

Simon Krenn, BSc

**Subregionale Ermittlung des Bruttoinlandsprodukts und  
der Wirtschaftssektoren mittels Deep Learning und  
Fusion von hochaufgelösten multimodalen Geo- und  
Fernerkundungsdaten in Brasilien**

**MASTERARBEIT**

zur Erlangung des akademischen Grades

Master of Science

Masterstudium

Geospatial Technologies

eingereicht an der

**Technischen Universität Graz**

**Betreuer**

Ao.Univ.-Prof. Mag. Dr.rer.nat. Wolfgang Sulzer

Institut für Geographie u. Raumforschung

der Karl-Franzens-Universität Graz

Graz, Mai 2026

# Vorwort

Die vorliegende Masterarbeit entstand im Rahmen des Studiums „Geospatial Technologies“ in Kooperation mit dem Deutschen Zentrum für Luft- und Raumfahrt e.V. (DLR). Das Ziel der Arbeit ist die Entwicklung eines geeigneten Modells zur Ermittlung des BIP, sowie der einzelnen Wirtschaftssektoren auf subregionaler Ebene auf Basis von Geo- und Fernerkundungsdaten.

Die Kombination aus Fernerkundung, Geodatenanalyse und Deep Learning gestützter Ableitung diverser Zielvariablen macht dieses Themenfeld sowohl wissenschaftlich als auch gesellschaftlich bedeutsam. Gleichzeitig ermöglicht mir die Arbeit, mich in jenem Bereich der Geotechnologien weiterzubilden und zu forschen, der mich bereits im Verlauf meines Studiums am meisten begeisterte: der Fernerkundung.

Ein besonderer Dank gilt dabei meinen Betreuern Wolfgang Sulzer und Michael Wurm, für ihre fachliche Unterstützung und die konstruktive Begleitung während des gesamten Entstehungsprozesses dieser Arbeit.

Ebenso danke ich dem DLR für die Bereitstellung relevanter Daten und Ressourcen. Insbesondere danke ich Thomas Stark für die Möglichkeit, diese Masterarbeit in Kooperation durchzuführen, für das entgegengebrachte Vertrauen, sowie für die fachlichen Einblicke und den wertvollen Wissensaustausch. Dieser hat mich weit über diese Arbeit hinaus bereichert.

Ein herzlicher Dank gilt darüber hinaus allen Personen in meinem engeren Umfeld, die mich während des Entstehungsprozesses dieser Arbeit begleitet und unterstützt haben und mir geholfen haben, auch in herausfordernden Phasen, das Ziel nicht aus den Augen zu verlieren.

Für punktuelle sprachliche Überarbeitungen einzelner Formulierungen wurde unterstützend ChatGPT (GPT-5, OpenAI) verwendet. Inhaltliche Konzeption, Analyse, Interpretation und Schlussfolgerungen wurden eigenständig erarbeitet.

## Kurzfassung

Konventionelle Wirtschaftsdaten wie etwa das Bruttoinlandsprodukt (BIP) oder die Wertschöpfung der einzelnen Wirtschaftssektoren liegen vor allem auf subregionaler Ebene häufig nur in räumlich aggregierter Form und mit zeitlicher Verzögerung vor. Dadurch wird die kleinräumige Erfassung wirtschaftlicher Aktivität erheblich erschwert. Aus diesem Grund ist das Ziel dieser Arbeit die Ermittlung des BIP, sowie der sektoralen Wirtschaftsleistung mithilfe eines Deep-Learning-Ansatzes, der multimodale Geo- und Fernerkundungsdaten fusioniert. Als Untersuchungsgebiet wurde dabei Brasilien gewählt.

Hierfür wurden verschiedene hochaufgelöste Datengrundlagen, wie etwa Sentinel-2, Black Marble Nachlichtdaten, Landbedeckungsinformationen und OpenStreetMap-Daten, in einem multimodalen Workflow kombiniert. Zur Merkmalsextraktion wurde ein ResNet-18-Encoder, unter dem Einsatz von Transfer Learning, verwendet. Diese extrahierten Merkmale werden mit einer Fusionsmethode kombiniert und ein MLP-Regressor übernimmt abschließend die Modellierung und Ermittlung der Zielvariable. Für die Fusion wurden vier unterschiedliche Methoden (Concatenation, Attention Pooling, Graph-based Fusion und Mixture of Experts) hinsichtlich ihrer Eignung zur Modellierung der Wirtschaftsleistung getestet.

Die Ergebnisse zeigen, dass der entwickelte Ansatz vielversprechende Resultate für die Fragestellung liefert. Insbesondere das BIP sowie der Dienstleistungssektor konnten mit hoher Genauigkeit modelliert werden ( $R^2 \approx 0,86$ , normalisiert) während für den primären Sektor gewisse Herausforderungen bestehen bleiben ( $R^2 \approx 0,55$ , normalisiert). Unter den getesteten Fusionsstrategien erwies sich Concatenation als der geeignetste Ansatz. Insgesamt verdeutlicht die Arbeit das Potenzial von multimodaler Datenfusion und Deep Learning zur kleinräumigen Ableitung ökonomischer Indikatoren und liefert Ansatzpunkte für weiterführende Forschung in diesem Bereich.

**Schlüsselwörter:** Fernerkundung, Geodaten, Deep Learning, Datenfusion, BIP, Wirtschaftssektoren, Brasilien

## Abstract

Conventional economic data, such as gross domestic product (GDP) or the value added of individual economic sectors, are often available only in spatially aggregated form and with a time lag, particularly at a subregional level. This makes it considerably more difficult to capture economic activity at a fine-scale level. For this reason, the goal of this work is to estimate GDP and sectoral economic output using a deep learning approach that fuses multimodal geospatial and remote sensing data. Brazil was selected as the study area.

To this end, various high-resolution datasets, such as Sentinel-2, Black Marble nighttime light data, land cover information and OpenStreetMap data, were combined in a multimodal workflow. A ResNet-18 encoder, utilizing transfer learning, was used for a feature extraction. The extracted features were then combined by means of a fusion method, while an MLP regressor performed the final modelling and prediction of the target variable. Additionally, for the fusion step, four different methods (concatenation, attention pooling, graph-based fusion and mixture of experts) were evaluated for their suitability in modeling economic performance.

The results show that the developed approach yields promising results for this research question. In particular, GDP and the services sector could be modeled with high accuracy ( $R^2 \approx 0,86$ , normalized), while significant challenges remain for the primary sector ( $R^2 \approx 0,55$ , normalized). Among the fusion strategies tested, concatenation proved to be the most suitable approach. Overall, this work highlights the potential of multimodal data fusion and deep learning for deriving economic indicators at a subregional level and provides starting points for further research in this area.

**Keywords:** Remote Sensing, Geospatial Data, Deep Learning, Data Fusion, GDP, Economic Sectors, Brazil

# Inhaltsverzeichnis

|  |    |
|--|----|
| Abbildungsverzeichnis .....  | 7  |
| Tabellenverzeichnis .....  | 9  |
| Formelverzeichnis .....  | 9  |
| 1. Einleitung .....  | 10 |
| 1.1 Zielsetzung und Forschungsfragen .....   | 10 |
| 1.2 Methodik und Aufbau der Arbeit .....   | 11 |
| 2. Grundlagen der Arbeit .....   | 12 |
| 2.1 Projektrahmen .....  | 12 |
| 2.2 Untersuchungsgebiet .....  | 13 |
| 2.3 Grundlagen des Deep Learning .....   | 15 |
| 2.3.1 Einführung in die KI .....   | 15 |
| 2.3.2 Einführung in Deep Learning .....  | 16 |
| 2.3.3 Einführung in CNN .....  | 19 |
| 3. State of the Art .....  | 21 |
| 3.1 Fernerkundungs- und Geodaten als Grundlage für die Ermittlung der<br>Wirtschaftsleistung ..... | 21 |
| 3.2 Datenfusion und Deep Learning zur BIP-Ermittlung .....   | 23 |
| 3.3 Herausforderungen und Fehlerquellen .....  | 25 |
| 4. Datengrundlagen .....   | 28 |
| 4.1 Referenzdaten .....  | 28 |
| 4.2 Sentinel-2 Daten .....   | 32 |
| 4.2.1 NDVI .....   | 33 |
| 4.2.2 GNDVI .....  | 34 |
| 4.2.3 NDWI .....   | 35 |
| 4.3 Land Cover .....   | 35 |
| 4.4 Black Marble Daten .....   | 39 |
| 4.5 OpenStreetMap .....  | 41 |
| 4.6 Weitere Datenvorbereitung .....  | 45 |
| 5. Modellentwicklung .....   | 47 |
| 5.1 Workflow .....   | 47 |
| 5.2 Datenverteilung und Normalisierung .....   | 48 |
| 5.2.1 Datenverteilung .....  | 48 |

|       |   |    |
|-------|---|----|
| 5.2.2 | Datennormalisierung und -aufbereitung für das Modell .....    | 51 |
| 5.3   | ResNet-18 Backbone als Encoder .....                          | 54 |
| 5.4   | Fusionsmethoden.....  | 55 |
| 5.4.1 | Concatenation.....  | 55 |
| 5.4.2 | Attention Pooling.....  | 56 |
| 5.4.3 | Graph-based Fusion .....                                      | 56 |
| 5.4.4 | Mixture of Experts.....                                       | 56 |
| 5.5   | MLP-Regressor .....   | 57 |
| 6.    | Ergebnisse .....  | 59 |
| 6.1   | Resultate der Fusionsmethoden.....                            | 60 |
| 6.1.1 | BIP .....   | 60 |
| 6.1.2 | Primärer Sektor.....  | 63 |
| 6.1.3 | Sekundärer Sektor .....                                       | 64 |
| 6.1.4 | Tertiärer Sektor .....  | 66 |
| 6.1.5 | Öffentlicher Sektor .....                                     | 68 |
| 6.2   | Vergleichende Analyse von Ergebnissen und Referenzdaten ..... | 71 |
| 6.3   | Detailanalyse anhand ausgewählter Beispiele .....             | 75 |
| 6.3.1 | Analyse des Großraumes São Paulo - Rio de Janeiro .....       | 76 |
| 6.3.2 | Best- und Worst Case Szenario.....                            | 79 |
| 7.    | Diskussion.....   | 83 |
| 7.1   | Interpretation der Ergebnisse .....                           | 83 |
| 7.1.1 | Beantwortung der Forschungsfragen.....                        | 83 |
| 7.1.2 | Diskussion der Fusionsmethoden.....                           | 85 |
| 7.1.3 | Vergleich mit bestehenden Forschungsergebnissen .....         | 86 |
| 7.2   | Datenqualität .....   | 88 |
| 7.3   | Limitationen.....   | 92 |
| 7.3.1 | Modelllimitationen .....                                      | 92 |
| 7.3.2 | Feature Importance Analysis .....                             | 95 |
| 8.    | Fazit .....   | 97 |
|       | Literaturverzeichnis .....                                    | 99 |

# Abbildungsverzeichnis

|  |    |
|--|----|
| Abb. 1: Die Verteilung der Untersuchungsgebiete innerhalb von Brasilien. ....  | 14 |
| Abb. 2: Eine schematische Darstellung eines einfachen Deep-Learning-Modells in Bezug auf Fernerkundungsdaten. ....                   | 17 |
| Abb. 3: Ein beispielhafter Aufbau eines CNN in Bezug auf Fernerkundungsdaten... ..   | 20 |
| Abb. 4: Der Prozess der Disaggregation der ökonomischen Referenzwerte schematisch dargestellt. ....                                  | 29 |
| Abb. 5: Ein Vergleich der BIP-Werte vor und nach der Disaggregation am Beispiel der Großräume von São Paulo und Rio de Janeiro. .... | 31 |
| Abb. 6: Die WorldCover Daten als Beispiel innerhalb einer Zelle (Altstadt Porto Alegre). ....  | 38 |
| Abb. 7: Die disaggregierten Black Marble Daten als Beispiel innerhalb einer Zelle (Altstadt Porto Alegre).....                       | 41 |
| Abb. 8: Der Binärraster des Straßennetzwerkes der Altstadt von Porto Alegre. ....  | 43 |
| Abb. 9: Der Abstandsrafter der POIs der Altstadt von Porto Alegre.. ....   | 44 |
| Abb. 10: Der Dichteraster des Straßennetzwerkes der Altstadt von Porto Alegre. . .   | 45 |
| Abb. 11: Der zusammengefasste Workflow der Masterarbeit als Graphik dargestellt. ....  | 47 |
| Abb. 12: Das Verteilungsmuster der BIP-Werte des gesamten Datensatzes. ....  | 49 |
| Abb. 13: Das Verteilungsmuster der getesteten Sampling-Methoden der BIP-Werte als Boxplots dargestellt.....                          | 50 |
| Abb. 14: Die Verteilung der BIP-Werte vor und nach der Quadratwurzel-Transformation vergleichend gegenübergestellt. ....             | 51 |
| Abb. 15: Eine beispielhafte Darstellung der finalen Inputdaten vor der Merkmalsextraktion anhand einer Zelle. ....                   | 53 |
| Abb. 16: Regressionsplots der normalisierten BIP-Werte für alle Fusionsmethoden. ....  | 60 |
| Abb. 17: Regressionsplots der denormalisierten BIP-Werte für alle Fusionsmethoden. ....  | 62 |
| Abb. 18: Bester Regressionsplot der normalisierten Werte des primären Sektors (Fusionsmethode: Graph-based).....                     | 63 |
| Abb. 19: Bester Regressionsplot der denormalisierten Werte des primären Sektors (Fusionsmethode: Mixture of Experts).....            | 64 |

|   |    |
|---|----|
| Abb. 20: Bester Regressionsplot der normalisierten Werte des sekundären Sektors (Fusionsmethode: Graph-based).....                                      | 65 |
| Abb. 21: Bester Regressionsplot der denormalisierten Werte des sekundären Sektors (Fusionsmethode: Concatenation).....                                  | 66 |
| Abb. 22: Bester Regressionsplot der normalisierten Werte des tertiären Sektors (Fusionsmethode: Concatenation).....                                     | 67 |
| Abb. 23: Bester Regressionsplot der denormalisierten Werte des tertiären Sektors (Fusionsmethode: Concatenation).....                                   | 68 |
| Abb. 24: Bester Regressionsplot der normalisierten Werte des öffentlichen Sektors (Fusionsmethode: Concatenation).....                                  | 69 |
| Abb. 25: Bester Regressionsplot der denormalisierten Werte des öffentlichen Sektors (Fusionsmethode: Graph-based).....                                  | 70 |
| Abb. 26: Gegenüberstellung der Verteilung der Referenzdaten und der ermittelten BIP-Werte für das gesamte Untersuchungsgebiet.....                      | 73 |
| Abb. 27: Gegenüberstellung der Verteilung der Referenzdaten und der ermittelten Werte des primären Sektors für das gesamte Untersuchungsgebiet. ....    | 74 |
| Abb. 28: Gegenüberstellung der Verteilung der Referenzdaten und der ermittelten Werte des öffentlichen Sektors für das gesamte Untersuchungsgebiet..... | 75 |
| Abb. 29: Darstellung des relativen Fehlers des BIP pro Zelle in Prozent im Großraum São Paulo - Rio de Janeiro.....                                     | 77 |
| Abb. 30: Darstellung des relativen Fehlers des BIP pro Zelle (ohne BIP-Referenzwerte < 100) in Prozent im Großraum São Paulo - Rio de Janeiro. ....     | 78 |
| Abb. 31: Darstellung des relativen Fehlers des BIP pro Zelle (ohne BIP-Referenzwerte < 1.000) in Prozent im Großraum São Paulo - Rio de Janeiro.....    | 79 |
| Abb. 32: Negativbeispiel der Modellarchitektur anhand der Zelle mit der ID 392266 im Süden des Untersuchungsgebietes.....                               | 80 |
| Abb. 33: Positivbeispiel der Modellarchitektur anhand der Zelle mit der ID 458022 im Norden des Untersuchungsgebietes.....                              | 82 |
| Abb. 34: Der Vergleich der digitalisierten Schulen in OSM mit den Schulen aus der offiziellen Geobr-Datenbank in einem Teil von Porto Alegre. ....      | 90 |
| Abb. 35: Vergleich der Metriken der verwendeten Matching-Methoden im gesamten Stadtgebiet von Porto Alegre. ....  | 91 |
| Abb. 36: Regressionsplot der normalisierten und denormalisierten Werte des BIP mit einfacher Datenaugmentierung.....                                    | 94 |

|  |    |
|--|----|
| Abb. 37: Regressionsplot der normalisierten und denormalisierten Werte des primären Sektors mit einfacher Datenaugmentierung. .... | 95 |
|--|----|

## Tabellenverzeichnis

|   |    |
|---|----|
| Tab. 1: Übersicht der verwendeten Inputdaten .....  | 28 |
| Tab. 2: Werte der besten Fusionsmethode pro Zielvariable .....  | 71 |
| Tab. 3: Durchschnittswerte pro Zielvariable (über 4 Folds) mit der Fusionsmethode Concatenation. .... | 71 |

## Formelverzeichnis

|   |    |
|---|----|
| Formel 1: NDVI (Normalized Difference Vegetation Index).....          | 33 |
| Formel 2: GNDVI (Green Normalized Difference Vegetation Index).....   | 34 |
| Formel 3: NDWI (Normalized Difference Water Index) .....              | 35 |
| Formel 4: Relativer Fehler in Prozent je Zielvariable und Zelle ..... | 76 |

# 1. Einleitung

In einer zunehmend vernetzten Welt gewinnen ökonomische Informationen immer weiter an Bedeutung. Als zentrale Messgröße der wirtschaftlichen Leistungsfähigkeit eines Staates bildet das Bruttoinlandsprodukt (kurz BIP) die Grundlage für politische Entscheidungen, wirtschaftliche Strategien und wissenschaftliche Analysen. Somit beschreibt das BIP den Wert der im Inland produzierten Güter und Dienstleistungen und dient als zentrale Kennzahl zur Beurteilung der gesamtwirtschaftlichen Aktivität und des wirtschaftlichen Wohlstands (Dyner und Sheiner 2018). Da jedoch BIP-Daten in der Praxis meist in großräumigen Aggregationen und mit erheblicher temporaler Verzögerung veröffentlicht werden (Bolivar 2024), erschwert dies diverse Entscheidungsprozesse. Ebenso kann es vorkommen, dass diese Daten nicht verfügbar sind (Chen und Nordhaus 2019), vor allem auf regionaler Ebene. Insofern diese kleinräumlich vorhanden sind, wird ihre Verteilung oftmals als regelmäßig angenommen, wodurch weitere Problematiken entstehen können (Yue et al. 2014). Aufgrund dessen ist eine detaillierte und grenzunabhängige Erfassung wirtschaftlicher Aktivität auf lokaler Ebene für Planung, Politik und Forschung gleichermaßen bedeutend.

Eine Möglichkeit, dies zu erreichen, ist die Verwendung von Fernerkundungs- und Geodaten. Diese Datenquellen helfen dabei, objektive Untersuchungen vorzunehmen (Sun et al. 2020) und sind kostengünstiger oder kostenfrei, im Vergleich zu anderen Methoden, welche zur Ermittlung des BIPs eingesetzt werden (Price und Atkinson 2022). Eine zentrale Herausforderung bei der Nutzung von Fernerkundungs- und Geodaten in diesem Kontext besteht darin, das BIP in ländlich geprägten Regionen oder auf regionaler Ebene präzise zu ermitteln (Sun et al. 2020; Chen und Nordhaus 2019). Ebenso wird bei vielen Ansätzen in diesem Bereich nur eine einzelne oder eine sehr geringe Anzahl an Datenquellen verwendet. Um mehrere heterogene Datenquellen effektiv zu kombinieren, hat sich in den letzten Jahren insbesondere Deep Learning als vielversprechender Ansatz etabliert (Chen et al. 2022; Liu et al. 2022).

## 1.1 Zielsetzung und Forschungsfragen

Ausgehend davon beschäftigt sich die vorliegende Arbeit mit der Entwicklung eines Deep-Learning-Workflows, welcher multimodale Fernerkundungs- und Geodaten

fusioniert, um das BIP sowie die Wertschöpfung in den einzelnen Wirtschaftssektoren auf subregionaler Ebene zu ermitteln. Als Untersuchungsgebiet wurden dabei ausgewählte Regionen Brasiliens herangezogen. Der Grund hierfür liegt in der Heterogenität des Landes sowie einer bereits bestehenden Datengrundlage, worauf jedoch im weiteren Verlauf der Arbeit genauer eingegangen wird. Folglich ist das Ziel dieser Arbeit, die Wirtschaftsleistung, verstanden als „BIP inklusive sektoraler Wertschöpfung“, für das gesamte Untersuchungsgebiet mithilfe des entwickelten Modells subregional zu quantifizieren. Es soll eine Methodik geschaffen werden, die vielversprechende Datenquellen verknüpft und dadurch die Wirtschaftsleistung ermitteln kann, wobei bestehende Problematiken bestmöglich gemindert werden sollen. Somit lautet die zentrale Forschungsfrage der Arbeit wie folgt:

- Wie lassen sich das Bruttoinlandsprodukt und die Wirtschaftssektoren auf subregionaler Ebene in Brasilien durch Deep Learning mittels multimodaler Geodatenfusion ermitteln?

Im Zuge dieser Analyse sollen zusätzlich noch die folgenden beiden sekundären Forschungsfragen beantwortet werden:

- Wie zuverlässig kann das entwickelte Modell das BIP und die einzelnen Wirtschaftssektoren ermitteln?
- Welche Fusionsmethoden eignen sich für die Ermittlung der Wirtschaftsleistung auf subregionaler Ebene am besten?

## 1.2 Methodik und Aufbau der Arbeit

Die vorliegende Arbeit basiert auf einem datenbasierten Ansatz, bei dem verschiedene Fernerkundungs- und Geodatenquellen in einem multimodalen Deep-Learning-Modell kombiniert werden. Hierfür werden zunächst die Datengrundlagen aufbereitet und in ein gemeinsames Analyseformat überführt. Darauf aufbauend werden unterschiedliche Fusionsmethoden innerhalb der eingesetzten Deep-Learning-Architektur getestet, um ihre Eignung für die Ermittlung des BIP sowie der sektoralen Wirtschaftsleistung auf subregionaler Ebene zu prüfen. Die entwickelten Modelle werden im Anschluss validiert und miteinander verglichen, um folglich den für die zentrale Fragestellung dieser Arbeit geeignetsten Ansatz zu identifizieren.

Zur Beantwortung dieser Fragestellung werden zunächst einige Grundlagen der Arbeit erklärt, danach wird der derzeitige Stand der Forschung aufbereitet. Anschließend wird in Kapitel 4 genauer auf die Inputdaten eingegangen. Nach der Datenaufbereitung wird die Modellentwicklung erläutert, worin beispielsweise die Datennormalisierung oder die getesteten Fusionsmethoden beschrieben werden. In weiterer Folge werden die Ergebnisse präsentiert und in Kapitel 7 interpretiert. Im Fazit wird die Arbeit reflektiert und es werden mögliche Ausblicke aufgezeigt, die insbesondere bei weiterführenden Forschungen in diesem Bereich helfen sollen.

## 2. Grundlagen der Arbeit

Bevor nun genauer auf die Arbeit selbst eingegangen wird, werden zunächst ein paar Informationen bezüglich ihrer Grundlagen gegeben, damit der gesamte Kontext der Arbeit besser verstanden werden kann. Dabei wird zunächst der Projektrahmen beschrieben und anschließend wird die Wahl des Untersuchungsgebietes begründet. Darauf folgt das Unterkapitel „Grundlagen des Deep Learning“, worin für den weiteren Verlauf der Arbeit wichtige Begriffe und Prozesse erläutert werden.

### 2.1 Projektrahmen

Zu Beginn wird in diesem Unterkapitel der Rahmen dieser Arbeit beschrieben, um ihren Kontext besser verstehen zu können. Diese wurde unter anderem in Kooperation mit dem Deutschen Zentrum für Luft- und Raumfahrt (DLR) geschrieben und basiert auf einer weiteren Masterarbeit. Diese wurde von Martin Wiltsche im Jahre 2023 verfasst und wurde ebenfalls vom DLR begleitet. Dabei stand, ähnlich wie in dieser Arbeit, die Ableitung des BIPs in Brasilien auf Basis von Fernerkundungsdaten im Mittelpunkt. An diese Erkenntnisse soll nun die vorliegende Arbeit anknüpfen. In seiner Arbeit beschreibt Wiltsche mögliche fortsetzende Untersuchungen wie folgt: „Weiterführende Forschung in diesem Bereich wäre eine Disaggregation der BIP-Daten nach Wirtschaftssektoren unter Einbindung von Landbedeckungs- und OpenStreetMap Daten, sowie eine weitere Optimierung des Machine Learning Modells hinsichtlich ruraler Räume“ (Wiltsche 2023, S. 2). Die genannten Punkte stellen einen zentralen Teil der Arbeit dar. Folglich wird nicht nur das BIP, sondern auch die einzelnen Wirtschaftssektoren subregional ermittelt. Es werden zusätzliche Geodaten implementiert und der vorliegende Ansatz wurde weiterentwickelt, sowie verschiedene Fusionsmethoden getestet. Ebenfalls besteht das Untersuchungsgebiet nicht nur aus

urbanen Zentren, sondern wird deutlich erweitert, worauf im folgenden Unterkapitel genauer eingegangen wird.

## 2.2 Untersuchungsgebiet

Die genaue Abgrenzung des Untersuchungsgebiets ist ein entscheidender Faktor, um sinnvolle Ergebnisse zu erhalten. Besonders in einem räumlich heterogenen Land wie Brasilien ist eine ausgeglichene Wahl entscheidend, um sowohl die ökonomische als auch die ökologische Vielfalt des Landes zu repräsentieren. Gleichzeitig müssen technische Aspekte wie die Größe der Datenmengen berücksichtigt werden, um den Rechenaufwand in einem vertretbaren Rahmen zu halten (Goodfellow et al. 2016).

Daher wurden für diese Arbeit Gebiete ausgewählt, die zusammen rund 90 % des BIPs von Brasilien abdecken, jedoch nur etwa 10 % der brasilianischen Landesfläche einnehmen. Die Orientierung am BIP stellt sicher, dass der wirtschaftliche Fokus erhalten bleibt, dabei wurde jedoch auch gezielt darauf geachtet, rurale Regionen miteinzubeziehen. Somit berücksichtigt die Auswahl auch die Heterogenität der Analyse, da sowohl dicht bebaute urbane Räume als auch spärlich besiedelte rurale Gebiete abgebildet werden. Abbildung 1 veranschaulicht nun die Verteilung der Untersuchungsgebiete:

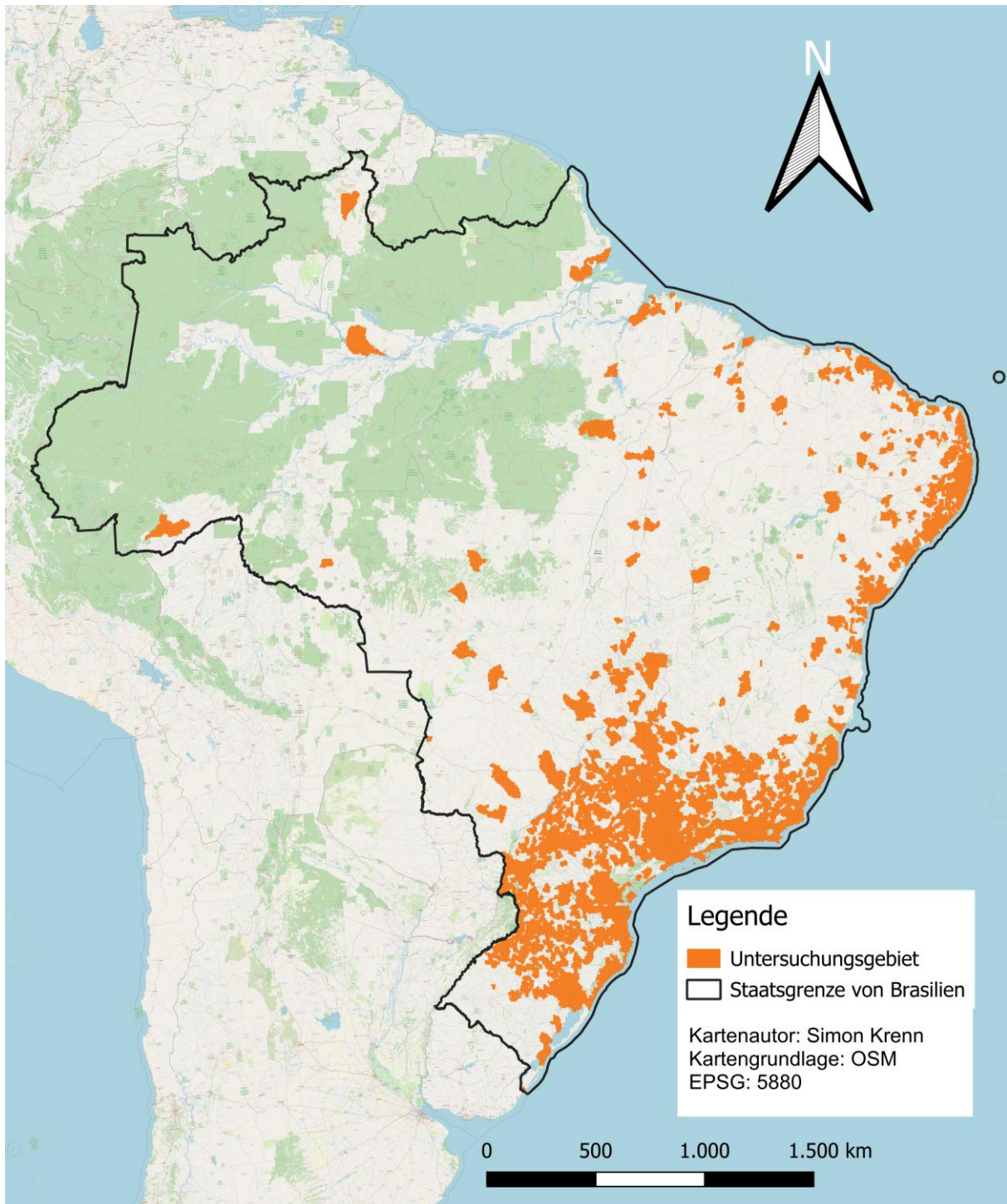


Abb. 1: Die Verteilung der Untersuchungsgebiete innerhalb von Brasilien. Quelle: Eigener Entwurf.

Wie in Abbildung 1 ersichtlich wird, liegt die Mehrheit der Untersuchungsgebiete (in orange dargestellt) in den dicht besiedelten Regionen an der Ost- und Südostküste, wo sich auch die größten urbanen Agglomerationen wie São Paulo oder Rio de Janeiro befinden. Ergänzend wurden ländliche und landwirtschaftlich geprägte Gebiete in anderen Landesteilen miteinbezogen, um eine heterogene Basis zu schaffen. Diese Vielfalt ist entscheidend, um diverse Verzerrungen bei der späteren Modellierung zu

vermeiden und die Generalisierbarkeit der Ergebnisse zu erhöhen (Goodfellow et al. 2016).

Die Entscheidung, nicht das gesamte brasilianische Staatsgebiet zu verwenden, ist vor allem auf die Größe des Landes zurückzuführen. Mit mehr als 8,5 Mio. km<sup>2</sup> ist Brasilien das flächenmäßig fünftgrößte Land der Erde. Eine vollständige Abdeckung hätte zu sehr hohen Prozessierungszeiten, sowohl bei der Datenaufbereitung als auch beim Training der Deep-Learning-Modelle, geführt, was den Rahmen dieser Arbeit überschritten hätte. Aufgrund des verwendeten Ansatzes bleibt der Umfang realisierbar, ohne an Repräsentativität zu verlieren.

Ein weiterer Grund für die Wahl Brasiliens liegt in der bereits vorhandenen Datenbasis. In der Masterarbeit von Wiltche wurden Sentinel-2-Aufnahmen für das gesamte Land aufbereitet. Diverse zeitaufwändige Prozessierungsschritte dieser Inputdaten, wie z.B. radiometrische Korrekturen, mögliche notwendige Georeferenzierungen oder die Erstellung von Wolkenmasken konnten dadurch gänzlich vermieden werden und der Fokus konnte stärker auf die Modellentwicklung selbst gelegt werden.

## 2.3 Grundlagen des Deep Learning

Bevor nun der derzeitige Stand der Forschung im Bereich der geodatenbasierten BIP-Ermittlung dargelegt wird, muss zunächst der Begriff des Deep Learning, ein zentraler Aspekt dieser Arbeit, erläutert werden. Dies ist wichtig, um in weiterer Folge ähnliche Literatur sowie den gesamten Workflow der Arbeit verstehen zu können. Deshalb erfolgt in Unterkapitel 2.3.2 eine genauere Betrachtung dieses Themenbereichs. Zuvor wird jedoch dessen Einordnung in den übergeordneten Kontext der Künstlichen Intelligenz beschrieben.

### 2.3.1 Einführung in die KI

Künstliche Intelligenz (KI) bezeichnet Algorithmen und Systeme, die mithilfe von großen Datenmengen lernen, um Aufgaben zu übernehmen, die traditionell menschliche Intelligenz erfordern. Sie sind in der Lage, aus Erfahrungen zu lernen und ihr Entscheidungsverhalten selbstständig für die jeweilige Aufgabe zu optimieren. Beispiele hierfür sind Mustererkennung, Entscheidungsfindung oder Prognosen. In Bezug auf die Fernerkundung wird KI oftmals zur Verarbeitung großer Datenmengen, zur Automatisierung von Analysen und Arbeitsprozessen oder zur Ableitung diverser Informationen, die mit klassischen Methoden nur schwer oder mit großem Aufwand

durchführbar wären, verwendet. Dabei könnte es sich z.B. um Datenfusion und Modellentwicklungen, diverse Klassifikationen oder Rauschminderung in den Datengrundlagen handeln (Janga et al. 2023). Der Hauptgrund für den Einsatz von KI in der Fernerkundung ist jedoch die hohe Leistungsfähigkeit, die diverse Modelle bei der Erkennung komplexer Muster oder bei der Vorhersage gewisser Zielvariablen erreichen (Zhu et al. 2017).

Gleichzeitig bringt die Verwendung von KI aber auch gewisse Anforderungen mit sich. Modelle benötigen ausreichende und repräsentative Trainingsdaten, die sorgfältig aufbereitet werden müssen. Fehler in den Inputdaten können sich über den gesamten Workflow hindurchziehen und folglich zu stark fehlerhaften Ergebnissen führen. Auch weitere, bereits in den Inputdaten vorherrschende Verzerrungen, können einen erheblichen Einfluss auf die Güte des Modells haben. Deshalb sollte der/die Anwender:in sich immer im Klaren sein, welche Vor- und Nachteile die verwendeten Daten besitzen und wie sich diese auswirken könnten. Ein Beispiel hierfür wäre eine ungleiche Abdeckung der Datengrundlagen (Janga et al. 2023).

### 2.3.2 Einführung in Deep Learning

Deep Learning ist ein Teilgebiet des maschinellen Lernens und der Künstlichen Intelligenz, das sich auf den Einsatz künstlicher neuronaler Netze mit vielen Schichten konzentriert (Goodfellow et al. 2016). Im Mittelpunkt steht dabei das automatische Lernen von Datenrepräsentationen, um komplexe Muster und Strukturen in großen Datenmengen zu erkennen. Der Begriff „deep“ verweist auf den Aufbau bzw. die Tiefe der Modelle.

Ein typisches Deep-Learning-Modell ist das sogenannte Multilayer Perceptron (MLP). Es besteht aus einem Netzwerk von Neuronen, die in Schichten organisiert sind. Dabei unterscheidet man zwischen der Eingabeschicht (visible layer), den verborgenen Schichten (hidden layers) und einer Ausgabeschicht (output layer). Die Eingaben werden dabei schrittweise verarbeitet, wodurch immer abstraktere Merkmale extrahiert werden können. Ein beispielhafter Aufbau für ein Deep-Learning-Modell wird in Abbildung 2 dargestellt:

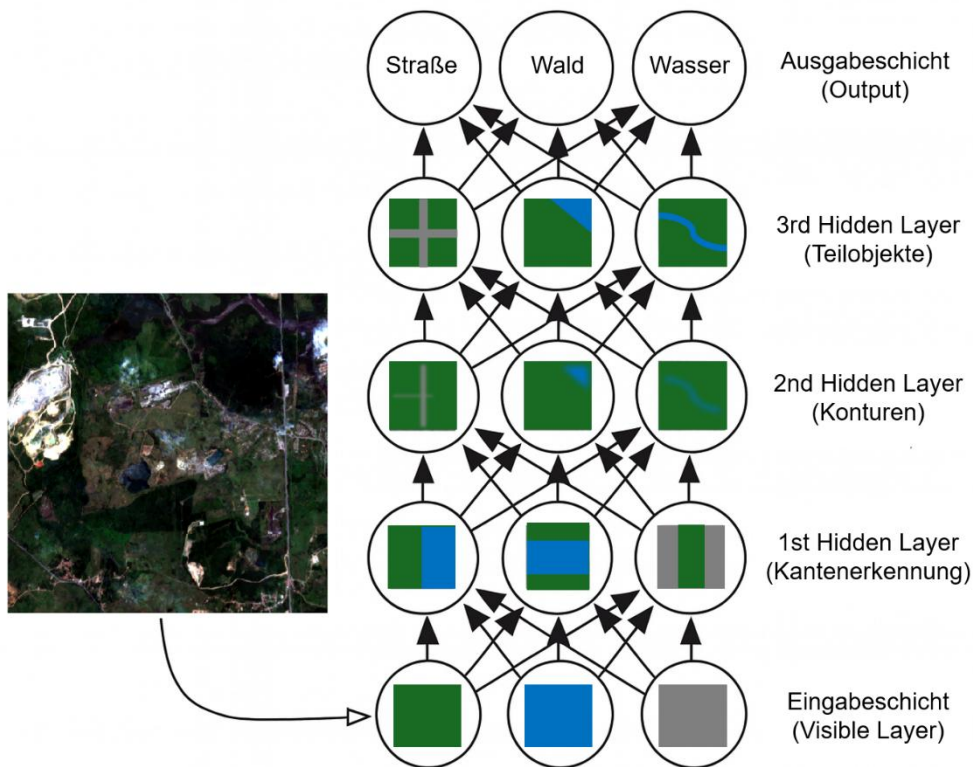


Abb. 2: Eine schematische Darstellung eines einfachen Deep-Learning-Modells in Bezug auf Fernerkundungsdaten. Quelle: Eigener Entwurf nach Goodfellow et al. 2016, S. 6.

