Deep-Learning-Bildverarbeitung von 3D-Laserscanner-Aufnahmen für ein Weichenstopf-Assistenzsystem

Deep Learning Image Processing of 3D Laser Scanner Recordings for a Turnout Tamping Assistance System

Dr. techn. Gerald Zauner, Dipl.-Ing. Martin Bürger, Dipl.-Ing. Dr. techn. Florian Auer,
Dipl.-Ing. Tobias Müller, Linz (Österreich)

Zusammenfassung

Sicheres und bequemes Reisen im Zug ist nur auf Gleisen mit einer exakten geometrischen Gleislage möglich. Gleisstopfmaschinen dienen dieser Instandhaltungsaufgabe. Gerade der Weichenstopf-Vorgang ist ein komplexes Verfahren zur Verbesserung und Stabilisierung der Gleislagequalität in Weichen. In der Regel wird diese Tätigkeit deshalb auch von erfahrenen Bedienern durchgeführt. Im Beitrag wird der aktuelle Stand der Entwicklung eines 3D-Scanner-basierten Sensorsystems für ein neues Weichenstopf-Assistenzsystem beschrieben, das den Bediener in komplexen Stopfbereichen unterstützen und entlasten soll. Eine zentrale Aufgabe übernimmt in diesem Zusammenhang die digitale Bildverarbeitung, die auf der Basis von 3D-Scannerdaten die sogenannte semantische Segmentierung (basierend auf Deep-Learning-Algorithmen) durchführt, um wesentliche und kritische Schienenbereiche vollautomatisch zu erkennen. Es werden die grundlegenden Prinzipien von künstlichen neuronalen Netzwerken im Anwendungsbereich der Bildverarbeitung dargestellt und anhand der konkreten Problemstellung präsentiert.

Abstract

Trains can only travel safely and comfortably if the track geometry is exact. Tamping machines ensure this. Turnout tamping improves and stabilises the track geometry quality in turnouts. Being a complex task, it is mostly carried out by experienced operators. In this article the current status of development of a sensor system based on a 3D scanner for a new turnout tamping assistance system will be described. It will support and relieve operators when tamping complex areas. In this context, digital image processing is key. Using 3D scanner data, it carries out semantic segmentation (based on deep-learning-algorithms) to allow the detection of important and critical rail areas fully automatically. This article details the basic principles of artificial neural networks in the application field of image processing.

1 Einleitung

1.1 Der Stopfvorgang

Wenn ein Zug über eine Strecke fährt, erzeugt er enorme dynamische Kräfte. Das gesamte Gleis, bestehend aus dem Gleisrost und dem Schotterbett, ist ein elastisches System, das die eingebrachten Kräfte elastisch aufnimmt und über das Schotterbett in den Untergrund abträgt. Zusätzlich kommt es durch die Betriebsbelastung verursacht auch zu Setzungen im Schotterbett und im Untergrund. Die ideale Geometrie der Gleise ist dann nicht mehr gewährleistet, im fortgeschrittenen Stadium müssen in diesen Bereichen Geschwindigkeitsbeschränkungen auferlegt werden. Um eine solche Situation zu vermeiden, müssen Gleise in regelmäßigen Abständen instandgehalten werden. In diesem Zusammenhang ist das Gleisstopfverfahren die wichtigste Instandhaltungsmaßnahme für Eisenbahngleise.

Eine der Hauptaufgaben von Stopfmaschinen ist der Hebe- und Richtvorgang, bei dem der Gleisrost in seine Solllage zurückgehoben und seitlich verschoben wird. Dann folgt der eigentliche Stopfvorgang, der der Wiederverfüllung samt Verdichtung des Hohlraums unter der Gleis-Schwelle dient. Die Stopfpickel dienen als Verdichtwerkzeuge, ein Großteil dieser Verdichtung wird über die dynamische Vibration erzeugt. Jahrzehntelange Forschungen und Erfahrungen haben dazu beigetragen, die Arbeitsparameter während des Verdichtvorgangs zu optimieren. Eine Pickelfrequenz von 35 Hz und eine Vibrationsamplitude von 5 mm haben sich dabei als ideal erwiesen. Diese Optimierung des Stopfvorgangs hat weltweit zu einer deutlich dauerhafteren Gleislage beigetragen. Damit ist das klassische Schottergleis selbst für Hochgeschwindigkeitsgleise bestens geeignet; es stellt in jedem Fall die wirtschaftlichste Lösung für den Eisenbahn-Fahrweg dar [1].

