

Samplingstrategien zur Trainingsdatenerzeugung für tiefe Segmentierungsmodelle

Bachelorarbeit von Joël Bender

Leitung: Prof. Dr. Lorenz Hurni

Betreuer: Dr. Magnus Heitzler

27.05.2020

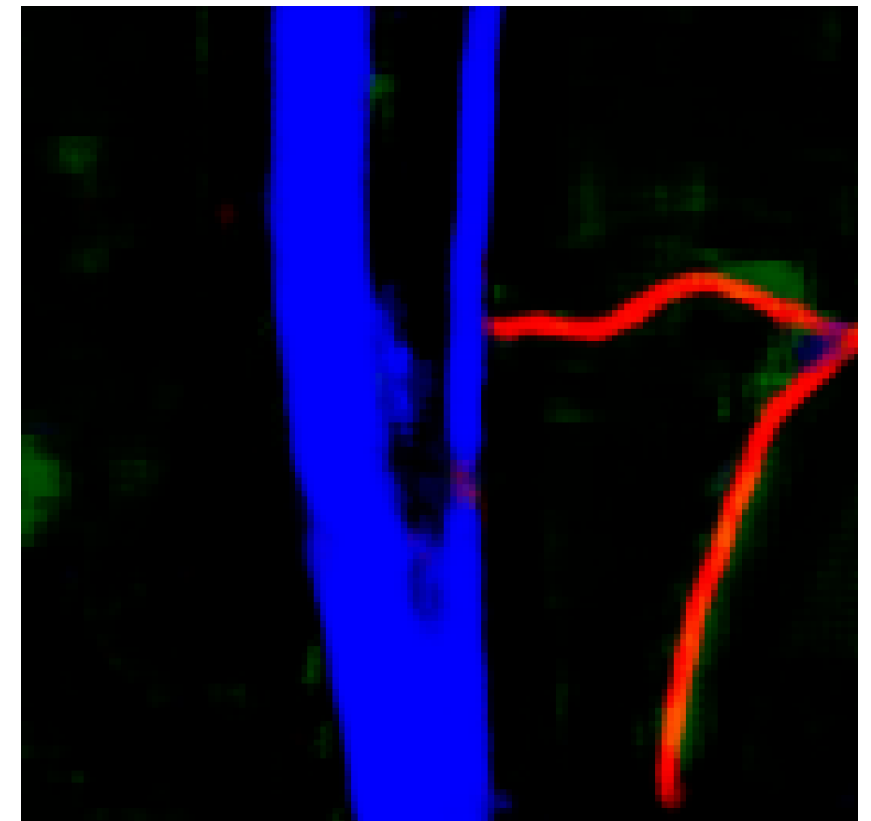
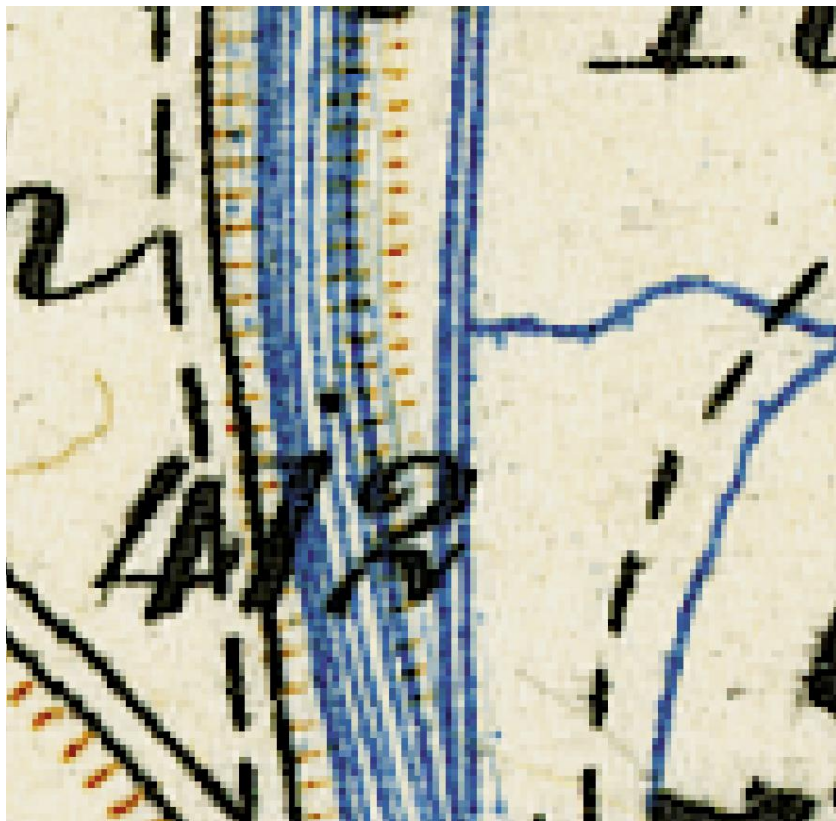
Inhaltsverzeichnis

1. Problemstellung
2. Grundlagen
3. Methode und Vorgehen
4. Ergebnisse
5. Diskussion und Ausblick

Problemstellung

Problemstellung

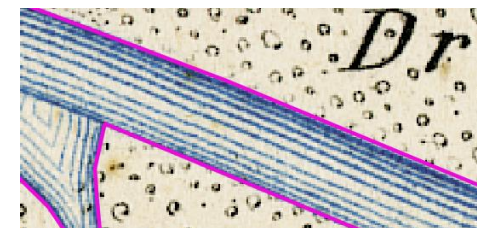
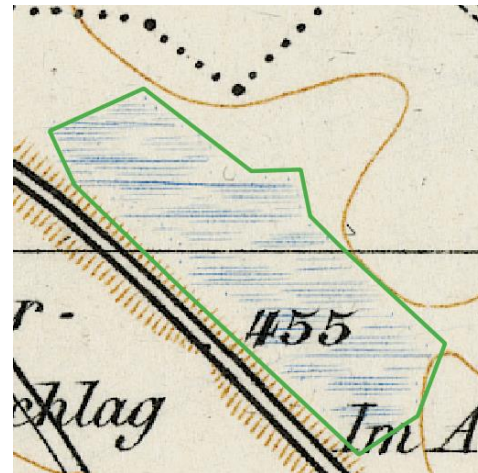
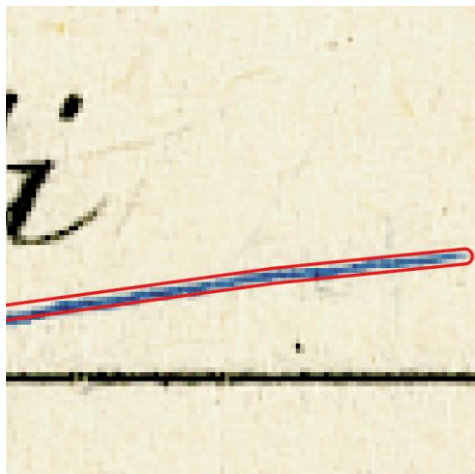
- Extraktion von hydrologischen Featureklassen aus der Siegfriedkarte → neuronales Netzwerk
- Workflows generieren, um Trainingsdaten zu sampeln



Grundlagen

Grundlagedaten

- Kartenblätter der Siegfriedkarte für die Jahre 1879 (10 Blätter) und 1880 (17 Blätter)
- Groundtruthdaten: Vektorisierter Datensatz über die Featureklassen «Stream», «Wetland» und «Riverlake» für die Jahre 1879 und 1880