Die Inputdaten, in diesem Fall verschiedene Pixel eines Orthophotos, werden in den ersten Layer „eingespeist“ und verarbeitet. Diese Daten werden dann in weiterer Folge an die erste verborgene Schicht weitergegeben. In dieser Schicht können bereits einfache Merkmale wie Kanten bzw. Intensitätsübergänge repräsentiert werden, wodurch erste Eigenschaften der Inputdaten extrahiert werden können. Anschließend werden diese Informationen von der ersten verborgenen Schicht an die zweite verborgene Schicht gesendet, welche in diesem Beispiel verschiedene Konturen aus den erhaltenen Informationen ableiten kann. Beispielsweise können zuvor detektierte Kantenkombinationen zu komplexeren Merkmalen kombiniert werden, wie etwa zu linearen Straßenkonturen, geschwungenen Flussverläufen oder Flächengrenzen von Wäldern. Anschließend erfolgt der Übergang von der zweiten zur dritten verborgenen Schicht, in der basierend auf den zuvor extrahierten Informationen einzelne Teilobjekte wie Straßenkreuzungen oder Flussabschnitte identifiziert werden können. Abschließend folgt die Ausgabeschicht, in der einzelne Bildbereiche als Straße, Wald oder Wasserfläche klassifiziert werden. Die finale Zuordnung erfolgt dabei typischerweise anhand der von den Ausgabeneuronen berechneten

Klassenwahrscheinlichkeiten. Die Klasse mit der höchsten Wahrscheinlichkeit wird ausgewählt.

Damit das Modell lernt, wie es diese Entscheidungen treffen soll, wird es anhand vieler Beispiele trainiert. Es vergleicht seine Vorhersage mit der richtigen Antwort und berechnet einen Fehlerwert über eine Kostenfunktion, die die Qualität der Vorhersage misst. Dieser Fehler wird anschließend mittels Backpropagation zurück durch das Netzwerk gerechnet. Dabei wird für jede Verbindung im Netzwerk bestimmt, wie stark sie zum Fehler beigetragen hat. Auf dieser Basis werden diese Gewichte Schritt für Schritt angepasst, meist mithilfe eines Gradientenverfahren, damit das Modell beim nächsten Durchlauf verbesserte Vorhersagen treffen kann (Goodfellow et al. 2016).

Dieser Lernprozess wiederholt sich über viele sogenannte Epochen hinweg. Mit jeder Runde passt sich das Modell weiter an die Trainingsdaten an, bis es die Zusammenhänge zuverlässig erkennt. Damit vermieden wird, dass das Modell die Trainingsdaten nur auswendig lernt, auch Overfitting genannt, werden meist zusätzliche Techniken eingesetzt, bei denen beispielsweise zufällig ausgewählte Neuronen während des Trainings vorübergehend ausgeschaltet werden. Das hilft dem Netzwerk, allgemeinere Muster zu lernen, die auch auf neue, unbekannte Daten anwendbar sind (Goodfellow et al. 2016).

Durch die Kombination großer Datenmengen, leistungsfähiger Hardware (GPUs) und effizienter Trainingsverfahren hat Deep Learning in den letzten Jahren entscheidende Fortschritte in diversen Forschungsbereichen erzielt. Auch in der Fernerkundung ermöglicht es neue und stark verbesserte Ansätze, vor allem, um komplexe und hochaufgelöste Fernerkundungsdaten zu verarbeiten (Adegun et al. 2023).

Während die genannten Methoden wie z.B. ein einfacher MLP jedoch eher für kleinere und tabellarische Datensätze geeignet sind, stoßen sie bei sehr großen Eingaben wie hochauflösenden Bildern schnell an ihre Grenzen, da jedes Neuron mit allen Eingabepixeln verbunden wäre. Um diese Herausforderung zu lösen, werden bei Bilddaten häufig Convolutional Neural Networks (CNNs) eingesetzt. CNNs nutzen Faltungsschichten, die lokale Bildbereiche verarbeiten und dadurch deutlich weniger Parameter benötigen. Außerdem können sie Merkmale wie Kanten, Texturen oder Formen automatisch und sehr effizient extrahieren (Goodfellow et al. 2016). Auf diese

Netzwerke wird im nächsten Unterkapitel näher eingegangen, da sie auch die Grundlage für das in dieser Arbeit eingesetzte Modell bilden.

### 2.3.3 Einführung in CNN

Wie bereits erwähnt, stoßen MLPs bei der Verarbeitung sehr großer Eingaben wie hochauflösende Bilder schnell an ihre Grenzen, da die Anzahl der zu lernenden Parametern enorm anwächst. CNNs wurden entwickelt, um dieses Problem zu lösen und sind heute der Standard in der (fernerkundungsbasierten) Bildverarbeitung (Adegun et al. 2023).

Der zentrale Baustein eines CNN ist die Faltungsschicht (Convolutional Layer). Anstatt jedes Neuron mit allen Eingabepixeln zu verbinden, wird ein kleiner Filter (Kernel) über das Eingabebild „geschoben“. Dieser Filter erkennt lokale Muster, beispielsweise Kanten oder Texturen, und erzeugt daraus eine sogenannte „Feature Map“. Dabei handelt es sich um eine Merkmalsdarstellung, die die erkannten, lokalen Strukturen des Bildes widerspiegelt. Mehrere Filter pro Schicht ermöglichen es so dem Netzwerk, verschiedene Merkmale gleichzeitig zu erfassen (Goodfellow et al. 2016).

Ein typischer Aufbau einer CNN-Schicht umfasst die folgenden drei Schritte:

- **Faltung:** Mehrere Filter berechnen lineare Kombinationen der Eingaben in kleinen, lokal begrenzten Bereichen.
- **Nichtlinearität:** Das Ergebnis wird durch eine Aktivierungsfunktion wie ReLU (Rectified Linear Unit) geschickt. Diese Funktion lässt positive Werte unverändert, setzt negative Werte jedoch auf null und ermöglicht dem Netzwerk so, komplexe, nichtlineare Muster zu lernen.
- **Pooling:** Eine Pooling-Schicht fasst benachbarte Werte der Feature Map zusammen, meist durch Auswahl des größten Werts (Max-Pooling). Dadurch wird die Darstellung robuster gegenüber kleinen Verschiebungen im Bild (Translationen) und die Datenmenge für die nächsten Schichten wird reduziert.

Mehrere Faltungs- und Pooling-Schichten können aufeinander folgen, sodass das Netzwerk immer komplexere Merkmale erkennt, von Kanten über Formen bis hin zu ganzen Objekten. Nach den Faltungsschichten wird das Ergebnis oft „flach“ gemacht (Flattening) und an vollständig verbundene Schichten weitergegeben, die auf Basis der extrahierten Merkmale eine Klassifikation durchführen (Goodfellow et al. 2016).

Dieser Aufbau soll durch Abbildung 3 nochmals verdeutlicht werden:

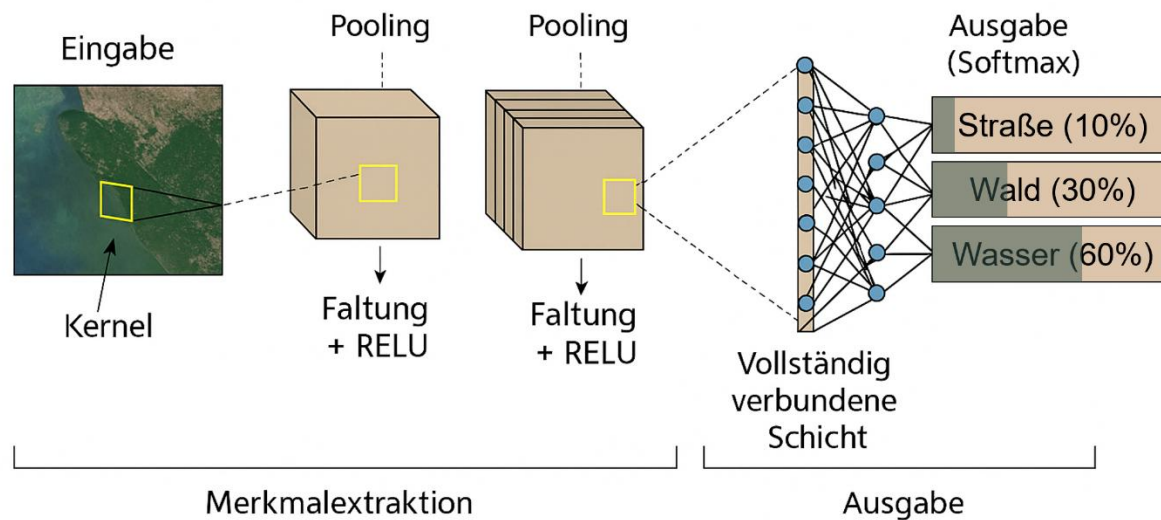


Abb. 3: Ein beispielhafter Aufbau eines CNN in Bezug auf Fernerkundungsdaten. Quelle: Eigener Entwurf, teilweise erstellt mit ChatGPT (GPT-5, OpenAI).

Die Abbildung zeigt den schematischen Aufbau eines CNNs am Beispiel eines Orthophotos. Links ist der Eingabebereich dargestellt, der aus den Pixeln des Bildausschnitts besteht. Mit Hilfe eines Kernels werden zunächst lokale Merkmale extrahiert. In der Abbildung ist dies durch die gelben Kästchen angedeutet.

Die resultierenden Feature Maps werden anschließend durch eine Aktivierungsfunktion transformiert und in der Pooling-Schicht weiter verdichtet. Dabei wird die räumliche Auflösung reduziert, sodass nur die wichtigsten Informationen erhalten bleiben. Dieser Vorgang wiederholt sich über mehrere Faltungs- und Pooling-Schritte hinweg, wodurch in tieferen Schichten zunehmend komplexere Strukturen erfasst werden.

Am Ende der Merkmalsextraktion werden die Feature Maps „abgeflacht“ und in eine vollständig verbundene Schicht eingespeist. Diese kombiniert alle erkannten Merkmale und übergibt sie an die Ausgabeschicht. In dieser Schicht sorgt die sogenannte Softmax-Funktion schließlich dafür, dass die Ergebnisse als Wahrscheinlichkeiten interpretiert werden können. So kann das Modell entscheiden, ob der betrachtete Bildausschnitt mit höherer Wahrscheinlichkeit eine Straße, eine Wald- oder eine Wasserfläche darstellt.

CNNs bilden die Grundlage für das in dieser Arbeit entwickelte Modell. Bevor jedoch die konkrete Modellentwicklung beschrieben wird, folgt zunächst ein Überblick über den aktuellen Stand der Forschung im Bereich der fernerkundungs- und geodatenbasierten BIP-Ermittlung.

### 3. State of the Art

Nachdem nun die wichtigsten Grundlagen definiert wurden, wird im folgenden Kapitel der aktuelle Stand der Forschung beschrieben. Dies ist wichtig, um einen Überblick diverser Ansätze in diesem Forschungsbereich zu erhalten sowie mögliche Fehlerquellen in Erfahrung zu bringen. Hierfür wird zunächst auf grundlegende Literatur im Bereich der fernerkundungs- und geodatenbasierten BIP-Ermittlung eingegangen. Anschließend werden vielversprechende Methoden zur Datenfusion und zur Ermittlung der Wirtschaftsleistung mit Deep-Learning-Ansätzen angeführt. Abschließend werden immer wieder in der Literatur beschriebene Herausforderungen und Fehlerquellen genannt, um diese bestmöglich im weiteren Verlauf der Arbeit vermeiden zu können.

#### 3.1 Fernerkundungs- und Geodaten als Grundlage für die Ermittlung der Wirtschaftsleistung

Ein bedeutender Fortschritt im Zuge dieses Forschungsbereiches war die Nutzung von satellitengestützten Nachlichtdaten als Indikator für die Wirtschaftsleistung. Elvidge et al. (1997) zeigten erstmals, dass beleuchtete Flächen in Nachlichtdaten mit wirtschaftlicher Produktivität zusammenhängen. Chen und Nordhaus (2011) entwickelten darauf aufbauend eine Methode, um Nachlichtdaten zur Ermittlung von BIP-Werten zu verwenden. Sie fanden dabei heraus, dass dieser Ansatz vor allem in Ländern mit unzuverlässiger Datenlage als vielversprechend angesehen werden kann. Dabei ist jedoch zu beachten, dass Straßen und andere Infrastrukturen in verschiedenen Staaten unterschiedlich stark beleuchtet sein können, wodurch die Interpretation von Nachlichtdaten als Indikator für wirtschaftliche Aktivität entsprechend differenziert betrachtet werden muss.

Diese Ansätze wurden über die Jahre hinweg durch mehrere Studien bestätigt und ausgebaut, darunter Chen und Nordhaus (2019), die VIIRS-Daten zur Schätzung des BIP auf Stadt- und Staatsebene verwendeten. Sie fanden heraus, dass Nachlichtdaten eine bessere Vorhersagekraft für das BIP in städtischen Regionen

haben als in ländlichen Gebieten, wo wirtschaftliche Aktivitäten wie Land- oder Forstwirtschaft oft nicht durch künstliche Beleuchtung sichtbar werden (Chen und Nordhaus 2019). Trotz ihrer Nützlichkeit weisen Nachtlichtdaten Einschränkungen auf. In sehr hellen, urbanen Zentren tritt eine Sättigung auf, die Unterschiede zwischen Metropolen verzerrt (Li et al. 2020). In ländlichen Regionen wiederum können relevante Aktivitäten wie Land- oder Forstwirtschaft unbeleuchtet bleiben, was zu Unterschätzungen führt. Saisonale Effekte, temporäre Beleuchtungen oder Störungen können weitere Fehlerquellen darstellen (Wu et al. 2013; Chen und Nordhaus 2019).

Neben Nachtlichtdaten können aber auch weitere Fernerkundungsdaten eingesetzt werden. Beispiele hierfür sind Landbedeckungsdaten, erstellt mithilfe von MODIS oder Sentinel-2 Aufnahmen, wodurch eine genauere Differenzierung zwischen urbanen, industriellen und landwirtschaftlichen Gebieten ermöglicht wird. Diesen Ansatz wählten etwa auch Sun et al. (2020) in ihrem Paper: „Estimation of GDP Using Deep Learning With NPP-VIIRS Imagery and Land Cover Data at the County Level in CONUS“. Darin kombinierten sie VIIRS-Nachtlichtdaten mit Landbedeckungsinformationen, welche aus MODIS Aufnahmen erstellt wurden. Die BIP-Schätzung erfolgte auf Kreisebene in den USA und ihr Modell zeigte, dass die Resultate durch diese Kombination deutlich verbessert wurden. Des Weiteren beschreiben sie Land Cover Daten als nützlichen Zusatz zu Nachtlichtdaten, vor allem in Regionen mit niedriger Lichtemission, wie etwa in ländlichen und von der Landwirtschaft geprägten Gebieten. Jedoch sind auch hier teilweise komplexe Wirtschaftsstrukturen vorhanden, bei denen die erzielten Ergebnisse als ausbaufähig eingestuft werden (Sun et al. 2020).

Eine weitere Möglichkeit, um potenziell noch bessere Resultate zu generieren, wäre das Hinzufügen zusätzlicher Inputdaten. Hierfür bietet sich beispielsweise die Verwendung diverser Indizes aus den Satellitenbildern oder grundlegend andere Geodaten an. Yue et al. (2014) verfolgen in ihrer Studie einen ähnlichen Ansatz und kombinieren Nachtlichtdaten und Land Cover Daten zusätzlich mit dem EVI (Enhanced Vegetation Index). Mit diesen Inputdaten wurde ein Modell entwickelt, um das BIP in Zhejiang, China, zu schätzen. Dabei verwendeten sie einen auf Gridzellen basierenden Ansatz, um einerseits die räumliche Heterogenität widerzuspiegeln und die Integration rasterbasierter Daten zu erleichtern und andererseits, um Fehler bzw. eine Beeinflussung durch administrative Grenzen zu vermeiden. Sie fanden dabei heraus, dass diese Kombination die Genauigkeit, insbesondere für den primären

Sektor, erhöht. Dabei haben sie mithilfe des EVI landwirtschaftliche und „nicht-landwirtschaftliche“ Gridzellen erstellt, woraus sie dann anschließend Zellen mit dem gesamten BIP berechneten. Des Weiteren wird erwähnt, dass diverse Nachteile bzw. Fehlerquellen der Nachlichtdaten, wie die niedrige räumliche Auflösung oder das Auftreten von „blooming“ oder „overflow“, so gut wie möglich korrigiert werden müssen, damit bestmögliche Resultate erzielt werden können (Yue et al. 2014).

Um das Potenzial dieser Methoden weiter auszuschöpfen, wird zunehmend die Fusion heterogener (Geo-)Datensätze erforscht. Eine Möglichkeit ist z.B. die Integration von OSM-Daten, in Form von Straßennetzwerken oder Points of Interest (POI). Dadurch sollen zusätzliche Informationen über beispielsweise für das BIP bedeutsame Infrastruktur erlangt werden, um folglich die Resultate weiter zu verbessern. Wu et al. (2024) entwickelten ein Modell, das POI-Daten von Gaode Maps API, eine chinesische Kartierungs- und Geokodierungsplattform, mittels Fernerkundungsdaten fusioniert. Bei den POIs handelte es sich um Punkte, die stark mit dem sekundären oder tertiären Sektor in Verbindung stehen wie etwa „Shopping“ oder „Finance“ und wurden als Density-Map in das Modell integriert. Sie zeigten, dass ihre Methode herkömmliche Modelle übertrifft und es gelang, eine hochaufgelöste Kartierung des BIPs zu erstellen. (Wu et al. 2024).

Eine weitere Möglichkeit, um zusätzliche Geodaten zu integrieren, ist die Nutzung von Straßennetzwerken. Faktoren wie die Straßenlänge oder die Dichte des Netzwerkes, korrelieren ebenfalls mit dem BIP und können als geeigneter Input verwendet werden (Han et al. 2024).

Es ist jedoch nicht nur wichtig, welche Daten überhaupt in Frage kommen, sondern auch, wie diese kombiniert werden können und wie die Architektur des Modells selbst aussehen sollte. Um diese Thematiken weiter zu vertiefen, beschreibt das nächste Unterkapitel verschiedenste Ansätze zur Datenfusion sowie diverse Deep-Learning-Methoden im Bereich der fernerkundungs- und geodatenbasierten BIP-Ermittlung.

### 3.2 Datenfusion und Deep Learning zur BIP-Ermittlung

Datenfusion beschreibt die Integration mehrerer, häufig heterogener Datenquellen mit dem Ziel, eine konsistentere und genauere Datengrundlage zu schaffen, als es mit einzelnen Datensätzen möglich wäre. Ghamisi et al. (2019) verstehen Datenfusion im Kontext der Fernerkundung als ein umfassendes Rahmenwerk, das räumliche,

spektrale, zeitliche und radiometrische Informationen miteinander verknüpft. Sie unterscheiden dabei drei zentrale Ansätze. Auf Pixelebene werden Rohdaten unterschiedlicher Sensoren direkt kombiniert, um eine verbesserte Darstellung der zugrundeliegenden Szene zu erhalten. Auf Merkmalsebene hingegen werden zunächst relevante Eigenschaften aus den verschiedenen Quellen extrahiert und anschließend zu einem gemeinsamen Merkmalsraum zusammengeführt, bevor die eigentliche Klassifikation oder Regression erfolgt. Die Entscheidungsebene schließlich fasst die Resultate mehrerer separater Modelle oder Klassifikatoren zu einer finalen Vorhersage zusammen (Ghamisi et al. 2019).

Diese Einteilung zeigt, dass Datenfusion nicht nur auf der Ebene der Eingangsdaten, sondern auch in späteren Verarbeitungsschritten erfolgen kann. Gerade für komplexe Anwendungen, wie der Ermittlung des BIPs auf regionaler Ebene, bietet sie die Möglichkeit, die jeweiligen Stärken unterschiedlicher Datenquellen zu nutzen und die Schwächen einzelner Datensätze auszugleichen.

Eine Möglichkeit zur Datenfusion ist die Verwendung von Deep-Learning-Modellen. Dabei werden separate Modelle für jede Datenquelle trainiert. Diese extrahieren Merkmale, welche folglich zu einem gemeinsamen Input (Feature-Set) kombiniert werden. Der fusionierte Input kann dann beispielsweise von einem zusätzlichen Modell weiterverarbeitet werden. Diese Kombination ermöglicht es, die komplementären Informationen aus verschiedenen Datenquellen optimal zu nutzen, selbst in komplexen Szenen oder bei fehlenden Daten aus einem der Inputs. Ein Beispiel hierfür wäre ein Informationsverlust durch Bewölkung in Sentinel-2 Daten, der jedoch durch Land Cover Daten aufgefangen wird (Hong et al. 2021).

Ebenso kann für die BIP-Ermittlung Transfer Learning zur Datenfusion verwendet werden. Dabei wird ein bereits vortrainiertes Modell genutzt, um aussagekräftige Merkmale extrahieren zu können. Diese Merkmale werden anschließend mithilfe von Dimensionsreduktion (z. B. PCA) verdichtet. Dieser Ansatz ermöglicht es, die räumlichen Informationen zu nutzen, ohne ein Modell vollständig neu trainieren zu müssen (Price und Atkinson 2022).

Ein weiteres Beispiel für den erfolgreichen Einsatz von Datenfusion und Deep Learning in der BIP-Ermittlung liefern Li et al. (2021). Sie entwickelten einen zweistufigen Transfer Learning Workflow, der Nachlichtdaten als Proxy für wirtschaftliche Aktivität

nutzt. Zunächst wurde ein vortrainiertes Netzwerk (VGG-16) mithilfe von Satellitenbildern und Nachlichtdaten feinjustiert, um wirtschaftsrelevante Merkmale aus den Bildern zu extrahieren. Anschließend wurden diese hochdimensionalen Merkmalsvektoren mittels PCA (Hauptkomponentenanalyse) dimensionsreduziert. Dabei (PCA) handelt es sich um ein nicht parametrisches Verfahren, wodurch die Merkmalsvektoren als sogenannte Hauptkomponenten „zusammengefasst“ werden. Diese beinhalten die gesamte Varianz der Daten, wobei die ersten Komponenten den größten Anteil der Varianz erklären. Beispielsweise könnten die ersten 6 Komponenten ca. 80% der gesamten Varianz abdecken, während die restlichen Komponenten nur noch geringe Anteile besitzen. Danach werden die dimensionsreduzierten Daten durch statistische Kenngrößen wie Mittelwert oder Standardabweichung auf „county-level“ zusammengefasst, bevor sie in das Regressionsmodell integriert wurden. Die daraus resultierenden Ergebnisse erfüllten die Erwartungen mit  $R^2$ -Werten von 0,69-0,71 und liefern wertvolle Erkenntnisse für weiterführende Forschungen (Li et al. 2021).

Für derartige Modelle werden in den meisten Arbeiten CNNs verwendet, wie es etwa auch Chen et al. (2022) in ihrem Paper: „Mapping Gridded Gross Domestic Product Distribution of China Using Deep Learning With Multiple Geospatial Big Data“ machen. Ihr Modell, GDPnet, basiert auf einem 1D-CNN mit residualen Verbindungen und kombiniert dabei verschiedene Inputquellen wie z.B. Nachlichtdaten, Landbedeckungsdaten oder die Straßennetzwerkdichte, die jeweils sektorenspezifisch kombiniert werden. Die Besonderheit dabei war, dass ihr Ansatz höhere  $R^2$  Werte als Vergleichsmodelle aufwies und dass trotz einer effizienteren bzw. kürzeren Trainingszeit (Chen et al. 2022).

All diese Beispiele unterstreichen das Potenzial von multimodaler Datenfusion, Transfer Learning und CNNs für die BIP-Ermittlung auf subregionaler Ebene. Trotz dieser Potenziale gibt es aber immer noch einige zentrale Herausforderungen, die wiederholt in diversen Quellen genannt werden. Ausgewählte Beispiele werden im folgenden Unterkapitel nun genauer erläutert.

### 3.3 Herausforderungen und Fehlerquellen

Diverse Herausforderungen oder Fehlerquellen können dabei während des gesamten Workflows auftreten. Dies beginnt bereits bei den Datengrundlagen. Ihre Abdeckung und Qualität haben einen starken Einfluss auf die generelle Güte des Modells, ebenso

wie verschiedene (notwendige) Aufbereitungsschritte der Inputdaten, um sie weiterverarbeiten zu können. Des Weiteren gibt es auch bei der Modellentwicklung selbst einige Aspekte, die berücksichtigt werden müssen. Nachfolgend werden einige Beispiele für diese Problemstellungen aufgezählt.

Eine Herausforderung liegt in der Verfügbarkeit und Qualität von Trainingsdaten. Diese besitzen einen enormen Einfluss auf die Qualität des Modells und müssen folglich sorgfältig aufbereitet werden (Donaldson und Storeygard 2016). Ebenso liegen die als Referenzwerte genutzten BIP-Werte oftmals nur auf aggregierter Verwaltungsebene vor, wodurch das Problem modifizierbarer räumlicher Bezugseinheiten (modifiable Areal Unit Problem, kurz MAUP) entsteht. Das kann zu Verzerrungen führen, wenn die Wirtschaftsleistung innerhalb der administrativen Grenzen als homogen angenommen wird (Chen et al. 2022). Hinzu kommt die Heterogenität der Datenquellen. Wirtschaftsdaten liegen als Polygone vor, welche anhand administrativer Grenzen erstellt werden. Fernerkundungsdaten sind jedoch meist rasterbasiert. Folglich ist eine räumliche Disaggregation sowie eine Datenfusion notwendig, wodurch zusätzliche Verzerrungen entstehen können (Yue et al. 2014). Ebenso treten auch bei der sektoralen Differenzierung Schwierigkeiten auf, da Landwirtschaft, Industrie und Dienstleistungen unterschiedlich stark mit verfügbaren Inputdaten korrelieren (Wu et al. 2024).

Auch die einzelnen Datenquellen bringen spezielle Herausforderungen mit sich. Nachtlichtdaten (NTL) neigen in hellen urbanen Zentren zu Sättigungseffekten oder bilden unbeleuchtete wirtschaftliche Aktivitäten, wie Landwirtschaft, nur unzureichend ab (Chen und Nordhaus 2019). Sentinel-2 Daten müssen sehr zeitintensiv aufbereitet werden (Wolkenmaske oder diverse Korrekturen/Anpassungen) und die Qualität weiterer (frei-verfügbarer) Geodaten wie OSM-Daten unterscheidet sich stark nach Region (Han et al. 2024).

Ein weiteres Problem sind die (möglichen) zeitlichen Unterschiede diverser Datenquellen. Während Satellitenbilder und Nachtlichtdaten eine immer höhere temporale Auflösung aufweisen, liegen offizielle Wirtschaftsstatistiken oft mit mehrmonatiger Verzögerung vor. Verwendbare Referenzdaten für derartige Projekte können teilweise noch weitaus älter sein. Dies erschwert die Kalibrierung und Validierung der Modelle zusätzlich. Die Herausforderungen der einzelnen

Datenquellen können zwar durch multimodale Datenfusion gemindert werden, zur Gänze sind diese Problematiken jedoch nicht zu beheben.

Bei der Modellentwicklung selbst besteht die Gefahr von Overfitting, wenn das Modell auf nicht repräsentativen Trainingsdaten beruht. Dabei kann die Sampling-Methode helfen, um robustere Resultate zu gewährleisten (Sun et al. 2020). Hinzu kommen ressourcenbezogene Herausforderungen wie der hohe Rechen- und Datenaufwand für das Training komplexer CNN-Modelle, basierend auf hochaufgelösten Inputdaten (Chen et al. 2022).

Schließlich bleibt auch die Übertragbarkeit von Modellen ein kritischer Punkt. Modelle, die in einem Land trainiert wurden, liefern nicht zwangsläufig verlässliche Ergebnisse in anderen Regionen mit abweichenden Beleuchtungsintensitäten, Urbanisierungsgraden oder wirtschaftlichen Strukturen, was die Notwendigkeit von Transfer-Learning Ansätzen unterstreicht (Price und Atkinson 2022).

Damit die beschriebenen Fehlerquellen möglichst geringgehalten werden, ist eine sorgfältige Datenaufbereitung unerlässlich. Die Güte des Modells hängt unmittelbar mit der Qualität der verwendeten Daten zusammen. Bereits kleine Fehler können sich im weiteren Verlauf der Arbeit kumulieren und so die Ergebnisse erheblich verzerren. Im folgenden Kapitel werden daher die in dieser Arbeit verwendeten Datenquellen inklusive notwendiger Aufbereitungsschritte genauer beschrieben.

## 4. Datengrundlagen

Aus der Literaturrecherche ergeben sich die folgenden fünf zentralen Datenquellen, die als Input für diese Arbeit dienen und in Tabelle 1 dargestellt werden:

*Tab. 1: Übersicht der verwendeten Inputdaten. Quelle: Eigene Darstellung.*

| Datenquelle               | Inhalt   | Erfassungsjahr | Beispielstudien mit vergleichbaren Datenquellen |
|---------------------------|--|----------------|---|
| <b>IBGE-Referenzdaten</b> | BIP und sektorale Wirtschaftsleistung                        | 2010, 2016     | Chen et al. 2022;<br>Stark et al. 2025          |
| <b>Sentinel-2</b>         | RGB + NIR-Bänder und abgeleitete Indizes (NDVI, GNDVI, NDWI) | 2019-2021      | Bolivar et al. 2024;<br>Wu et al. 2024          |
| <b>NASA Black Marble</b>  | Nachtlichtdaten  | 2016           | Wu et al. 2013;<br>Chen und Nordhaus 2019       |
| <b>WorldCover</b>         | Landbedeckungsklassen  | 2021           | Sun et al. 2020;<br>Yue et al 2014              |
| <b>OpenStreetMap</b>      | Straßennetz und POIs   | Stand: 2025    | Han et al. 2024;<br>Wu et al. 2024              |

Sämtliche Datensätze werden dabei in das gemeinsame Bezugssystem SIRGAS 2000/Brazil Polyconic (EPSG: 5880) überführt und auf ein regelmäßiges 5x5 km-Grid angepasst, worauf im folgenden Unterkapitel 4.1 genauer eingegangen wird. Ebenso werden alle Inputdaten im Zuge der Vorprozessierung als Raster aufbereitet, um sie effizient fusionieren und weiterverarbeiten zu können.

### 4.1 Referenzdaten

Bei den Referenzdaten handelt es sich um diverse ökonomische Daten im Shapefile-Format, welche vom brasilianischen Institut für Geographie und Statistik (IBGE) herausgegeben wurden. Die Daten sind dabei in zwei Teile unterteilt. Einerseits liegen die BIP- und Sektordaten auf kommunaler Ebene vor und stammen aus dem Jahr 2016, andererseits werden noch Einkommensdaten verwendet, die auf Nachbarschaftslevel vorliegen und 2010 herausgegeben wurden. Der Unterschied des Detailgrades kann dabei in Abbildung 4 genauer erkannt werden. Neben dem BIP beinhaltet der gröber aufgelöste Datensatz sektorale Informationen zum primären- (Land- und Forstwirtschaft), sekundären- (Industrie), tertiären- (Dienstleistungen) sowie öffentlichen Sektor (staatliche Einnahmen). Daraus ergeben sich 5 Zielvariablen für die Arbeit, welche mithilfe der verwendeten Modelle ermittelt werden sollen. Damit die Referenzdaten nun für den weiteren Verlauf der Arbeit passend aufbereitet werden

können, werden die BIP- und Wirtschaftsdaten mithilfe der Einkommensdaten disaggregiert. Die Disaggregation ähnelt methodisch dem Ansatz von Wiltche 2023, wird hier jedoch etwas verkürzt beschrieben und mit einem 5x5 km Raster durchgeführt.

Hierzu werden BIP- und Einkommenspolygone mit einem 5x5km Grid verschnitten. Dieses 5x5km Grid überspannt das gesamte Untersuchungsgebiet und besteht aus über 44.000 Zellen. Es sorgt dafür, dass alle Datengrundlagen auf eine gleichmäßige Fläche (eine 5x5km Zelle) zugeschnitten werden können. Die resultierenden Teilflächen aus der Verschneidung der Referenzdaten werden flächenmäßig neu berechnet und die Einkommenswerte werden proportional zur Flächenänderung skaliert und als Variable „neues Einkommen“ abgespeichert. Zur Verknüpfung werden die geschnittenen Einkommenspolygone als Zentroide (Punkte) überführt und mittels räumlicher Join-Operation den BIP-/Sektorpolygonen zugewiesen. Die Disaggregation erfolgt schließlich über ein gewichtetes Verfahren, bei dem das Einkommen als Verteilungsmaßstab dient. Jeder Punkt erhält einen BIP-Anteil, der seinem Einkommensanteil an der jeweiligen Gesamtfläche entspricht. Analog werden Landwirtschaft, Industrie, Dienstleistungen und der öffentliche Sektor disaggregiert. Im letzten Schritt werden die Punktwerte wieder auf die zuvor erstellten Gridzellen aggregiert, sodass die disaggregierten BIP- und Sektordaten in einheitlicher räumlicher Auflösung vorliegen. Dieser Prozess soll in Abbildung 4 veranschaulicht werden:

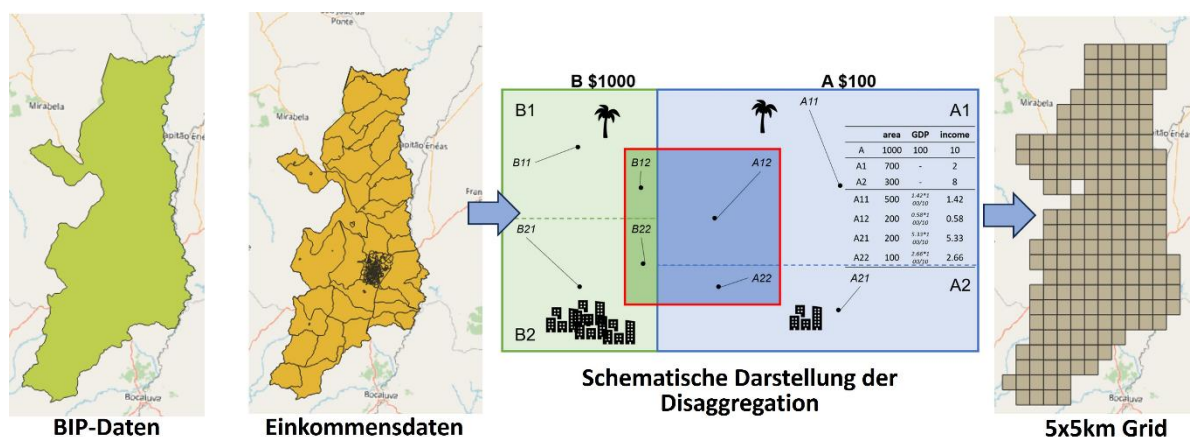


Abb. 4: Der Prozess der Disaggregation der ökonomischen Referenzwerte schematisch dargestellt. Quelle: Eigener Entwurf.

Durch diese Visualisierung wird der Prozess der Disaggregation verdeutlicht. Ausgangspunkt ist die Aggregationsebene, in der die BIP-Daten vorliegen (links).

Ergänzend werden die Einkommensdaten auf Nachbarschaftsebene herangezogen, wodurch die Dissaggregation ermöglicht wird. Im Mittelteil wird gezeigt, wie die Flächen mit den Einkommenspolygonen von einer Gridzelle überlagert werden und das aggregierte BIP proportional anhand der Einkommensinformationen auf kleinere Teilflächen verteilt wird. Dabei soll die Tabelle (im blauen Bereich in der Mitte der Abbildung) als Beispiel für eine mögliche Verteilung der Werte dienen:

Es gibt zwei größere Gebiete (A und B), wobei die Region A dazu verwendet wird, die Verteilung zu erläutern. A besitzt eine Gesamtfläche von 1000, einen BIP-Wert von 100 \$ und ein gesamtes Einkommen von 10 \$. Da der Einkommensdatensatz einen höheren Detailgrad aufweist, liegen die einzelnen Einkommenswerte kleinräumiger vor. Die Teilregion A1 besitzt dabei ein Einkommen von 2 \$ und das Gebiet A2 den Wert von 8 \$. Nun wird Region A mit dem Grid, in rot dargestellt, verschnitten, wodurch sich insgesamt vier Teilgebiete bilden (A12, A11, A22 und A21). Diese besitzen jetzt eine bestimmte Fläche des ursprünglichen Teilgebietes (A1 mit 700 wird zu A12 (200) und A11 (500)). Auf diesem Anteil basierend wird folglich auch das ursprüngliche Einkommen der Teilregion aufgeteilt, was anschließend für die Berechnung des relativen BIP-Wertes verwendet wird. Hierfür wird dieser Wert mit dem gesamten BIP der Region A multipliziert und folgend durch den gesamten Einkommenswert der Region A dividiert. Dadurch wird das gesamte BIP der Region A proportional auf die vier neu entstandenen Teilgebiete verteilt, der ihrem relativen Einkommensanteil entspricht, ohne dass ein Informationsverlust der BIP-Werte stattfindet.