I **Bild. 1: Stopfmaschine mit einem am Dach montierten 3D-Laserscanner**

1.2 Autonomes Stopfen (Assistenzsystem)

Langjährige Erfahrung und Geschicklichkeit des Bedieners sind notwendig, um ein qualitativ hochwertiges Stopfergebnis in Weichen zu erzielen. Der Stopfvorgang, insbesondere in Weichenbereichen, ist durch eine Vielzahl von manuellen Eingriffen gekennzeichnet. Hohe Arbeitsgeschwindigkeiten, verbunden mit größtmöglicher Arbeitsqualität, werden nur von einem erfahrenen oder gut ausgebildeten Bedienpersonal erreicht. Aus diesem Grund sind umfangreiche Kenntnisse in verschiedenen Fachbereichen (Gleisvermessung, Maschinenbau, Oberbautechnik, Regelwerke) für den idealen Stopfprozess notwendig. Zu den wichtigsten Aufgaben der beiden Bediener einer Stopfmaschine gehören etwa: Steuerung des Stopfaggregats einschließlich Dreh- und Pickelspreizung oder Pickelschwenkung, Steuerung des Hebe- und Richtaggregats, Bedienung der Zusatzhebung im Bereich des abzweigenden Gleisstrangs etc.

Bei der Entwicklung des Weichenstopf-Assistenten wurde nun das Ziel verfolgt, ein automatisches Assistenzsystem gemäß Level 3 der Norm SAE J3016 (ursprünglich zur Charakterisierung des autonomen Fahrens von straßengebundenen Kraftfahrzeugen definiert) zu entwickeln. Der Fokus liegt auf der automatisierten Unterstützung von Stopfarbeiten in schwierigen Umgebungen wie Weichen und Kreuzungen. Bei diesem Automatisierungsgrad erstellt das System Aktionsempfehlungen, die der Bediener vor der jeweiligen Aktion bestätigen kann. Ziel ist es, den Bediener zu entlasten, die Arbeitsgeschwindigkeit zu erhöhen und die Arbeitsqualität auf konstant hohem Niveau zu stabilisieren. Grundsätzlich ist der Stopfassistent aber auch für höhere Autonomiegrade geeignet [2 bis 4].

1.3 3D-Umgebung und relevante Bildinformationen

Die Umgebung (hauptsächlich der Oberbau unmittelbar vor der Stopfmaschine) wird mit einem 3D-Linienscanner abgetastet, der vorderseitig auf dem Dach der Stopfmaschine montiert ist *(Bild 1)*. Die einzelnen Scans liefern präzise 3D-Daten, die mit Informationen wie der Position oder dem Zeitpunkt der Aufzeichnung rechnerisch zusammengeführt werden. Gleichzeitig werden die Position der Maschine und die Position jedes einzelnen Aggregats ständig aktualisiert. Auf diese Weise werden Positionsdaten aller relevanten Maschinenteile in einem 3D-Gesamtmodell zusammengeführt, das in der Folge die Grundlage für jede weitere Steuerungsentscheidung während des Betriebs bildet.

Die zentrale Aufgabe der digitalen Bildverarbeitung ist die exakte Erfassung der relevanten Infrastrukturbestandteile unmittelbar vor dem Zug (Schotter, Schwellen, Schienen, Verbindungselemente etc.), die für den Stopfvorgang korrekt durch das Bildverarbeitungssystem kategorisiert

werden müssen. Zum Beispiel dürfen die hydraulischen Stopfpickel nur in geeignete Schotterbereiche eindringen, niemals in Schienen- oder Schwellenbereiche, die ernsthaft beschädigt werden könnten. Auf der anderen Seite sollten nichtkritische Bereiche (z.B. Pflanzen) keine Unterbrechung des Arbeitsprozesses verursachen, indem sie ein Pseudohindernis anzeigen.

Deshalb wurde entschieden, modernste Deep-Learning-Methoden (siehe Abschnitt 3.1) für dieses anspruchsvolle Problem der Bildanalyse anzuwenden, da diese sich in vielen anderen Bereichen (etwa beim autonomen Fahren im Straßenverkehr) als äußerst robust erwiesen haben und weil sie andere bildgestützte Segmentierungsansätze weit übertreffen. Darüber hinaus sollte das System flexibel genug sein, um es durch einen einfachen Lernprozess leicht an neue Szenarien anpassen zu können (z.B. unterschiedliche Formen von Vegetation oder auch verschieden gestaltete Gleis-Infrastruktur, die weltweit durch großen Variantenreichtum charakterisiert ist).