Methode und Vorgehen

Methode

- Rasterdaten digitalisieren
- Undersampling: Aus Kartenausdehnung werden nur einzelne Stücke verwendet

- Grund:

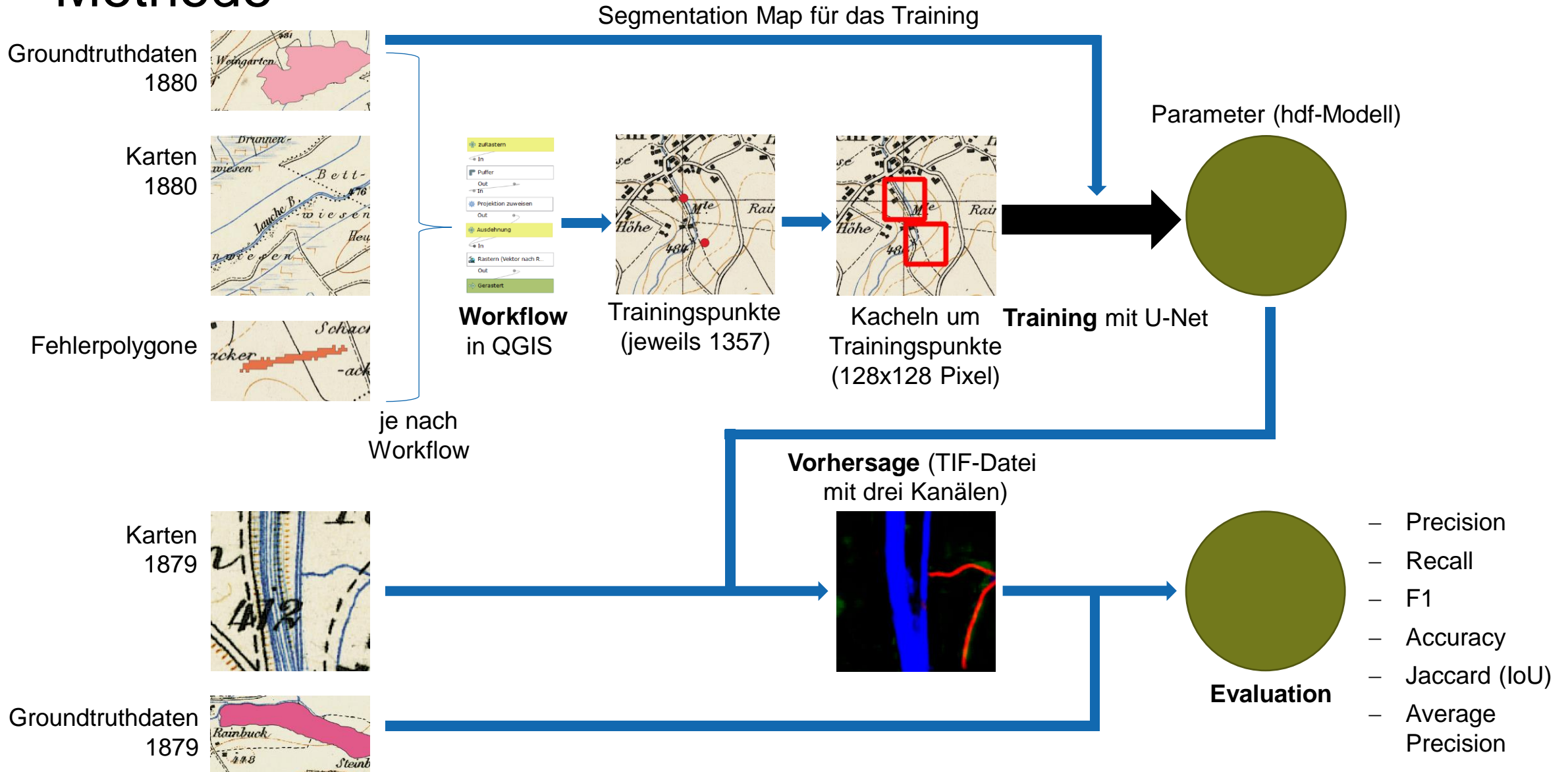
- Klassenungleichgewicht (Johnson & Khoshgoftaar, 2019):

Klasse	Stream	Wetland	Riverlake	Nichts davon	Gesamt
Fläche [km ²]	4.607	6.108	9.607	872.272	892.5

- Trainingszeit

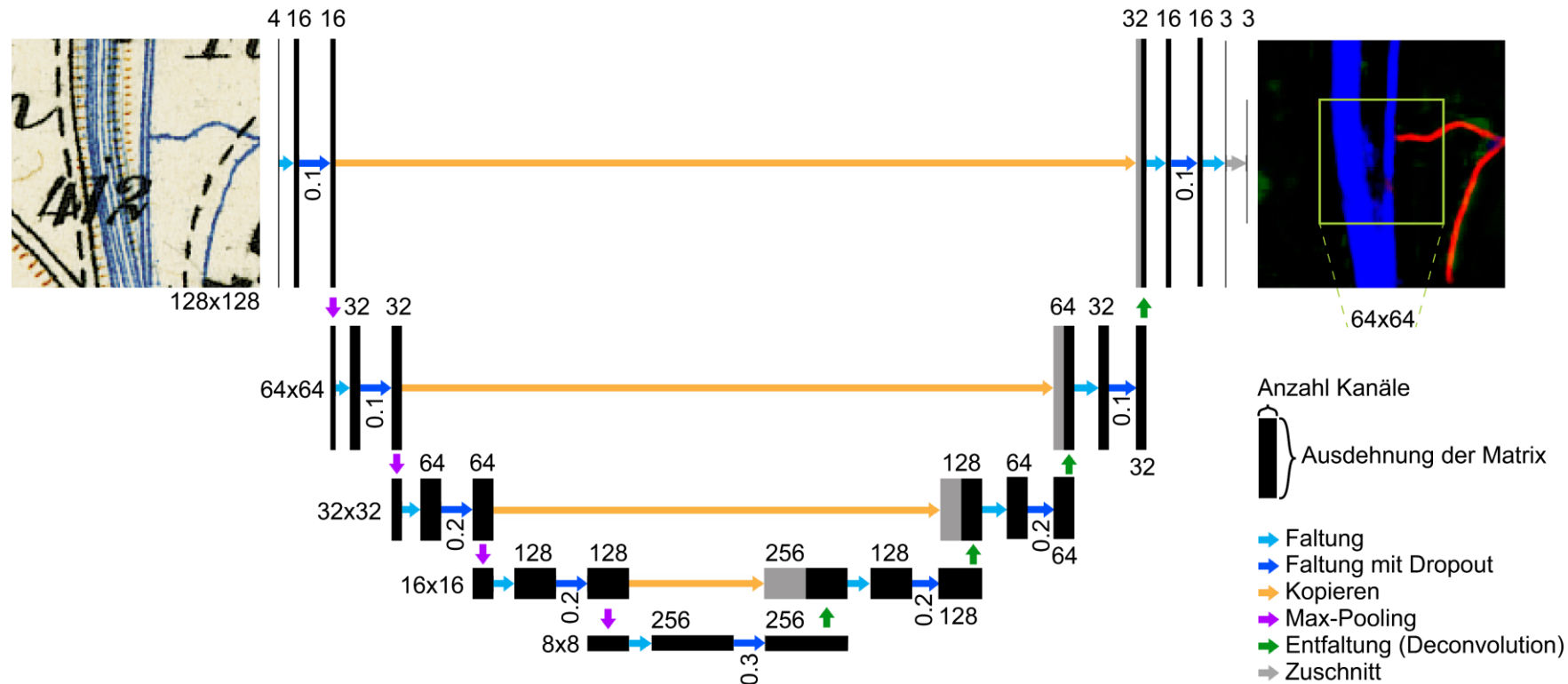
- **Ziel der Arbeit: Workflows für Undersampling erstellen und bewerten**