Im letzten Schritt erfolgt nun die Zusammenführung dieser disaggregierten Werte pro Zelle, damit eine einheitliche räumliche Bezugsgröße für das gesamte Projekt geschaffen werden kann. Zur Validierung dieser Methode werden die Werte in den Zellen nach diesem Prinzip rückaggregiert. Dabei wurde ersichtlich, dass der Informationsverlust der ökonomischen Werte durch diese Aggregation bei unter 1 % liegt. Da zusätzlich das Einkommen stark mit dem BIP korreliert, die Daten auf Nachbarschaftslevel eine hochaufgelöste Disaggregation ermöglichen und ähnliche „Top-Down“ Methoden in diverser Literatur verwendet werden (Stark et al. 2025; Murakami und Yamagata 2019) erweist sich dieser Ansatz als robust und methodisch vertretbar. Das durch die Disaggregation resultierende Grid wird abschließend in Abbildung 5 mit den Referenzdaten am Beispiel der Großräume von São Paulo und Rio de Janeiro verglichen.

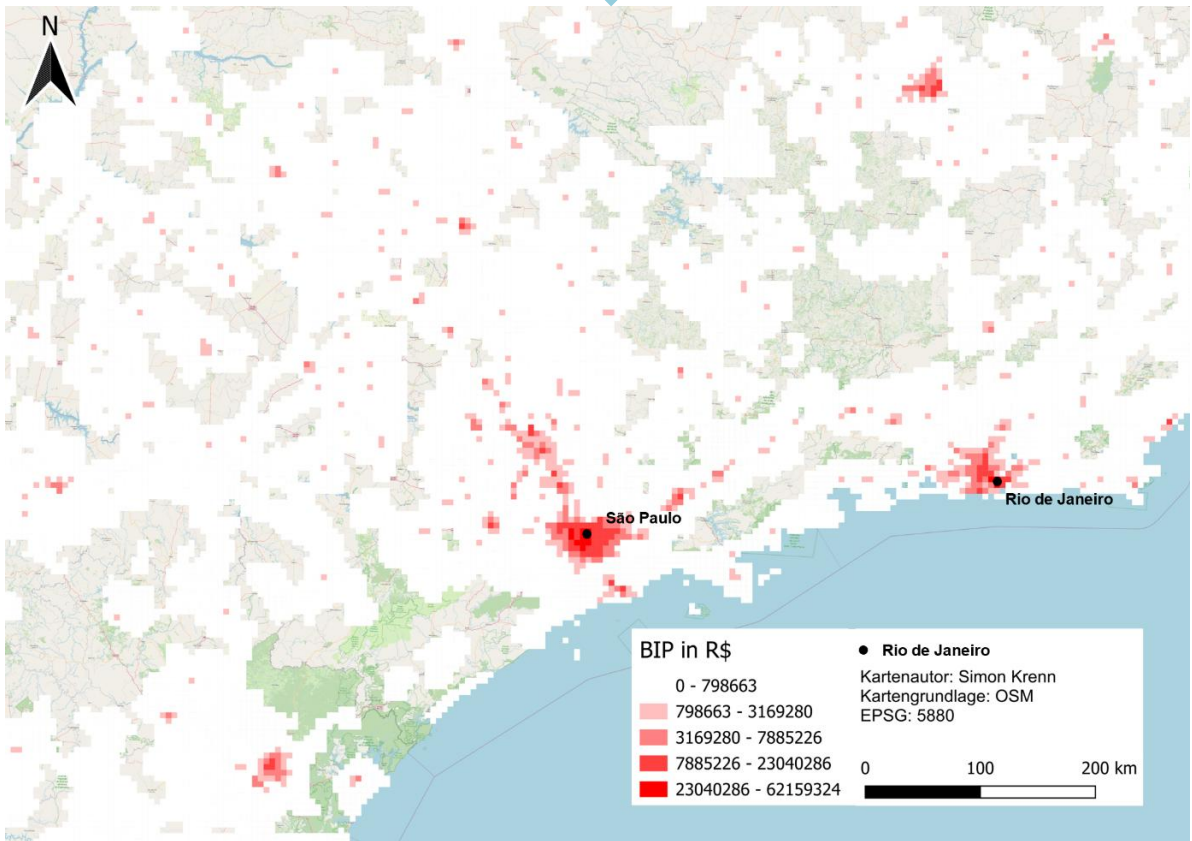
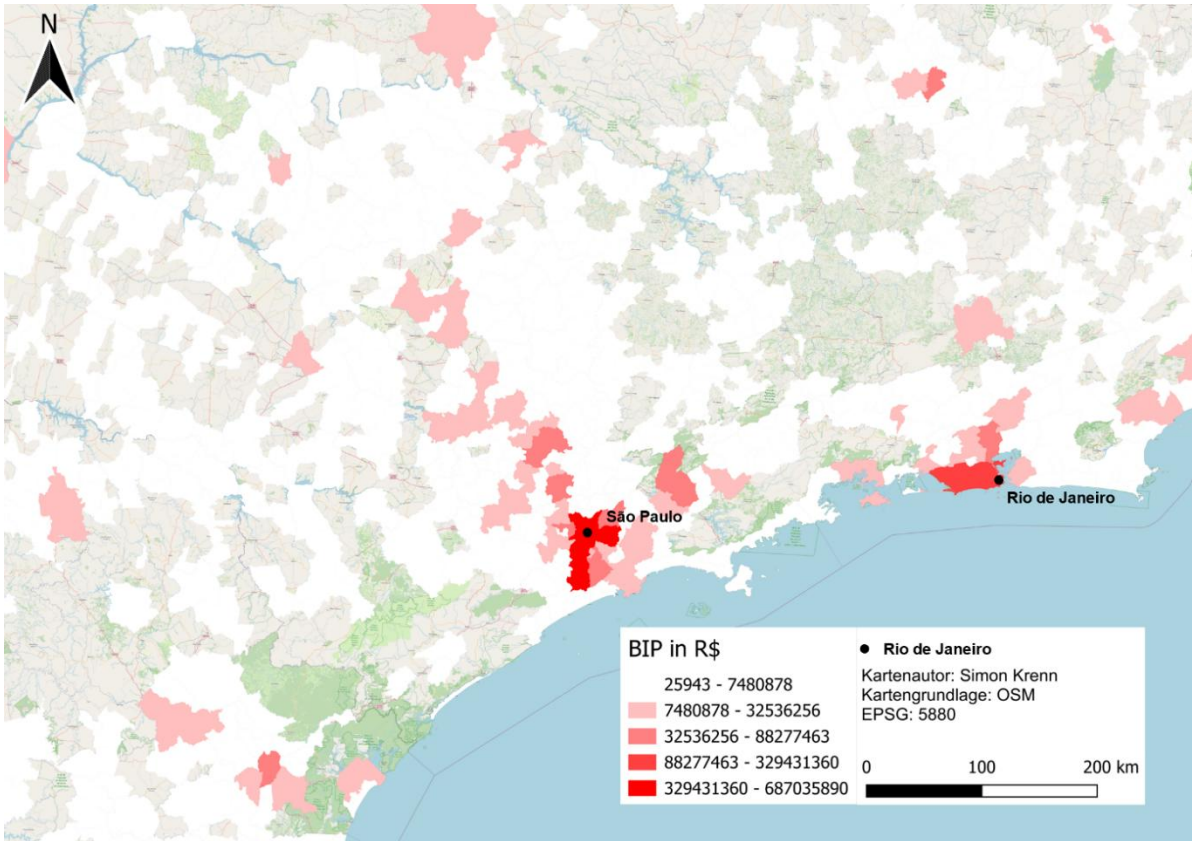


Abb. 5: Ein Vergleich der BIP-Werte vor (oben) und nach der Disaggregation (unten) am Beispiel der Großräume von São Paulo und Rio de Janeiro.  
 Quelle: Eigener Entwurf

Wie hierbei gut zu erkennen ist, ermöglicht die Disaggregation der Referenzdaten eine weitaus feineräumigere Analyse. Ebenfalls wird deutlich, dass der Einfluss von administrativen Grenzen deutlich geringer wird. Dies ist vor allem im Stadtgebiet von São Paulo gut ersichtlich. In den Referenzdaten erstreckt sich das Polygon der Stadt weit in den Süden und besitzt als Ganzes einen hohen BIP-Wert. Anhand der Disaggregation wird aber deutlich, dass dieser Bereich deutlich niedrigere BIP-Werte aufweist als das restliche Stadtgebiet. Ein ähnliches Phänomen ist auch im westlichen Bereich von Rio de Janeiro erkennbar, wodurch aufgezeigt wird, welchen Einfluss die Homogenisierung von ökonomischen Werten innerhalb von administrativen Grenzen mit sich bringt. Die Unterteilung der BIP-Werte innerhalb der Karte in die 5 Klassen erfolgte basierend auf natürlichen Unterbrechungen nach Jenks. Dadurch soll die Varianz innerhalb der Klassen minimiert und außerhalb der Klassen maximiert werden (Jenks 1967).

## 4.2 Sentinel-2 Daten

Sentinel-2 ist Teil des europäischen Copernicus-Programms und besteht aus zwei baugleichen Satelliten, die in einem sonnensynchronen Orbit in etwa 786 km Höhe operieren. Die Mission erfasst multispektrale Daten in 13 Bändern, die den sichtbaren, nahinfraroten sowie kurzwelligen infraroten Bereich des elektromagnetischen Spektrums abdecken. Vier dieser Bänder (B2, B3, B4 und B8) besitzen eine räumliche Auflösung von 10 m, sechs weitere Bänder werden mit 20 m und drei atmosphärische Bänder mit 60 m aufgezeichnet. Durch die Kombination aus hoher räumlicher und spektraler Auflösung sowie den frei zugänglichen Daten eignet sich Sentinel-2 besonders für (detaillierte) Analysen im Bereich der Fernerkundung (Esch et al. 2015; Wurm et al. 2019).

Für die in dieser Arbeit verwendeten Sentinel-2 Aufnahmen wurden, wie bereits erwähnt, die vorangehend aufbereiteten Daten benutzt, die vom DLR zur Verfügung gestellt wurden. Diese stammen aus dem Zeitraum von 2019-2021 und decken das gesamte Untersuchungsgebiet ab. Sie beinhalten alle Sentinel-2 Bänder, welche eine räumliche Auflösung von 10 m besitzen und können, aufgrund der bereits stattgefundenen Vorprozessierung, direkt weiterverarbeitet werden.

Bei der weiteren Verarbeitung dieser Daten war vor allem zu beachten, dass in manchen Aufnahmen nicht alle Bänder zur Verfügung stehen, sowie NaN Werte

vorkommen können. Diese sind in weiterer Folge nicht von der gewählten Modellarchitektur verarbeitbar. Aus diesem Grund wurden fehlerhafte Pixel sowie fehlende Bänder mit 0-Werten aufgefüllt. Aufgrund der multimodalen Datenfusion stellt dies jedoch keine Problematik dar, da genügend Inputdaten vorhanden sind, um diese „fehlenden“ Informationen auszugleichen (Li et al. 2022).

Somit dienen diese vier aufbereiteten Bänder als direkten Input für das Modell. Des Weiteren werden aus den Sentinel-2 Daten noch die folgenden Indizes berechnet:

- NDVI
- GNDVI
- NDWI

Diese dienen dazu, weitere Informationen aus den Sentinel-2 Daten zu extrahieren und werden auch in ähnlichen Ansätzen verwendet und beispielsweise mit Nachlichtdaten kombiniert, um das GDP zu ermitteln (Bolivar et al. 2024; Wu et al. 2024).

#### 4.2.1 NDVI

Der Normalized Difference Vegetation Index (NDVI) ist einer der am häufigsten verwendeten Fernerkundungsindizes und dient zur Quantifizierung von Vegetationsbedeckung, -dichte und -vitalität (Rattanakaew 2025). Er wird aus den Reflexionswerten des nahen Infrarots (NIR) und des roten Bandes berechnet, typischerweise nach der folgenden Formel 1:

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

*Formel 1: NDVI (Normalized Difference Vegetation Index)*

Für Sentinel-2 Daten mit 10 m räumlicher Auflösung werden hierfür das NIR-Band (Band 8) und das rote Band (Band 4) verwendet (Tassopoulos et al. 2021).

Der Wertebereich dieses Index liegt normalerweise zwischen -1 bis 1, wobei sehr hohe Werte auf eine gesunde und dichte Vegetation hinweisen und niedrigere Pixelwerte kranke oder fehlende Vegetation abbilden. Der Grund dafür ist, dass Flächen mit (gesunder) Vegetation höher im nahen Infrarotbereich reflektieren als im roten Spektralbereich, wodurch ein positiver NDVI-Wert entsteht. Diese Reflexionsunterschiede entstehen durch das Chlorophyll in den Pflanzen, das rote

Licht stark absorbiert, während es grünes Licht reflektiert und im nahinfraroten Bereich (NIR) besonders stark zurückstrahlt. Brachliegende Flächen hingegen reflektieren in beiden Bereichen ähnlich, wodurch der NDVI in diesen Gebieten gegen 0 geht. Wasserflächen können dann beispielsweise negative NDVI-Werte aufweisen, da hier der rote Spektralbereich höhere Reflexionswerte aufweist als der nahe Infrarotbereich (Rattanakaew 2025).

Basierend auf diesen Reflexionswerten kann mithilfe des NDVI beispielsweise die Gesundheit oder Phänologie von Vegetation beschrieben werden (Tassopoulos et al. 2021). So lassen sich Rückschlüsse auf diverse landwirtschaftliche Anbauflächen ziehen, die einen wesentlichen Anteil an der Wirtschaftsleistung haben (Bolivar et al. 2024). Ebenso dient er als geeignete Abgrenzung zwischen verbautem und unverbautem Gebiet.

#### 4.2.2 GNDVI

Der Green Normalized Difference Vegetation Index (GNDVI) ist eine Variante des NDVI, der anstelle des roten das grüne Band nutzt (Formel 2):

$$GNDVI = \frac{NIR - GREEN}{NIR + GREEN}$$

*Formel 2: GNDVI (Green Normalized Difference Vegetation Index)*

Im Zuge dieser Arbeit wird für den GNDVI das NIR-Band (Band 8) und das grüne Band (Band 3) von Sentinel-2 verwendet (Tassopoulos et al. 2021).

Dieser Index misst die „Grünheit“ der Vegetation und besitzt einen höheren Sättigungspunkt als der NDVI und ist im Vergleich zu diesem auch sensitiver gegenüber verschiedenen Chlorophyllgehalten der Pflanzen. Deshalb ist er auch vor allem für dichte Vegetation geeignet. Der Wertebereich liegt ebenfalls zwischen -1 und 1, wobei positive Werte erneut vegetative Flächen abbilden und Pixelwerte zwischen -1 bis 0 beispielsweise Wasser- oder Brachflächen entsprechen (Rattanakaew 2025).

Der GNDVI eignet sich daher besonders zur Analyse landwirtschaftlicher Flächen mit hoher Vegetationsdichte und stellt somit einen wichtigen ergänzenden Indikator für das Modell dar.

### 4.2.3 NDWI

Der Normalized Difference Water Index (NDWI) dient der Erkennung von Wasserflächen und der Abschätzung von Wassergehalt in der Vegetation. In dieser Arbeit wird die klassische Form nach McFeeters (1996) eingesetzt, die aus den Reflexionswerten des grünen und des NIR-Bandes berechnet wird (Formel 3):

$$NDWI = \frac{GREEN - NIR}{GREEN + NIR}$$

*Formel 3: NDWI (Normalized Difference Water Index)*

Für die Berechnung des NDWI werden somit ebenfalls Band 3 und Band 8 von Sentinel-2 verwendet (Sentinel Hub o.J.).

Der NDWI hebt insbesondere Wasserflächen hervor, da diese eine geringe Reflexion im nahen Infrarotbereich aufweisen und in Kombination mit dem grünen Spektralbereich herausstechen. NDWI-Werte über 0.5 stellen daher typischerweise Wasserflächen dar, während niedrigere Werte meist vegetative oder verbaute Flächen repräsentieren (Sentinel Hub o.J.). Im Kontext dieser Arbeit liefert er somit wertvolle Zusatzinformationen über z.B. Wasserverfügbarkeit und trägt, wie auch in ähnlichen Ansätzen gezeigt wurde (Wu et al. 2024), zur Modellierung des primären Sektors bei.

## 4.3 Land Cover

Wie bereits in Kapitel 3 beschrieben, dienen Landbedeckungsdaten als geeigneter Input für die Ermittlung des BIPs. Der dafür erforderliche Zeitaufwand, um diese Inputdaten selbst zu generieren, ist jedoch nicht zu vernachlässigen. Um speziell für das Untersuchungsgebiet angepasste Klassifikationen zu erstellen, wäre die Entwicklung eines eigenen Deep-Learning-Modells notwendig, damit qualitativ hochwertige Datengrundlagen produziert werden können. Aufgrund des damit verbundenen hohen Zeitaufwandes werden in dieser Arbeit die Land Cover Daten von WorldCover verwendet. Dabei handelt es sich um ein globales Produkt der ESA (European Space Agency), welches 2021 mit einer räumlichen Auflösung von 10 m herausgegeben wird und primär auf Sentinel-2 Daten basiert. Für die Erstellung werden Zeitreihen von Sentinel-2 (optisch) sowie ergänzend Sentinel-1 (Radar) herangezogen, die nach einer Wolkenmaskierung zu 10-Tages-Composites verarbeitet werden. Aus diesen Daten werden spektrale, zeitliche und topographische Merkmale abgeleitet, welche anschließend in ein Klassifikationsmodell einfließen.

Unter Einbezug zusätzlicher Referenzdatensätze, sowie diverser expertenbasierter Regeln, entsteht so ein konsistentes globales Landbedeckungsprodukt (Zanaga et al. 2022).

Im Datensatz werden die folgenden 11 Klassen differenziert (Van De Kerchove et al. 2022):

### Tree cover

Zur Klasse „Tree cover“ zählen alle Flächen mit mindestens 10 % Baumanteil, einschließlich Aufforstungen und Plantagen (z. B. Ölpalme, Oliven). Unterwuchs (Sträucher, Kräuter), Bebauung oder Wasserflächen können vorhanden sein, solange die Baumdeckung dominiert. Auch saisonal oder dauerhaft überflutete Flächen mit Baumanteil sind eingeschlossen, mit Ausnahme von Mangroven.

### Shrubland

„Shrubland“ beinhaltet Flächen mit mindestens 10 % Strauchanteil (unter 5 m Höhe). Unterwuchs aus Kräutern oder Gras ist zulässig, solange Sträucher dominieren. Dazu gehören natürliche Strauchflächen sowie semiaride Gebiete mit Strauchvegetation.

### Grassland

Mit „Grassland“ werden Flächen mit mindestens 10 % Grasanteil beschrieben. Dazu zählen offene Wiesen, Savannen, Steppen und Weideflächen. Sträucher oder einzelne Bäume können vorhanden sein, solange die Flächen durch Gras dominiert werden.

### Cropland

Unter „Cropland“ werden alle landwirtschaftlich genutzten Flächen, zusammengefasst die innerhalb von zwölf Monaten aktiv bewirtschaftet werden (Aussaat, Anbau, Ernte). Darunter fallen einjährige Kulturen, Dauerkulturen und Plantagen, sofern sie überwiegend agrarisch genutzt werden. Gewächshäuser werden der folgenden Kategorie „Built-Up“ zugeordnet.

## Built-up

In der „Built-Up“ Klasse werden alle künstlichen Flächen, die durch menschliche Aktivitäten geprägt sind, wie etwa Wohngebiete, Industrieanlagen oder Verkehrsflächen zusammengefasst. Urbane Grünflächen wie z.B. Parks sind jedoch nicht inkludiert. Ebenso wie Mülldeponien, welche als Brachfläche („Bare/sparse vegetation“) definiert werden.

## Bare/sparse vegetation

Die Klasse „Bare/sparse vegetation“ umfasst Flächen mit weniger als 10 % Vegetationsbedeckung. Dazu gehören beispielsweise Felsen, Geröll, Sandflächen, oder Gebiete mit sehr spärlicher Vegetation.

## Snow and ice

„Snow and Ice“ beinhalten dauerhaft oder saisonal von Schnee oder Gletschern bedeckten Flächen.

## Permanent water bodies


„Permanent Water Bodies“ umfasst alle Süß- und Salzwasserflächen, die mindestens neun Monate im Jahr von Wasser bedeckt sind. Dazu zählen etwa Flüsse, Stauseen oder Meere.

## Herbaceous wetland

„Herbaceous wetland“ enthält alle Flächen, die regelmäßig oder dauerhaft überschwemmt sind und von krautiger Vegetation dominiert werden.

## Mangroves

Unter der Klasse der „Mangroves“ versteht man salztolerante Bäume und Sträucher, die entlang von Küsten, Flussmündungen oder Gezeitenzonen wachsen.

 Moss and lichen

Die letzte Klasse „Moss and lichen“ beinhaltet Flächen, die überwiegend von Moosen und Flechten bedeckt sind, wie etwa in kargen Regionen.

Um diese Datengrundlage noch etwas verständlicher darzulegen, wird diese am Beispiel einer Gridzelle im Stadtgebiet von Porto Alegre in Abbildung 6 dargestellt:

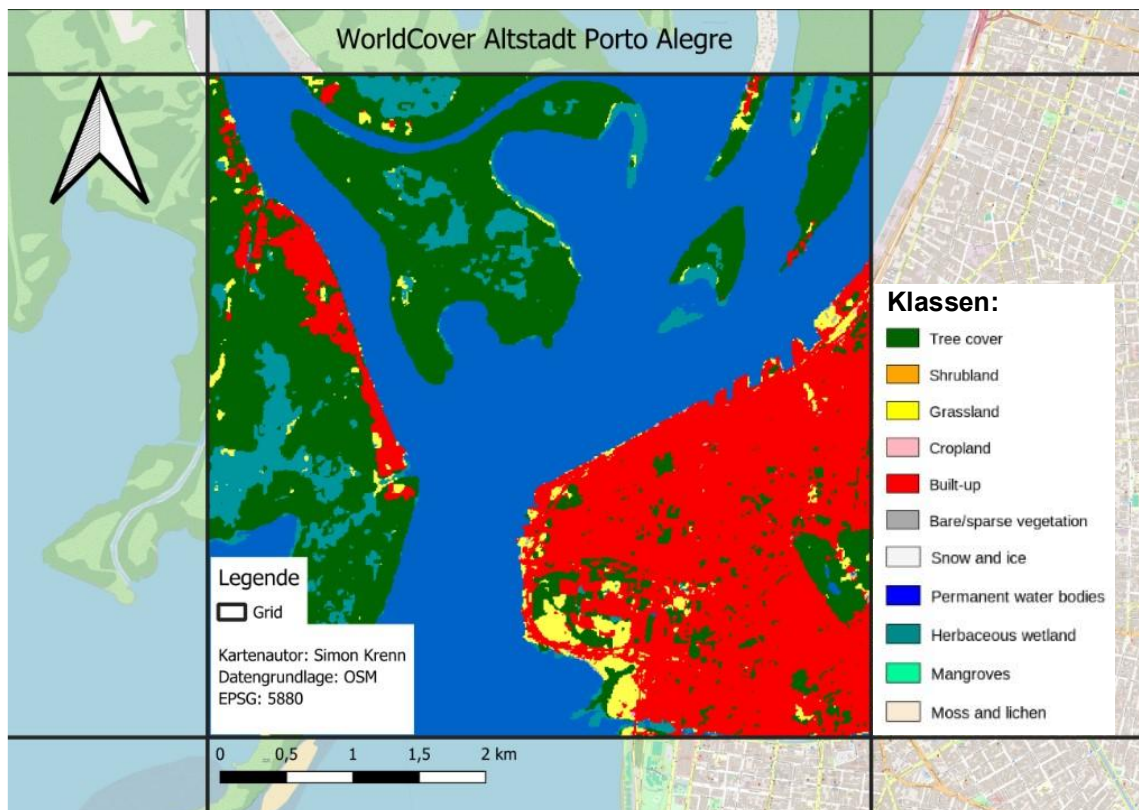


Abb. 6: Die WorldCover Daten als Beispiel innerhalb einer Zelle (Altstadt Porto Alegre). Quelle: Eigener Entwurf.

Dieser Ausschnitt der WorldCover Daten befindet sich im Bereich der Altstadt von Porto Alegre. Man kann hier gut die verschiedenen vorhandenen Landbedeckungsklassen unterscheiden. Die schwarzen Linien zeigen die Grenzen der einzelnen Zellen, womit verdeutlicht werden soll, wie die einzelnen Daten für die jeweilige Zelle aufbereitet werden. Zur weiteren Veranschaulichung ist der linke Bildbereich außerhalb der hervorgehobenen Zelle ebenfalls mit WorldCover Daten hinterlegt, im rechten Bildbereich dient OSM als Kartengrundlage.

Durch die farbliche Kodierung können nun die verschiedenen Klassen der WorldCover Daten gut unterschieden werden. Versiegelte Flächen („Built-Up“) wie z.B. Siedlungsgebiete sind in Rot dargestellt, „Tree Cover“ in Grün, permanente

Wasserflächen („permanent water bodies“) in Blau usw. Die Darstellung hebt die deutliche Trennung zwischen urbanen Strukturen im Osten und wasser- bzw. von Vegetation dominierten Flächen im mittleren und westlichen Bereich des Kartenausschnitts hervor.

Ein Nachteil dieser Daten ist jedoch, dass es sich um ein globales Produkt handelt und folglich die Datenaufbereitung von Seiten der ESA nicht für einen einzelnen Staat, wie in diesem Fall für Brasilien, optimiert werden kann. Dadurch sind gewisse Homogenisierungen zu erwarten. Eine Möglichkeit, die Datenqualität hierbei zu beschreiben, ist die „Overall Accuracy“. Dabei handelt es sich um ein weitverbreitetes Messinstrument, welches die grundlegende Genauigkeit der Klassifikation wiedergibt (Salah 2017). Da die generelle Overall Accuracy der Daten für Südamerika  $77.9 \pm 1.1\%$  beträgt (Van De Kerchove et al. 2022, S. 25) handelt es sich dabei um eine für dieses Projekt ausreichende Genauigkeit, vor allem in Kombination mit der damit verbundenen Zeitersparnis.

#### 4.4 Black Marble Daten

Die verwendeten Nachlichtdaten stammen aus dem NASA Black Marble-Datensatz, der global verfügbare Informationen zur nächtlichen Lichtemission bereitstellt (Román et al. 2018). Die Daten basieren auf Messungen des Visible Infrared Imaging Radiometer Suite (VIIRS). Das VIIRS befindet sich an Bord des Suomi-NPP-Satelliten und verfügt über 22 Spektralbänder von Ultra violett bis zum mittleren Infrarot. Dazu gehört auch das Day-Night-Band (DNB), ein panchromatisches Band im sichtbaren und nahinfraroten Bereich, das auch Nachlichter erfassen kann, was als Grundlage für diese Arbeit verwendet wird und vom DLR zur Verfügung gestellt wurde.

Die Daten besitzen dabei eine räumliche Auflösung von etwa 452x452m und stammen aus dem Jahr 2016. Für die Weiterverarbeitung wurden diese reprojiziert und auf 500m resampled, um sie besser mit den restlichen, auf 10m aufgelösten Daten, fusionieren zu können. Der ursprüngliche Wertebereich der Pixel lag zwischen 0 und 255 (global), nach der Aufbereitung zwischen 0 und 215. Diese Pixelwerte sind dabei relativ zu interpretieren und stellen keine direkt quantifizierbaren Werte realer Lichtemission dar. Höhere Pixelwerte weisen auf vergleichsweise stärkere Lichtintensitäten hin, niedrigere Werte auf geringere Intensitäten.

Die räumliche Auflösung der Daten hatte hierbei einen direkten Einfluss auf die Wahl der Gridsize für den gesamten Workflow. Es wurde ein 5x5km Grid gewählt, um einen Kompromiss zwischen einer möglichst hohen räumlichen Auflösung und einer hinreichenden Repräsentativität der Nachlichtdaten innerhalb jeder Zelle zu finden. Kleinere Gridzellen würden zu Zellen mit nur wenigen Black-Marble-Pixeln führen und nachfolgend die Struktur dieser Daten innerhalb einer Zelle entfernen, während größere Zellen den subregionalen Detailgrad mindern würden. Die gewählte Zellgröße ermöglicht somit eine räumlich differenzierte Erfassung der Wirtschaftsleistung, ohne die Datenqualität wesentlich zu beeinträchtigen.

Um diesen Input nun bestmöglich aufzubereiten, werden sie zusätzlich mithilfe der WorldCover Daten disaggregiert. Hierbei wurden ausschließlich jene Pixelwerte behalten, die laut den Land Cover Daten der Klasse „Built-up“ entsprechen. Dadurch wird sichergestellt, dass Lichtemissionen aus nicht bebauten bzw. nicht versiegelten Flächen, etwa durch temporäre Lichtquellen, Mischpixel, oder Artefakte im Bild, nicht miteinbezogen werden. Die so disaggregierten NTL-Daten erreichen dadurch auch die räumliche Auflösung des WorldCover Datensatzes (10m) und erlauben somit eine deutlich präzisere Zuordnung von Lichtemissionen zu Siedlungsstrukturen. In Abbildung 7 wird der disaggregierte Datensatz, erneut am Beispiel von Porto Alegre, dargestellt. Dabei ist gut zu erkennen, in welchen Stadtbereichen die stärkste Lichtemission stattfindet, aber auch, wie nicht bebaute Flächen mit Lichtemission durch die Disaggregation entfernt wurden.

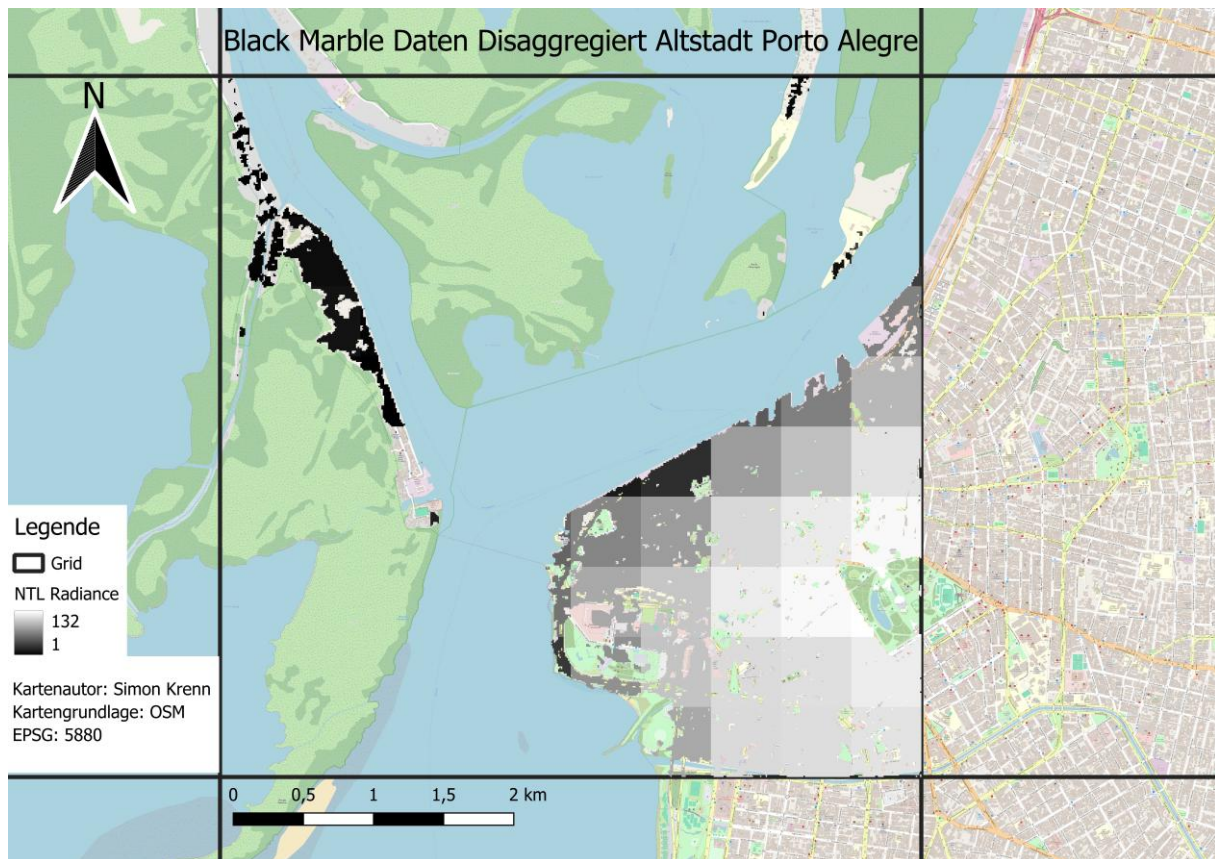


Abb. 7: Die disaggregierten Black Marble Daten als Beispiel innerhalb einer Zelle (Altstadt Porto Alegre). Quelle: Eigener Entwurf.

## 4.5 OpenStreetMap

Als letzte Inputdatenquelle wird OpenStreetMap (OSM) verwendet. Auch diese (oder ähnliche) Daten wurden bereits in der Literaturrecherche beschrieben und erwiesen sich als geeigneter Zusatz zu den bisher genannten Inputs (Wu et al. 2024; Han et al. 2024).

OSM ist ein kooperatives, frei verfügbares Kartenprojekt, das 2004 gestartet wurde und heute zu den bekanntesten Formen freiwillig erstellter Geodaten zählt. Die Plattform ermöglicht es Nutzer:innen weltweit, geografische Objekte wie Straßen, Gebäude oder Landnutzungsflächen zu erfassen und kontinuierlich zu aktualisieren. Dadurch entsteht eine umfangreiche, frei zugängliche Geodatenbank, deren räumliche Abdeckung und Aktualität vielerorts mit amtlichen Daten vergleichbar ist (Vargas-Muñoz et al. 2021; Han et al. 2024). OSM-Daten liegen meist im Polygonformat vor, die mit beschreibenden Tags (Kategorien) versehen werden, um Informationen zu den räumlichen Objekten bereitzustellen. Die Datenqualität kann dabei je nach Region stark variieren, da sie von der Anzahl und dem Wissen der beitragenden Personen abhängt. Dennoch hat sich OSM als wertvolle Grundlage für verschiedenste

Anwendungen in den Geowissenschaften etabliert (Moradi et al. 2022; Camboim et al. 2015).

Im Zuge dieser Arbeit wurden zwei Datenschwerpunkte gewählt. Einerseits vordefinierte POIs (basierend auf den Kategorien von OSM), welche einen direkten oder indirekten Bezug auf das BIP besitzen, andererseits das in OSM implementierte Straßennetzwerk von Brasilien.

Die OSM-Daten, sowohl die POIs als auch das Straßennetzwerk, wurden dabei als je drei unterschiedliche Inputraster aufbereitet:

- Binärraster
- Proximityraster
- Densityraster

Die Daten wurden dabei automatisiert mit einem Pythonskript heruntergeladen und für jede Gridzelle aufbereitet. Es wird zunächst ein 250m Puffer um die jeweilige Zelle gebildet, um Randeffekte bei den Distanz- und Dichteberechnungen zu vermeiden. Die Datenabfrage erfolgt über die Pythonbibliothek „osmnx“, wobei hier doppelte Geometrien entfernt werden und nur für die weitere Prozessierung relevante Attributspalten behalten werden. Die Geometrien werden in das EPSG:5880 projiziert und auf eine Pixelgröße von 10x10m rasterisiert. Das daraus resultierende Raster wird anschließend auf die exakte Zellengeometrie zugeschnitten. Je nach Rastertyp erfolgen unterschiedliche Aufbereitungsschritte:

Der Binärraster ist dabei die einfachste Form und gibt an, ob sich an einer Position/Pixel ein POI bzw. eine Straße befindet oder nicht (siehe Abb. 8). Hierfür werden die via Pythoncode akquirierten OSM-Daten direkt verwendet und mit 0 oder 1 versehen.