1.4 Deep Learning in der Gleis-Instandhaltung

Nach Kenntnisstand der Autoren ist dies der erste publizierte Versuch einer semantischen Segmentierung, basierend auf Deep Learning mit 3D-Laserscanner-Tiefenbildern im Anwendungsfeld der automatisierten Gleisstopfung. In der Literatur werden verschiedene Deep-Learning-Ansätze beschrieben, die ausschließlich auf der Analyse von Farbbildern (RGB-Bildern) beruhen. Meist werden bei den dort beschriebenen Verfahren Anwendungen im Bereich der Schieneninspektion bzw. Verwaltung der Gleisinfrastruktur behandelt [5 bis 9].

2 3D-Tiefenbild-Aufnahme mittels Laserscanner

Der hier beschriebene Bildanalyseprozess verwendet ausschließlich 3D-Tiefenbilder als Eingabedaten *(Bild 2, rechts)* von einem realen Gleisbereich (Bild 2, links). Die unterschiedlichen Grauwerte entsprechen dabei verschiedenen Abständen zum Sensor (d.h. je heller der Bildpixel, desto geringer ist der Abstand). Der Vorteil der Verwendung von Tiefenbildern besteht darin, dass das Aussehen der Objekte nicht variiert (z.B. aufgrund unterschiedlicher Lichtbedingungen), wie dies bei Standard-RGB-Kameras der Fall ist, wobei das 3D-Scannersystem verständlicherweise teurer ist. Da der Scanner jedoch auch für andere Zwecke während des Maschinenbetriebs verwendet wird, fällt dieser Nachteil in der Praxis nicht weiter ins Gewicht.

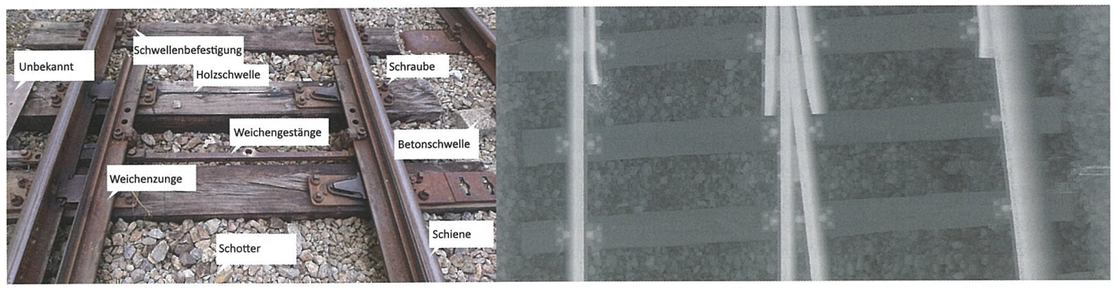
Die 3D-Tiefenbilder werden von einem rotierenden Laserscanner bereitgestellt. Der Scanner selbst liefert einzelne Scans mit Millimeter-Tiefengenauigkeit, die kontinuierlich zu einem vollen High-Resolution-Bild (hochauflösendes Bild) mit einer typischen örtlichen Auflösung von ca. 4.000 x 1.000 Pixel zusammengeführt werden. Die Arbeitsgeschwindigkeit beim Weichenstopfen liegt bei ca. 1 km/h, was zu einer lateralen Scanauflösung von ca. 2 mm führt. Dies reicht aus, um detaillierte Scanbilder zu erstellen, wobei auch eine Darstellung von sehr kleinen Objekten möglich ist (Schwellenbefestigungen etc.). Da der Messkopf des Scanners direkt vor dem Zug und das eigentliche Stopfaggregat ungefähr in der Mitte der Maschine montiert sind und somit das Scanning eines Bereichs und der eigentliche Stopfvorgang an dieser Stelle etwas zeitversetzt erfolgen, bleibt für die Bildverarbeitung ein Zeitfenster von etwa 10 s für alle notwendigen Analysen.