Methode



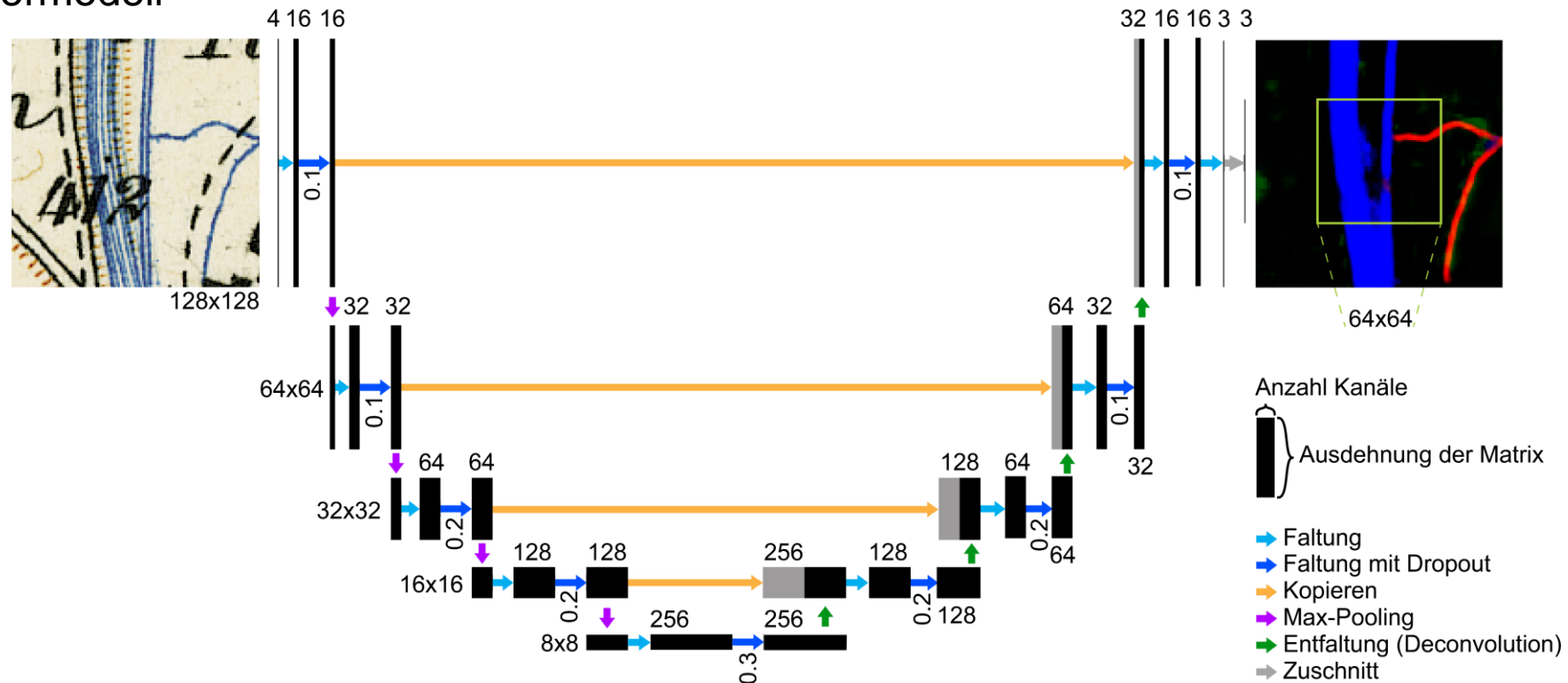
Training: U-Net-Framework

- Fully Convolutional Neural Network: Neuronales Netzwerk, welches vollständig auf Faltungen beruht
- 1 941 283 zu trainierende Parameter in den Faltungen



Training: U-Net-Framework

- Inputdaten der Grösse 128 x 128 Pixel rund um Trainingspunkte (160 x 160 m)
- Ergibt jeweils Segmentation Map von 64 x 64 Pixel
- Parametermodell