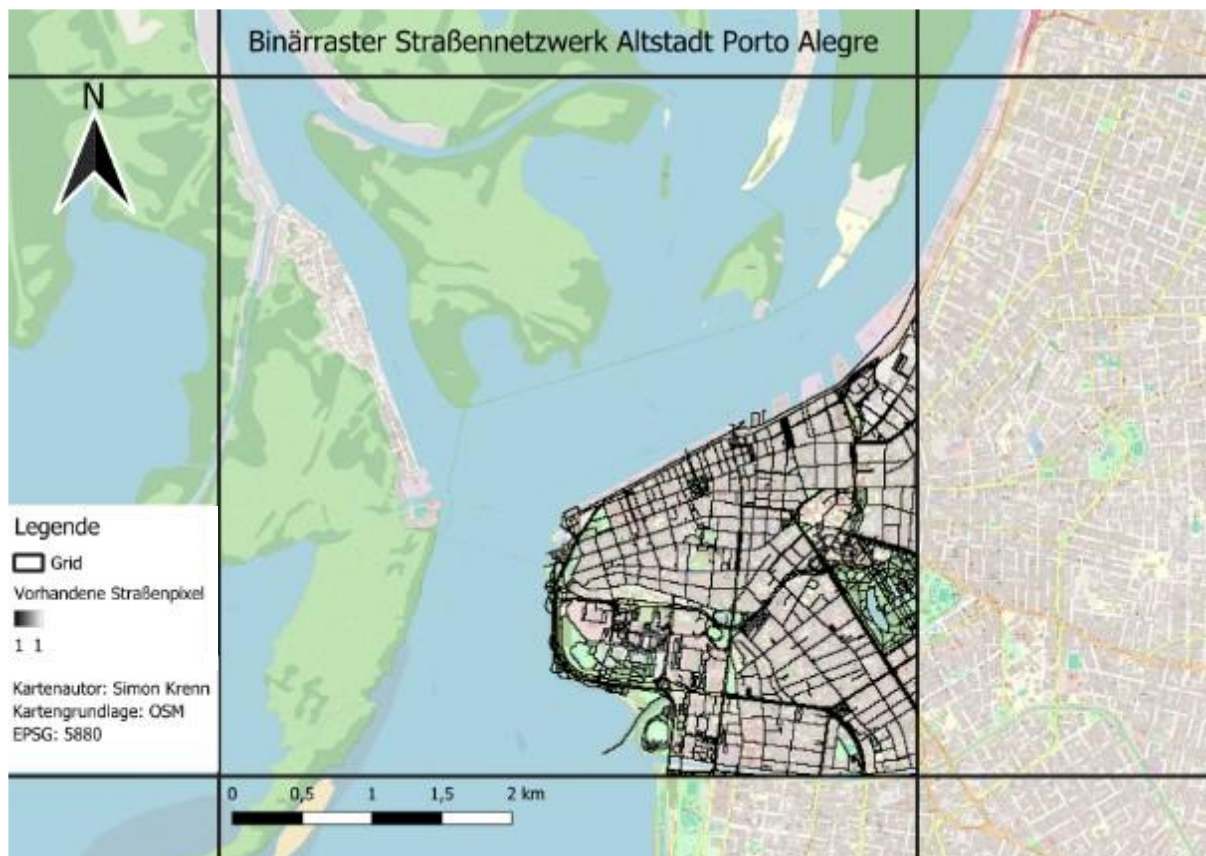


Abb. 8: *Der Binärraster des Straßennetzwerkes der Altstadt von Porto Alegre. Quelle: Eigener Entwurf.*

Beim Proximityraster wird mithilfe von `scipy`, einer weiteren Pythonbibliothek, (mit der Funktion `distance_transform_edt`) eine Erreichbarkeitsanalyse durchgeführt. Ausgangspunkt ist hierfür der bereits beschriebene Binärraster, auf den eine euklidische Distanztransformation angewandt wird. Daraus resultiert für jedes Pixel der Abstand zum nächstgelegenen Zielpixel (POI bzw. Straßenpixel). Abschließend wird noch eine Logarithmustransformation angewandt, um die Ergebnisse zu glätten. Diese Resultate (Abb. 9) dienen als weiteren Input für das entwickelte Modell.

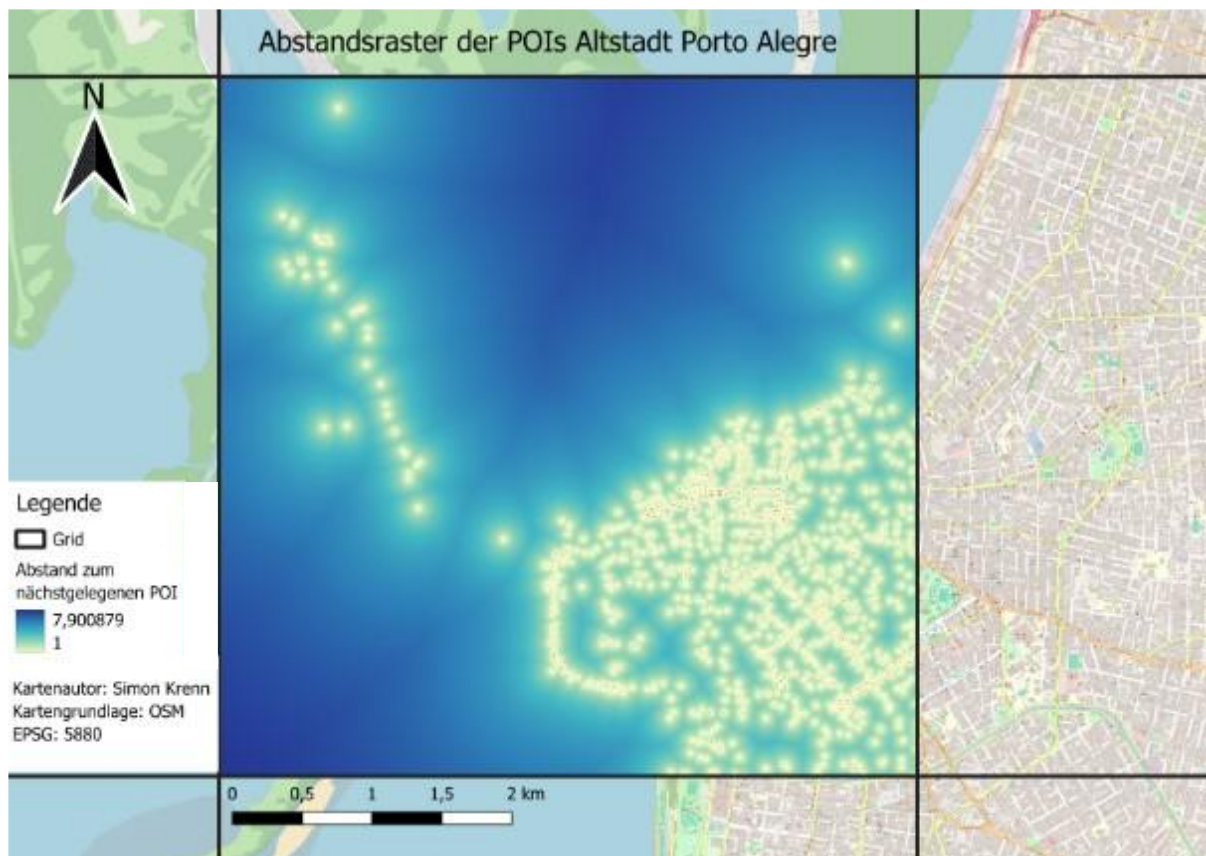


Abb. 9: Der Abstandsraaster der POIs der Altstadt von Porto Alegre. Quelle: Eigener Entwurf.

Beim Dichteraster werden den OSM-Objekten zunächst ordinale Gewichte (0–5) zugewiesen, wobei wirtschaftsnahe Kategorien (z. B. Industrie, Handel, Banken) höher und Kategorien ohne klaren Wirtschaftsbezug niedriger gewichtet werden (Wu et al. 2024). Für Straßendaten erfolgt eine analoge Gewichtung nach highway-Klassen (z.B. werden Autobahnen/Primärstraßen höher und Siedlungsstraßen niedriger bewertet). Die Gewichtungen dienen als generalisierte Näherung und sollen trotz möglicher Unterschiede zwischen Wirtschaftssystemen einen global übertragbaren Workflow ermöglichen. Anschließend werden die gewichteten Geometrien rasterisiert, wobei überlappende Objekte innerhalb eines Pixels additiv zusammengeführt werden. Zur Reduktion lokaler Schwankungen wird das Ergebnis mit einem Gauß-Filter geglättet (`scipy.ndimage.gaussian_filter`,  $\sigma = 2$ ) und anschließend erneut mit einer Logarithmustransformation gedämpft. Die so entstehende Dichtekarte dient als weiterer Input für das entwickelte Modell und wird beispielhaft anhand einer Zelle in Abbildung 10 dargestellt.

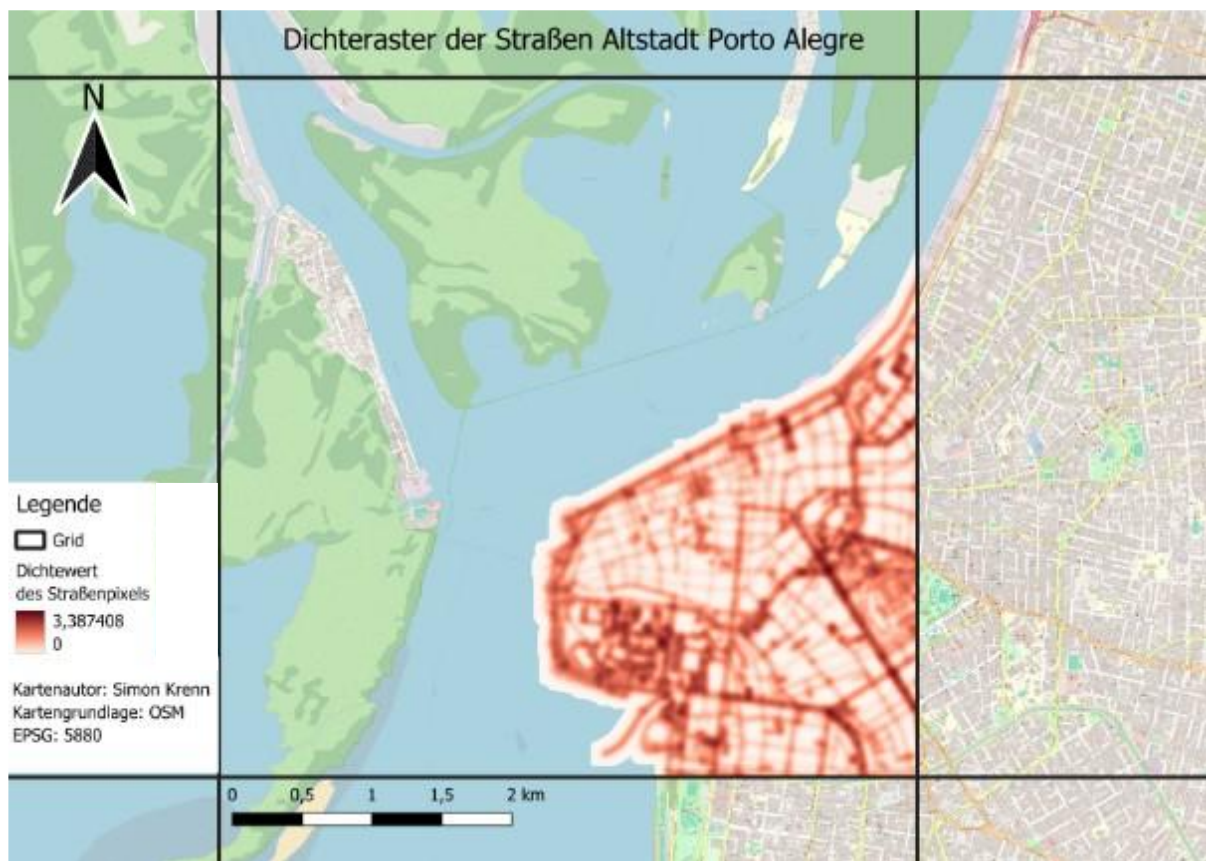


Abb. 10: Der Dichteraster des Straßennetzwerkes der Altstadt von Porto Alegre. Quelle: Eigener Entwurf.

Mithilfe der angewandten Gewichtung und Filterung wird in der Karte deutlich, wo innerhalb der Zelle eine hohe Dichte von ökonomisch relevanten Straßenkategorien vorliegt. Basierend auf diesen Werten lassen sich Rückschlüsse auf potenzielle Verkehrsknotenpunkte sowie stark frequentierte Bereiche ziehen.

## 4.6 Weitere Datenvorbereitung

Nachdem nun alle Inputquellen einzeln aufbereitet wurden, erfolgt im nächsten Schritt eine räumliche Angleichung, um eine pixelgenaue Überlagerung der Datenquellen pro Zelle zu erhalten. Zunächst wird mit einer Bounding Box nach allen passenden Eingabedateien (pro Datengrundlage) für eine Zelle gesucht. Wird eine Zelle nicht zur Gänze durch eine einzelne Eingabedatei abgedeckt, wird aus den möglichen Datenquellen ein Mosaik erstellt. Liegen mehrere Dateien übereinander, z.B. bei größeren Sentinel-2 Aufnahmen, wird der höchste vorliegende Pixelwert verwendet. Damit soll sichergestellt werden, dass wichtige Pixelinformationen nicht durch 0-Werte überlagert werden.

Weitere mögliche Problemquellen wie unterschiedliche Projektionen oder Pixelgrößen wurden bereits behoben, damit nun jedoch die unterschiedlichen Inputdaten exakt aneinander angepasst werden können, wird als Resampling Methode Nearest Neighbor verwendet (Tassopoulos et al. 2021; Thapa et al. 2019). Das bedeutet nun, das anschließend beispielsweise das Pixel in Zeile 6, Spalte 7 der Zelle 403301 im Binärraster exakt dem entsprechenden Pixel in Zeile 6, Spalte 7 des WorldCover-Rasters derselben Zelle entspricht. Dies ist entscheidend, damit die Inputs korrekt vom Modell verarbeitet werden können und bestmögliche Ergebnisse erzielt werden. Ebenso muss jede Zelle dieselbe Anzahl an Inputdaten mit der gleichen Anzahl an Pixeln besitzen, um im weiteren Verlauf keine Fehlermeldungen zu erhalten. Folglich wurden für Zellen, bei denen manche der Datengrundlagen nicht verfügbar waren, wie z.B. OSM-POIs in sehr ländlichen Gebieten ohne Siedlungsstruktur, Dummy-Dateien erstellt. Diese bestehen nur aus 0-Werten und dienen damit als „Platzhalter“, damit das Modell später die Daten problemlos verarbeiten kann. Aufgrund der Datenfusion der verschiedenen Inputs stellen die Dummy-Dateien jedoch kein Problem dar.

Somit ist die Datenvorbereitung abgeschlossen und die Datenquellen können in das Modell integriert werden. Im folgenden Kapitel 5 wird nun genauer darauf eingegangen. Dabei wird der durchgeführte Workflow und die Modellentwicklung detailliert beschrieben.

## 5. Modellentwicklung

Das folgende Kapitel beschäftigt sich mit den wichtigsten Aspekten der Modellentwicklung und beschreibt die dafür notwendigen Schritte. Zunächst wird der gesamte Workflow der Arbeit nähergebracht, damit der Kontext der bisherigen, aber auch noch folgenden Abschnitte besser verstanden werden kann. Danach folgen Unterkapitel bezüglich der Aufteilung des Datensatzes sowie weiterer Normalisierungsschritte. Als nächstes werden diverse Eigenschaften der implementierten Modellarchitektur beschrieben. Für die Merkmalsextraktion der einzelnen Inputdaten wird ResNet-18 verwendet. Darauffolgend werden die getesteten Fusionsmethoden erläutert, womit die extrahierten Informationen kombiniert werden. Abschließend erfolgt eine Beschreibung des MLP-Regressors, womit das Modelltraining und die Ermittlung der Wirtschaftsleistung durchgeführt werden.

### 5.1 Workflow

Um die aufeinander aufbauenden Arbeitsschritte des Projektes besser verstehen zu können, wird in diesem Unterkapitel der gesamte Workflow zusammengefasst. Dieser ist in der folgenden Abbildung 11 dargestellt:

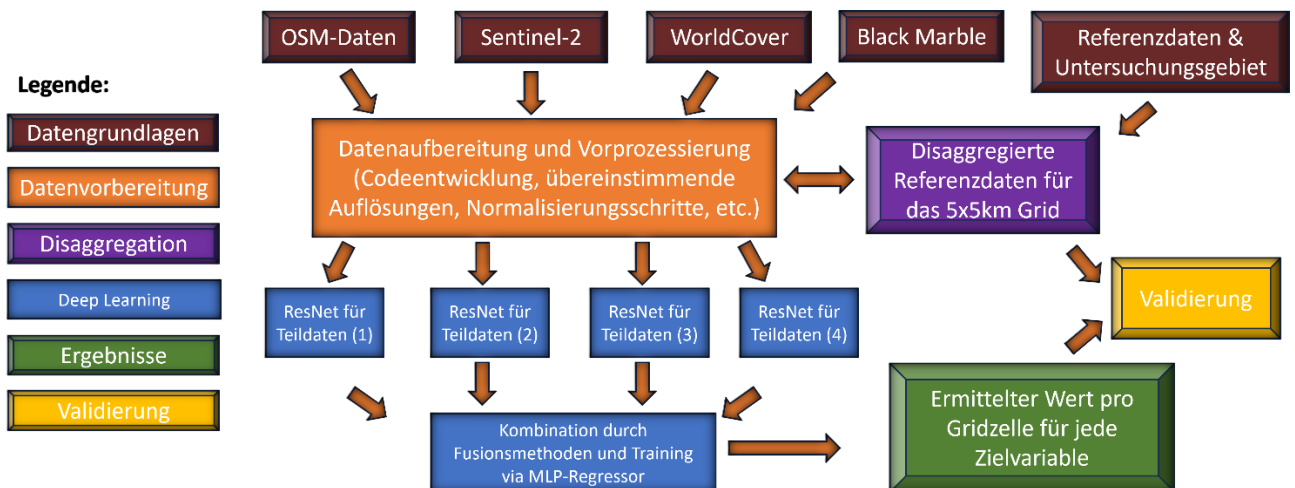


Abb. 11: Der zusammengefasste Workflow der Masterarbeit als Graphik dargestellt. Quelle: Eigener Entwurf.

Zu Beginn werden die bereits erklärten Datengrundlagen (in dunkelrot dargestellt) aufbereitet. Diese werden heruntergeladen und anschließend, teilweise basierend auf der bereits beschriebenen Literatur, weiterverarbeitet. Während der Datenvorbereitung werden alle erforderlichen Schritte durchgeführt (orange dargestellt), sodass die Datensätze anschließend mit dem ResNet-Encoder verarbeitet werden können.

Ebenfalls werden die Referenzdaten wie beschrieben disaggregiert (lila Rahmen). An diesem Punkt befindet sich der derzeitige Stand der Arbeit, am Übergang von der Datenvorbereitung zum Deep Learning. Die Datenaufbereitung, aber auch die Modellentwicklung erfolgte fast ausschließlich in Python (mithilfe diverser Packages wie Rasterio, Scipy, Pytorch, usw.), für visuelle Überprüfungen sowie zur Kartenerstellung wurde QGIS verwendet. Die aufbereiteten Daten werden anschließend im ResNet-Encoder verarbeitet, wodurch Teildatensätze extrahiert werden, die relevante Merkmale aus den einzelnen Datengrundlagen beinhalten. Darauffolgend werden diese mithilfe einer Fusionsmethode kombiniert und zum Schluss erfolgt die Ermittlung der Wirtschaftsleistung. Mithilfe eines MLP-Regressors werden die zuvor erstellten Teildaten sowie die disaggregierten Referenzdaten prozessiert, um als Ergebnis des Workflows einen ermittelten Wert in R\$, für alle Zielvariablen in jeder Zelle des Untersuchungsgebietes zu erhalten.

## 5.2 Datenverteilung und Normalisierung

Um nun jedoch die Datengrundlagen mit der erstellten Modellarchitektur weiterverarbeiten zu können, sind weitere Aufbereitungsschritte notwendig. Zunächst wird der Datensatz in Trainings-, Validierungs- und Testgruppen unterteilt und anschließend normalisiert. Diese Aufteilung ist notwendig, um eine robuste Modellentwicklung und eine praxisnahe Bewertung des Modells zu gewährleisten. In Unterkapitel 5.2.1 wird die Datenverteilung charakterisiert und die verwendeten Aufteilungsvarianten beschrieben. Die abschließenden Normalisierungsschritte werden in Kapitel 5.2.2 erläutert.

### 5.2.1 Datenverteilung

Ein entscheidender Aspekt bei der Entwicklung und Validierung von Deep-Learning-Modellen ist die sinnvolle Aufteilung des verfügbaren Datensatzes. Die Wahl der verwendeten Methode ist dabei ausschlaggebend. Während in vielen Anwendungsfällen einfache Aufteilungen wie 70/30 zwischen Trainings- und Testdaten üblich sind, können derartige Ansätze bei stark unausgeglichenen Datenverteilungen zu Problemen führen (Xu und Goodacre 2018).

Visualisiert man die Wertaufteilung dieses Datensatzes anhand der BIP-Werte pro Zelle, erhält man folgendes Verteilungsmuster (Abb.12):

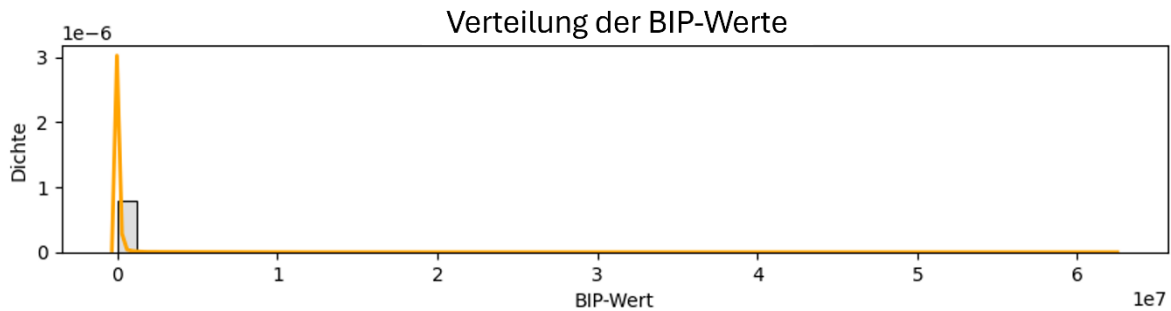


Abb. 12: *Das Verteilungsmuster der BIP-Werte des gesamten Datensatzes. Quelle: Eigener Entwurf.*

Hierbei ist gut in Orange zu erkennen, dass die Verteilung der BIP-Werte eine starke Asymmetrie aufweist. Dies soll durch das ebenfalls erstellte Histogramm (in grau) verdeutlicht werden, welches nur im Bereich der niedrigen BIP-Werte sichtbar wird. Das vorliegende Muster entspricht dabei einer typischen „Long-Tail-Distribution“. Bei dieser Art von Verteilung handelt es sich um ein statistisches Muster, bei dem eine große Anzahl kleiner Werte und nur wenige, aber sehr große Werte das Gesamtbild prägen (He und Garcia 2009).

Diese Problematik hat direkte Auswirkungen auf die Leistungsfähigkeit von Modellen. Werden diese ausnahmslos mit den kleinen Werten trainiert, tendieren sie dazu, ihre Vorhersagen an diesen häufigen Mustern auszurichten (Overfitting), während sie aber Extremwerte unzureichend repräsentieren können (Liu et al. 2019). Um diese Effekte zu vermindern, können verschiedene Strategien verwendet werden. Beispielsweise kann mithilfe von Data Rebalancing (Oversampling seltener Werte oder Untersampling häufiger Werte) die Datenstruktur ausgeglichen werden oder durch diverse Sampling-Methoden kann ein gleichmäßigeres Verteilungsmuster erzielt werden (Kang et al. 2020).

Im vorliegenden Projekt wurde der Ansatz gewählt, das Problem bereits bei der Datenaufteilung anzugehen. Folglich soll der gesamte Wertebereich des Datensatzes möglichst vollständig in allen 4 Teilgruppen abgedeckt sein, damit die zuvor beschriebenen Probleme vermindert werden können. Im Zuge der Arbeit wurden hierfür zwei Methoden getestet. Eine auf k-Fold Cross-Validation basierende Methode sowie ein zyklischer Ansatz, bei dem der Datensatz gleichmäßig aufgeteilt wird. Als Zielvariable bzw. Wertebereich wird hierfür das BIP verwendet.

Bei der getesteten Variante der k-Fold Cross-Validation wird der Datensatz zunächst in k-Teilmengen (Folds) unterteilt. Dabei soll jeder Fold die Verteilung der Zielvariable möglichst gut abdecken (Kohavi 1995). Bei diskreten Klassen kann dies bedenkenlos verwendet werden, kontinuierliche Wertebereiche, wie in diesem Fall, müssen zuvor jedoch in einzelne Kategorien unterteilt werden. Hierfür werden im vorliegenden Projekt die BIP-Werte in gleich große Quantile aufgeteilt, um eine pseudo-stratifizierte Verteilung zu erreichen. Es wurden 20 Quantile verwendet, da dies zur stabilsten Verteilungsstruktur innerhalb der Folds führte. Dabei gibt es jedoch auch Einschränkungen. Extremwerte können aufgrund der Stratifizierung nicht in jeder Teilgruppe ausreichend repräsentiert sein, aber auch die generelle Spannweite der Folds kann variieren. In dieser Arbeit wird ein Deep-Learning-Modell trainiert, das jedoch den gesamten Wertebereich kennen muss, um zuverlässige Ergebnisse zu liefern. Deshalb wird zusätzlich eine zyklische Verteilung getestet.

Der Ablauf dieser Variante ist dabei wie folgt: Die Daten werden zunächst anhand der BIP-Werte sortiert und anschließend werden sie in zyklischer Reihenfolge den vier Teilmengen zugewiesen. Beispielsweise wird der höchste BIP-Wert dem ersten Fold zugewiesen, der zweithöchste dem zweiten Fold usw., bis der gesamte Datensatz in vier verschiedene Teilmengen aufgeteilt ist. Dadurch wird sichergestellt, dass jeder Fold den gesamten Wertebereich abdeckt und eine möglichst gleichmäßige Verteilung erzielt wird.

In der folgenden Abbildung 13 wird nun das Verteilungsmuster der beiden Varianten gegenübergestellt:

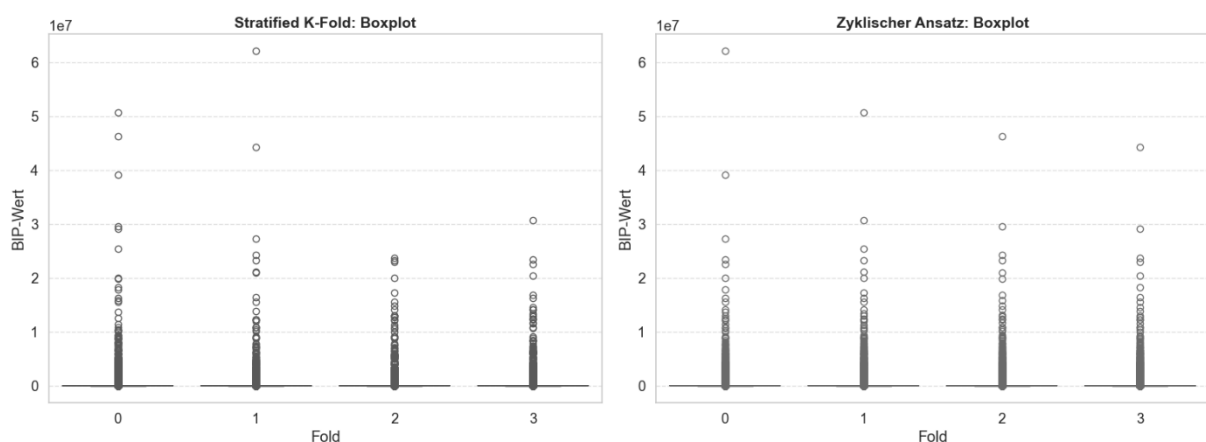


Abb. 13: Das Verteilungsmuster der getesteten Sampling-Methoden der BIP-Werte als Boxplots dargestellt.  
Quelle: Eigener Entwurf.

Darin ist gut zu erkennen, dass mithilfe der zyklischen Verteilung der gesamte Wertebereich bestmöglich auf alle Folds aufgeteilt wird. Dies ist jedoch bei der Stratified k-Fold Variante nicht in diesem Ausmaß gegeben. Beispielsweise befinden sich in Fold zwei keine der höchsten BIP-Werte, wodurch diese Teilmenge nicht vollständig repräsentativ ist, was zu Verzerrungen führen kann.

Alles in allem stellt der zyklische Ansatz eine praxisnahe Lösung dar, um die Datensätze bestmöglich für das Modell aufzubereiten und denkbare Probleme wie Overfitting zu mindern, weshalb diese Variante für den weiteren Verlauf der Arbeit gewählt wurde.

### 5.2.2 Datennormalisierung und -aufbereitung für das Modell

Nachdem der Datensatz nun in vier Folds unterteilt wurde, müssen diese noch für die Modelle aufbereitet werden. Zunächst werden die Zielvariablen vor der Modellierung per Quadratwurzel-Transformation normalisiert, um den Einfluss von Ausreißern zu reduzieren und die Long-Tail Distribution abzuschwächen. Die Auswirkungen dieser Transformation auf die Verteilung ist in Abbildung 14 dargestellt.

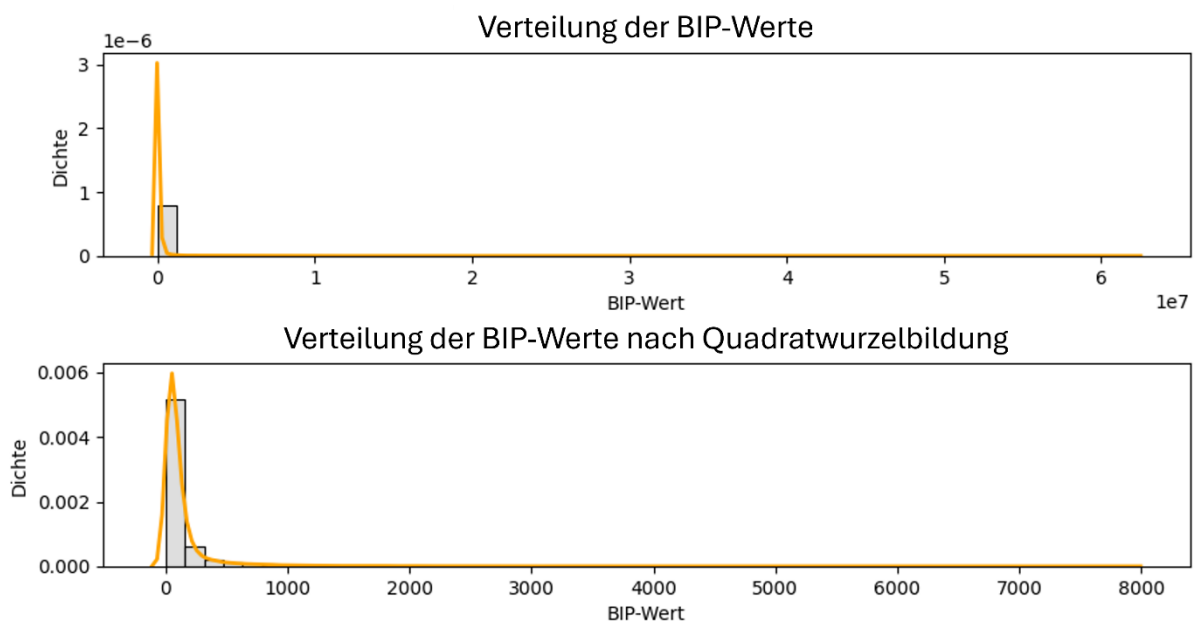


Abb. 14: Die Verteilung der BIP-Werte vor und nach der Quadratwurzel-Transformation vergleichend gegenübergestellt. Quelle: Eigener Entwurf.

Die Visualisierung zeigt, dass die Quadratwurzel-Transformation die große Spannweite des Wertebereiches deutlich reduziert und die zuvor stark ausgeprägte Long-Tail Distribution sichtbar abgeschwächt wird.

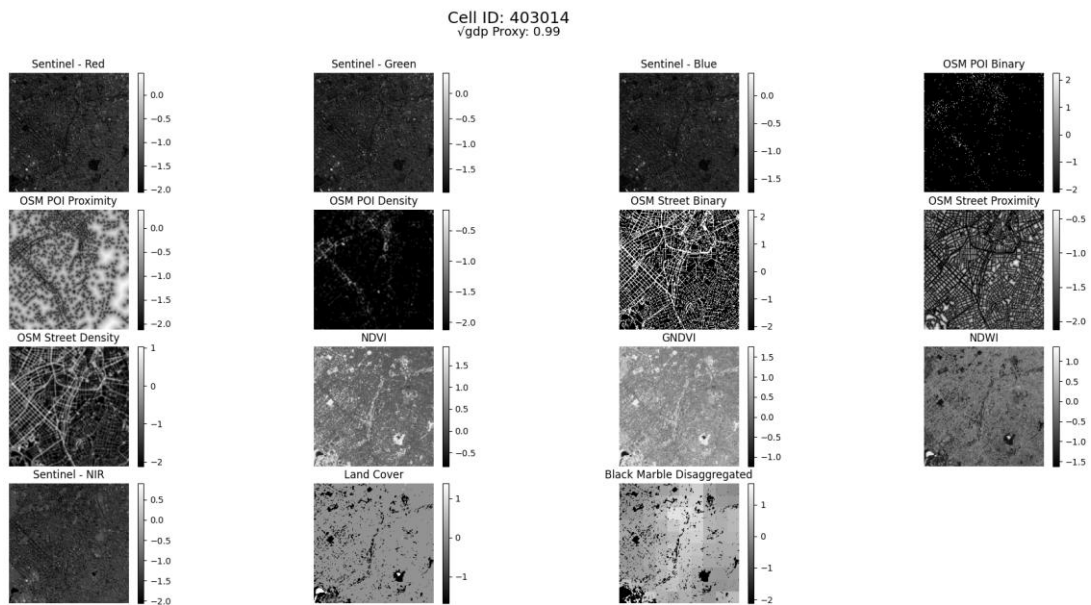
Anschließend werden die Werte auf den Bereich von 0-1 beschränkt, indem alle Werte durch den höchsten Wert pro Zielvariable geteilt werden, um dieser Problematik weiter entgegenzuwirken. Das bedeutet, dass z.B. beim BIP der höchste Wert nach der Wurzeltransformation knapp unter 8000 liegt, (gut in Abb. 14 ersichtlich) und folglich alle Werte durch 8000 dividiert werden. Danach liegen die höchsten BIP-Werte nahe eins und die niedrigsten gehen gegen null. Dasselbe Prinzip wird in weiterer Folge auch für die restlichen Zielvariablen angewandt.

Ebenfalls werden alle Daten auf eine einheitliche Auflösung von  $224 \times 224$  Pixel gebracht, um die Gewichte des Transfer Learning (basierend auf ImageNet, welches mit  $224 \times 224$  Pixeln vortrainiert wurde) optimal nutzen zu können. Die einzelnen Inputs werden dann als Eingabestapeln ("Stacks") zusammengefügt, damit sie vom Modell verarbeitet werden können. Jeder dieser Stacks muss aus 3 Inputquellen bestehen, weshalb folgende Stapel erstellt wurden:

- RGB-Bänder von Sentinel-2
- Indizes-Stack, bestehend aus NDVI, GNDVI und NDWI
- POI-Stack, basierend auf den einzelnen POI-Rastern der OSM-Daten
- Straßennetz-Stack mit den einzelnen Straßennetzrastern der OSM-Daten
- Stack, bestehend aus Nachlichtdaten, Landbedeckungsdaten und dem NIR-Kanal von Sentinel-2

Diese Stacks wurden dann noch weiter unter Bezugnahme auf ImageNet-Statistiken normalisiert, um die Nutzung und den Lerneffekt durch das vortrainierte Modell weiter zu verbessern.

Wie nun die finalen Inputdaten aufgebaut sind, soll durch Abbildung 15 verdeutlicht werden:



*Abb. 15: Eine beispielhafte Darstellung der finalen Inputdaten vor der Merkmalsextraktion anhand einer Zelle.  
Quelle: Eigener Entwurf*

Hierbei ist exemplarisch die Zelle mit der ID 403014 dargestellt, die mit einem finalen BIP-Wert von 0,99 den höchsten Wert dieser Zielvariable im gesamten Datensatz aufweist. Die einzelnen Ausschnitte zeigen die unterschiedlichen Datenquellen, die in weiterer Folge zu den Stapeln zusammengeführt werden und in die Modellierung einfließen. Dazu zählen, wie bereits erwähnt, die multispektralen Sentinel-2-Bänder (Rot, Grün, Blau, NIR), die daraus abgeleiteten Indizes (NDVI, GNDVI, NDWI) sowie die Informationen aus den OSM-Daten für die POIs und das Straßennetzwerk. Ergänzt werden diese noch mit den Landbedeckungs- und Nachlichtdaten.

Der nach der finalen Normalisierung sichtbare Wertebereich, der im ersten Augenblick etwas ungewöhnlich für Fernerkundungs- und Geodaten wirkt, spiegelt die Anpassung der Daten an die ImageNet-Statistiken wider. So besitzt beispielsweise das rote Sentinel-2 Band Werte zwischen ca. -2 und 0,5. Die Abbildungen verdeutlichen jedoch, dass die Struktur der einzelnen Datensätze auch nach dieser Anpassung erhalten bleibt.

Ebenfalls wird in den dargestellten Inputdaten deutlich, weshalb diese Zelle den höchsten BIP-Wert aufweist. Es handelt sich um ein dicht besiedeltes Stadtgebiet mit einem ausgeprägten Straßennetz, einer hohen POI-Dichte, starker Nachlichtemission und folglich niedrige NDVI-Werte. Konkret umfasst die visualisierte Zelle einen Teil des Stadtgebietes von São Paulo.