3 Neuronale Netzwerke in der Bildverarbeitung

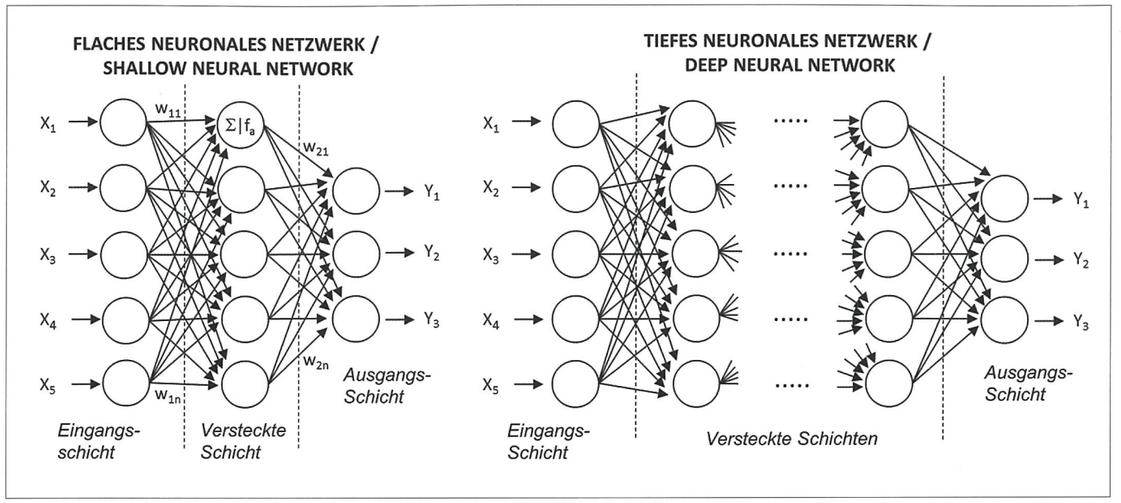
3.1 Neuronale Netzwerke und Deep Learning

Deep Learning ist ein neuer Name für eine bereits sehr lange bekannte Technik aus dem Bereich der künstlichen Intelligenz, die mit dem Begriff neuronale Netzwerke (NN) umschrieben werden kann und die in ihrer Grundform bereits mehr als 70 Jahre bekannt ist. Künstliche NN sind dabei vom Gehirn und dessen grundlegenden biologischen Vorgängen inspiriert. Erstmals wurden NN 1944 von Warren McCulloch und Walter Pitts vorgeschlagen, zwei Forschern der University of Chicago (später am MIT), die auch als Begründer der Kognitionswissenschaften gelten.

Es handelt sich bei Deep Learning um eine Form des sogenannten überwachten maschinellen Lernens, bei dem einem Computer ein Trainingssatz von Beispielen bereitgestellt wird (Lerndaten), um zuvor ungesehene Beispiele korrekt verarbeiten zu können. Bildklassifikation ist in diesem Zusammenhang eine typische Anwendung – als Eingangsgröße wird ein Bild eines Objekts präsentiert und der Rechner soll entscheiden, welches Objekt dargestellt ist. Dabei ist ein zentraler Aspekt dieser Algorithmen die Fähigkeit zur Generalisierung, d.h. nicht jedes Entscheidungskriterium muss im Vorfeld penibel vom Menschen vorprogrammiert werden, sondern der Algorithmus kann aufgrund seiner Lernerfahrung auch neue, bis dato ungesehene Objekte einer zuvor trainierten Klasse korrekt identifizieren.



I Bild 2: Relevanter Gleisbereich (links); typisches 3D-Scan-Tiefenbild (rechts)



I Bild 3: Grundlegender Aufbau des flachen neuralen Netzwerks (links); tiefgehendes neurales Netzwerk mit vielen versteckten Schichten (rechts)

Bild 3 stellt das Grundprinzip eines NN dar, das aus miteinander verbundenen Neuronen besteht. Dabei ist jedes Neuron mit allen Neuronen der darauffolgenden Schicht verbunden (fully connected – voll vernetzt). Für die konkrete Aufgabe der Bildklassifikation werden beispielsweise am sogenannten Eingangs-Layer (input layer) die Grauwerte der einzelnen Pixel als Eingangswerte angelegt (Werte X1, X2, X3 …). Diese bestimmen in der Folge die Neuronenaktivität. Jede darauffolgende Verbindung hat eine Gewichtung (w11, w12, w13 …), die das in die nächste Schicht weitergegebene Summensignal festlegt und die Neuronenaktivität in der neuen Schicht (hidden layer – versteckte Schicht) bestimmt. Dabei muss dieser Wert vor der Weitergabe noch über eine spezielle nicht-lineare Aktivierungsfunktion fa normiert werden. Die Aktivitäten der Ausgangsschicht (output layer) können schließlich als Wahrscheinlichkeiten für die korrekte Erkennung der entsprechenden Klasse interpretiert werden (Y1, Y2 oder Y3). Für das Training werden diese Gewichte zunächst zufällig initialisiert. Am Eingangs-Layer werden nacheinander die Trainingsbilder angelegt und die berechnete Aktivität wird am Ausgang mit jenen der gewünschten Soll-Klasse verglichen und vorhandene Abweichungen werden einem Lern-Algorithmus, wie etwa dem Backpropagation-Algorithmus, zur Adaptierung der Gewichte übergeben. Dieser Vorgang wird mit allen Trainingsbildern mehrfach wiederholt (Epochen), bis der Erkennungsfehler minimal wird.