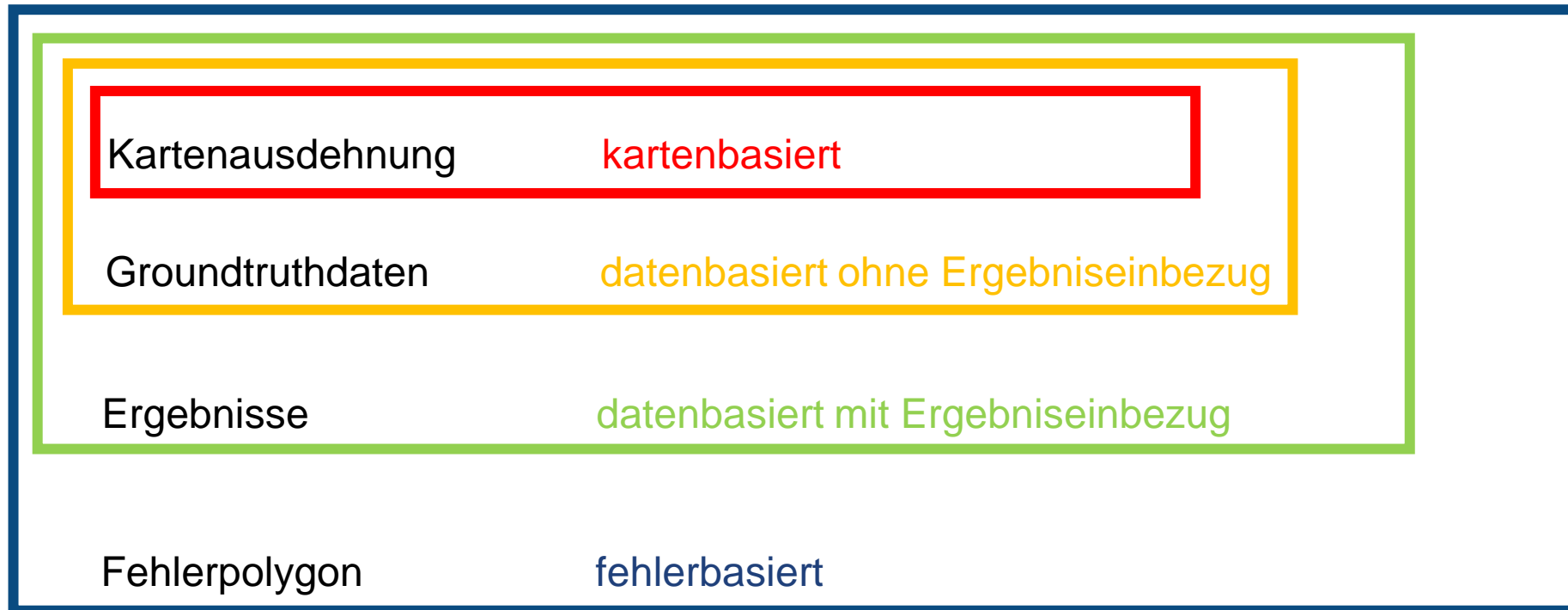


Ergebnisse

Ergebnisse

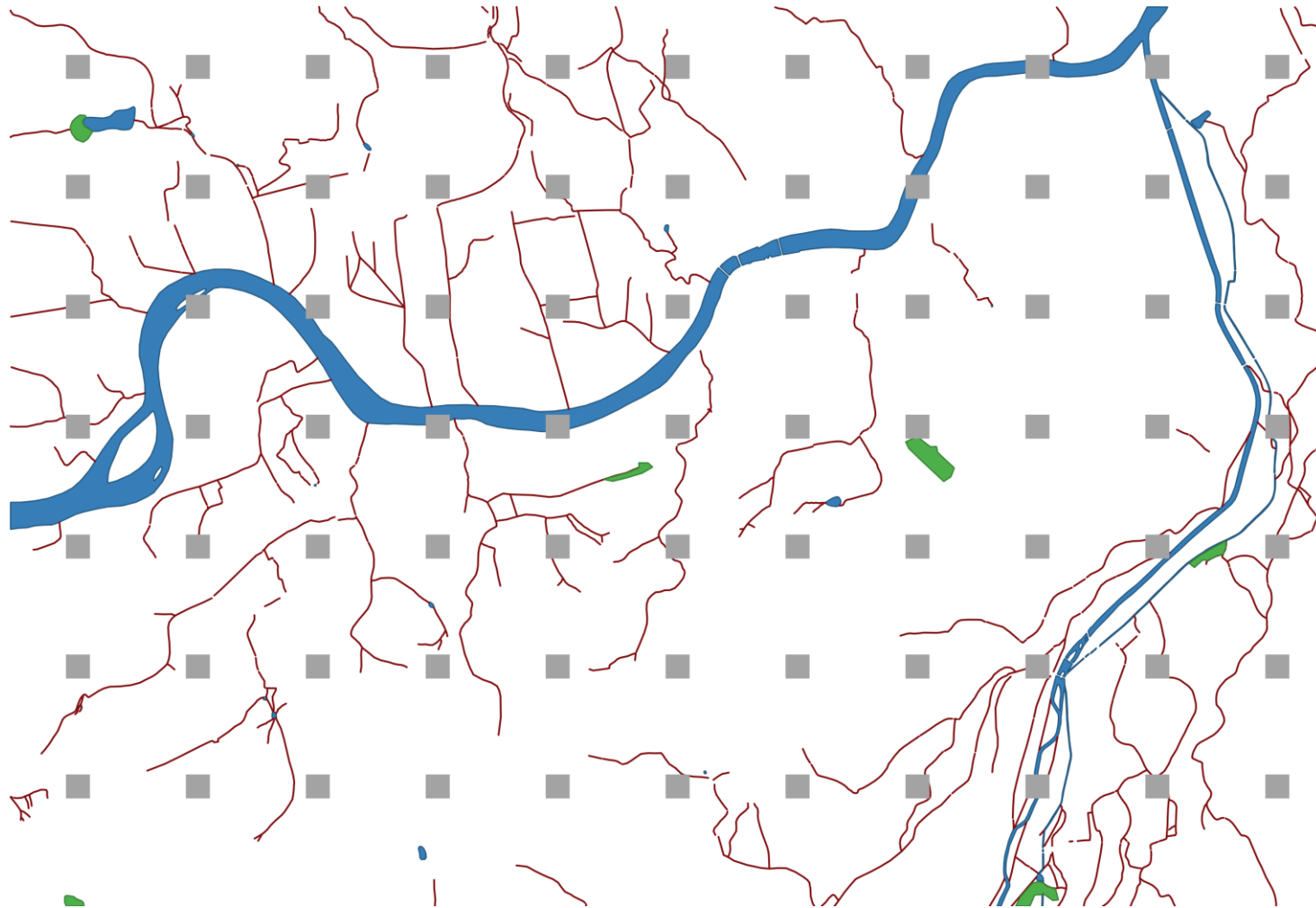
- Workflows
- Trainierte Modelle und daraus berechnete Vorhersagen
- Güteparameter für die Workflows
- Empfehlungsrahmen

Workflows zur Trainingsdatenerzeugung



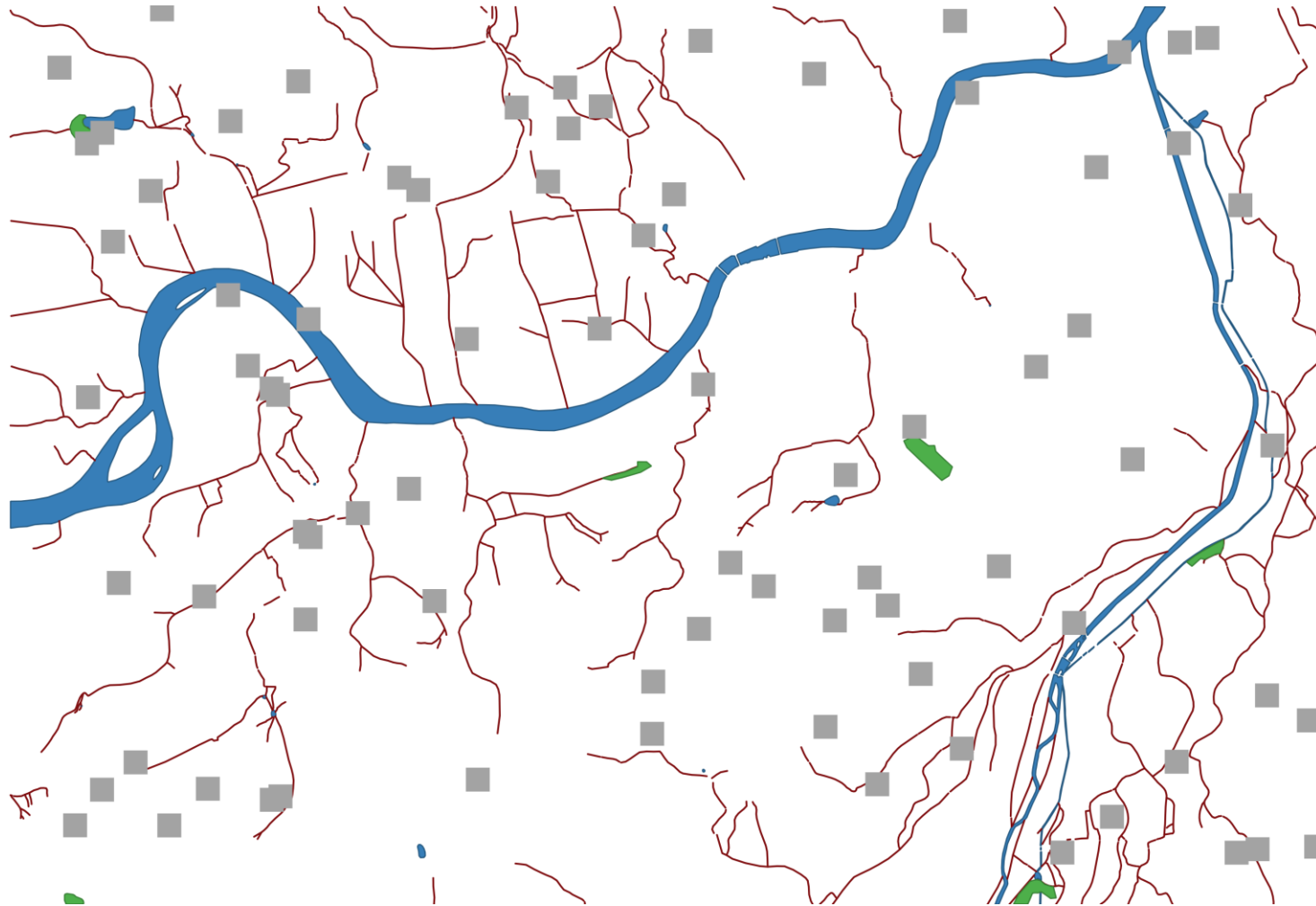
Zusätzlich: Mischung aus mehreren Instanzen