Somit sind die Inputs vollständig für das Modell aufbereitet und können von diesem verarbeitet werden. Dabei wird zunächst mithilfe eines Encoders die Merkmalsextraktion eines jeden Inputstacks durchgeführt, worauf im folgenden Kapitel genauer eingegangen wird.

### 5.3 ResNet-18 Backbone als Encoder

Jeder der prozessierten Stapel wird nun durch einen Encoder verarbeitet. Dieser wird dazu verwendet, die Stacks mithilfe von Transfer Learning, also einem bereits vortrainierten Modell, in kompakte Merkmalsrepräsentationen umzuwandeln. Anstatt Klassen oder ähnliches vorherzusagen, liefert der Encoder pro Stack einen Feature-Vektor, der anschließend fusioniert und weiterverarbeitet werden kann. Im Zuge dieser Arbeit wird ein ResNet-18-Backbone als Encoder verwendet. Der Begriff Backbone beschreibt dabei das Kernnetzwerk eines Modells, welches zur Extraktion der Merkmale verwendet wird.

Klassische CNN-Architekturen stoßen mit zunehmender Tiefe an Grenzen, da sogenannte verschwindende oder explodierende Gradienten das Training sehr tiefer Netze erschweren. Residual Networks (ResNets) nach He et al. (2016) adressieren dieses Problem durch die Einführung von Skip- oder Shortcut-Connections. Diese ermöglichen es, die Ausgabe einer früheren Schicht unverändert an eine spätere weiterzugeben, sodass Informationen und Gradienten ohne Verlust durch das Netzwerk fließen können. Anstatt dass jede Schicht eine vollständig neue Transformation des Eingangssignals lernt, wird der Eingang unverändert an die Ausgabe der Schicht weitergeleitet und addiert. Die Schicht lernt somit nur die Veränderung, also die „Residuen“, die notwendig sind, um das gewünschte Ergebnis zu erreichen. Dieser Ansatz vereinfacht das Lernen, da er es den Schichten erlaubt, die Identität abzubilden, falls keine Veränderung nötig ist. Dadurch können sehr tiefe Netze trainiert werden, ohne dass ihre Genauigkeit leidet.

ResNet-18, die in dieser Arbeit eingesetzte Variante, umfasst insgesamt 18 gewichtete Schichten, die zu acht Residual-Blöcken zusammengefasst sind. Es zählt damit zu den leichteren Modellen der ResNet-Familie, bietet jedoch ausreichend Kapazität, um komplexe Muster in den verwendeten Datenquellen zu erfassen (He et al. 2016). Innerhalb dieser Arbeit wird die PyTorch-Implementierung von ResNet-18 aus dem torchvision-Modul verwendet. Diese wurde bereits auf dem ImageNet-Datensatz

vortrainiert (Transfer Learning), bei dem bereits gelernte Bildmerkmale wie z.B. Kanten für eine neue Aufgabe wiederverwendet werden. Der finale Klassifikationslayer von ResNet, der standardmäßig für ImageNet-Klassen vorgesehen ist, wird in diesem Fall jedoch entfernt. Dadurch gibt das Netzwerk nicht mehr eine Klassenzuordnung, sondern einen 512-dimensionalen Feature-Vektor aus, der im weiteren Verlauf als Eingabe für die multimodale Datenfusion dient. ResNet wird dabei parallel genutzt, also die verschiedenen Inputstacks werden separat verarbeitet, wodurch für jede Instanz ein eigener Feature-Vektor erstellt wird. Diese können anschließend kombiniert werden, wofür unterschiedliche Fusionsmethoden verwendet werden können.

## 5.4 Fusionsmethoden

Nachdem der Encoder die Merkmalsextraktion durchgeführt hat, müssen diese Merkmale anschließend sinnvoll kombiniert werden. Die vorliegende Arbeit untersucht dabei vier verschiedene Strategien zur Merkmalsfusion, die jeweils unterschiedliche Stärken und Schwächen besitzen. Die Fusion dient dazu, die komplementären Informationen der verschiedenen Datenquellen passend zu kombinieren, damit der abschließende Regressor bestmögliche Resultate erzielen kann.

### 5.4.1 Concatenation

Die einfachste Form der Fusion ist die Concatenation, bei der die einzelnen Merkmalsvektoren aneinandergehängt werden. Dadurch entsteht ein großer, kombinierter Vektor, der die Informationen aller Inputs enthält (Chango et al. 2022). Dieser Vektor wird anschließend vom Regressor weiterverarbeitet, um die verwendete Zielvariable, beispielsweise das BIP, zu ermitteln. Diese Methode macht keinerlei Annahmen über Zusammenhänge zwischen den einzelnen Inputs, sondern überlässt es dem Regressor, diese selbst zu lernen. In dieser Arbeit erfolgt die Fusion mit Concatenation durch eine einfache `torch.cat`-Operation, wodurch der Vektor eine Länge von 512 Dimensionen mal der Anzahl an Stacks besitzt.

Der Vorteil dieser Variante ist, dass alle Datenquellen gleich zur Vorhersage beitragen und einfach kombiniert werden können.

### 5.4.2 Attention Pooling

Eine weitere Methode ist das Attention Pooling. Anstatt alle Inputs gleich zu behandeln, wird für jede Eingabe ein Aufmerksamkeitsgewicht berechnet, das angibt, wie wichtig die jeweilige Grundlage für die aktuelle Vorhersage ist. Hierfür wird zunächst ein Query-Vektor gebildet, indem der Mittelwert über alle Modalitätsrepräsentationen berechnet wird. Dieser wird anschließend mittels einer Multi-Head-Attention mit den Modalitätsvektoren verknüpft. Das resultierende gewichtete Mittel hebt stärker gewichtete Modalitäten hervor, während weniger relevante Informationen abgeschwächt werden (Liu et al. 2018). Der daraus entstehende Fusionsvektor besitzt eine Länge von 512 Dimensionen.

Allerdings verursacht Attention Pooling durch die zugrunde liegende Multi-Head-Attention einen deutlich höheren Rechenaufwand als klassische (Pooling-) Operationen. Zudem kann der Nutzen dieser Methode vor allem bei weniger tiefen Architekturen begrenzt sein kann (Chen et al. 2022).

### 5.4.3 Graph-based Fusion

Bei der graphbasierten Fusion werden die Modalitäten als Knoten eines Graphen interpretiert, zwischen denen Beziehungen modelliert werden können. In der hier verwendeten Implementierung geschieht dies mithilfe eines kleinen mehrschichtigen MLP, das auf die gestapelten Merkmalsvektoren angewendet wird. Diese werden anschließend gemittelt, sodass ein gemeinsamer Fusionsvektor entsteht.

Dieser Ansatz eignet sich besonders dann, wenn man annimmt, dass zwischen den Modalitäten nicht nur additive, sondern komplexere nichtlineare Abhängigkeiten bestehen (D'Souza et al. 2024). Auch hier besitzt der Ergebnisvektor 512 Dimensionen. Ein Nachteil dieser Fusionsmethode ist jedoch, dass sie stark von der Qualität und der Vollständigkeit aller Modalitäten abhängig ist. Besitzen diese eine geringe Qualität oder fehlen Daten, kann dies die Resultate dieser Methode verfälschen (D'Souza et al. 2024; Zhang et al. 2024).

### 5.4.4 Mixture of Experts

Die Mixture of Experts-Methode berechnet für jede Eingabe einen Gate-Wert, der angibt, wie stark die jeweilige Modalität in die Fusion einfließen soll (Liu et al. 2025). Alle Merkmalsvektoren werden zunächst zu einem langen Vektor zusammengefügt, der durch ein kleines neuronales Netz, das sogenannte Gating-Netzwerk, geleitet wird.

Dieses liefert für jede Modalität ein Gewicht, das mithilfe einer Softmax-Funktion normalisiert wird. Anschließend werden die Merkmalsvektoren gemäß diesen Gewichten aufsummiert. Der finale Ergebnisvektor besitzt auch bei dieser Methode wieder 512 Dimensionen.

Ein Vorteil dieser Methode besteht darin, dass diese Modelle nur die jeweils relevanten Experten oder Modalitäten aktivieren und dadurch ihre Kapazität effizient nutzen (vgl. Shazeer et al. 2017). Ein Nachteil ist jedoch, dass sie häufig zu unbalancierter Expertennutzung neigen, wodurch einzelne Experten überbeansprucht und andere kaum trainiert werden (Chen et al. 2023).

## 5.5 MLP-Regressor

Nachdem nun die Inputdaten mithilfe des ResNet-Encoders verarbeitet und mit einer Fusionsmethode kombiniert wurden, erfolgt abschließend die Ermittlung der Wirtschaftsleistung. Dies geschieht durch die Verwendung eines MLP-Regressors. Als einen Regressor bezeichnet man ein neuronales Netz, welches trainiert wird, um kontinuierliche (numerische) Werte vorherzusagen. In der Ausgabeschicht wird dabei im Vergleich zu CNNs keine Softmax-Funktion verwendet, sondern es wird ein reeller Wert ausgegeben. Ebenfalls unterscheiden sie sich noch durch die sogenannte Loss-Funktion, bei der meist Fehlermaße wie z.B. der MSE (Mean Squared Error) verwendet werden. Diese gibt den Unterschied zwischen den vorhergesagten Werten des Modells und den Referenzwerten an und dient dem Modell als Grundlage, um seine Gewichte während des Trainings anzupassen (Goodfellow et al. 2016).

Der in dieser Arbeit eingesetzte Regressor besteht aus zwei vollständig verbundenen Schichten. Zunächst wird der fusionierte Feature-Vektor über eine lineare Transformation auf 128 Dimensionen abgebildet. Im Anschluss wird eine ReLU-Aktivierungsfunktion angewendet, um nichtlineare Abhängigkeiten zwischen den extrahierten Merkmalen zu erfassen (Nair und Hinton, 2010). Die zweite lineare Schicht reduziert diesen Zwischenvektor schließlich auf einen einzelnen Skalar, der die vorhergesagte Zielvariable (z. B. das BIP) repräsentiert.

Die Modellarchitektur wurde dabei mit dem Deep-Learning-Framework PyTorch entwickelt und unter Einsatz von PyTorch Lightning strukturiert, trainiert und validiert. Durch die Aufteilung des Datensatzes in vier Folds werden zunächst zwei dieser Teilmengen für das Training und eine für die Validierung des Modells verwendet. Das

Training verläuft dabei epochenweise. Eine Epoche entspricht dabei einem vollständigen Durchlauf durch diese Folds. Pro Lauf sind bis zu 50 Epochen vorgesehen. Nach jeder Trainingsepoche erfolgt sofort eine Validierung, bei der der Validierungsverlust berechnet wird. Die dabei erlernten Gewichte werden jedoch nicht vor einer neuen Epoche zurückgesetzt, sondern basierend auf diesen wird das Training mit einer neuen Epoche fortgesetzt und diese folglich optimiert (die Gewichte bleiben also weiter trainierbar). Zur Optimierung der Netzgewichte wird dabei der Adam-Optimierer mit einer Lernrate von  $1 \times 10^{-4}$  eingesetzt. Das beste Modell (niedrigster Validierungsverlust) wird abschließend abgespeichert. Verbessert sich der Validierungsverlust des Modells jedoch über 15 aufeinanderfolgende Epochen nicht um mindestens 0.1, wird das Training frühzeitig gestoppt.

Anschließend wird das beste Modell mithilfe des letzten, ungenutzten Folds getestet (sogenannter „Real-World“-Test). Auf Basis der gelernten Gewichte ermittelt das Modell nun für bisher unbekannte Daten eine Zielvariable pro Zelle. Die daraus resultierenden Werte werden den Referenzdaten gegenübergestellt. Hierfür werden RMSE und  $R^2$ , sowohl für die normalisierten als auch für die rücktransformierten Werte berechnet. Die Rücktransformation erfolgt dabei Gegengleich zur Normalisierung. Zunächst werden die Werte, z.B. beim BIP mit 8000 multipliziert, um den Wertebereich von 0-1 zu verlassen und abschließend werden die Ergebnisse quadriert. Somit befindet man sich wieder in der Wertespanne der ursprünglichen Referenzwerte. Die hiermit erzielten Resultate werden tabellarisch in einer Textdatei gespeichert und in Form eines Regressionsplots visualisiert. Die Plots zeigen die ermittelten Werte gegenüber den echten Werten mit Idealgerade sowie RMSE- und  $R^2$ -Werten. Diese Ergebnisdarstellungen werden in Kapitel 6 genauer erläutert.

Für die Reproduzierbarkeit der Methode wird ein Global-Seed gesetzt (42) und der Trainer läuft deterministisch, wodurch Zufallsparameter kontrolliert werden, sodass bei gleichen Einstellungen identische Ergebnisse entstehen. Innerhalb des Trainings wird ebenfalls die Reihenfolge des Datensatzes gemischt, um Korrelationen in der Abfolge zu vermeiden und die Generalisierung zu verbessern. In der Validierung und dem Real-World-Test bleibt die Reihenfolge aber gleich, damit die Auswertungen vergleichbar und stabil bleiben. Ein gesamter Durchlauf, vom Trainingsbeginn bis zum finalen Regressionsplot, dauert ca. einen Tag und es wird dabei immer nur eine Zielvariable ermittelt.

Abschließend lässt sich die gesamte Modellstruktur nun wie folgt zusammenfassen:

Der Encoder (ResNet-18) extrahiert Merkmale aus den Datengrundlagen, die verwendete Fusionsmethode kombiniert die verarbeiteten multimodalen Daten und der MLP-Regressor führt, basierend auf diesem Input, den Trainingsprozess sowie die Ermittlung des Zielwertes durch. Aufgrund der Trennung der einzelnen Schritte ist es möglich, die Fusionsstrategie oder den Regressor anzupassen und experimentell zu vergleichen, ohne Änderungen an den Encodern durchführen zu müssen. Die Resultate des Modells werden nun im folgenden Kapitel 6 dargelegt und in Kapitel 7 interpretiert.

## 6. Ergebnisse

In diesem Kapitel werden nun die erzielten Ergebnisse präsentiert. Zunächst werden ausgewählte Regressionsplots dargestellt und anschließend werden die ermittelten Zielwerte zu einem Datensatz zusammengefasst und den Referenzwerten gegenübergestellt. Zum Schluss werden ausgewählte Beispiele nähergebracht, die die Grenzen und potenziellen Schwachpunkte der Methodik veranschaulichen.

Damit die verschiedenen Resultate sinnvoll miteinander verglichen werden können, benötigt man diverse Metriken, damit die Qualität der Ergebnisse verifiziert werden kann. Im Zuge dieser Arbeit wurden hierfür vor allem der RMSE und  $R^2$  verwendet. Die Regressionsplots selbst bestehen aus den sogenannten „Predictions“, den innerhalb der Plots dargestellten Werte, die eine Zelle des Testdatensatzes repräsentieren. Dabei befindet sich der ermittelte Wert auf der Y-Achse und der Realwert auf der X-Achse. Die orange Linie spiegelt die Idealgerade wider und darunter sind in der Legende der RMSE- und  $R^2$  Wert angegeben.

Im folgenden Unterkapitel werden nun die Ergebnisse der einzelnen Fusionsmethoden präsentiert. Um diesen Teil der Arbeit jedoch im Rahmen zu halten, werden nur die Ergebnisse der BIP-Ermittlung genauer beschrieben. Bei den Resultaten der restlichen Zielwerte werden nur die besten Modelle erläutert, um dieses Kapitel nicht unnötig aufzublasen.

## 6.1 Resultate der Fusionsmethoden

### 6.1.1 BIP

Für jeden Wirtschaftswert können die ermittelten Ergebnisse in zwei Kategorien unterteilt werden. Die normalisierten Werte sowie die denormalisierten Werte im ursprünglichen Wertebereich. Abbildung 16 zeigt zunächst die Resultate der normalisierten BIP-Werte für die vier betrachteten Fusionsmethoden.

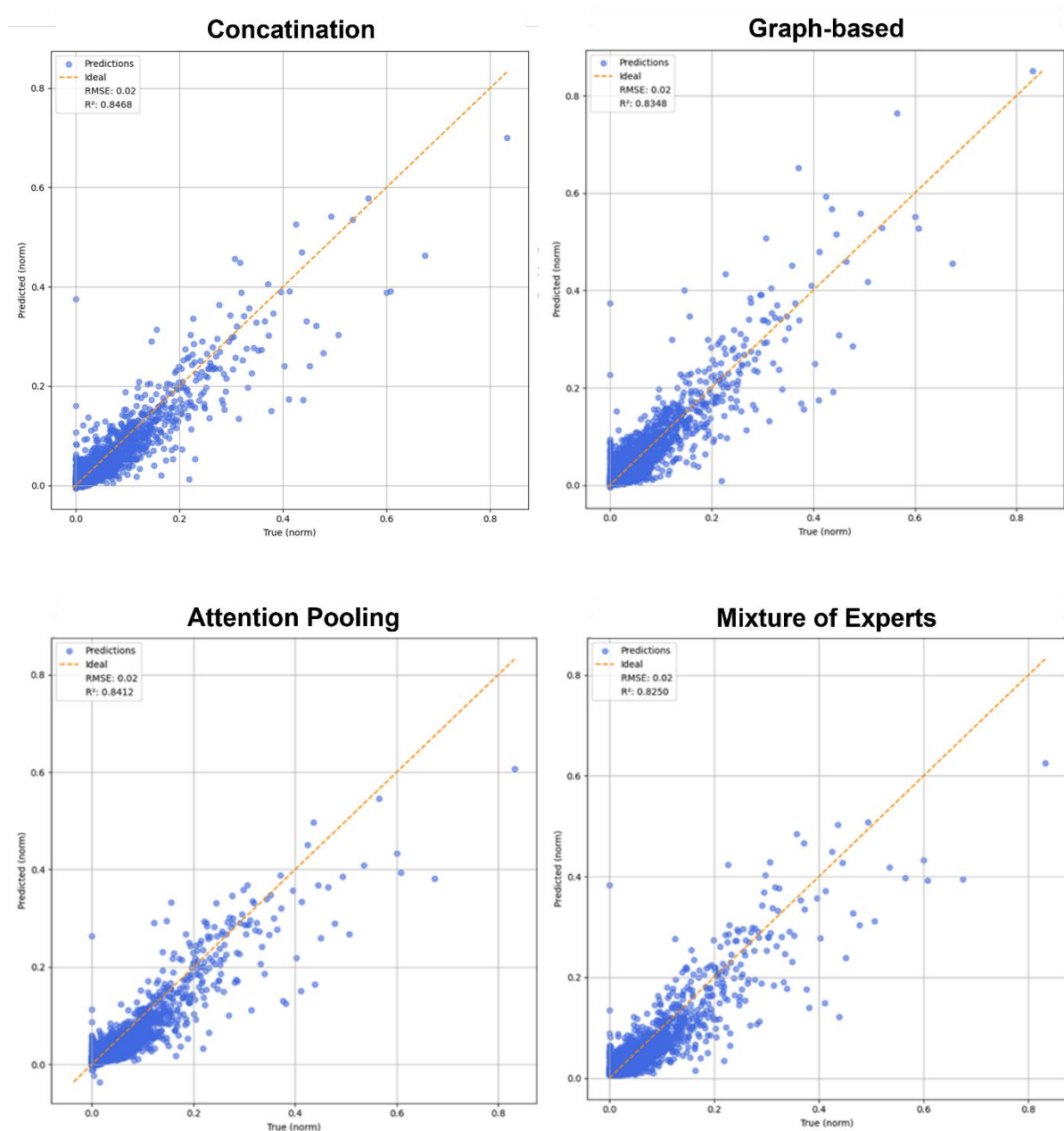


Abb. 16: Regressionsplots der normalisierten BIP-Werte für alle Fusionsmethoden. Quelle: Eigener Entwurf.

Unter den getesteten Ansätzen erzielt das Concat-Modell mit einem  $R^2$  von 0,8468 und einem RMSE von 0,02 die beste Modellgüte. Die Vorhersagen liegen eng entlang der Idealgeraden, was auf geringe Abweichungen zwischen Modelloutput und Referenzwerten hinweist. Das graphbasierte Modell erreicht mit einem  $R^2$  von 0,8348 leicht schwächere Ergebnisse und zeigt vor allem im mittleren Wertebereich eine größere Streuung. Das Attention Pooling Modell bewegt sich mit einem  $R^2$  von 0,8412 nahezu auf dem Niveau von Concatenation, während das Mixture-of-Experts Modell mit einem  $R^2$  von 0,8250 die geringste Modellgüte aufweist.

Als nächstes werden die Ergebnisse der Realwerte des BIP (Abb. 17) präsentiert:

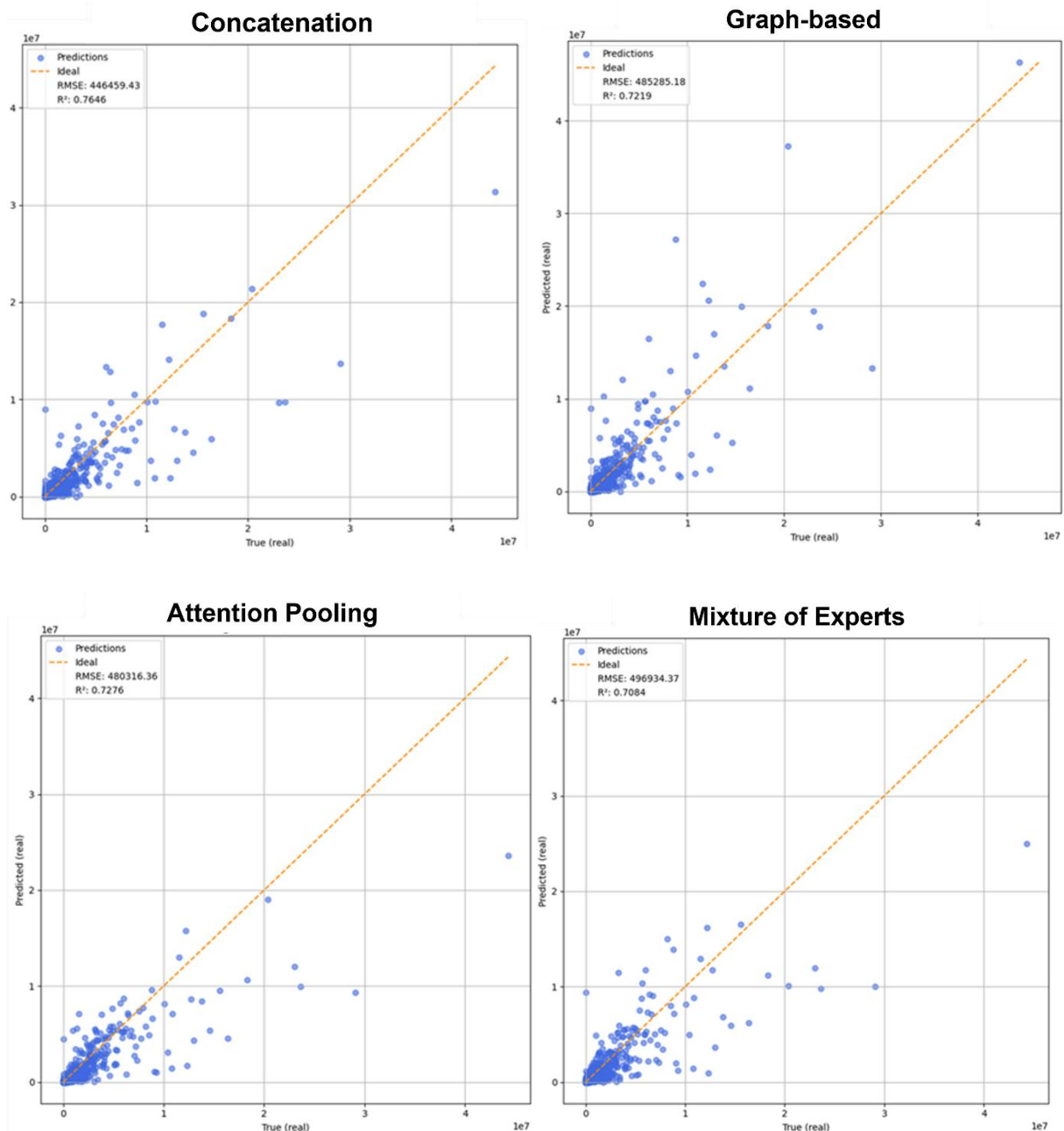


Abb. 17: Regressionsplots der denormalisierten BIP-Werte für alle Fusionsmethoden. Quelle: Eigener Entwurf.

Auch bei den denormalisierten BIP-Werten liefert Concatenation die besten Resultate (RMSE = R\$ 446.459,43;  $R^2 = 0,7646$ ). Die Punktwolke zeigt eine enge Orientierung an der Idealgeraden. Graph-based (RMSE = R\$ 485.285,18;  $R^2 = 0,7219$ ) und Attention Pooling (RMSE = R\$ 480.316,36;  $R^2 = 0,7276$ ) erzielen vergleichbare, leicht schwächere Werte. Mixture of Experts weist mit einem RMSE von R\$ 496.934,37 und einem  $R^2$  von 0,7084 erneut die geringste Modellgüte auf.

### 6.1.2 Primärer Sektor

Die Ergebnisse des primären Sektors (Abb. 18 und Abb. 19) unterscheiden sich deutlich von denen des BIP. Sowohl der Wertebereich als auch die Verteilung dieser Zielvariable weichen stark vom BIP, aber auch von den übrigen Sektoren ab, was sich auch in den erzielten Metriken widerspiegelt.

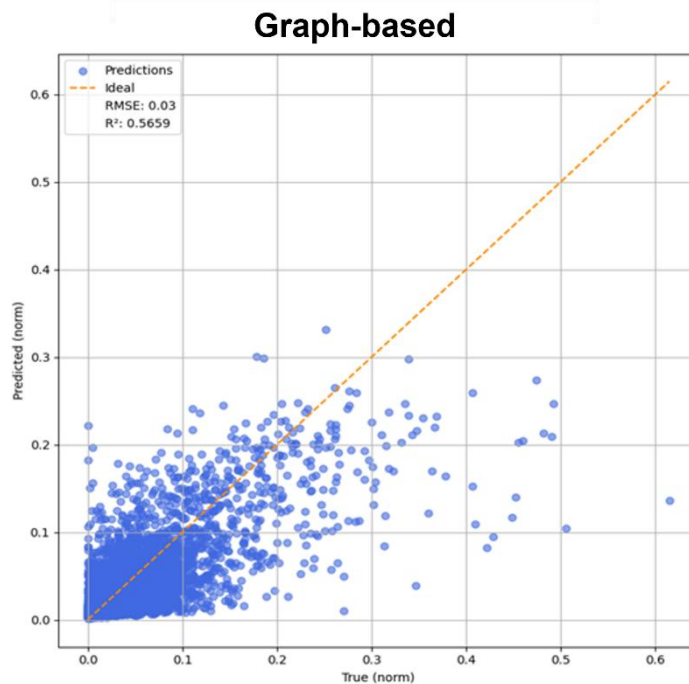


Abb. 18: Bester Regressionsplot der normalisierten Werte des primären Sektors (Fusionsmethode: Graph-based).  
Quelle: Eigener Entwurf.

Für die normalisierten Werte erzielt das graphbasierte Modell die besten Resultate (RMSE = 0,03;  $R^2 = 0,5659$ ). Die ermittelten Werte befinden sich größtenteils im Bereich der Idealgeraden, weisen jedoch eine erhebliche Streuung auf.

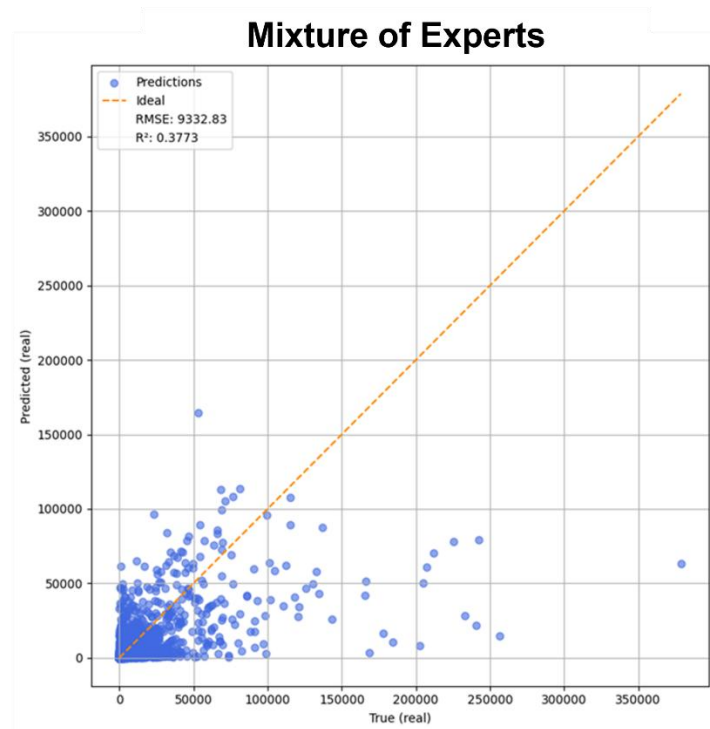


Abb. 19: Bester Regressionsplot der denormalisierten Werte des primären Sektors (Fusionsmethode: Mixture of Experts).  
Quelle: Eigener Entwurf.

Bei den realen Werten zeigt das Mixture-of-Experts-Modell die höchste Modellgüte (RMSE = R\$ 9.332,83;  $R^2 = 0,3773$ ). Im Scatterplot zeigt sich eine deutlich breitere Streuung sowie ein klarer Rechtstrend, der auf eine systematische Unterschätzung des Modells hinweist.

### 6.1.3 Sekundärer Sektor

Beim sekundären Sektor (Industrie) nähert sich die Verteilung der „Predictions“ deutlich stärker der des BIP an, was sich auch in der Modellperformance widerspiegelt (Abb. 20 und Abb. 21).

## Graph-based

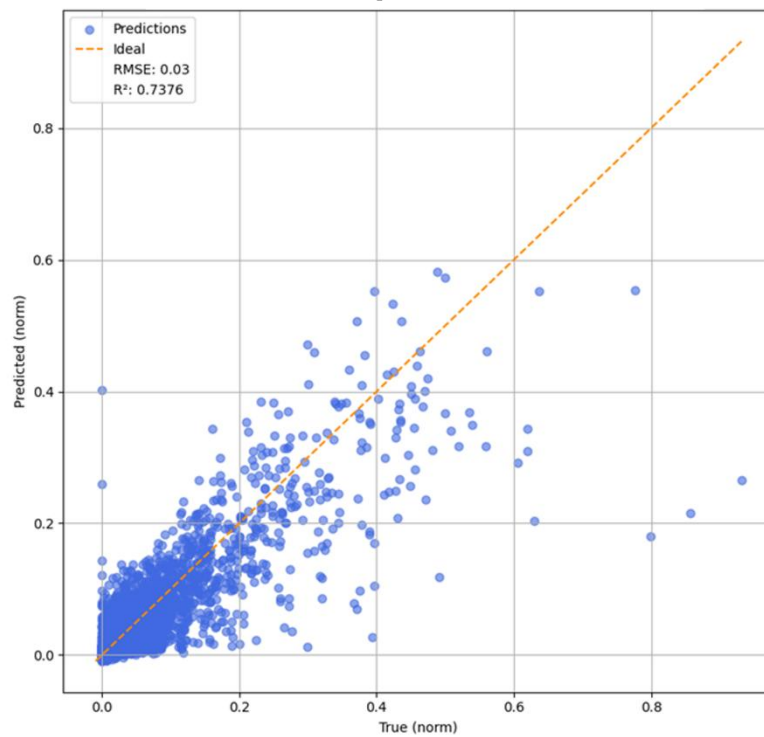


Abb. 20: Bester Regressionsplot der normalisierten Werte des sekundären Sektors (Fusionsmethode: Graph-based).  
Quelle: Eigener Entwurf.

Für die normalisierten Werte liefert das graphbasierte Modell die besten Ergebnisse. Mit einem RMSE von 0,03 und einem  $R^2$ -Wert von 0,7376 liegen die ermittelten Werte meist nahe der Idealgeraden, wobei eine moderate Streuung bei höheren Werten sichtbar ist.

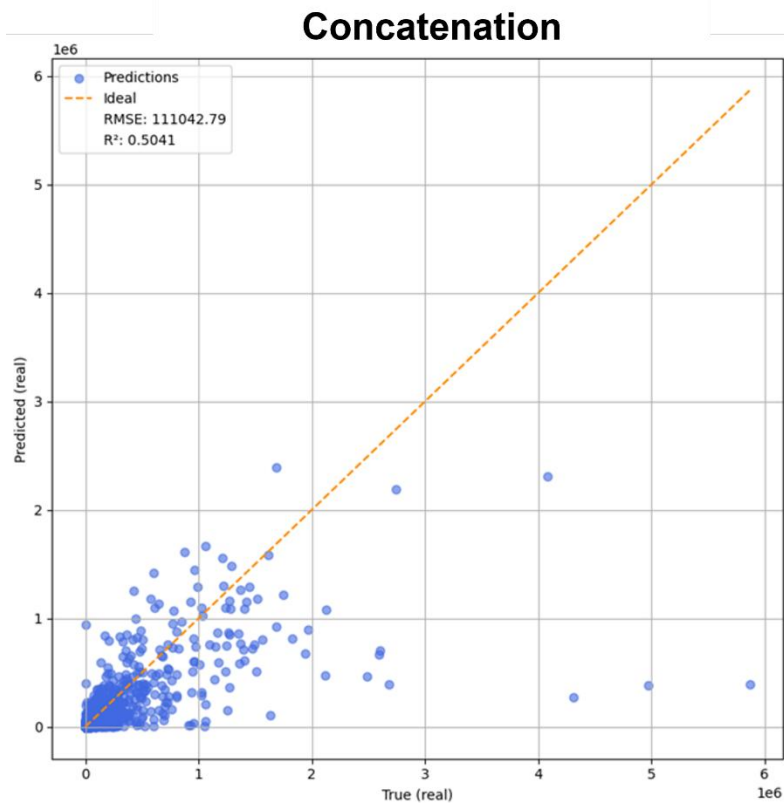


Abb. 21: Bester Regressionsplot der denormalisierten Werte des sekundären Sektors (Fusionsmethode: Concatenation).  
Quelle: Eigener Entwurf.

Bei den Realwerten erzielt Concatenation die beste Modellgüte (RMSE = R\$111.042,79;  $R^2 = 0,5041$ ). Die Punktwolke folgt dem allgemeinen Trend der Idealgeraden, zeigt jedoch bei großen Werten eine deutliche Rechtsneigung, was erneut auf eine Unterschätzung hinweist.

#### 6.1.4 Tertiärer Sektor

Die Regressionsplots des tertiären Sektors (Dienstleistungen) ähneln dem des BIP am meisten, sowohl bei der Verteilung der „Predictions“, als auch bei den erzielten Metriken (Abb. 22 und Abb. 23).

## Concatenation

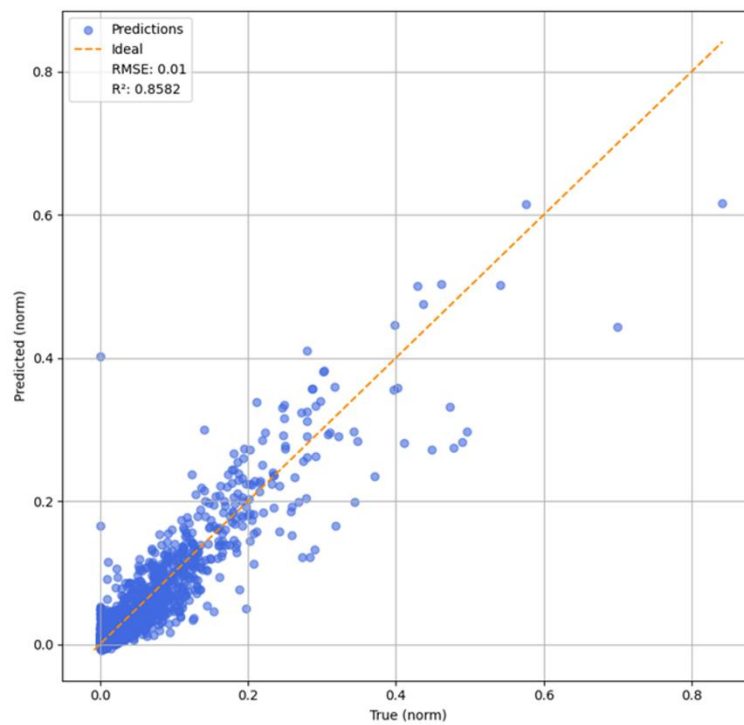


Abb. 22: Bester Regressionsplot der normalisierten Werte des tertiären Sektors (Fusionsmethode: Concatenation).  
Quelle: Eigener Entwurf.

Für die normalisierten Werte liefert das Concat-Modell die besten Resultate. Mit einem RMSE von 0,01 und einem  $R^2$ -Wert von 0,8582 liegen die Vorhersagen eng um die Idealgerade, wodurch eine hohe Übereinstimmung zwischen vorhergesagten und tatsächlichen Werten erkennbar ist.