In den 1990er-Jahren ließ das Interesse an dieser Form von Lern-Algorithmen stark nach. Neben theoretischen Problemen war insbesondere die notwendige Rechnerleistung für ein effektives Training noch nicht verfügbar. Außerdem standen alternative Algorithmen zur Verfügung, die vergleichbar gute Ergebnisse produzierten, wie jene von NN zu dieser Zeit. Mit der allgemeinen Verfügbarkeit von GPUs (graphical processing units – leistungsstarke Grafikkarten) und entsprechenden Lerndaten kam es jedoch 2012 zu einem Wiederaufleben des Interesses an NN. Deep Learning (d.h. NN mit vielen hidden layers, im Gegensatz zu Shallow Neural Networks (flache neurale Netzwerke) mit nur einer versteckten Schicht *(Bild 3, links)* und besonderen Netzwerkarchitekturen – sogenannte Convolutional Neural Networks (CNN – faltendes neuronales Netzwerk, siehe Abschnitt 3.2.) – führten zu einer regelrechten Revolution im Bereich der Bildklassifikation und haben seitdem enorme Fortschritte, insbesondere im Bereich der Computer Vision ermöglicht. Die visuellen Erkennungsfähigkeiten künstlicher Systeme kommen hier jenen von Menschen schon sehr nahe oder übertreffen diese bereits in gewissen Anwendungsfeldern. Im Bereich der Bildklassifikation erzielen etwa Netzwerke wie AlexNet [10], VGG [11] oder GoogLeNet [12] zuvor unerreichbare Ergebnisse.

Deep Learning ermöglicht eine neue Art des Programmierens. Beim Training eines Lernmodells werden Parameter (Netzwerkgewichte) so angepasst, dass eine bestimmte Eingabe (z.B. ein Bild) einer bestimmten Ausgabe (Klasse) zugeordnet wird – die Kriterien dafür müssen nicht mehr explizit vom Programmierer vorgegeben werden. D.h. der Bildverarbeiter muss nicht mehr im Vorfeld überlegen, anhand welcher Kriterien bzw. anhand welcher Bild-Features (Bild-Merkmale) er ein Bildklassifikationsproblem lösen kann. Der Deep-Learning-Algorithmus erzeugt jetzt selbstständig geeignete und spezifisch an die jeweilige Problemstellung angepasste Features, auf deren Basis in der Folge Entscheidungen getroffen werden können.

1943: Warren McCulloch und Walter Pitts erstellen ein Rechenmodell für neuronale Netze auf der Grundlage von Mathematik und Algorithmen, das als Schwellenlogik bezeichnet wird.

1958: Frank Rosenblatt erstellt das Perzeptron, einen Algorithmus zur Mustererkennung, basierend auf einem zweischichtigen neuronalen Netzwerk mit einfacher Addition und Subtraktion. Er schlug auch zusätzliche Schichten mit mathematischen Notationen vor, die jedoch erst 1975 verwirklicht werden sollten.

1980: Kunihiko Fukushima schlägt das Neocognitron vor, ein hierarchisches, mehrschichtiges künstliches neuronales Netzwerk, das für die Erkennung von Handschriften und andere Mustererkennungsprobleme verwendet wurde.

1998: CNN (Convolutional Neural Network) – die Standardreferenz für CNNs – stammt aus den Jahren 1998/1999, erstellt von Yann LeCun et al.: „Object Recognition with Gradient-Based Learning" (http://yann.lecun.com/exdb/publis/pdf/lecun-99.pdf) abgerufen im März 2019.

2012: Deep Learning im Bereich Bildklassifikation übertrifft erstmals klassische Algorithmen – dank der Verfügbarkeit entsprechend leistungsfähiger Hardware (GPUs) und Lerndaten.