Workflows zur Trainingsdatenerzeugung (Beispiele)



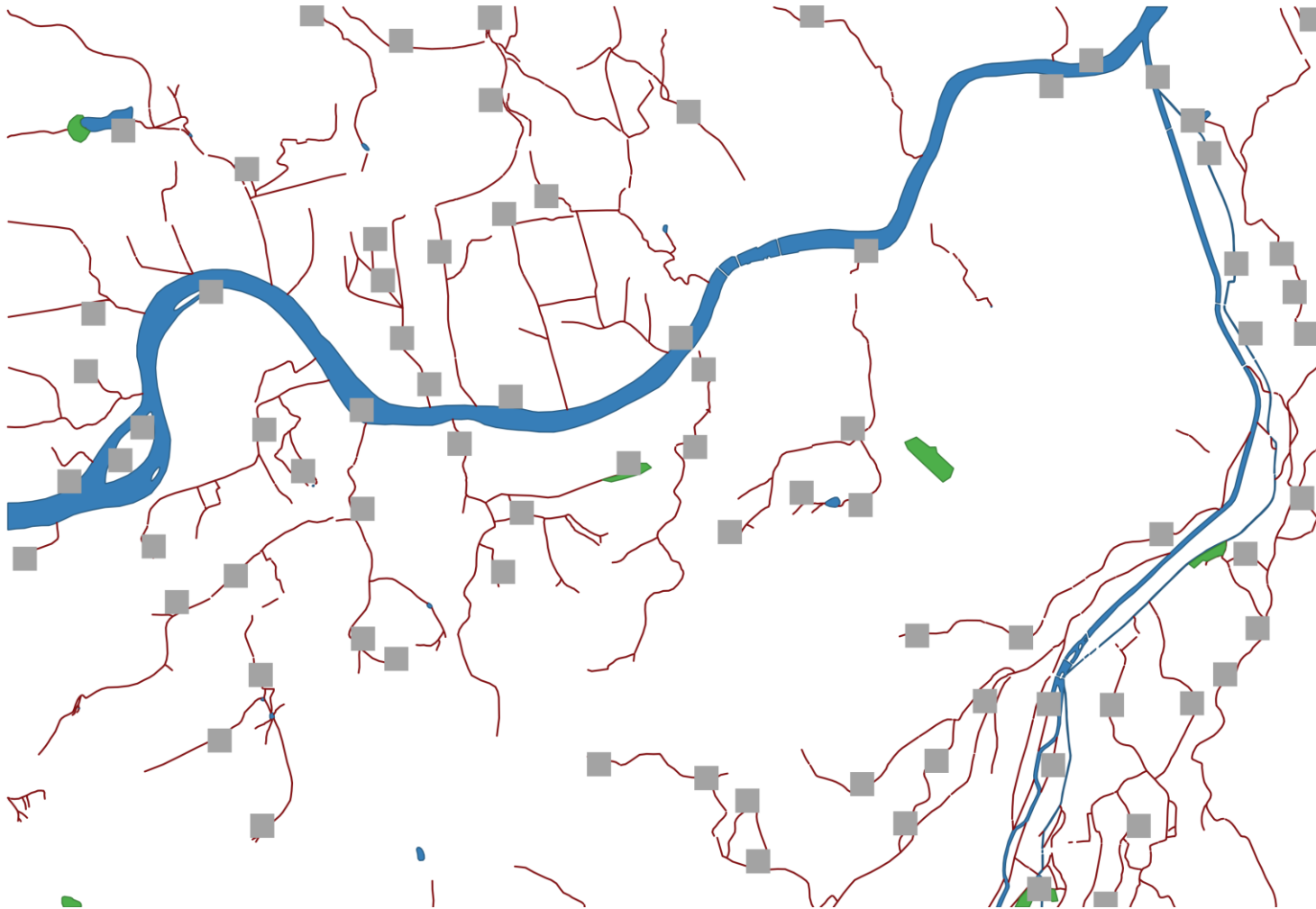
Kartenbasiert: Gitterpunkte

Workflows zur Trainingsdatenerzeugung (Beispiele)



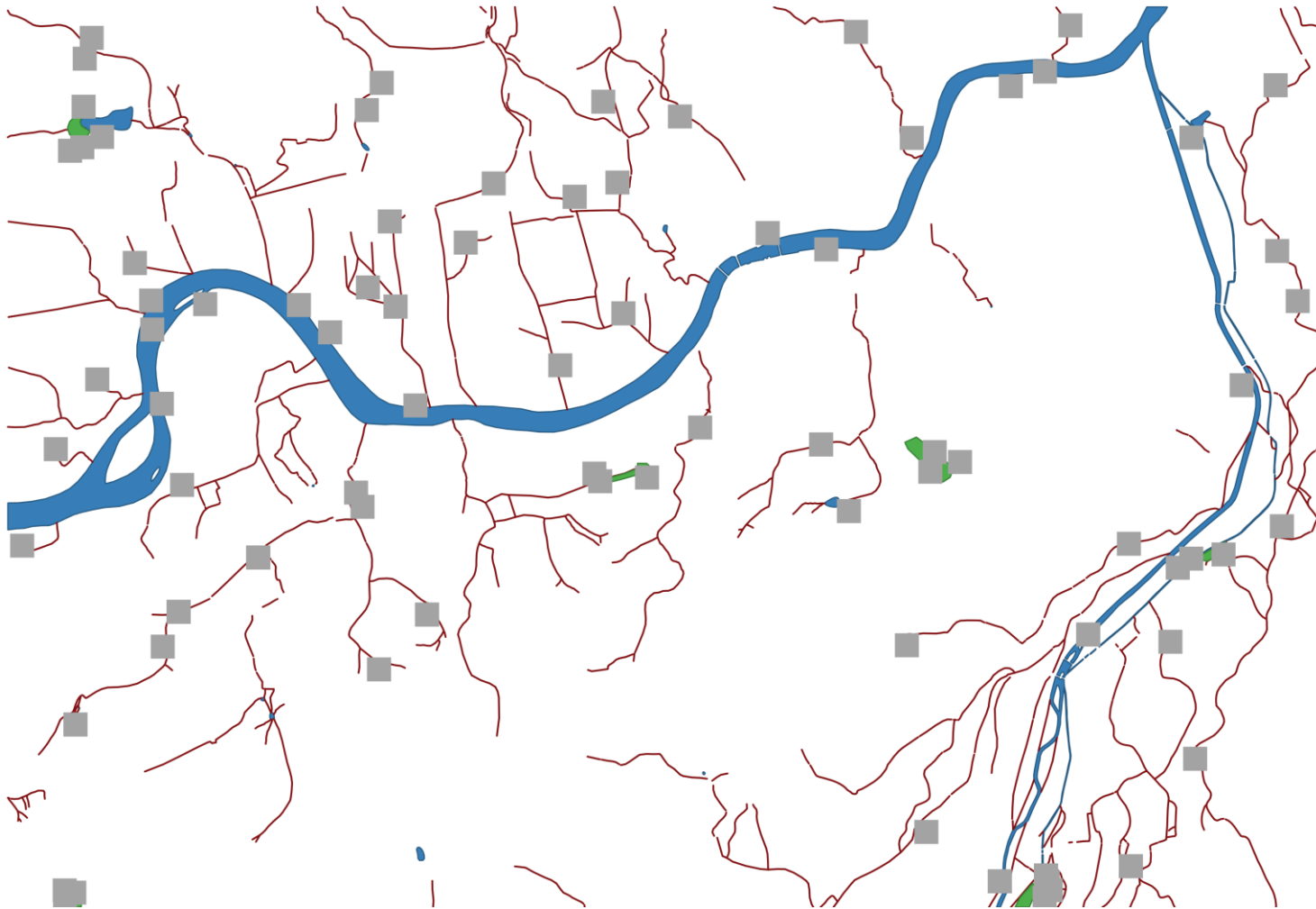
Kartenbasiert: Zufällige Punkte

Workflows zur Trainingsdatenerzeugung (Beispiele)