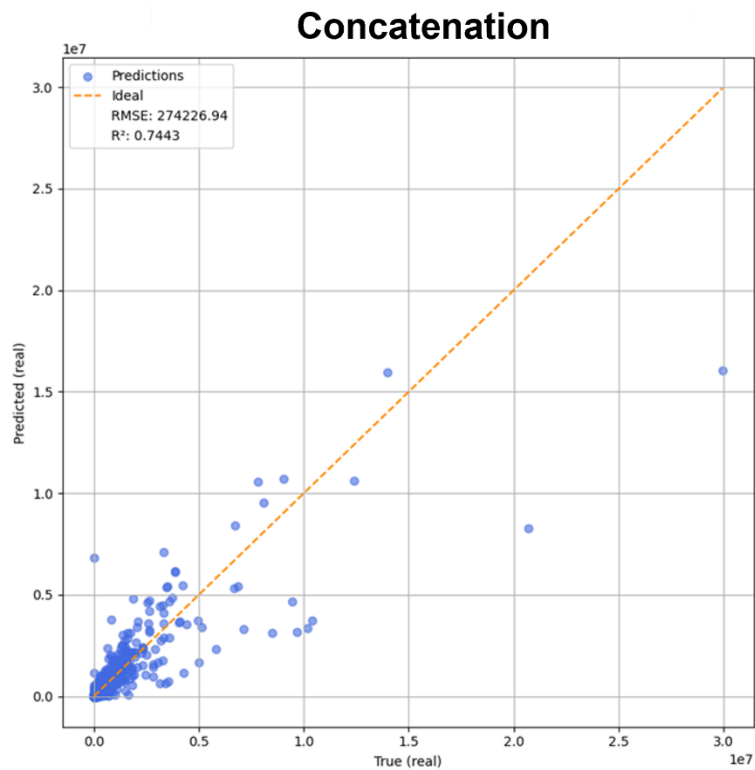


Abb. 23: Bester Regressionsplot der denormalisierten Werte des tertiären Sektors (Fusionsmethode: Concatenation).  
Quelle: Eigener Entwurf.

Auch bei den realen Werten erzielt Concatenation die höchste Güte (RMSE = R\$274.226,94;  $R^2 = 0,7443$ ). Die Streuung nimmt im Vergleich zur Normalisierung leicht zu, ist jedoch insgesamt weiterhin moderat und folgt dem erwarteten Trend.

#### 6.1.5 Öffentlicher Sektor

Zum Schluss werden in diesem Unterkapitel noch die Ergebnisse des öffentlichen Sektors dargestellt (Abb. 24 und Abb. 25). Bei diesen Resultaten ist am auffälligsten, wie stark sich die normalisierten Werte von den denormalisierten „Predictions“ unterscheidet.

## Concatenation

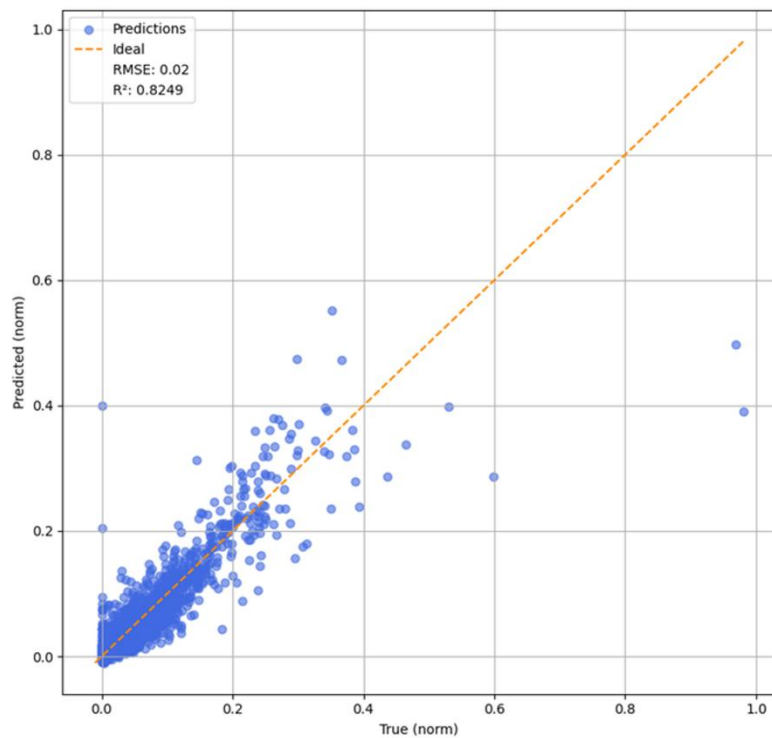


Abb. 24: Bester Regressionsplot der normalisierten Werte des öffentlichen Sektors (Fusionsmethode: Concatenation).  
Quelle: Eigener Entwurf.

Für die normalisierten Werte liefert Concatenation die besten Ergebnisse. Mit einem RMSE von 0,02 und einem R<sup>2</sup>-Wert von 0,8249 liegen die Vorhersagen überwiegend nahe der Idealgeraden, wobei nur zwei stärkere Ausreißer im höheren Wertebereich erkennbar sind.

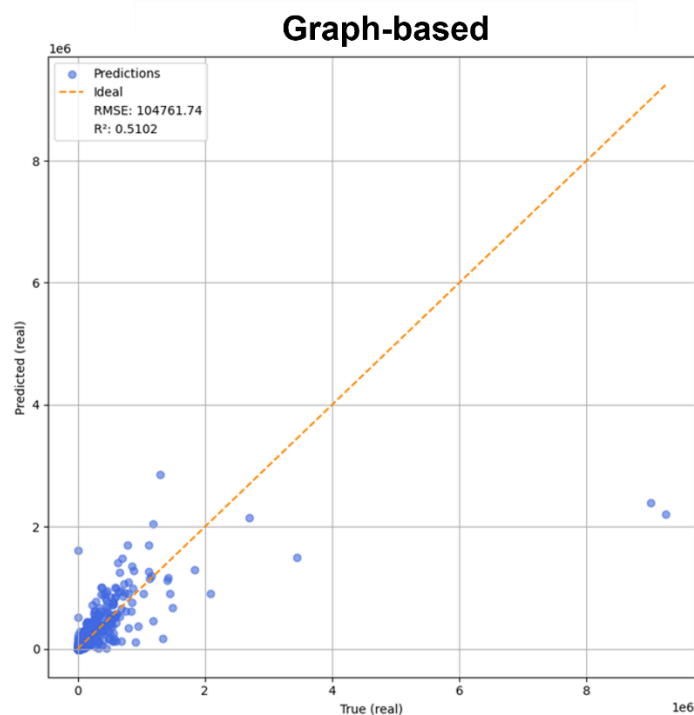


Abb. 25: Bester Regressionsplot der denormalisierten Werte des öffentlichen Sektors (Fusionsmethode: Graph-based).  
Quelle: Eigener Entwurf.

Für die realen Werte erzielt das graphbasierte Modell die höchste Modellgüte. Es erreicht einen RMSE von R\$ 104.761,74 und einen  $R^2$ -Wert von 0,5102. Die Punktwolke folgt dem Trend der Idealgeraden, zeigt jedoch eine deutliche Streuung, insbesondere bei den höchsten Zielwerten.

Die Auswertung der Modelle für alle fünf Zielvariablen zeigt deutliche Unterschiede in der Vorhersagegüte zwischen den Fusionsmethoden. Während Concatenation insbesondere beim BIP und im Dienstleistungssektor die besten Resultate liefert, weist die graphbasierte Variante Vorteile im primären und sekundären Sektor auf. Die nachfolgende Tabelle 2 fasst die erzielten Metriken der besten Modellvarianten für eine verbesserte Übersicht nochmal zusammen:

**Tab. 2: Werte der besten Fusionsmethode pro Zielvariable.**  
**Quelle: Eigene Berechnung und Darstellung**

| Zielvariable            | Beste Fusionsmethode (normalisiert) | R <sup>2</sup> (norm) | RMSE (norm) | Beste Fusionsmethode (real) | R <sup>2</sup> (real) | RMSE in R\$ (real) |
|-------------------------|-------------------------------------|-----------------------|-------------|-----------------------------|-----------------------|--------------------|
| <b>BIP</b>              | Concat                              | 0,8468                | 0,02        | Concat                      | 0,7646                | 446.459,43         |
| <b>Landwirtschaft</b>   | Graph-based                         | 0,5659                | 0,03        | Mixture of Experts          | 0,3773                | 9.332,83           |
| <b>Industrie</b>        | Graph-based                         | 0,7376                | 0,03        | Concat                      | 0,5041                | 111.042,79         |
| <b>Dienstleistungen</b> | Concat                              | 0,8582                | 0,01        | Concat                      | 0,7443                | 274.226,94         |
| <b>Staatseinnahmen</b>  | Concat                              | 0,8249                | 0,02        | Graph-based                 | 0,5102                | 104.761,74         |

**Legende:** RMSE = Root Mean Square Error, R<sup>2</sup> = Bestimmtheitsmaß, norm = normalisierter Wertebereich, real = Realwertbereich, R\$ = Brasilianischer Real, Concat = Concatenation.

## 6.2 Vergleichende Analyse von Ergebnissen und Referenzdaten

Da die Verwendung von Concatenation als Fusionsmethode die vielversprechendsten Resultate lieferte, wird im weiteren Verlauf der Arbeit nur noch diese Variante eingesetzt. Damit nun die ermittelten Werte noch besser mit den Referenzwerten verglichen werden können, werden hierfür Vergleichskarten von ganz Brasilien angefertigt.

Um diese Ergebnisdarstellungen anfertigen zu können, müssen zunächst Resultate für alle vier Folds ermittelt werden. Hierfür wird jeder Teildatensatz einmal für das Training und einmal für den Real-World Test verwendet, um keine Verzerrungen zu erhalten und den gesamten Datensatz gleichmäßig zu berücksichtigen. Die dabei erzielten Werte sind in Tabelle 3 dargestellt.

**Tab. 3: Durchschnittswerte pro Zielvariable (über 4 Folds) mit der Fusionsmethode Concatenation.**  
**Quelle: Eigene Berechnung und Darstellung**

| Zielvariable            | RMSE (norm) | R <sup>2</sup> (norm) | RMSE in R\$ (real) | R <sup>2</sup> (real) |
|-------------------------|-------------|-----------------------|--------------------|-----------------------|
| <b>BIP</b>              | 0,0150      | 0,8554                | 468.350,0409       | 0,7669                |
| <b>Dienstleistungen</b> | 0,0132      | 0,8577                | 286.616,0192       | 0,7575                |
| <b>Industrie</b>        | 0,0265      | 0,7479                | 98.235,2495        | 0,5804                |
| <b>Landwirtschaft</b>   | 0,0282      | 0,5512                | 9.783,0682         | 0,3347                |
| <b>Staatseinnahmen</b>  | 0,0159      | 0,8119                | 97.169,4141        | 0,4309                |

**Legende:** RMSE = Root Mean Square Error, R<sup>2</sup> = Bestimmtheitsmaß, norm = normalisierter Wertebereich, real = Realwertbereich, R\$ = Brasilianischer Real.

Die Ergebnisse verdeutlichen erneut, dass die Modelleistung zwischen den Zielvariablen deutlich variiert. Die besten Resultate werden für das BIP und den tertiären Sektor erzielt, die beide hohe  $R^2$ -Werte in der Normalisierung (0,8554 bzw. 0,8577) sowie moderat hohe  $R^2$ -Werte im Realwertbereich aufweisen. Die niedrigste Modellgüte tritt im primären Sektor auf. Der sekundäre- und der öffentliche Sektor weisen beide bei den normalisierten Werten gute Ergebnisse auf, die ermittelten Realwerte besitzen jedoch noch Verbesserungspotenzial. Ebenfalls muss in Bezug auf diese Tabelle noch erwähnt werden, dass der RMSE der Realwerte nur zur Vervollständigung dient und nicht als Vergleichswert zwischen den Variablen herangezogen werden kann. Der Grund dafür ist der sich zueinander stark unterscheidende Wertebereich der einzelnen Zielvariablen.

Nachdem nun Werte für alle fünf Zielvariablen für das gesamte Untersuchungsgebiet ermittelt wurden, können hiermit Ergebniskarten erstellt werden. Diese werden mit den Referenzdaten im selben Maßstab gegenübergestellt, um herauszufinden, wie sehr sich die räumliche Struktur der Referenzdaten und der ermittelten Werte ähnelt. Damit diese vergleichbar gegenübergestellt werden können, werden die Resultate in je 5 Klassen eingeteilt. Die Klassengrenzen werden dabei mittels des Jenks-Natural-Breaks-Algorithmus definiert. Dieser Algorithmus teilt die Daten so, dass die Varianz innerhalb der Klassen minimiert und die Varianz zwischen den einzelnen Klassen maximiert wird (Jenks 1967). Damit soll eine unverfälschte Darstellung der Datenstruktur sichergestellt werden. Gleichzeitig dient diese Einteilung dazu, die Reproduzierbarkeit der Ergebnisse zu erhöhen und subjektive Eingriffe zu vermeiden. Ein Nachteil dieser Methode sind die unregelmäßigen Klassengrenzen, dies wird jedoch zugunsten einer besseren Vergleichbarkeit der Karten akzeptiert.

In der folgenden Abbildung 26 werden nun die ermittelten Werte des BIP sowie die dazugehörigen Referenzwerte verglichen:

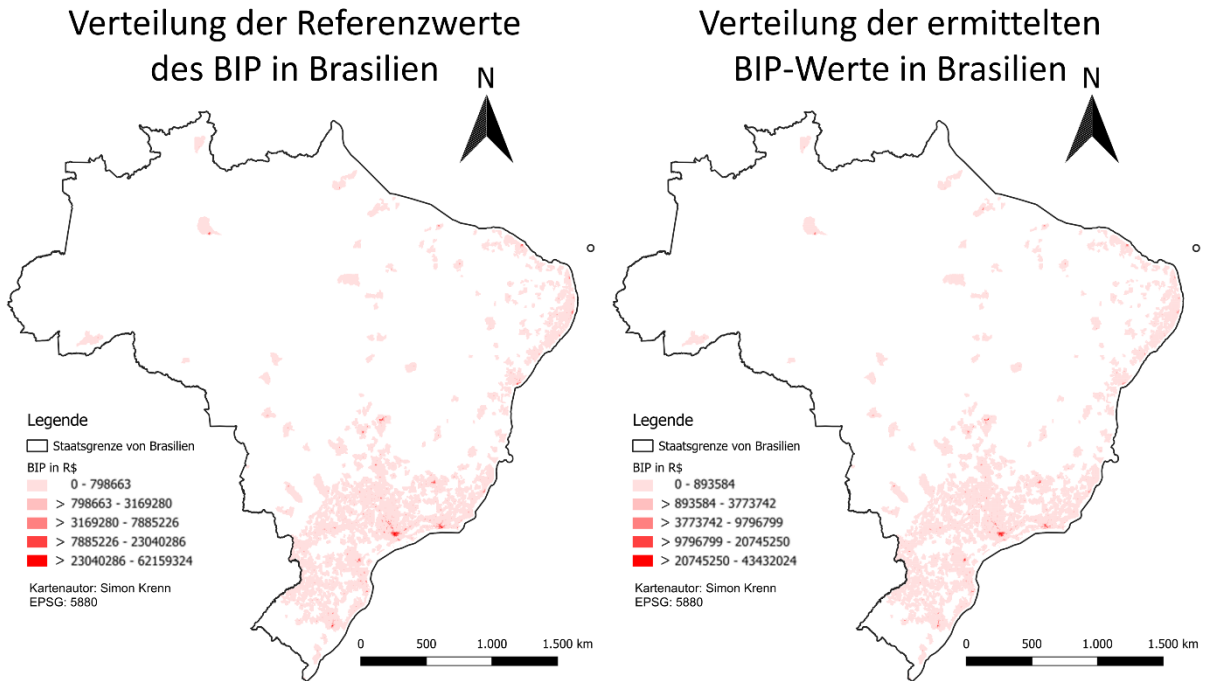


Abb. 26: Gegenüberstellung der Verteilung der Referenzdaten und der ermittelten BIP-Werte für das gesamte Untersuchungsgebiet. Quelle: Eigener Entwurf.

In beiden Darstellungen ist gut zu erkennen, dass die höchsten BIP-Werte in den wirtschaftlich stark entwickelten Regionen des Südostens und Südens auftreten, insbesondere im Bereich der Metropolregionen São Paulo und Rio de Janeiro. Auch entlang der Küstenregionen im Nordosten sind vereinzelt höhere Werte sichtbar, während das Landesinnere und weite Teile des Nordens und Westens vergleichsweise niedrige BIP-Werte aufweisen.

Ebenfalls wird deutlich, dass sich die grundlegende Struktur und Verteilung der Werte der beiden Karten stark ähnelt. Hotspots befinden sich in beiden Visualisierungen an denselben Positionen und auch die generelle Werteverteilung ist ähnlich. Ein paar Unterschiede sind dennoch anzutreffen. Die Referenzdaten weisen eine breitere Spannweite, vor allem bei den höchsten BIP-Werten, auf, bzw. die Extremwerte werden tendenziell etwas unterschätzt (Maximum: R\$ 62.159.324 in den Referenzdaten und R\$ 43.432.024 beim Modell). Zudem sind einige kleinere Cluster hoher BIP-Werte in der Modellkarte nicht reproduziert, die jedoch in der Referenzkarte sichtbar sind.

Insgesamt lässt sich aus der visuellen Gegenüberstellung ableiten, dass das Modell großräumige Muster bzw. die generelle Struktur der BIP-Werte zuverlässig erfasst hat, jedoch Schwierigkeiten bei (lokalen) Extremwerten besitzt.

Der visuelle Vergleich der modellierten Daten mit den Referenzwerten weist beim sekundären, aber auch beim tertiären Sektor ähnliche Muster auf. Beim Dienstleistungssektor bleibt ebenfalls die generelle Struktur der Daten erhalten bzw. konnte erfolgreich ermittelt werden. Aber auch hier ist das generelle Wertespektrum der Modelldaten, vor allem bei Extremwerten, niedriger als in den Referenzdaten. Beim Industriesektor ist der Unterschied der Wertespanne etwas stärker ausgeprägt, aber auch hier ähnelt das Gesamtbild der ermittelten Daten stark den Referenzwerten.

Wie bereits bei den Fusionsmethoden unterscheidet sich auch hier der primäre Sektor stark von den restlichen Wirtschaftswerten, was in Abbildung 27 verdeutlicht wird:

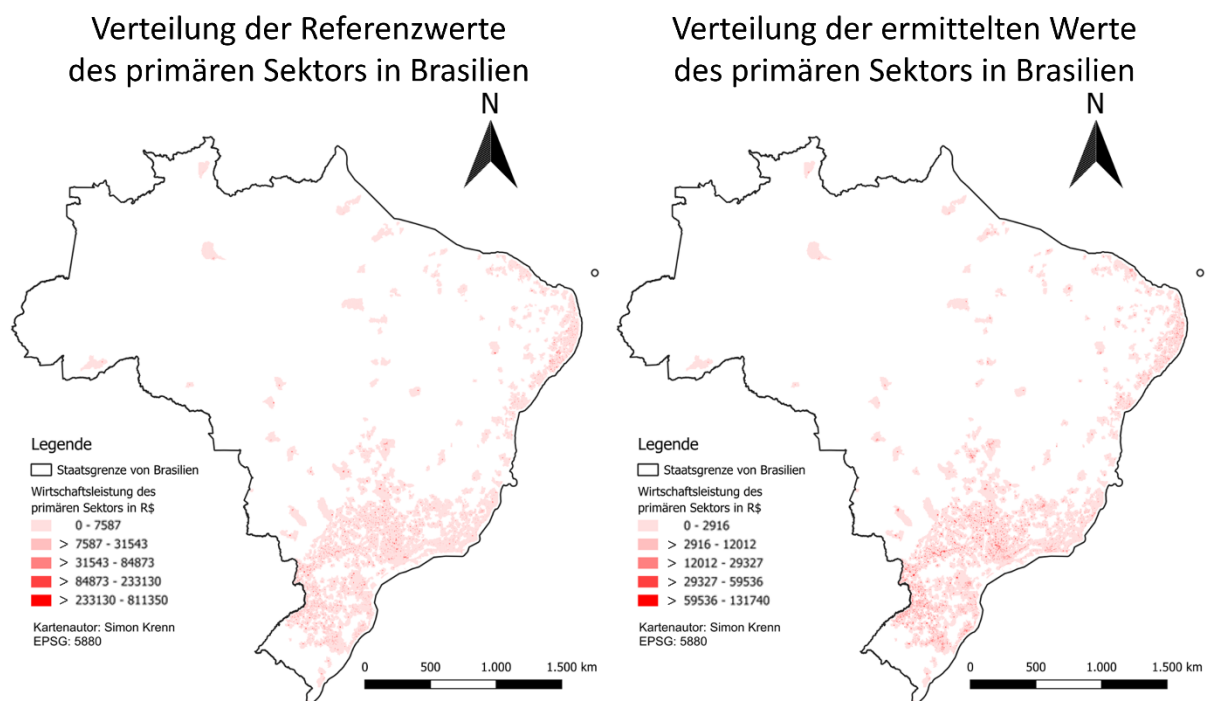


Abb. 27: Gegenüberstellung der Verteilung der Referenzdaten und der ermittelten Werte des primären Sektors für das gesamte Untersuchungsgebiet.   
 Quelle: Eigener Entwurf.

Dieser Sektor weist die höchste wirtschaftliche Aktivität vor allem in (landwirtschaftlich geprägten) Regionen im Süden des Landes auf, dabei jedoch eher im Landesinneren und nicht wie zuvor in den Küstenregionen. Die Verteilung dieses Teildatensatzes weist jedoch deutlich andere Merkmale als die restlichen Zielvariablen auf. In manchen Bereichen des Untersuchungsgebietes nimmt die Verteilung stark alternierende Muster an, die sowohl bei den Referenzwerten, aber auch in der Modellierung sichtbar sind. Die beiden Karten unterscheiden sich jedoch stark voneinander. Auch hier unterschätzt die Modellierung die höchsten Werte der Zielvariable, aber auch die Klassengrenzen selbst unterscheiden sich beim primären Sektor deutlich stärker

voneinander als bei den restlichen Wirtschaftswerten. Die Struktur der Referenzdaten konnte hier eher suboptimal wiedergegeben werden.

In den Karten des öffentlichen Sektors (Abb. 28) wird deutlich, dass sich die Hotspots breit über das gesamte Untersuchungsgebiet verteilen und eine starke Verdichtung in urbanen Zentren und regionalen Verwaltungsstandorten aufweisen.

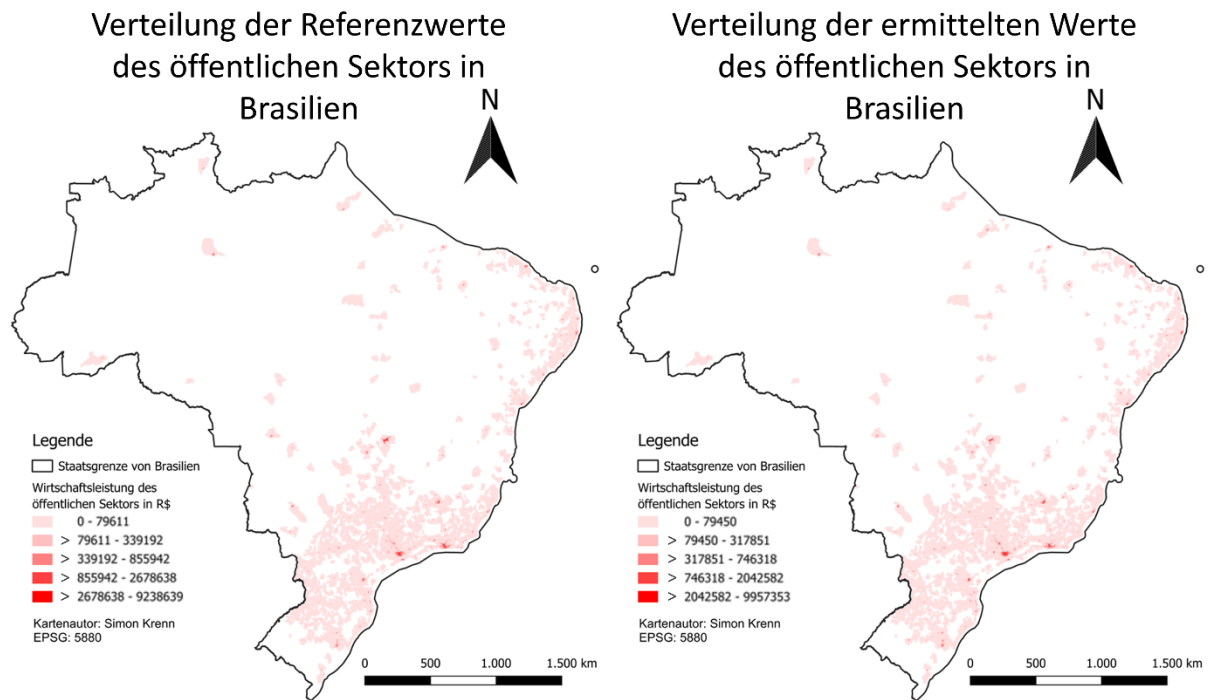


Abb. 28: Gegenüberstellung der Verteilung der Referenzdaten und der ermittelten Werte des öffentlichen Sektors für das gesamte Untersuchungsgebiet. Quelle: Eigener Entwurf.

Spannend ist hierbei, dass der Bereich rund um die brasilianische Hauptstadt Brasília im Zentrum des Landes erstmals als stark ausgeprägter Hotspot in den Vordergrund tritt, wodurch die Verwaltungsfunktion der Stadt unterstrichen wird. Die modellierten Werte ähneln in diesem Fall den Referenzwerten stark. Die Struktur der Daten unterscheidet sich nur gering und auch die Wertebereiche der einzelnen Klassen weichen nur gering voneinander ab.

Abschließend ist es jedoch auch noch wichtig, nicht nur die gesamte Struktur der Daten zu analysieren und zu beschreiben, sondern auch kleinräumige und folglich detailliertere Analysen durchzuführen.

### 6.3 Detailanalyse anhand ausgewählter Beispiele

Aus diesem Grund werden innerhalb dieses Kapitels ausgewählte Beispiele nähergebracht. Hierfür wird zunächst der Großraum von São Paulo und Rio de Janeiro

genauer analysiert, um eine der Hauptproblematiken des gewählten Workflows näherbringen zu können. Zum Abschluss werden noch zwei ausgewählte Zellen näher beschrieben, um ein „Worst-Case“ und Best-Case“ Szenario der Arbeit darzulegen.

### 6.3.1 Analyse des Großraumes São Paulo - Rio de Janeiro

Analysiert man die erzielten Resultate weiter wird deutlich, dass die Zielvariablen vor allem bei sehr niedrigen Referenzwerten stark überschätzt werden. Beispielsweise gibt es einige Zellen, bei denen der Referenzwert nahe 0 liegt (mit beispielsweise R\$ 17), der durch das Modell ermittelte Wert wird dann jedoch auf rund R\$ 1.700 geschätzt. Um dies zu verdeutlichen, wurde die folgende Kartenserie (Abb. 29-31) erstellt. Hierfür wurde der relative Fehlerwert für jede Zielvariable pro Zelle abgebildet. Dieser wurde anhand der folgenden Formel 4 berechnet:

$$\text{Relativer Fehler in Prozent} = \left( \frac{|Zielvariable_{pred} - Zielvariable_{ref}|}{Zielvariable_{ref}} \right) \times 100$$

*Formel 4: Relativer Fehler in Prozent je Zielvariable und Zelle*

Es wird der Absolutwert aus der Subtraktion des ermittelten Wertes und des Referenzwertes berechnet, welcher dann durch den Referenzwert dividiert wird. Abschließend erfolgt eine Multiplikation mit 100, um das Ergebnis in Prozent zu erhalten. Dieser Fehlerwert ist nun in der folgenden Kartenserie des Großraumes von São Paulo und Rio de Janeiro dargestellt. Die erste Karte enthält alle Resultate, die zweite Visualisierung zeigt alle Ergebnisse mit einem Referenzwert größer als R\$ 100 und die letzte Karte beinhaltet nur noch Zellen, die einen Referenzwert von mehr als R\$ 1.000 besitzen.

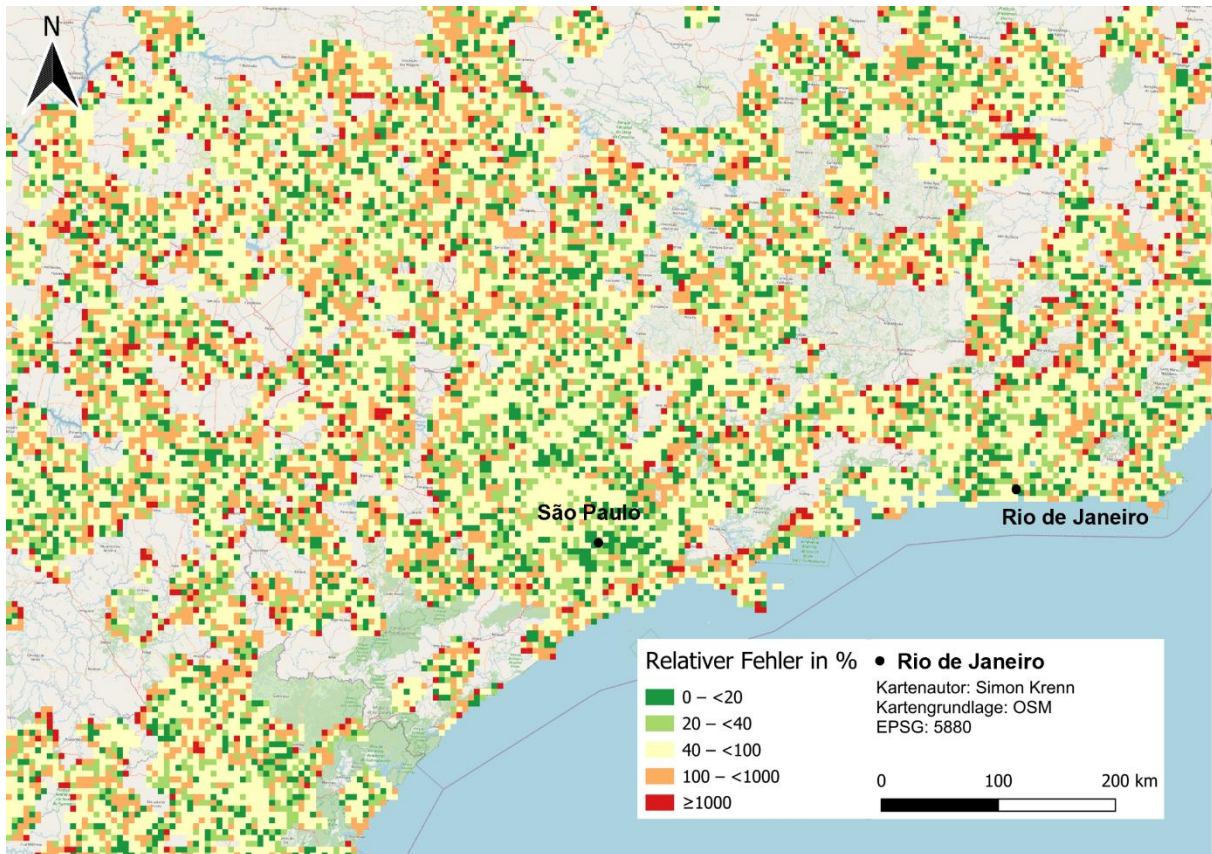


Abb. 29: Darstellung des relativen Fehlers des BIP pro Zelle in Prozent im Großraum São Paulo - Rio de Janeiro.  
Quelle: Eigener Entwurf

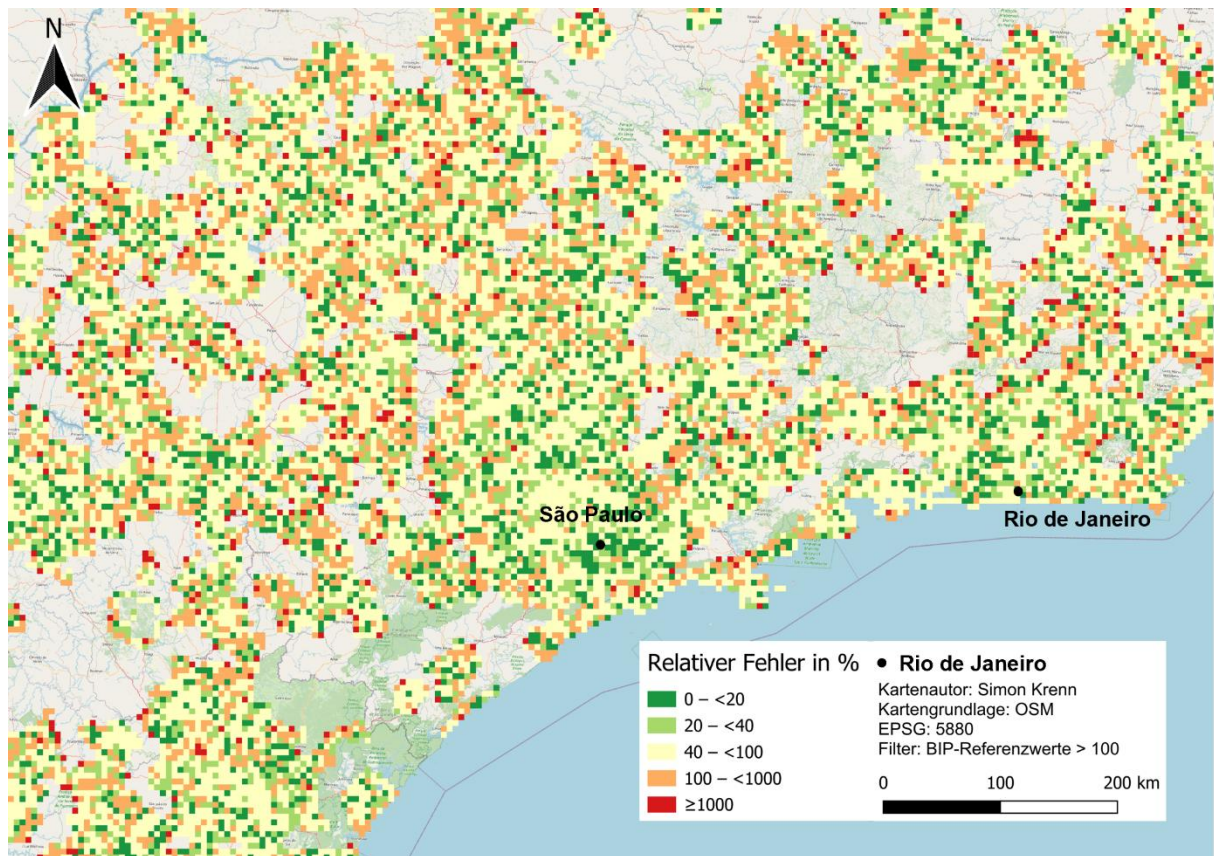


Abb. 30: Darstellung des relativen Fehlers des BIP pro Zelle (ohne BIP-Referenzwerte < 100) in Prozent im Großraum São Paulo - Rio de Janeiro.  
Quelle: Eigener Entwurf

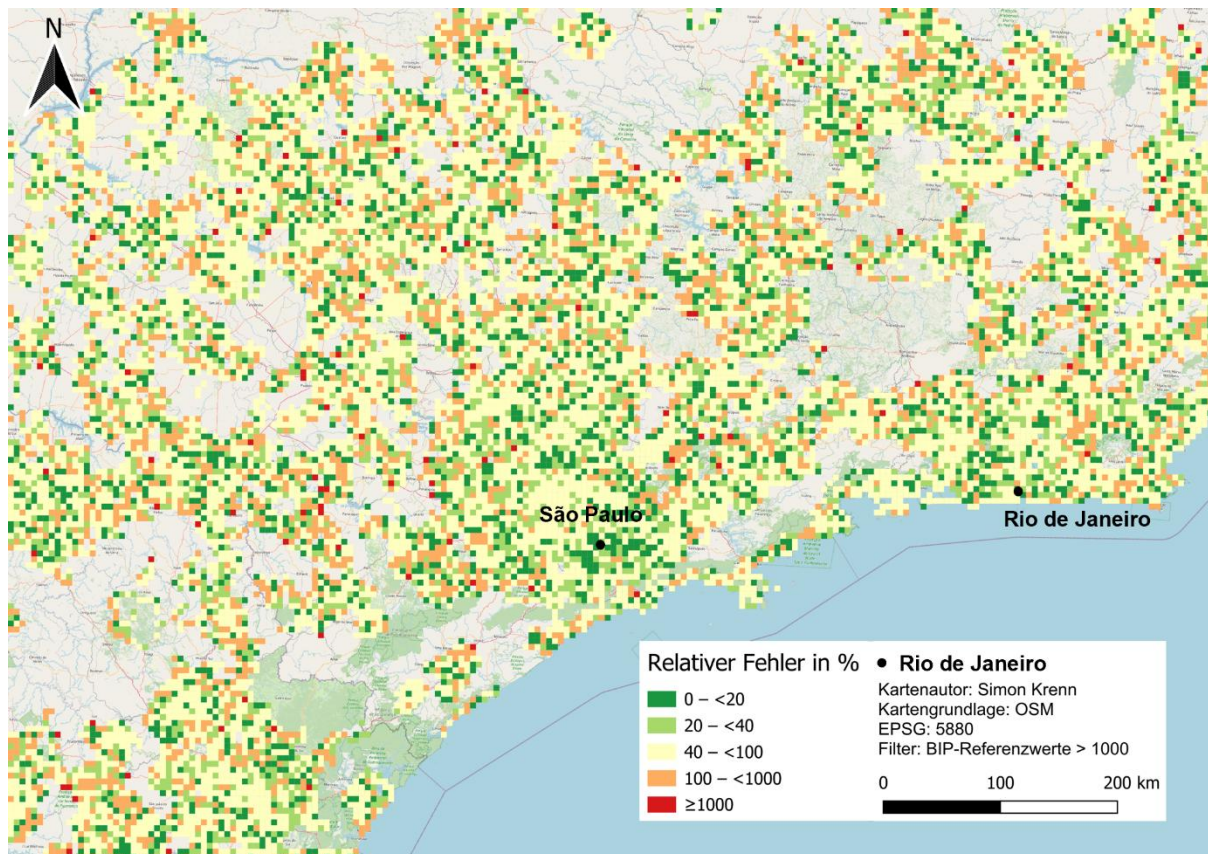


Abb. 31: Darstellung des relativen Fehlers des BIP pro Zelle (ohne BIP-Referenzwerte < 1.000) in Prozent im Großraum São Paulo - Rio de Janeiro.  
Quelle: Eigener Entwurf

Vergleicht man die drei Karten miteinander, wird deutlich, dass durch die schrittweise Filterung zunehmend jene Zellen entfernt werden, die besonders große Abweichungen aufweisen. Auffällig ist zudem, dass sich diese stark fehlerbehafteten Zellen überwiegend am Rand des Untersuchungsgebiets befinden. Dadurch zeigt sich, dass die allgemeine Modelleistung ohne diese Randzellen deutlich positiver ausfällt. Da sich dieses Muster konsistent über alle Zielvariablen hinweg beobachten lässt, beschränkt sich die exemplarische Darstellung bewusst nur auf das BIP.