I Tabelle 1: Kurzer Abriss der historischen Entwicklung neuronaler Netzwerke

Tabelle 1 zeigt einen kurzen Abriss der historischen Entwicklung neuronaler Netzwerke.

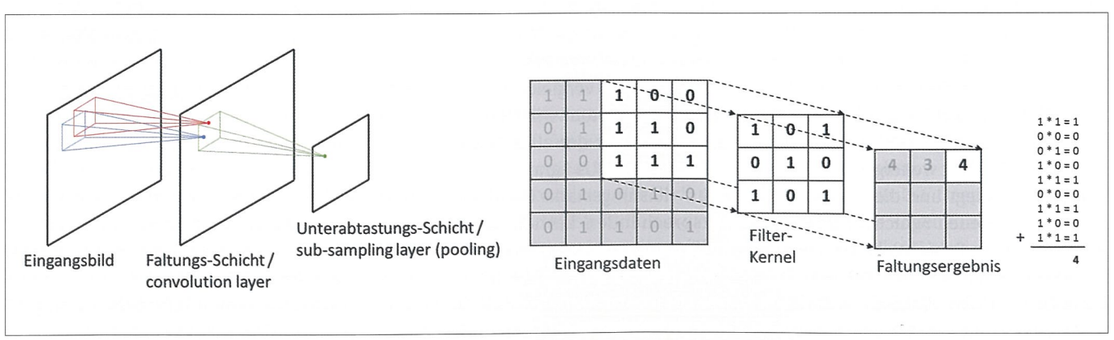
3.2 Convolutional Neural Networks zur Bildklassifikation

Die spezielle Architektur eines CNN verarbeitet die Eingangsdaten in einer Abfolge von Faltungs- sowie Subsampling-Operationen (convolution – falten, pooling - unterabtasten) und nichtlinearen Normierungen (z.B. mittels sogenannter ReLu-Aktivierungsfunktionen) *(Bild 4* und *Bild 5)*. Vergleichbar mit dem Kernel (Filterkern) eines linearen Filters kommt hier eine frei wählbare Anzahl von Filtern definierter Größe (z.B. 5x5) zum Einsatz, deren Filterwerte den Gewichten in einem NN entsprechen.

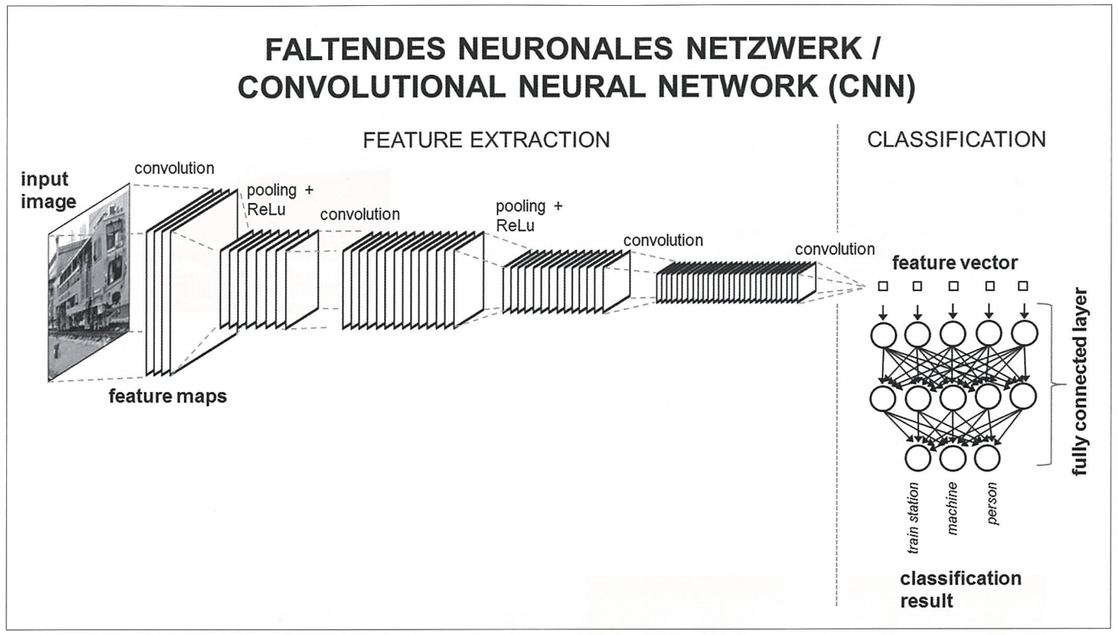
Die Weitergabe von Neuronenaktivitäten geschieht durch eine Faltungsoperation des jeweiligen Kernels mit den Eingangsdaten der vorherigen Schicht (d.h. hier sind auch Faltungen über Volumen möglich). Diese Faltungsergebnisse werden auch als feature maps (Merkmals-Karten) bezeichnet. Der Faltungsvorgang wird in diesem Zusammenhang als Merkmalsdetektor eingesetzt. Als Eingang für die Faltungsoperation können Rohdaten (z.B. eine Bildmatrix) dienen, aber auch die feature map eines vorangegangenen Faltungsprozesses. Diese Operation kann dann als Filtervorgang interpretiert werden, bei dem der Filter-Kernel die Eingangsdaten für spezifische Zwecke filtert – z.B. lässt ein Kanten-Kernel nur Informationen über Kanten in einem Bild passieren, andere Bildstrukturen werden hingegen unterdrückt. *Bild 4*, rechts zeigt, wie der Kernel über die Eingabedaten geschoben wird, um die gefalteten (Output-)Daten zu erzeugen. Bei jedem Schritt wird der Kernel mit den Eingabewerten innerhalb seiner Grenzen multipliziert und summiert, wodurch ein einzelner Eintrag in der Ausgabe-feature-map erstellt wird. In der Praxis ist der Ausgabewert immer dann groß, wenn das Merkmal, nach dem gesucht wird, auch in der Eingabe erkannt wird.