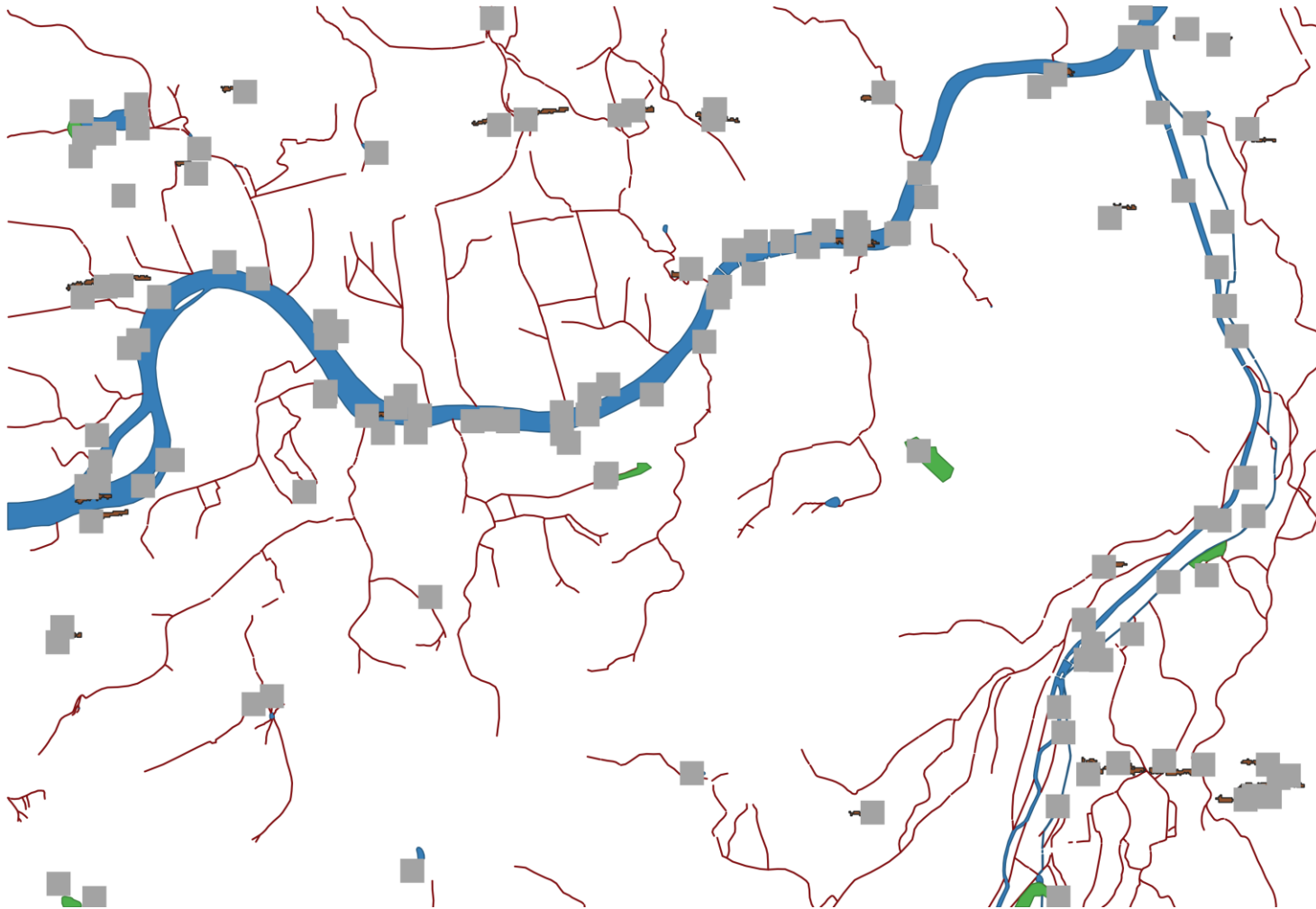
Datenbasiert (o.E.): Pufferpunkte
mit Mindestabstand

Workflows zur Trainingsdatenerzeugung (Beispiele)



Datenbasiert (m.E): Wetlandpunkte

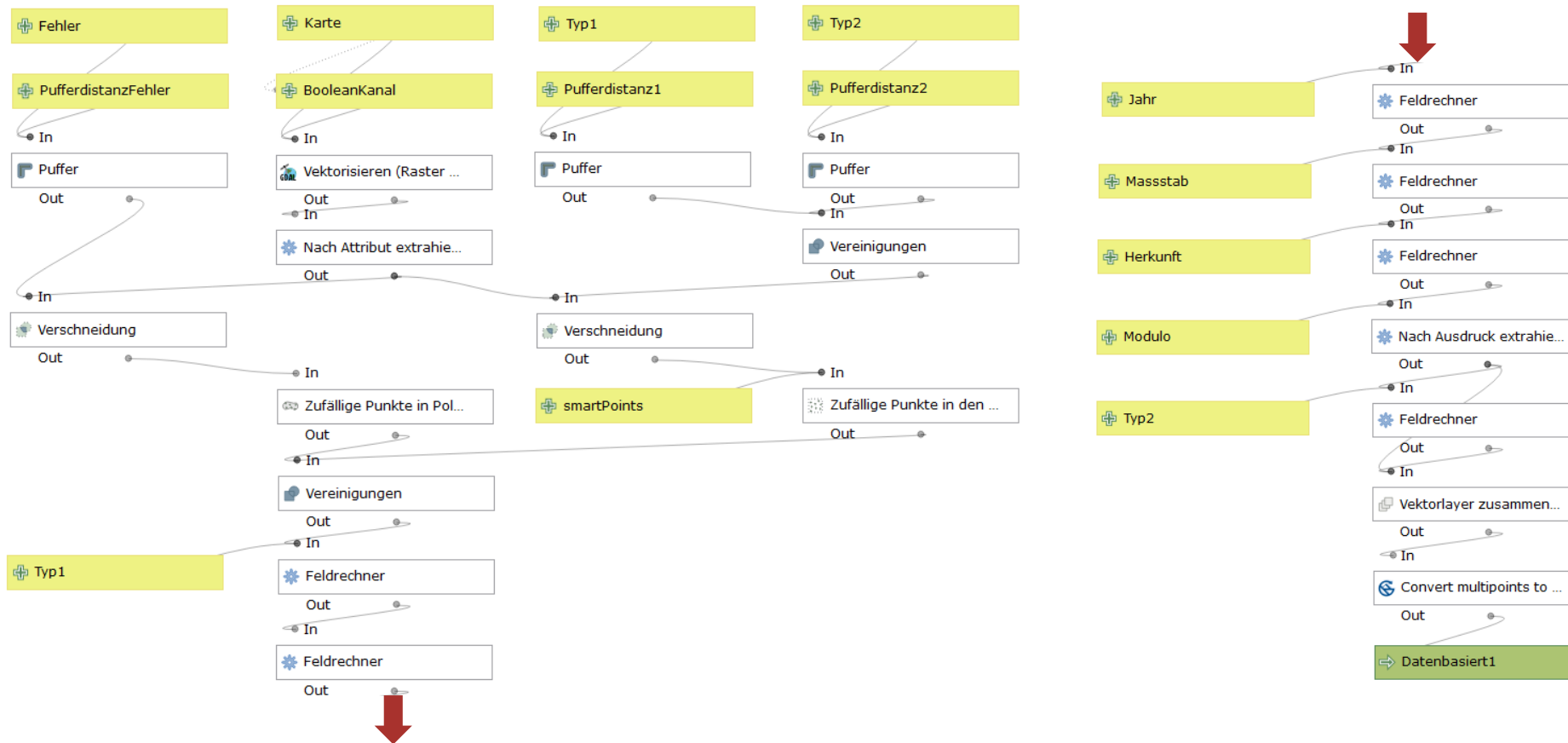
Workflows zur Trainingsdatenerzeugung (Beispiele)