Eine genauere Analyse dieser Randzellen zeigt, dass die hohen Abweichungen überwiegend auf Randeffekte der Referenzdaten zurückzuführen sind. Das Modell erhält dadurch fehlerhafte oder unvollständige Referenzwerte, was zu einer leichten Verzerrung des Modells führen könnte. Diese grundlegende Problematik soll nun durch das folgende Worst Case Szenario verdeutlicht werden.

### 6.3.2 Best- und Worst Case Szenario

Abschließend werden in diesem Unterkapitel noch zwei Zellen etwas genauer beschrieben. Dabei handelt es sich einmal um ein Resultat mit ausgezeichneten

Werten, das in allen Sektoren (außer dem öffentlichen Sektor) sehr gute Ergebnisse vorzuweisen hat und einmal um einen Fall, der durchwegs unzufriedenstellende Resultate aufweist. Da dieses „Negativbeispiel“ an das vorherige Unterkapitel anknüpft, wird dieses zuerst erläutert.

Dabei handelt es sich um die Zelle mit der ID 392266, welche sich südlich von der Kleinstadt Pedro de Toledo, in der Nähe von São Paulo im Süden des Landes, befindet (Abb. 32).

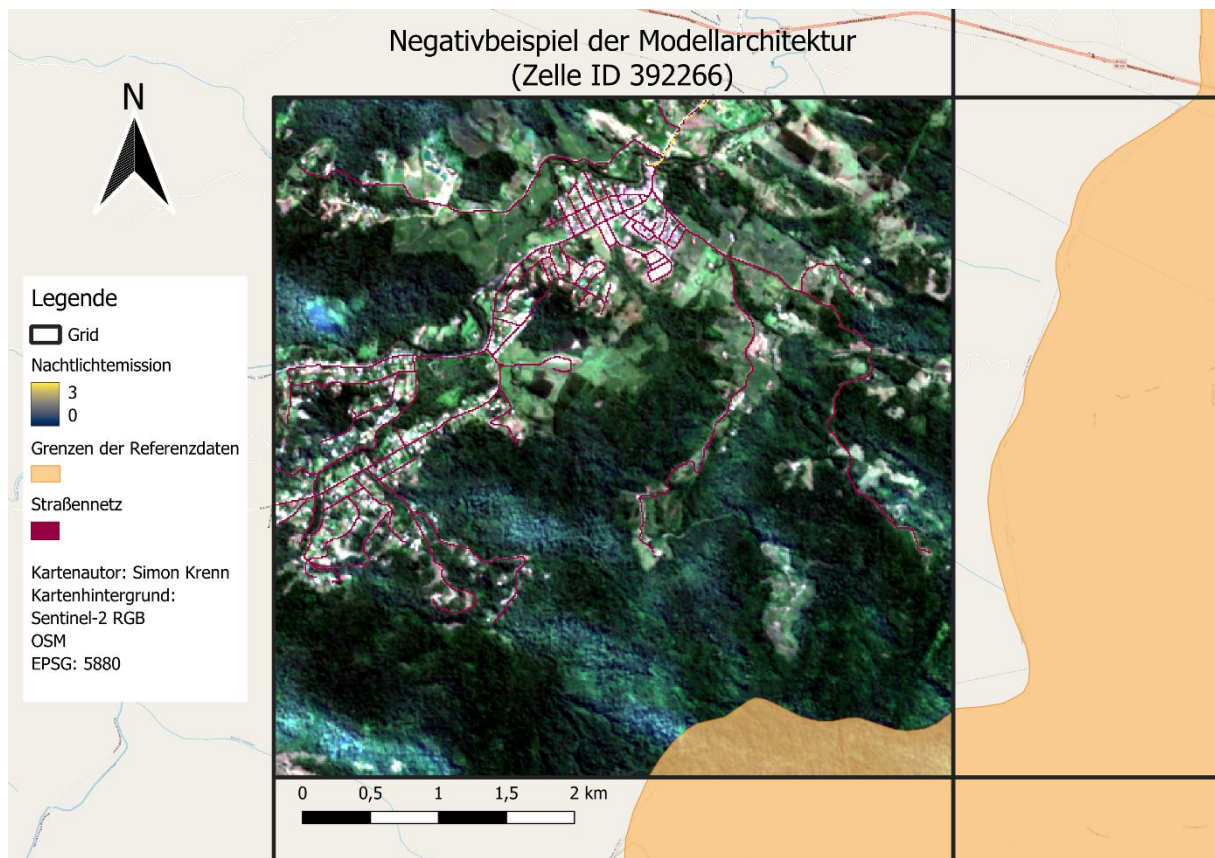


Abb. 32: Negativbeispiel der Modellarchitektur anhand der Zelle mit der ID 392266 im Süden des Untersuchungsgebietes. Quelle: Eigener Entwurf

In dieser Abbildung ist nun die zuvor beschriebene Zelle dargestellt. Um die Situation besser verstehen zu können, sind einerseits ausgewählte Inputdaten der Zelle visualisiert (Sentinel-2 RGB, Straßennetz und die Nachtlichtemission), aber auch die Grenzen der ursprünglichen Referenzdaten vor der Disaggregation. Zu erkennen ist ein eher dünn besiedeltes Gebiet, worin dennoch einige Straßen anzutreffen sind, sowie ein paar Pixel mit etwas nächtlicher Emission im nördlichen Bereich der Zelle.

Entscheidend für dieses Szenario sind jedoch die Grenzen der Referenzdaten. Wie in der Karte verdeutlicht wird, befinden sich nur sehr kleine Teile davon innerhalb der

Zelle. Aufgrund dieser geringen Fläche werden der Zelle, während der Disaggregation, sehr niedrige Referenzwerte zugewiesen. Im weiteren Verlauf des Workflows wird jedoch auf die Daten der gesamten Zelle zugegriffen. Folglich treten dadurch Verzerrungen innerhalb des Modells auf, da die Zelle trotz einer leichten Besiedelung, einem moderat ausgeprägten Straßennetz usw. sehr niedrige Referenzwerte vorzuweisen hat. Am Beispiel des BIP besitzt diese Zelle einen relativen Fehlerwert von 2.726,58%, der zugewiesene Referenzwert beträgt R\$ 876,46, der durch das Modell ermittelte BIP-Wert liegt jedoch bei R\$ 24.773,85.

Dieses Beispiel deckt sich auch mit den in Kapitel 6.3.1 beschriebenen suboptimalen Resultaten an den Randbereichen. Eine vertiefende Analyse dieser Problematik erfolgt in Kapitel 7.3.

Beim folgenden „Positivbeispiel“ (Abb. 33) ist jedoch das Gegenteil der Fall. Hierbei konnten durchgehend positive Ergebnisse erzielt werden (relativer Fehler beim BIP von 0.09% oder beim sekundären Sektor 2.82%), außer beim öffentlichen Sektor (Relativer Fehler von etwas über 100%). Dies ist möglicherweise auf nur wenige, innerhalb dieses Stadtteiles liegende, administrative Einrichtungen zurückzuführen.

Bei diesem Beispiel handelt es sich um die Zelle mit der ID 458022, welche sich im Westen der Großstadt São Luís im Norden des Landes befindet.

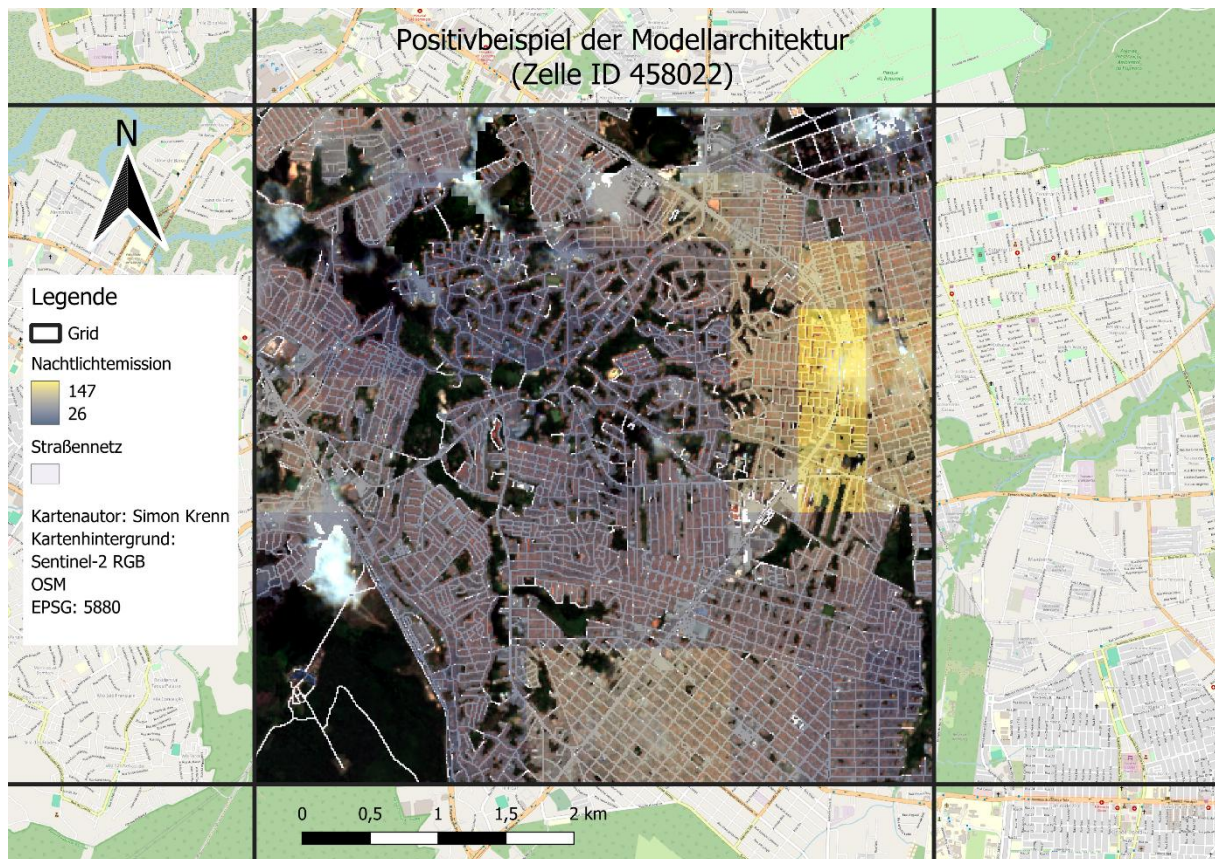


Abb. 33: Positivbeispiel der Modellarchitektur anhand der Zelle mit der ID 458022 im Norden des Untersuchungsgebietes. Quelle: Eigener Entwurf

In dieser Darstellung ist nun gut zu erkennen, dass es sich hierbei um ein dicht besiedeltes Stadtgebiet mit einigen Grünflächen handelt. Es wird hiermit verdeutlicht, dass die höchste Lichtemission im östlichen und südlichen Bereich der Zelle anzutreffen ist und dabei einen Maximalwert von 147 aufweist. Folglich ist eine starke nächtliche Lichtemission innerhalb der Zelle anzutreffen. Ebenfalls wird die Disaggregation der Black Marble Daten gut erkennbar, sowie das dichte Straßennetz der Zelle. Dadurch liegen genügend Parameter vor, um die eher hohen Referenzwerte dieser Zelle (Referenzwert des BIP = R\$ 5.104.254,55, ermittelter BIP-Wert = R\$ 5.108.970,5 oder Referenzwert des sekundären Sektors = R\$ 996.761,89, ermittelter Wert des sekundären Sektors = R\$ 968.682) präzise zu ermitteln.

Im folgenden Kapitel werden die präsentierten Ergebnisse eingehender interpretiert und hinsichtlich ihrer Bedeutung für die zugrunde liegenden Forschungsfragen diskutiert. Ebenfalls wird näher auf die Datenqualität eingegangen sowie gewisse Limitationen des Modells beschrieben.

## 7. Diskussion

Dieses Kapitel beschäftigt sich nun ausführlich mit der Interpretation und kritischen Reflexion der erzielten Ergebnisse. Dabei werden Rückschlüsse mithilfe bestehender Forschung gezogen und die Forschungsfragen der Arbeit beantwortet. Anschließend erfolgt eine differenzierte Auseinandersetzung mit den verwendeten Fusionsmethoden sowie der Qualität der Datengrundlagen (vor allem OSM). Der letzte Schwerpunkt des Kapitels liegt auf der Diskussion der Limitationen dieser Arbeit.

### 7.1 Interpretation der Ergebnisse

In diesem Unterkapitel werden zunächst die Forschungsfragen beantwortet sowie die Resultate der Fusionsmethoden interpretiert. Anschließend werden die Resultate mit bestehenden Forschungsergebnissen verglichen, um Übereinstimmungen, Unterschiede und mögliche Erklärungsansätze herauszuarbeiten.

#### 7.1.1 Beantwortung der Forschungsfragen

Nachfolgend werden nun die Forschungsfragen der Arbeit nacheinander diskutiert. Jede Frage wird anhand der erzielten Resultate beantwortet. Die zentrale Forschungsfrage der Arbeit lautete wie folgt:

- Wie lassen sich das Bruttoinlandsprodukt und die Wirtschaftssektoren auf subregionaler Ebene in Brasilien durch Deep Learning mittels multimodaler Geodatenfusion ermitteln?

Basierend auf den in Kapitel 6 präsentierten Ergebnissen kann eindeutig gesagt werden, dass der verwendete Ansatz vielversprechende Resultate erzielen konnte. Mithilfe der verwendeten Datengrundlagen und des implementierten Modells konnten die Wirtschaftswerte für das gesamte Untersuchungsgebiet ermittelt werden. Die wichtigsten Schritte dabei waren die ausführliche Vorprozessierung und Aufbereitung der Datengrundlagen sowie die Modellentwicklung und die Wahl einer geeigneten Fusionsmethode. Die im Rahmen dieser Arbeit verwendeten Datenquellen konnten effizient integriert und weiterverarbeitet werden und stellten somit eine geeignete Grundlage für das Modell dar. Die Disaggregation der ursprünglichen Referenzwerte ermöglichte, auf administrative Grenzen beschränkte, Referenzdaten für das Training eines subregionalen Modells anwendbar aufzubereiten. Die in Kapitel 6.3 erwähnten Randproblematiken werden in Kapitel 7.3 näher diskutiert.

Die Performance des entwickelten Deep-Learning-Modells entsprach den Erwartungen und trug maßgeblich zu den erzielten Resultaten bei. Insbesondere Transfer Learning erwies sich als geeignete Methode, um auch unter zeitlichen Einschränkungen oder bei begrenzten personellen Ressourcen praktikable Projekte innerhalb der gegebenen Rahmenbedingungen zu realisieren. Durch den Einsatz dieser Methode konnten die relevanten Merkmale der Datengrundlage so extrahiert werden, dass der Regressor konsistente und verlässliche Ergebnisse auf subregionaler Ebene lieferte. Daran anknüpfend kann die nächste Forschungsfrage beantwortet werden.

- Wie zuverlässig kann das entwickelte Modell das BIP und die einzelnen Wirtschaftssektoren ermitteln?

Dabei ist zu erwähnen, dass die Leistungsfähigkeit des Modells stark von der untersuchten Zielvariable abhängt. Wie die Ergebnisse in Tabelle 3 zeigen, konnten insbesondere das BIP und der Dienstleistungssektor sehr präzise ermittelt werden. Die entsprechenden Metriken weisen, trotz der Größe und Heterogenität des Untersuchungsgebiets, hohe Werte auf. Auch der sekundäre und der öffentliche Sektor erzielten bei den normalisierten Werten zufriedenstellende Ergebnisse. Im Vergleich dazu wiesen die denormalisierten Werte dieser Zielvariablen jedoch noch Verbesserungspotenzial auf. Dies könnte durch zusätzliche Entwicklungsschritte im Modell adressiert werden, worauf in Kapitel 7.3 näher eingegangen wird. Im Gegensatz dazu fällt die Modellleistung des primären Sektors deutlich schwächer aus, was insbesondere auf die hohe räumliche und strukturelle Heterogenität dieser Zielvariable sowie die eingeschränkte direkte Abbildbarkeit durch die verwendeten Geodaten zurückgeführt werden kann.

Diese hohe Variation der Vorhersagegüte (je nach Sektor) deckt sich jedoch mit Erkenntnissen aus der Forschung. Darin wird aufgezeigt, dass Deep-Learning-Modelle bestimmte Zielvariablen wesentlich stabiler abbilden können als Variablen mit hoher natürlicher oder struktureller Variabilität, wie es in dieser Arbeit beim primären Sektor der Fall ist (Yeh et al. 2020). Dies verdeutlicht, dass sektorale Heterogenität einen zentralen Einfluss auf die Modellgenauigkeit ausübt. Hinzu kommt, dass wirtschaftliche Aktivitäten des primären Sektors, etwa Land- oder Forstwirtschaft, häufig nur indirekt oder unvollständig durch die verwendeten Inputdaten repräsentiert werden, was die Modellierung zusätzlich erschwert.

Des Weiteren ist für die Qualität der Ergebnisse eine sorgfältige Aufbereitung der Inputdaten, insbesondere ihre sinnvolle Aufteilung auf Trainings-, Validierungs- und Testdatensätze sowie eine konsistente Normalisierung, entscheidend. Diese Schritte beeinflussen die Modellperformance maßgeblich, da sie Skalierungseffekte reduzieren, Ausreißer abmildern und zu einem stabileren Trainingsprozess führen (Huang et al. 2020).

Des Weiteren ist der Vergleich der normalisierten und denormalisierten Werte genauer zu untersuchen. Während die Modelle auf normalisierten Daten durchgängig höhere Metriken und geringere Streuungen erreichen, nahm die Modellgüte nach der Denormalisierung (teilweise deutlich) ab. Dies ist jedoch nicht überraschend, da in den realen Werten eine höhere Varianz und eine stärkere Wirkung von Ausreißern vorliegt, was die Modellanpassung erschwert. In der Literatur wird ebenfalls darauf hingewiesen, dass Rücktransformationen die Sensitivität gegenüber Extremwerten erhöhen können und dadurch die Modellgüte (scheinbar) verschlechtern (Zhao et al. 2024).

Im Zuge der Ermittlung der Wirtschaftsleistung wurde auch deutlich, dass die verwendete Fusionsmethode einen starken Einfluss auf die Qualität der Resultate besitzt. Folglich soll nun auch in Kapitel 7.1.2 die letzte Forschungsfrage beantwortet werden, die wie folgt lautet:

- Welche Fusionsmethoden eignen sich für die Ermittlung der Wirtschaftsleistung auf subregionaler Ebene am besten?

### 7.1.2 Diskussion der Fusionsmethoden

Dieses Unterkapitel beschäftigt sich nun mit der Interpretation der Resultate der unterschiedlichen Fusionsmethoden. Hierzu ist es wichtig, sich nochmal die Ergebnisse aus Tabelle 2 genauer anzusehen. Concatenation erzielte beim BIP und Dienstleistungssektor die besten Resultate, die graphbasierte Variante wies Vorteile im primären und sekundären Sektor auf. Die Mixture of Experts Fusionsmethode konnte lediglich bei den denormalisierten Werten des primären Sektors Spitzenwerte erreichen, wohingegen Attention Pooling durchgehend eine geringere Modellgüte lieferte.

Diese Beobachtung deckt sich mit Ergebnissen aus der Literatur, die darauf hinweist, dass einfache Fusionsmethoden bei klar strukturierten oder begrenzt variierenden

Daten häufig konkurrenzfähig oder sogar überlegen sein können (Chen et al. 2025). Insbesondere Concatenation sticht dabei heraus, da keinerlei Annahmen über Interaktionen oder Gewichtungen der Modalitäten getroffen werden. Die Merkmalsvektoren werden also unverändert weitergegeben, weshalb der Regressor in weiterer Folge auf das gesamte Informationsspektrum zugreifen kann.

Komplexere Verfahren wie Attention Pooling oder Mixture of Experts treffen zusätzliche Annahmen, indem sie explizite Gewichtungen oder Entscheidungen über die Relevanz einzelner Modalitäten treffen. Dies kann bei manchen Modellen zu Informationsverlust, aber auch Überanpassung führen, ein Effekt, der auch in mehreren Studien beschrieben wird (Stahlschmidt et al. 2022; Pawłowski et al. 2023). Besonders die Mixture of Experts Variante neigt bei unpassender Regularisierung zu unausgeglichener Expertennutzung, was die Leistung einzelner Experten beeinträchtigen kann (Chen et al. 2023).

Die graphbasierte Fusionsmethode erzielte insbesondere im primären und sekundären Sektor zufriedenstellende Ergebnisse. Dies deutet darauf hin, dass diese Variante vor allem dann Vorteile bringt, wenn die Zielvariablen stärker variieren oder weniger homogen sind. Gleichzeitig zeigen die Resultate jedoch, dass dieser Vorteil stark vom jeweiligen Kontext sowie von der Qualität und Vollständigkeit der zugrunde liegenden Datenquellen abhängt (D'Souza et al., 2024; Zhang et al., 2024).

Zusammenfassend lässt sich sagen, dass Concatenation in diesem Setting die robusteste und geeignetste Fusionsmethode ist. Diese Variante verzichtet auf zusätzliche Gewichtungen oder Annahmen, wodurch sie gegenüber komplexeren Ansätzen weniger anfällig für Verzerrungen oder Überanpassung ist. Solange komplexere Ansätze nicht gezielt optimiert und detailliert an die verwendeten Daten angepasst werden, erweist sich Concatenation als die präziseste Fusionsmethode.

### 7.1.3 Vergleich mit bestehenden Forschungsergebnissen

Vergleicht man nun die in dieser Arbeit erzielten Resultate mit bestehenden Forschungsarbeiten wird deutlich, dass der entwickelte multimodale Deep-Learning-Ansatz einige Überschneidungen und äquivalente Ergebnisse aufweist, zugleich jedoch Besonderheiten des brasilianischen Kontexts hervorhebt.

Die erzielten Modellmetriken liegen in einem ähnlichen Bereich, wie es auch vergleichbare Studien erreicht haben oder erzielten sogar bessere Resultate (Stark et

al. 2025; Wu et al. 2024). Diverse Forschungsergebnisse belegen außerdem, dass die Wirtschaftsleistung vor allem im urbanen Raum zuverlässig erfasst werden kann, während ländlich und landwirtschaftlich geprägte Regionen deutlich schwieriger zu modellieren sind und in diesen Gebieten die Modellgüte stark abnimmt (Wang et al. 2019; Chen und Nordhaus 2019). Die Gründe hierfür sind meist geringe oder nicht vorhandene Werte in den Inputdaten wie z.B. bei der Lichtemission, oder OSM-POIs, eine hohe saisonale Variabilität (Chen und Nordhaus 2019) sowie sehr heterogene Verteilungsmuster. Die in dieser Arbeit beobachtete schwache Modellgüte für den primären Sektor deckt sich damit vollständig mit bestehenden Forschungsergebnissen und auch die in diesem Projekt verwendete Modellarchitektur konnte keine verbesserten Resultate erzielen.

Des Weiteren stellt der in dieser Arbeit verfolgte multimodale Ansatz generelle Entwicklungen in diesem Forschungsbereich wieder. Eine Vielzahl aktueller Veröffentlichungen betont, dass die Kombination unterschiedlicher Datenquellen die Modellgüte signifikant steigert und der Verwendung einer einzelnen Inputquelle vorzuziehen ist (Li et al. 2022; Stark et al. 2025). Weitere Arbeiten wie Sun et al. (2020), Yue et al. (2014) oder Wu et al. (2024) zeigen ebenfalls, dass die Fusion multimodaler Geodaten wie NTL, Land Cover oder POI-Daten komplementäre Informationen verknüpft und damit die Modellqualität messbar steigert. Dieses Muster spiegelt sich auch in den erzielten Resultaten der Arbeit wider und sollte in zukünftigen Arbeiten entsprechend berücksichtigt werden.

Damit verdeutlichen die Vergleichsstudien, dass die Ergebnisse dieser Arbeit sowohl methodisch als auch inhaltlich gut mit der aktuellen Forschung übereinstimmen und zudem aufzeigen, wo Herausforderungen in diesem Themenbereich liegen.

Einen starken Einfluss auf die Modellqualität hat jedoch auch die zugrundeliegende Datenqualität. Ein Modell kann nur so gut sein wie es ihm seine Inputdaten erlauben. In diversen Studien wird gezeigt, dass die Qualität und Vollständigkeit der Inputdaten einen erheblichen Einfluss auf die Modellgüte haben. Beispielsweise können Unterschiede in der OSM-Datenabdeckung die Leistungsfähigkeit der Modelle stark beeinflussen. Aus diesem Grund wird im folgenden Kapitel genauer auf die Qualität der verwendeten Datengrundlagen eingegangen.

## 7.2 Datenqualität

Dieses Unterkapitel beschäftigt sich nun mit der Qualität der Inputdaten, wobei das Hauptaugenmerk dabei auf den OSM-Daten liegt. Besonders in Gebieten mit wenig verfügbaren Daten oder ungleichmäßiger Verteilung können die Ergebnisse stark von der Datenqualität beeinflusst sein. Da die grundlegenden Sentinel-2 und Black Marble Daten vom DLR zur Verfügung gestellt wurden und es sich auch bei den Land Cover Daten um Datensätze der ESA handelt, kann hier von einer gewissen Grundgenauigkeit ausgegangen werden. Das größte Problem im Hinblick auf die Datenqualität stellt somit OSM dar. Da es sich hier um ein kooperatives und frei verfügbares Kartenprojekt handelt, müssen zu den damit einhergehenden Vorteilen wie der Datenverfügbarkeit in Regionen mit wenigen amtlichen Geodaten auch die Nachteile berücksichtigt werden. Da Nutzer:innen weltweit Geodaten erfassen, bearbeiten und aktualisieren können ist generell darauf zu achten, dass verschiedene Personen anders interpretieren. Folglich kann es zu inkonsistenten Daten kommen, beispielsweise bei der Benennung oder Zuweisung verschiedener Kategorien, aber auch durch unterschiedliche Detailgrade bei der Digitalisierung (Han et al. 2024, S. 20).

Diverse Studien über die Datenqualität von OpenStreetMap (OSM) zeigen, dass die Qualität der Kartendaten auch stark von der regionalen Wirtschaftslage beeinflusst ist. So lassen sich beispielsweise in China deutliche Unterschiede zwischen dem wirtschaftlich stärker entwickelten Osten und dem strukturschwächeren Westen festhalten. Östlich gelegene Großstädte wie Peking oder Shanghai weisen dabei eine deutlich höhere Datenqualität auf. Gleichzeitig zeigen lokale Analysen auf Rasterebene, dass auch in Gebieten mit generell niedriger Datenqualität vereinzelt Zellen mit hoher Genauigkeit vorkommen können. Diese lokalen Abweichungen deuten darauf hin, dass sich die Datenqualität auf kleineren Skalen differenzierter darstellt. (Han et al. 2024)

Die Untersuchung von Camboim et al. aus dem Jahre 2015 fokussiert sich auf die Vollständigkeit und Aktualität der OSM-Daten im brasilianischen Kontext. Dabei wird deutlich, dass die Abdeckung und Qualität der Daten stark mit dem Urbanisierungsgrad sowie sozioökonomischen Indikatoren korrelieren. Besonders städtische Gebiete wie z.B. São Paulo weisen eine merklich höhere Dichte an erfassten OSM-Objekten auf, was auf eine stärkere Nutzeraktivität in diesen Regionen

hinweist. Ländliche Regionen hingegen sind deutlich schwächer abgedeckt, was die Gefahr einer kartographischen Segregation verdeutlicht (Camboim et al. 2015).

Allerdings betonen die Autor:innen, dass OSM-Daten in Brasilien prinzipiell eine wertvolle Ergänzung oder sogar unumgängliche Alternative zu amtlichen Kartenwerken darstellen können. Der Grund hierfür sind die oft unzureichenden offiziellen Kartierungen auf lokaler Ebene oder die Verfügbarkeit offizieller Geodatenätze. Dabei sei jedoch entscheidend, die Unterschiede in der Datenqualität sichtbar zu machen, um eine fundierte Nutzung zu ermöglichen. Die Studie unterstreicht daher die Notwendigkeit, Qualität und Vollständigkeit der Daten systematisch zu analysieren (Camboim et al. 2015).

Aus diesem Grund werden die in dieser Arbeit verwendeten OSM-Daten stichprobenartig mit Referenzdaten verglichen. Die Validierung konzentrierte sich dabei auf den Vergleich von Punktdaten. Da kein vergleichbarer offizieller Datensatz für die verwendeten POIs gefunden werden konnte, wurden hierfür stellvertretend Bildungsstandorte analysiert. Zu diesem Zweck wurden OSM-Daten zu Bildungseinrichtungen mit amtlichen Schulstandorten aus der brasilianischen Geodatenbibliothek geobr (Stand 2023) verglichen. Als Untersuchungsgebiet diente dabei die Stadt Porto Alegre.

Für die Analyse wurden zunächst OSM-Punkte mit relevanten Tags wie: amenity=school, university, kindergarten oder education=\* aus dem Stadtgebiet extrahiert. Ergänzend wurde eine Namensfilterung durchgeführt, um zusätzliche, potenziell relevante, Objekte zu identifizieren. Hiermit sollen unvollständige oder inkonsistente Tags ebenfalls in die Analyse miteinbezogen werden. Für diese Filterung wurden Schlüsselwörter (portugiesisch) wie „escola“, „instituto“, etc. verwendet und vorhandene Polygon-Geometrien wurden zu Punktdaten konvertiert, um eine einheitliche Vergleichsbasis zu erhalten. Nach Entfernung redundanter Geometrien wurde abschließend eine kombinierte Punktmenge der OSM-Daten erstellt.

Der Vergleich mit den geobr-Punktdaten erfolgte dabei mit zwei unterschiedlichen Varianten. Bei der ersten Methode wurde ein Buffer-basierter spatial join (100 m) durchgeführt, wobei Überschneidungen zwischen beiden Punktmengen als Übereinstimmungen („True Positives“) gewertet wurden. Die zweite Testvariante verwendete eine KDTree-basierte Nearest-Neighbor-Analyse, bei dem für jeden Punkt

der jeweils nächstgelegene Punkt der anderen Datenquelle innerhalb von 100 m gesucht wurde.

Mithilfe des 100m Abstands/Buffer sollen mögliche Ungenauigkeiten innerhalb, aber auch zwischen den Datensätzen wie CRS-Konvertierungen, Digitalisierungsunterschiede und weitere Abweichungen ausgemerzt werden.

Für beide Methoden können klassische Qualitätsmetriken wie Precision, Recall oder der F1-Score berechnet werden, um sowohl die Abdeckung als auch die Genauigkeit der OSM-Daten gegenüber den Referenzdaten zu quantifizieren. Durch die Verwendung von zwei unterschiedlichen Varianten sollen ebenfalls mögliche Verzerrungen gemindert werden. Für ein besseres Verständnis dieser Validierung wurde die folgende Abbildung 34 angefertigt, die den Buffer-basierten Spatial Join verdeutlichen soll.

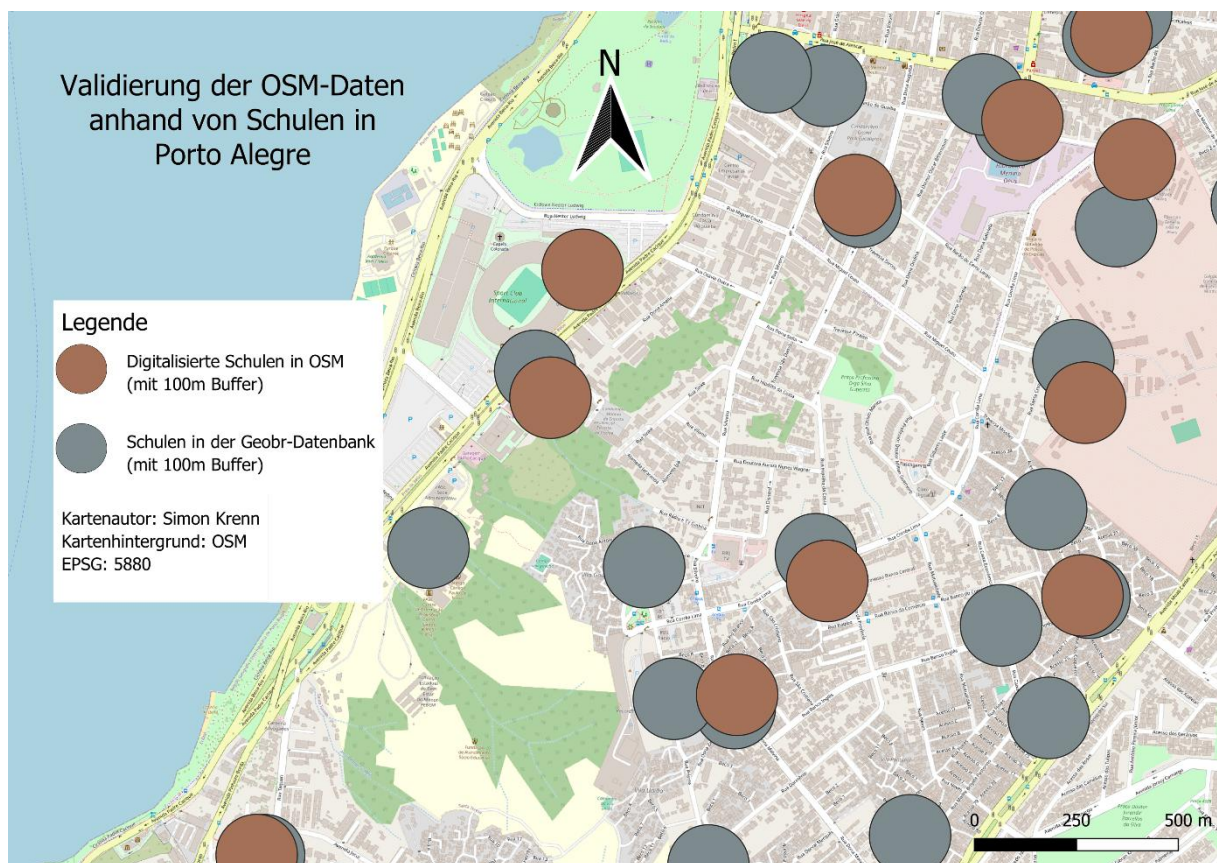


Abb. 34: Der Vergleich der digitalisierten Schulen in OSM mit den Schulen aus der offiziellen Geobr-Datenbank in einem Teil von Porto Alegre. Quelle: Eigener Entwurf

Die Karte zeigt die räumliche Verteilung der als Schulen erfassten Punktdaten aus OSM (braun dargestellt) und der geobr-Bibliothek (in grau visualisiert) in einem Teil von Porto Alegre. Deutlich erkennbar ist, dass ein Großteil der OSM-Schulen innerhalb

eines 100-m-Buffers mit den geobr-Standorten überlappt und somit korrekt erfasst wurde. Gleichzeitig wird sichtbar, dass die offizielle geobr-Datenbank insgesamt mehr Schulstandorte enthält als in OSM verfügbar sind.

Der abschließende Vergleich der Metriken der beiden getesteten Varianten für das gesamte Stadtgebiet von Porto Alegre ist nun in der folgenden Abbildung 35 dargestellt und ermöglicht eine Quantifizierung der Daten im Untersuchungsgebiet.