Um die Daten zusammenzuführen und um auch den Rechenaufwand zu reduzieren, kommt zusätzlich in manchen Schichten Unterabtastung (pooling) zur Anwendung. Besondere Bedeutung kommt der Aktivierungsfunktion zu, die die Gewichtswerte schließlich wieder der notwendigen nichtlinearen Transformation unterwirft (Normierung mittels sogenannter rectified linear units, ReLu – rektifizierte Lineareinheiten). Am Ende dieser Verarbeitungskette (dem sogenannten feature extraction-Netzwerk – Merkmal-Extraktions-Netzwerk) werden die verbliebenen Neuronen (der sogenannte feature vector – Merkmalsvektor) typischerweise mit einem klassischen Neuronalen Netzwerk verbunden (fully connected layer – vollständig verbundene Schicht), dessen Neuronenanzahl am Ausgang der Zahl der zu erkennenden Objektklassen entspricht (classification-Netzwerk – Klassifizierungs-Netzwerk).

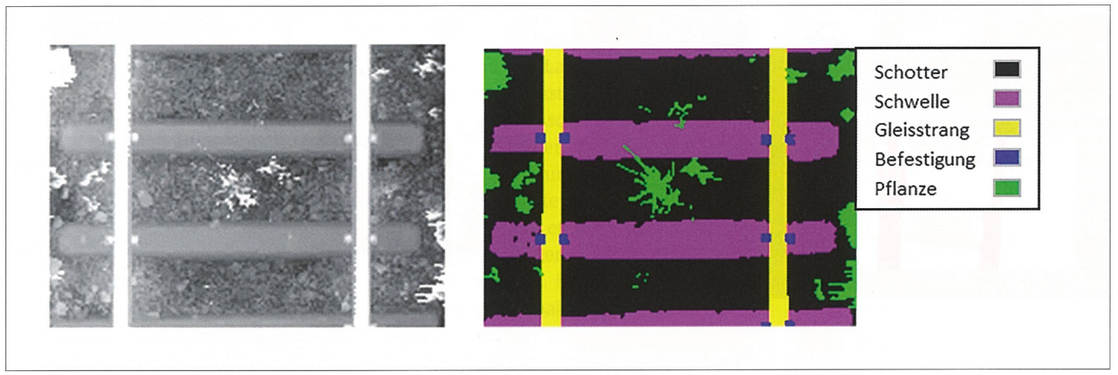
Während des Lernprozesses werden laufend alle Gewichte (Filterwerte) im gesamten Netzwerk mittels Backpropagation (Rückpropagierung) angepasst (auch bezeichnet als learning from scratch – von Grund auf lernen). Alternativ, da weniger rechenaufwendig, kann auch mit bereits vorberechneten Filtergewichten gearbeitet werden (d.h. diese werden einfach im feature extraction-Netzwerk hinterlegt). Nur die Gewichte im classification-Netzwerk werden neu trainiert und so an die eigene Problemstellung angepasst (transfer learning – Übertragen der Ergebnisse einer fertigen Problemlösung auf eine andere, aber verwandte Aufgabe).