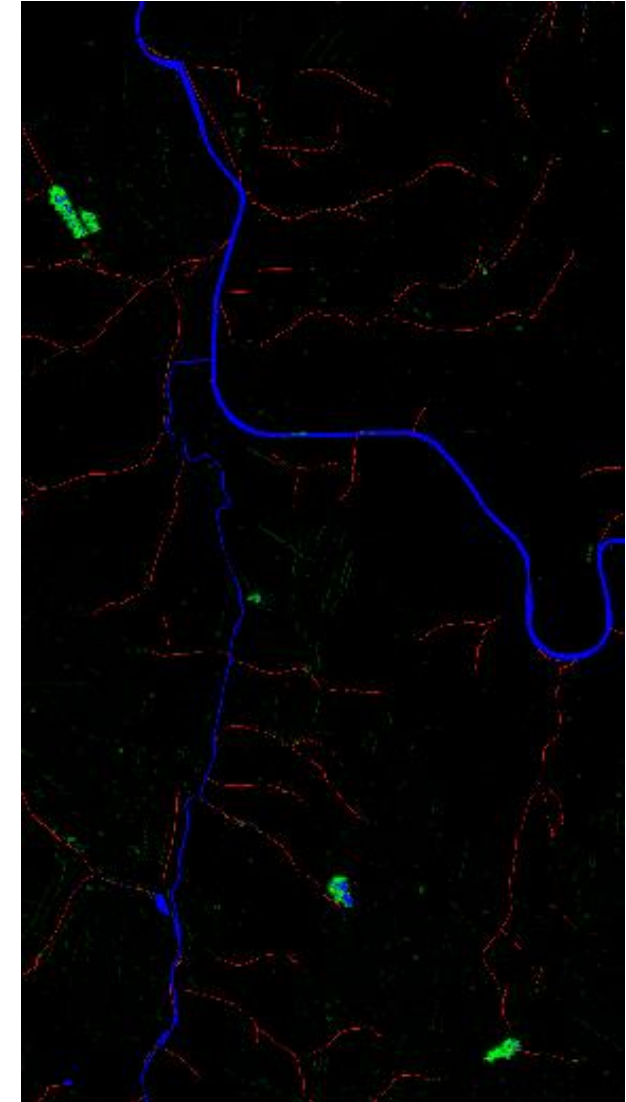
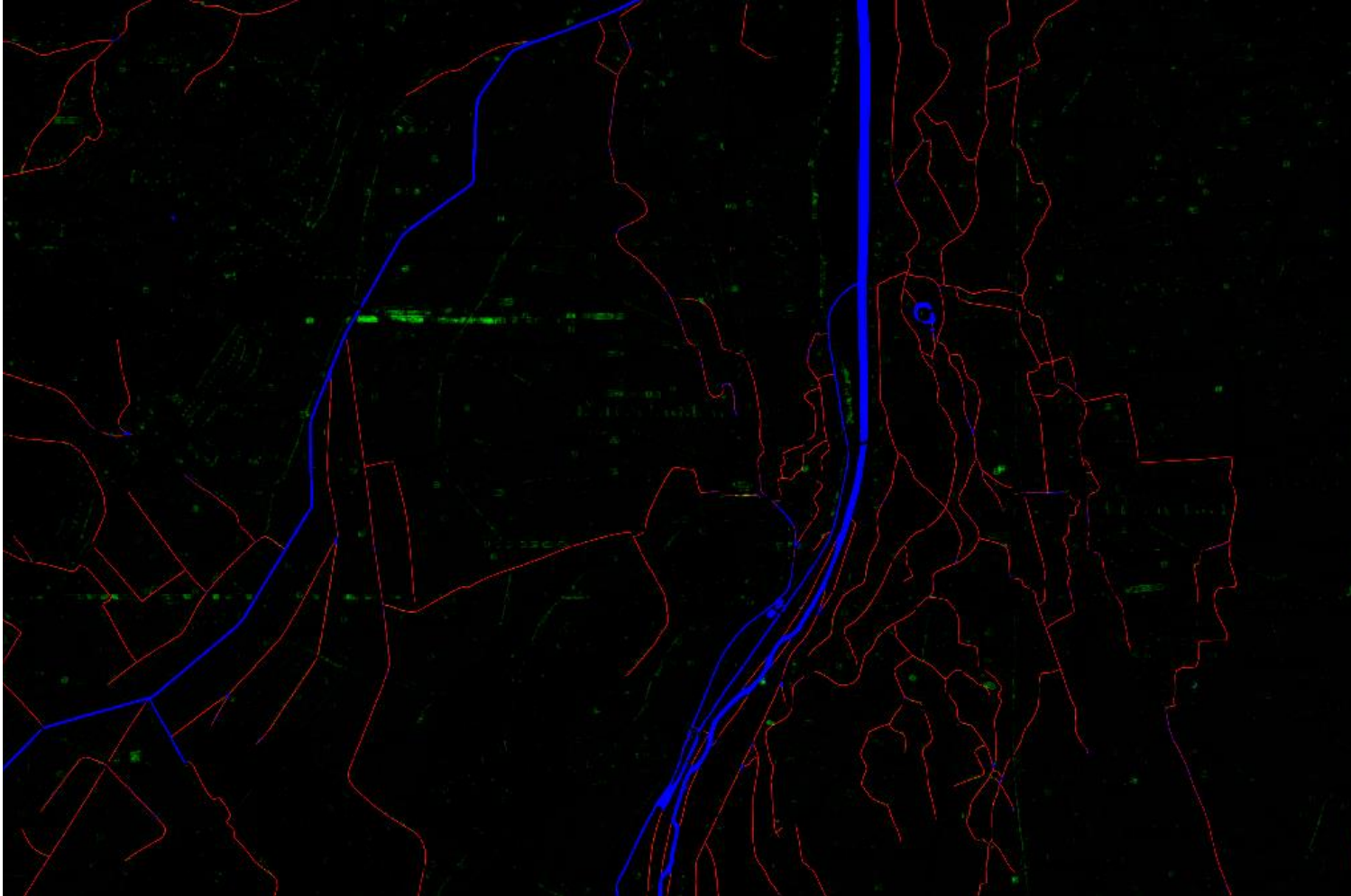
Fehlerbasiert: Fehlerbasierte Zwei-Feature Punkte

Workflows zur Trainingsdatenerzeugung

- Fehlerbasierte Zwei-Feature-Punkte:

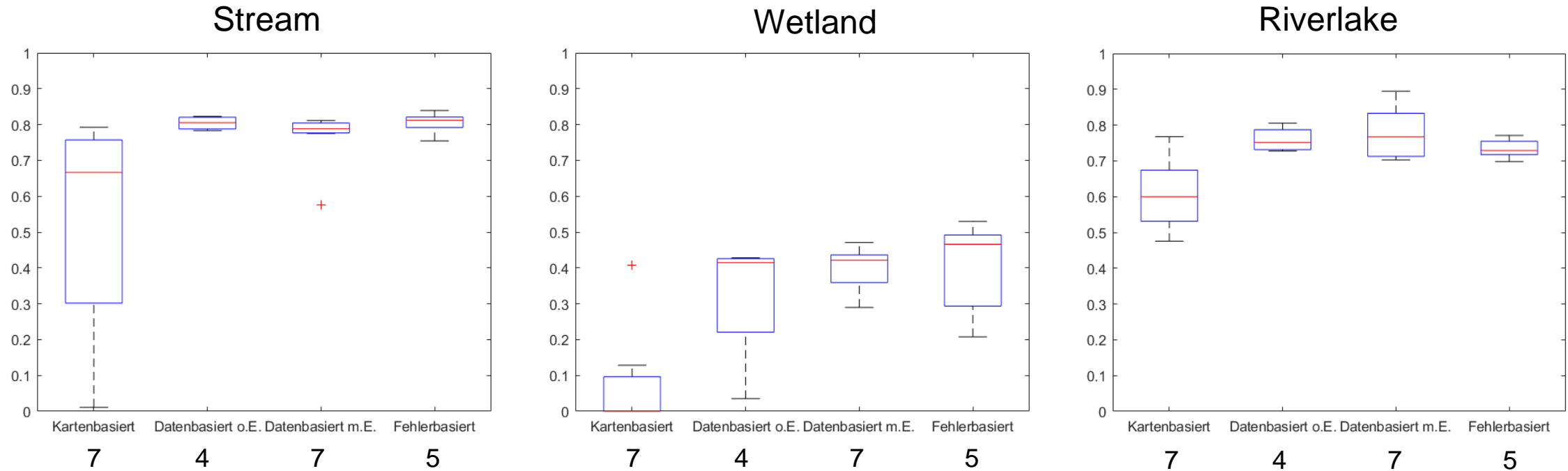


Trainierte Modelle / Vorhersagen



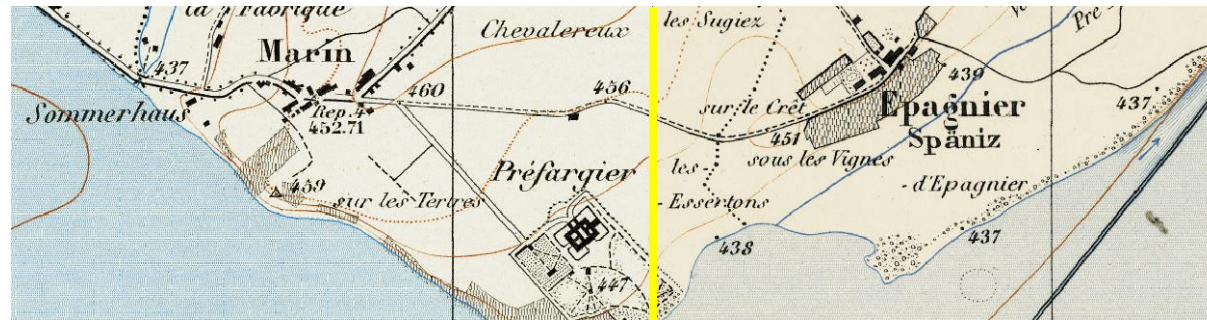
Güteparameter für die Vorhersagen

- F1-Score über Mittelwert aller Kartenblätter



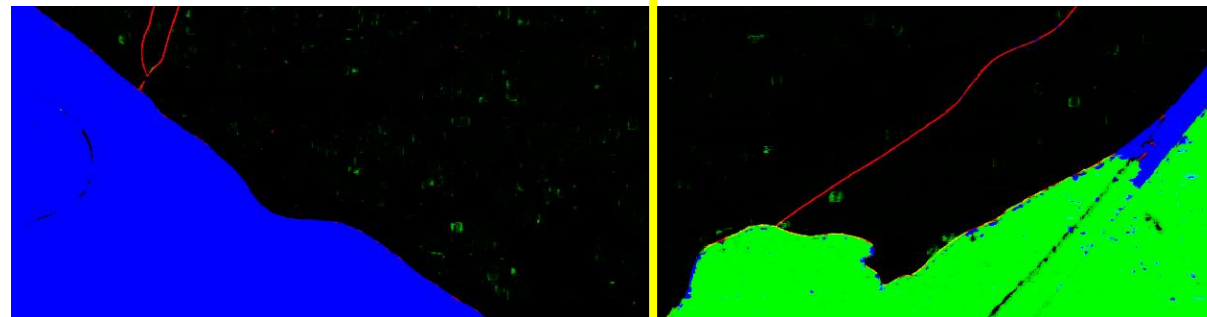
Empfehlungsrahmen

- Auf jeden Fall datenbasierte oder fehlerbasierte Workflows verwenden
- Fehlerbasierte Zwei-Feature-Punkte als bester Workflow (Mischung aus drei fehlerbasierten Instanzen)
- Trainingsdaten für alle Jahrgänge / Scans verwenden, sonst falsche Detektion:



Trainingsdaten (1880)

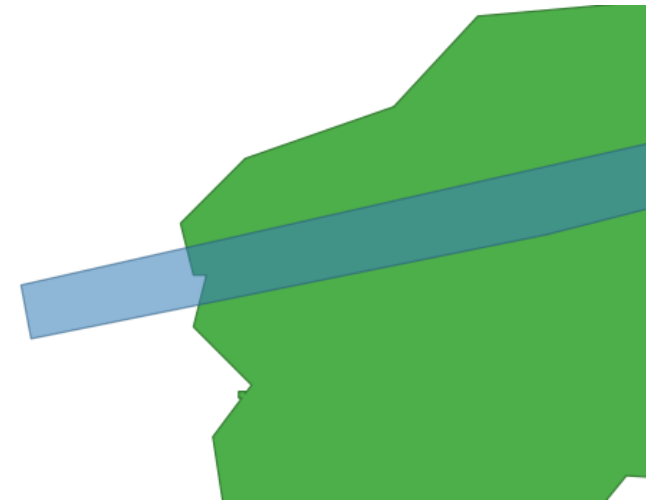
Testdaten (1879)



Diskussion und Ausblick

Diskussion

- Groundtruthdaten:
 - Unterscheidung zwischen Stream und Riverlake
 - Überschneidung mehrerer Features
- Workflows, Training und Evaluation:
 - QGIS-Funktion «Zufällige Punkte in den Layergrenzen»
 - Trainingsdauer und Anzahl Instanzen pro Workflow
 - Generelle Anwendbarkeit der Workflows



Ausblick

- Mehr Instanzen und Trainingsdaten
- Weiterverarbeitung der Daten nach der Extraktion
- Automatisierung der Workflows und Einbettung in ein QGIS-Plugin

Referenzen

- Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Freiburg.
- Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*(6).
- Die verwendeten Python-Frameworks und grundlegenden Geodaten als Basis für die Bilder wurden von Dr. Magnus Heitzler, IKG, ETH Zürich zur Verfügung gestellt.

Danke für die Aufmerksamkeit!

Fragen