slides/JosipJosifovski-Object_Recognition_SIFT_vs_Convolutional_Neural_Networks.pdf)
- Y. **LeCun**, Y. Bengio, G. Hinton. 2015. Deep learning. Nature 521, 436.
- D. G. **Lowe**: Object Recognition from Local Scale-Invariant Features. 1999. ICCV '99 Proceedings of the International Conference on Computer Vision. Volume 2, pp 1150–1157



- P. **Marcelino**. 2018. Transfer learning from pre-trained models. <https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751> [05.09.2019].
- W. S. **McCulloch**, W. Pitts. 1943. A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics 5, 115–133.
- A. F. **Meyer**, N. Lack, D. Jordan. 2019. Wildtier-Monitoring mit UAVs - Künstliche Intelligenz zur automatisierten Detektion von Infrarot-Signaturen, 39. Wissenschaftlich-Technische Jahrestagung der DGPF – Dreiländertagung der OVG, DGPF und SGPF in Wien – Publikationen der DGPF, Band 28, 2019
- A. **Ouaknine**. 2018. Review of Deep Learning Algorithms for Object Detection. <https://medium.com/zylapp/review-of-deep-learning-algorithms-for-object-detection-c1f3d437b852> [05.09.2019]
- F. **Rosenblatt**. 1958. The perceptron: a probabilistic model for information storage and organization in the brain. Psychological review 65, 386.
- M. **Schär**. 2018. Einsatz von Deep Learning zur Aktualisierung der Arealstatistik der Schweiz - erste Untersuchungen, 38. Wissenschaftlich-Technische Jahrestagung der DGPF und PFGK18 Tagung in München – Publikationen der DGPF, Band 27, 2018
- M. S. **Shahriar**, M. S. Rahman. 2015. Urban Sensing and Smart Home Energy Optimisations: A Machine Learning Approach. In Proceedings of the 2015 International Workshop on Internet of Things towards Applications (IoT-App '15). ACM, New York, NY, USA, 19-22. DOI=<http://dx.doi.org/10.1145/2820975.2820979>
- K. **Simonyan**, A. Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
- J. **Stallkamp**, M. Schlipsing, J. Salmen, C. Igel. 2011. The German Traffic Sign Recognition Benchmark: A multi-class classification competition, in: The 2011 International Joint Conference on Neural Networks. Presented at the The 2011 International Joint Conference on Neural Networks.
- V. **Sundaresan**, J. Lin. 1998. Recognizing Handwritten Digits and Characters.
- J. **Tang**, Q. Li. 2010. Fast template matching algorithm: Fast template matching algorithm. Journal of Computer Applications 30(6):1559-1561. DOI: 10.3724/SP.J.1087.2010.01559
- S. A. K. **Tareen**, Z. Saleem, "A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK," 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, 2018, pp. 1-10.
- M. **Waldrop**. 2019. What are the limits of deep learning? Proceedings of the National Academy of Sciences of the United States of America. 116 (4) pp. 1074-1077. doi: 10.1073/pnas.1821594116.
- A. **Waleed**. 2017. Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow. [https://github.com/matterport/Mask\\_RCNN](https://github.com/matterport/Mask_RCNN) [05.09.2019].
- J. **Yu**, Z. Wang, A. Majumdar, R. Rajagopal. 2018. DeepSolar: A Machine Learning Framework to Efficiently Construct a Solar Deployment Database in the United States. Joule Vol 2 Issue 12, P2605-2617. <https://doi.org/10.1016/j.joule.2018.11.021>