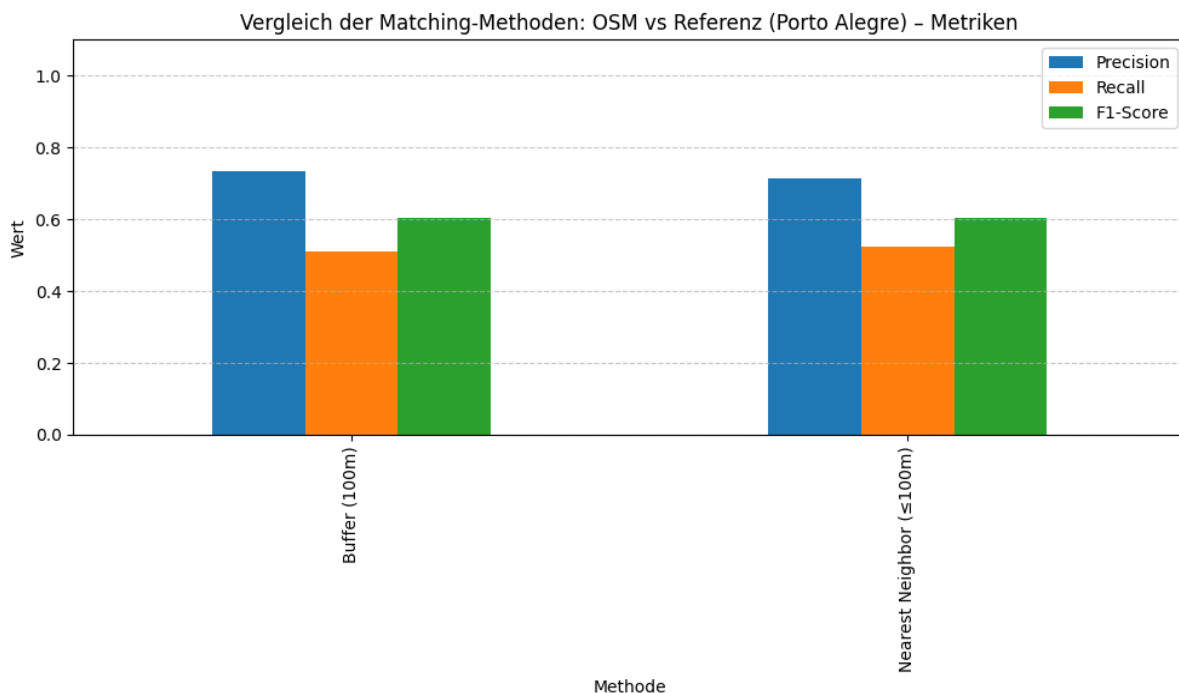


Abb. 35: Vergleich der Metriken der verwendeten Matching-Methoden im gesamten Stadtgebiet von Porto Alegre. Quelle: Eigener Entwurf.

Dabei wird erkennbar, dass beide Verfahren sehr ähnliche Ergebnisse aufweisen. Die Precision liegt jeweils bei rund 0,72, was bedeutet, dass etwa 72% der als übereinstimmend identifizierten OSM-Schulen tatsächlich einer geobr-Schule entsprechen. Der Recall beträgt bei beiden Methoden etwa 0,50, was bedeutet, dass rund die Hälfte der geobr-Schulen auch in OSM vorkommt. Der F1-Score liegt folglich bei ungefähr 0,60. Insgesamt zeigen die Ergebnisse, dass beide Matching-Ansätze vergleichbare Leistungen erbringen und einige der geobr-Schulen in OSM digitalisiert sind, jedoch auch viele Bildungsstandorte fehlen. Insgesamt wurden im gewählten Untersuchungsgebiet 601 OSM-Schulen gefunden, im geobr-Referenzdatensatz befinden sich innerhalb dieses Untersuchungsgebietes 843 Schulen.

Insgesamt lässt sich die Qualität der verwendeten OSM-Daten als solide, jedoch ausbaufähig einschätzen. Die Übereinstimmung der OSM-Punkte mit den amtlichen geobr-Daten zeigt, dass ein Großteil der in OSM erfassten Schulen korrekt digitalisiert und räumlich korrekt verortet ist. Gleichzeitig verdeutlichen die Metriken, dass OSM in Porto Alegre keine vollständige Abdeckung bietet und einige der offiziell verzeichneten Bildungsstandorte nicht enthält. Für den Einsatz in dieser Arbeit bedeutet dies, dass OSM-Daten als wertvoller Zusatz, aber nicht als alleinige flächendeckende Informationsquelle betrachtet werden sollten. Die Stärke der Daten liegt darin, vor allem ungenaue oder nicht vorhandene amtliche Geodaten wie die POIs oder das Straßennetz trotzdem flächendeckend verfügbar zu machen, die dabei vorhandenen Ungenauigkeiten müssen jedoch in Kauf genommen werden.

## 7.3 Limitationen

Im Letzten Unterkapitel der Diskussion wird abschließend auf Limitationen dieser Arbeit eingegangen. Dabei sollen diverse Grenzen des verwendeten Workflows aufgezeigt und mögliche Verbesserungsvorschläge geben werden.

### 7.3.1 Modelllimitationen

Zunächst werden Limitationen beschrieben, die einen direkten Einfluss auf das Modell selbst haben, sowie vielversprechende Ansätze erwähnt, die bei weiteren Forschungen in diesem Bereich angewandt werden können.

Ein limitierender Faktor betrifft die zeitlichen Unterschiede der verwendeten Input- und Referenzdaten. Aufgrund der Vielzahl heterogener Datensätze ist eine vollständige zeitliche Abstimmung der Daten in der Praxis kaum realisierbar. Derart detaillierte Referenzdaten werden selten erstellt und auch die verwendeten Inputdaten wie der vorprozessierte WorldCover Datensatz oder die implementierten OSM-Daten sind nur mit temporaler Verzögerung oder spärlicher Zeitangabe nutzbar.

Wie Chen und Nordhaus in ihrem Paper 2019 hervorheben, können diese unterschiedlichen Aufnahmezeitpunkte oder Verzögerungen in amtlichen Wirtschaftsstatistiken zu systematischen Verzerrungen führen und folglich die Modellgüte beeinträchtigen. Diese Problematik kann jedoch nur schwer komplett bereinigt werden. Künftige Arbeiten sollten daher prüfen, inwieweit eine zeitliche Abstimmung der Inputdaten umsetzbar ist bzw. diese so gering wie möglich halten, um die daraus resultierenden Limitationen so gut wie möglich zu reduzieren.

Ein weiterer Faktor ist die bereits angesprochene Disaggregation der Referenzdaten. Mithilfe dieser ist es möglich, die Daten für eine subregionale Analyse aufzubereiten, wie in der Arbeit jedoch deutlich wurde, sind die daraus resultierende Nachteile nicht außer Acht zu lassen.

Die disaggregierten Referenzwerte vielen häufig deutlich niedriger aus, als es die tatsächlichen Strukturen der Rasterzellen vermuten ließen. Dies liegt überwiegend an den in Abbildung 32 dargestellten Grenzen der ursprünglich verwendeten Referenzpolygone. Es befinden sich nur sehr kleine Teile des Referenzpolygons innerhalb der Zelle. Aufgrund dieser geringen Fläche werden der Zelle, während der Disaggregation, sehr niedrige Referenzwerte zugewiesen. Im weiteren Verlauf der Arbeit wird jedoch auf die Daten der gesamten Zelle zugegriffen. Der Rest der Zelle kann möglicherweise hohe Lichtemissionen aufweisen oder durch dicht besiedeltes Gebiet geprägt sein. Folglich treten dadurch Verzerrungen innerhalb des Modells auf. Die Grundproblematik beginnt dabei aber bereits bei den ursprünglichen Referenzdaten.

Wie Yue et al. (2014) hervorheben, führt die Disaggregation sozioökonomischer Daten häufig zu systematischen Verzerrungen. Die Daten beruhen meist auf groben administrativen Einheiten und es wird eine homogene Verteilung innerhalb dieser Flächen angenommen. Die reale räumliche Struktur wirtschaftlicher Aktivitäten ist jedoch meist stark heterogen. Auch Wurm und Taubenböck (2018) zeigen anhand einkommensbasierter Analysen in Brasilien, dass sozioökonomische Merkmale innerhalb formaler Verwaltungseinheiten stark variieren und daher nicht durch einheitliche Flächenwerte adäquat repräsentiert werden können.

Aus diesem Grund sollten sich ähnliche Forschungen in diesem Bereich eingehend mit der Disaggregation, aber auch den auftretenden „Randeffekten“ auseinandersetzen. Man könnte beispielsweise die verwendete Disaggregation so anpassen, dass nicht vollständig abgedeckte Zellen mit zusätzlichen Referenzwerten außerhalb des Untersuchungsgebietes abgedeckt werden (insofern diese vorhanden sind). Eine weitere Möglichkeit wäre die Hochskalierung der Referenzwerte in betroffenen Zellen, dies kann jedoch bei sehr ruralen Zellen (z.B. nur aus Wald- oder Brachflächen bestehend) zu zusätzlichen Verzerrungen führen. Deshalb sollte diese Variante nur unter Einbindung zusätzlicher Informationen verwendet werden, beispielsweise durch eine anteilsbasierte Verteilung anhand der Land-Cover-Klassen.

Allerdings führt die hierfür notwendige Definition passender Schwellenwerte zu zusätzlichen potenziellen Ungenauigkeiten. Ebenfalls könnten die bereits erwähnten Filter, basierend auf niedrigen Werten der Zielvariablen oder sehr kleinen Flächen der Ursprungspolygone, verwendet werden. Damit könnten „verzerrende“ Zellen von der gesamten Modellierung ausgeschlossen werden und somit die Modellgüte verbessern.

Es gibt jedoch auch noch weitere Limitationen, die innerhalb des Modells auftreten. Bezüglich des Encoders muss erwähnt werden, dass ResNet ursprünglich für dreikanalige RGB-Bilder aus dem ImageNet-Datensatz konzipiert wurde. Diese werden zwar in einigen Quellen in diesem Themenbereich eingesetzt (Adegun et al. 2023), dennoch könnte sich das Verwenden anderer Encoder positiv auf die Modellgüte auswirken.

Eine weitere Möglichkeit zur Verbesserung der Ergebnisse sind einfache Datenaugmentierungen wie beispielsweise spiegeln oder rotieren. Diese Transformationen variieren die Eingangsdaten geringfügig und fördern dadurch die Generalisierung des Modells und reduzieren das Risiko von Overfitting (Shorten und Khoshgoftaar 2019). Testweise wurden für das BIP und den primären Sektor einfache Datenaugmentierungen (spiegeln und rotieren) durchgeführt, womit bereits kleinere Verbesserungen der Modellgüte erreicht werden konnten (Abb. 36 und 37).

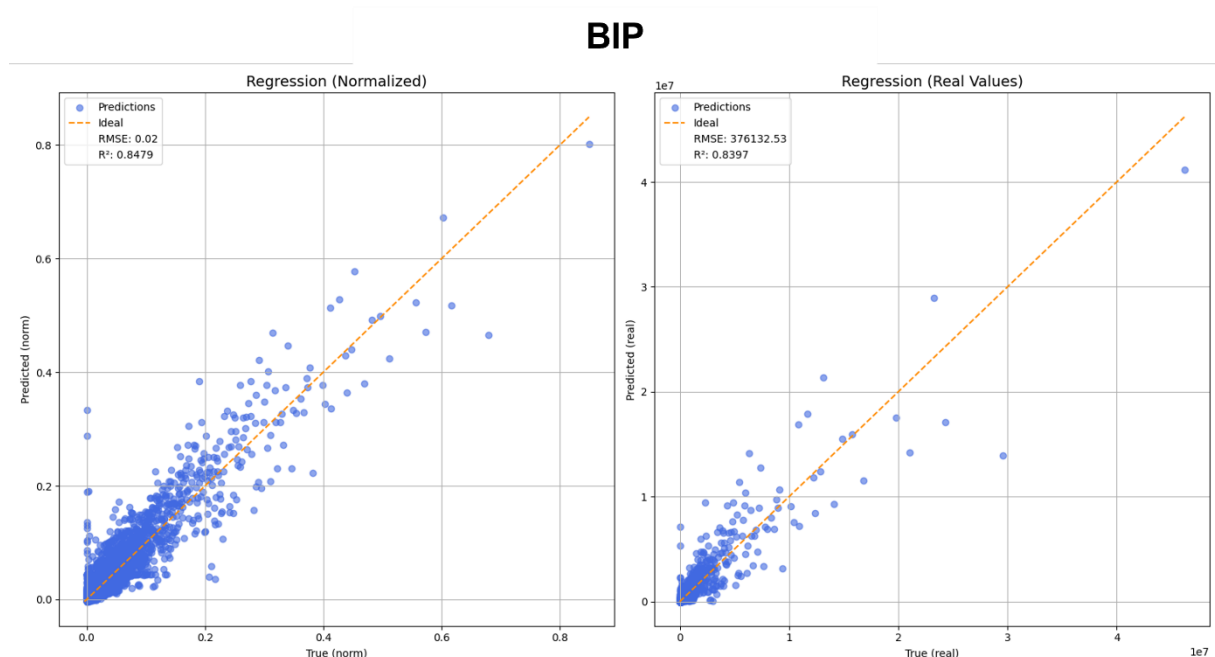


Abb. 36: Regressionsplot der normalisierten und denormalisierten Werte des BIP mit einfacher Datenaugmentierung. Quelle: Eigener Entwurf.

## Primärer Sektor

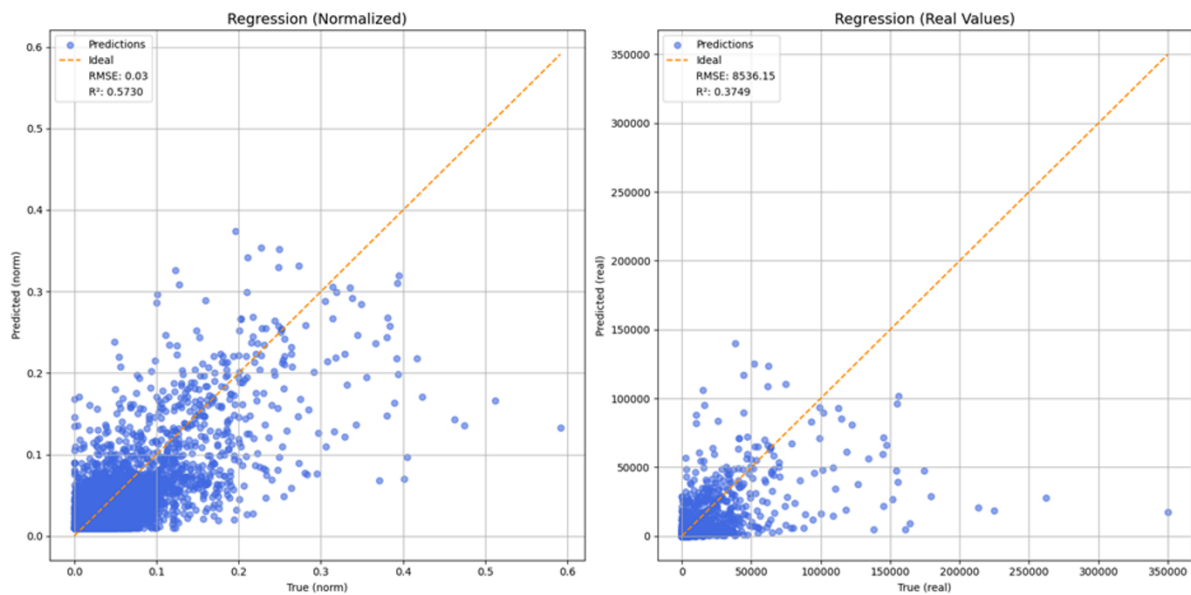


Abb. 37: Regressionsplot der normalisierten und denormalisierten Werte des primären Sektors mit einfacher Datenaugmentierung. Quelle: Eigener Entwurf.

Der Vergleich mit den Ergebnissen ohne Datenaugmentierung (siehe Kapitel 6.1) zeigt, dass bereits einfache Augmentierungsmethoden zu geringen Verbesserungen führen. Umso mehr ist zu erwarten, dass komplexere oder speziell auf die Daten zugeschnittene Augmentierungen die Modellgüte noch deutlich erhöhen könnten. Deshalb ist ihr Einsatz in zukünftigen Studien definitiv zu empfehlen.

### 7.3.2 Feature Importance Analysis

Ein weiterer Aspekt, wodurch die Resultate optimiert werden könnten, ist die Durchführung einer Feature Importance Analysis. Eine solche Analyse gewährt zusätzliche Einblicke in die relative Bedeutung der verwendeten Variablen und ermöglicht somit die Modellentscheidungen besser nachvollziehen zu können. Dieses Verfahren untersucht, welchen Beitrag einzelne Variablen zur Gesamtleistung eines Modells leisten und in welchem Ausmaß sie die Resultate beeinflussen. Durch diese Analyse wird sichtbar, welche Merkmale für die Generierung der Vorhersagen besonders relevant sind und welche nur eine untergeordnete Rolle spielen (Wurm et al. 2021).

Aus diesem Grund stellt die Integration einer Feature Importance Analysis eine spannende Möglichkeit für zukünftige Arbeiten dar, die Resultate noch eingehender analysieren zu können. Diese Methode könnte dazu beitragen, die Entscheidungen des Modells noch besser verstehen zu können und sie verdeutlicht, welche Datensätze

für das Modell besonders relevant sind und welche möglicherweise anders aufbereitet oder zusätzlich entfernt werden sollten.

Eine Feature-Importance-Analyse könnte ebenso aufzeigen, welche Datensätze (z. B. Straßennetz oder POIs) für welche Sektoren wichtiger sind oder ob Nachlichtdaten, wie in der Literatur beschrieben, vor allem beim BIP und dem tertiären Sektor entscheidend sind. Ebenso könnte man den primären Sektor weiter analysieren und es könnte getestet werden, ob beispielsweise nur die Verwendung der Sentinel-2 Daten sowie der WorldCover Daten vielversprechendere Resultate liefern würde. Dies wäre auch für Modelle mit begrenzter Rechenkapazität von Bedeutung, um nur relevante Daten pro Sektor zu verwenden und folglich die Modellkomplexität zu verringern.

Insgesamt trägt eine solche Analyse dazu bei, die Modellarchitektur präziser auf die Zielvariablen abstimmen zu können und die Interpretierbarkeit der Ergebnisse weiter zu erhöhen.

## 8. Fazit

Das Ziel dieser Arbeit war es, das Bruttoinlandsprodukt sowie die Wirtschaftssektoren auf subregionaler Ebene in Brasilien mithilfe multimodaler Geodatenfusion und eines Deep-Learning-Modells zu ermitteln. Die Ergebnisse zeigen, dass der entwickelte Workflow zuverlässige Resultate erzielen konnte. Insbesondere das BIP sowie der tertiäre Sektor konnten mit einer hohen Genauigkeit ermittelt werden, während der primäre Sektor, wie auch in zahlreichen vergleichbaren Studien beschrieben, deutlich stärkere Abweichungen aufweist.

Die Arbeit demonstriert damit den Mehrwert der multimodalen Datenfusion, um heterogene Daten wie Sentinel-2, Black Marble, WorldCover und OSM kombinieren zu können und damit wirtschaftliche Aktivität abzubilden. Gleichzeitig hat sich gezeigt, dass in diesem Fall weniger komplexe Fusionsmethoden wie Concatenation die besten und konstantesten Ergebnisse liefern, während komplexere Verfahren nur in bestimmten Fällen von Vorteil sind. Zukünftige Arbeiten könnten die verwendeten Fusionsmethoden jedoch noch weiterentwickeln oder zusätzliche Ansätze testen. Vor allem die komplexeren Methoden könnten noch stärker auf die vorhandene Datengrundlage angepasst werden und dadurch möglicherweise verbesserte Resultate erzielen.

Trotz der positiven Ergebnisse gibt es auch ein paar Limitationen und Verbesserungsmöglichkeiten innerhalb der Arbeit. Die Disaggregation der Referenzdaten, aber auch die Qualität und Abdeckung der Inputdaten (vor allem OSM) beeinflussen die Modellgüte und verdeutlichen die Notwendigkeit einer sorgfältigen Datenaufbereitung. Eine weitere Verbesserungsmöglichkeit liegt zudem bei der Anwendung einer Feature-Importance-Analyse. Diese könnte aufzeigen, welche Datenquellen tatsächlich die größten Beiträge zur Modellleistung liefern und in welchen Bereichen Inputdaten redundant sind.

Weitere Forschungen in diesem Bereich könnten sich ebenfalls mit der Integration zusätzlicher Daten beschäftigen. Einerseits wäre die Verwendung von weiteren sozioökonomischen Daten denkbar, andererseits gibt es noch weitere Geodaten, die implementiert werden könnten. Beispiele hierfür wären die Verwendung von weiteren aus OSM abgeleiteten Datensätzen wie etwa eine Density Map mit Gebäudedaten als Grundlage oder die Verwendung von Höhenmodellen (Xu et al. 2024).

Darüber hinaus wäre eine Validierung des Modells in anderen geographischen Regionen oder auf unabhängigen Datensätzen sinnvoll, um die Übertragbarkeit des Ansatzes zu prüfen. Die entwickelte Architektur bietet dafür eine solide Grundlage und könnte nach gezielten Anpassungen und Weiterentwicklungen auch auf größeren Ebenen, wie beispielsweise für mehrere Staaten oder globale Analysen, eingesetzt werden. Dabei müssen jedoch regionale Unterschiede wie Beleuchtungsinfrastruktur, landnutzungsspezifische Muster, usw. berücksichtigt werden.

Insgesamt zeigt die Arbeit das hohe Potenzial von Deep Learning basierten Ansätzen zur multimodalen Datenfusion und subregionalen Ermittlung der Wirtschaftsleistung. Darauf aufbauend können noch weiterführende Analysen durchgeführt werden, wie etwa die zuvor genannten Feature Importance Analysis. Gleichzeitig bietet der entwickelte Ansatz eine praxisnahe Grundlage, die sich durch gezielte Anpassungen auf andere Regionen, Staaten oder verwandte Fragestellungen übertragen lässt und dadurch wertvolle Impulse für zukünftige Forschungsarbeiten liefert.

## Literaturverzeichnis

- Adegun, A. A.; Viriri, S.; Tapamo, J. R. (2023): Review of deep learning methods for remote sensing satellite images classification: experimental survey and comparative analysis. In: *Journal of Big Data* 10, 93.
- Bolivar, O. (2024): GDP nowcasting: A machine learning and remote sensing data-based approach for Bolivia. In: *Latin American Journal of Central Banking* 5, 3, S. 100126.
- Camboim, S. P.; Bravo, J. V. M.; Sluter, C. R. (2015): An Investigation into the Completeness of, and the Updates to, OpenStreetMap Data in a Heterogeneous Area in Brazil. In: *ISPRS International Journal of Geo-Information* 4, 3, S. 1366-1388.
- Chango, W.; Lara, J. A.; Cerezo, R.; Romero, C. (2022): A review on data fusion in multimodal learning analytics and educational data mining. In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 12, 4, e1458.
- Chen, X.; Nordhaus, W. D. (2011): Using luminosity data as a proxy for economic statistics. In: *Proceedings of the National Academy of Sciences* 108, 21, S. 8589-8594.
- Chen, X.; Nordhaus, W.D. (2019): VIIRS Nighttime Lights in the Estimation of Cross-Sectional and Time-Series GDP. In: *Remote Sensing* 11, 9, S. 1057.
- Chen, F.; Datta, G.; Kundu, S.; Beerel, P. A. (2022): Self-Attentive Pooling for Efficient Deep Learning. arXiv preprint arXiv:2209.07659.
- Chen, Y.; Wu, G.; Ge, Y.; Xu, Z. (2022): Mapping Gridded Gross Domestic Product Distribution of China Using Deep Learning With Multiple Geospatial Big Data. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 15, S. 1791–1802.
- Chen, T.; Zhang, Z.; Jaiswal, A.; Liu, S.; Wang, Z. (2023): Sparse MoE as the New Dropout: Scaling Dense and Self-Slimmable Transformers. arXiv preprint arXiv:2303.01610.

- Chen, X.; Li, Y.; Zhang, W. (2025): Macroeconomic indicator selection for GDP forecasting using deep learning methods. In: *Sustainability* 17, 3, 843.
- D'Souza, N. S.; Wang, H.; Giovannini, A.; Foncubierta-Rodríguez, A.; Beck, K. L.; Boyko, O. B.; Syeda-Mahmood, T. (2024): Fusing modalities by multiplexed graph neural networks for outcome prediction from medical data and beyond. In: *Medical Image Analysis* 93, 103064.
- Donaldson, D.; Storeygard, A. (2016): The View from Above: Applications of Satellite Data in Economics. In: *Journal of Economic Perspectives* 30, 4, S. 171–198.
- Dynan, K. E.; Sheiner, L. M. (2018): *GDP as a Measure of Economic Well-Being*. Hutchins Center Working Paper No. 43. Washington, D.C.: The Brookings Institution.
- Elvidge, C. D.; Baugh, K. E.; Kihn, E. A.; Kroehl, H. W.; Davis, E. R.; Davis, C. W. (1997): Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. In: *International Journal of Remote Sensing* 18, 6, S. 1373-1379.
- Esch, T.; Heldens, W.; Metz, A. (2015): Die Erde im Bild – Satelliten als Werkzeug zur Beobachtung der Landoberfläche. In: Taubenböck, H.; Wurm, M.; Esch, T.; Dech, S. (Hg.): *Globale Urbanisierung: Perspektive aus dem All*. Berlin, Heidelberg: Springer Spektrum, S. 23-27.
- Ghamisi, P.; Rasti, B.; Yokoya, N.; Wang, Q.; Hofle, B.; Bruzzone, L.; Bovolo, F.; Chi, M.; Anders, K.; Gloaguen, R.; Atkinson, P. M.; Benediktsson, J.A. (2019): Multisource and Multitemporal Data Fusion in Remote Sensing: A Comprehensive Review of the State of the Art. In: *IEEE Geoscience and Remote Sensing Magazine* 7, 1, S. 6–39.
- Goodfellow, I.; Bengio, Y.; Courville, A. (2016): *Deep Learning*. Cambridge, MA: MIT Press.
- Han, C.; Lu, B.; Zheng, J.; Yu, D.; Zheng, S. (2024): Research on multiscale OpenStreetMap in China: data quality assessment with EWM-TOPSIS and GDP modeling. In: *Geo-spatial Information Science*, S. 1–25.

- He, H.; Garcia, E. A. (2009): Learning from imbalanced data. In: IEEE Transactions on Knowledge and Data Engineering 21, 9, S. 1263-1284.
- He, K.; Zhang, X.; Ren, S.; Sun, J. (2016): Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), S. 770-778.
- Hong, D.; Gao, L.; Yokoya, N.; Yao, J.; Chanussot, J.; Du, Q.; Zhang, B. (2021): More Diverse Means Better: Multimodal Deep Learning Meets Remote-Sensing Imagery Classification. In: IEEE Transactions on Geoscience and Remote Sensing 59, 5, S. 4340–4354.
- Huang, L.; Qin, J.; Zhou, Y.; Zhu, F.; Liu, L.; Shao, L. (2020): Normalization Techniques in Training DNNs: Methodology, Analysis and Application. arXiv preprint arXiv:2009.12836.
- Janga, B.; Agarwal, S.; Gupta, R. (2023): A review of practical AI for remote sensing in Earth observation: Trends, challenges, and applications. In: Remote Sensing 15, 16, 4112.
- Jenks, G. F. (1967): The Data Model Concept in Statistical Mapping. In: International Yearbook of Cartography 7, S. 186-190.
- Kang, B.; Xie, S.; Rohrbach, M.; Yan, Z.; Feng, J.; Yuille, A. (2020): Decoupling representation and classifier for long-tailed recognition. In: International Conference on Learning Representations (ICLR).
- Kohavi, R. (1995): A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI), S. 1137-1143.
- Li, X.; Zhou, Y.; Zhao, M.; Zhao, X. (2020): A harmonized global nighttime light dataset 1992-2018. In: Scientific Data 7, 168.
- Li, M.; Wang, F.; Jia, X.; Li, W.; Li, T.; Rui, G. (2021): Multi-source data fusion for economic data analysis. In: Neural Computing and Applications 33, 10, S. 4729–4739.

- Li, J.; Hong, D.; Gao, L.; Yao, J.; Zheng, K.; Zhang, B.; Chanussot, J. (2022): Deep learning in multimodal remote sensing data fusion: A comprehensive review. In: IEEE Transactions on Geoscience and Remote Sensing 60, S. 1-34.
- Liu, Y.; He, L.; Liu, W.; Liu, J. (2018): Exploring a Unified Attention-Based Pooling Framework for Speaker Verification. In: Proceedings of the International Symposium on Chinese Spoken Language Processing (ISCSLP), S. 149-153.
- Liu, Z.; Miao, Z.; Zhan, X.; Wang, J.; Gong, B.; Yu, S. X. (2019): Large-Scale Long-Tailed Recognition in an Open World. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, S. 2537-2546.
- Liu, P.; Li, J.; Wang, L.; He, G. (2022): Remote Sensing Data Fusion With Generative Adversarial Networks: State-of-the-art methods and future research directions. In: IEEE Geoscience and Remote Sensing Magazine 10, 2, S. 295–328.
- Liu, Y.; Wu, C.; Guan, M.; Wang, J. (2025): MixtureRS: A Mixture of Expert Network Based Remote Sensing Land Classification. In: Remote Sensing 17, 14, 2494.
- McFeeters, S. K. (1996): The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. In: International Journal of Remote Sensing 17, 7, S. 1425-1432.
- Moradi, M.; Roche, S.; Mostafavi, M. A. (2022): Exploring five indicators for the quality of OpenStreetMap road networks: A case study of Québec, Canada. In: Geomatica 75, 4, S. 178-208.
- Murakami, D.; Yamagata, Y. (2019): Estimation of gridded population and GDP scenarios with spatially explicit statistical downscaling. In: Sustainability 11, 7, 2106.
- Nair, V.; Hinton, G. E. (2010): Rectified Linear Units Improve Restricted Boltzmann Machines. In: Proceedings of the 27th International Conference on Machine Learning (ICML), S. 807-814.
- Pawłowski, M.; Wróblewska, A.; Sysko-Romańczuk, S. (2023): Effective Techniques for Multimodal Data Fusion. In: Sensors 23, 6, 3316.

- Price, N.; Atkinson, P.M. (2022): Global GDP Prediction With Night-Lights and Transfer Learning. In: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 15, S. 7128–7138.
- Rattanakaew, T. (2025): Using of Sentinel 2 Images for Tropical Economic Crops Growth Modeling. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLVIII-M-6-2025, S. 233-242.
- Román, M. O.; Wang, Z.; Sun, Q.; Kalb, V.; Miller, S. D.; Molthan, A.; Schultz, L.; Bell, J.; Stokes, E. C.; Pandey, B.; Seto, K. C.; Hall, D.; Oda, T.; Wolfe, R. E.; Lin, G.; Golpayegani, N.; Devadiga, S.; Davidson, C.; Sarkar, S.; Praderas, C.; Schmaltz, J.; Boller, R.; Stevens, J.; Ramos González, O. M.; Padilla, E.; Alonso, J.; Detrés, Y.; Armstrong, R.; Miranda, I.; Conte, Y.; Marrero, N.; MacManus, K.; Esch, T.; Masuoka, E. J. (2018): NASA's Black Marble nighttime lights product suite. In: Remote Sensing of Environment 210, S. 113-143.
- Salah, M. (2017): A survey of modern classification techniques in remote sensing for improved image classification. In: Journal of Geomatics 11, 1, S. 1-21.
- Sentinel Hub (Hg.) (o.J.): NDWI Normalized Difference Water Index. <https://custom-scripts.sentinel-hub.com/custom-scripts/sentinel-2/ndwi/>, zuletzt geprüft am 20.01.2026.
- Shazeer, N.; Mirhoseini, A.; Maziarz, K.; Davis, A.; Le, Q.; Hinton, G.; Dean, J. (2017): Outrageously Large Neural Networks: The Sparsely-Gated Mixture-of-Experts Layer. arXiv preprint arXiv:1701.06538.
- Shorten, C.; Khoshgoftaar, T. M. (2019): A survey on image data augmentation for deep learning. In: Journal of Big Data 6, 60.
- Stahlschmidt, S. R.; Ulfenborg, B.; Synnergren, J. (2022): Multimodal deep learning for biomedical data fusion: a review. In: Briefings in Bioinformatics 23, 2, bbab569.
- Stark, T.; Wurm, M.; Stokes, E.; Seto, K. C.; Taubenböck, H. (2025): GDP Estimation using a Deep Learning Fusion Model for Multi-Source Remote Sensing Data. In: 2025 Joint Urban Remote Sensing Event (JURSE 2025), S. 1-4.
- Sun, J.; Di, L.; Sun, Z.; Wang, J.; Wu, Y. (2020): Estimation of GDP Using Deep Learning With NPP-VIIRS Imagery and Land Cover Data at the County Level in

- CONUS. In: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 13, S. 1400–1415.
- Tassopoulos, D.; Kalivas, D.; Giovos, R.; Lougkos, N.; Priovolou, A. (2021): Sentinel-2 Imagery Monitoring Vine Growth Related to Topography in a Protected Designation of Origin Region. In: Agriculture 11, 8, 785.
- Thapa, S.; Chhetri, P. K.; Klein, A. G. (2019): Cross-Comparison between MODIS and VIIRS Snow Cover Products for the 2016 Hydrological Year. In: Climate 7, 4, 57.
- Van De Kerchove, R.; Zanaga, D.; Xu, P.; Tsendbazar, N. E.; Lesiv, M. (2022): WorldCover Product User Manual Version 2.0. Paris: European Space Agency. [https://esa-worldcover.s3.eu-central-1.amazonaws.com/v200/2021/docs/WorldCover\\_PUM\\_V2.0.pdf](https://esa-worldcover.s3.eu-central-1.amazonaws.com/v200/2021/docs/WorldCover_PUM_V2.0.pdf), zuletzt geprüft am 13.12.2025.
- Vargas-Muñoz, J.E.; Srivastava, S.; Tuia, D.; Falcão, A.X. (2021): OpenStreetMap: Challenges and Opportunities in Machine Learning and Remote Sensing. In: IEEE Geoscience and Remote Sensing Magazine 9, 1, S. 184–199.
- Wang, X.; Rafa, M.; Moyer, J. D.; Li, J.; Scheer, J.; Sutton, P. (2019): Estimation and Mapping of Sub-National GDP in Uganda Using NPP-VIIRS Imagery. In: Remote Sensing 11, 2, 163.
- Wiltsche, M. C. (2023): Ableitung des Bruttoinlandprodukts von Brasilien auf Basis von Nacht-Satellitenbildern und weiteren Geodaten durch Anwendung eines Deep-Learning-Modells. Masterarbeit. Technische Universität Graz, Graz.
- Wu, J.; Wang, Z.; Li, W.; Peng, J. (2013): Exploring factors affecting the relationship between light consumption and GDP based on DMSP/OLS nighttime satellite imagery. In: Remote Sensing of Environment 134, S. 111–119.
- Wu, N.; Yan, J.; Liang, D.; Sun, Z.; Ranjan, R.; Li, J. (2024): High-resolution mapping of GDP using multi-scale feature fusion by integrating remote sensing and POI data. In: International Journal of Applied Earth Observation and Geoinformation 129, S. 103812.

- Wurm, M.; Taubenböck, H. (2018): Detecting social groups from space – Assessment of remote sensing-based mapped morphological slums using income data. In: *Remote Sensing Letters* 9, 1, S. 41-50.
- Wurm, M.; Stark, T.; Zhu, X. X.; Weigand, M.; Taubenböck, H. (2019): Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 150, S. 59-69.
- Wurm, M.; Droin, A.; Stark, T.; Geiß, C.; Sulzer, W.; Taubenböck, H. (2021): Deep Learning-Based Generation of Building Stock Data from Remote Sensing for Urban Heat Demand Modeling. In: *ISPRS International Journal of Geo-Information* 10, 1, 23.
- Xu, Y.; Goodacre, R. (2018): On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. In: *Journal of Analysis and Testing* 2, 3, S. 249-262.
- Xu, N.; Zhang, S.; Jiang, S. (2024): Estimating GDP by Fusing Nighttime Light and Land Cover Data. In: *Canadian Journal of Remote Sensing* 50, 1, 2377641.
- Yeh, C.; Perez, A.; Driscoll, A.; Azzari, G.; Tang, Z.; Ferres, L.; Burke, M.; Lobell, D. B. (2020): Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. In: *Nature Communications* 11, 2583.
- Yue, W.; Gao, J.; Yang, X. (2014): Estimation of Gross Domestic Product Using Multi-Sensor Remote Sensing Data: A Case Study in Zhejiang Province, East China. In: *Remote Sensing* 6, 8, S. 7260–7275.
- Zanaga, D.; Van De Kerchove, R.; Daems, D.; De Keersmaecker, W.; Brockmann, C.; Kirches, G.; Wevers, J.; Cartus, O.; Santoro, M.; Fritz, S.; Lesiv, M.; Herold, M.; Tsendbazar, N. E.; Xu, P.; Ramoino, F.; Arino, O. (2022): ESA WorldCover 10 m 2021 v200. Zenodo. <https://doi.org/10.5281/zenodo.7254221>, zuletzt geprüft am 13.12.2025.

Zhang, Q.; Wei, Y.; Han, Z.; Fu, H.; Peng, X.; Deng, C.; Hu, Q.; Xu, C.; Wen, J.; Hu, D.; Zhang, C. (2024): Multimodal Fusion on Low-quality Data: A Comprehensive Survey. arXiv preprint arXiv:2404.18947.

Zhao, H.; Chen, R.; Li, P. (2024): Deep learning for multi-country GDP prediction with economic time series data. arXiv preprint arXiv:2409.02551.

Zhu, X. X.; Tuia, D.; Mou, L.; Xia, G.-S.; Zhang, L.; Xu, F.; Fraundorfer, F. (2017): Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. In: IEEE Geoscience and Remote Sensing Magazine 5, 4, S. 8-36.