I Bild 4: Ein faltendes neuronales Netzwerk (CNN) mit vielen alternierenden Faltungs- und Unterabtastungsschichten (links); Prinzip der 2D-Faltungsoperation (rechts)



I Bild 5: Prinzipieller Aufbau eines faltenden neuronalen Netzwerks



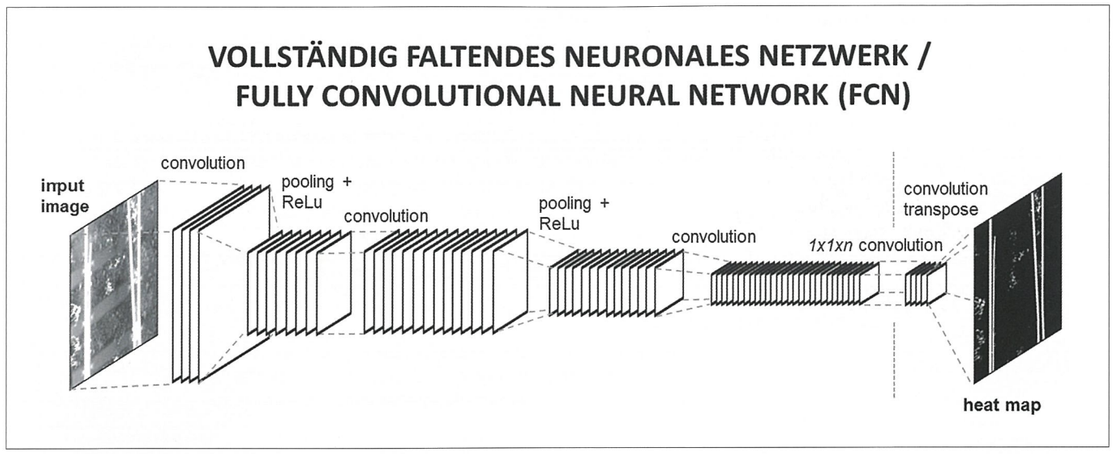
I Bild 6: 3D-Scannerbild (links), gewünschtes Ergebnisbild mit pixelgenau segmentierten Bereichen, die relevante Bildinhalte wie Schotter, Schienen, Pflanzen usw. darstellen, (rechts)

4 Semantische Segmentierung von 3D-Scannerdaten

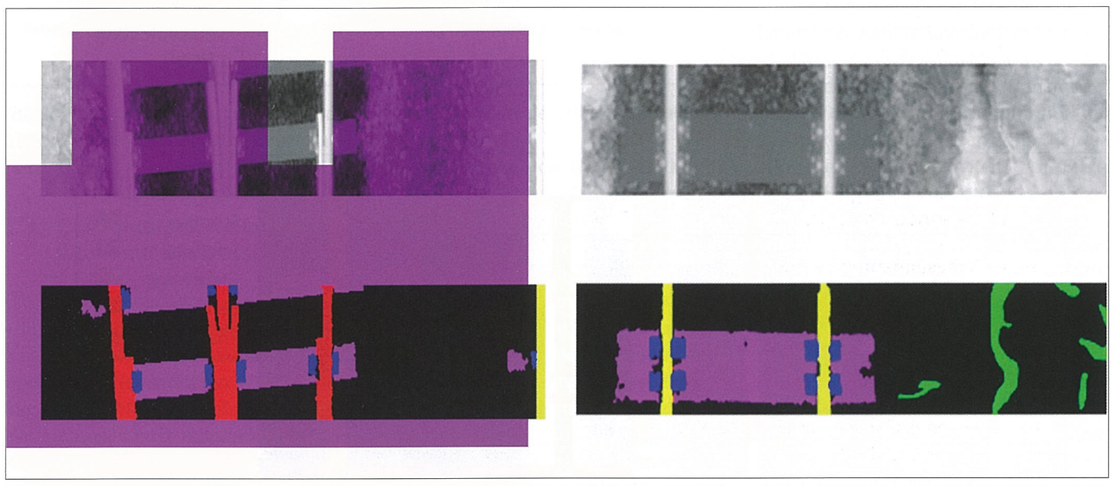
4.1 Fully Convolutional Neural Networks zur Bildsegmentierung

Eine Erweiterung der oben dargestellten Bildklassifikation stellt die semantische Bildsegmentierung (semantic segmentation) dar. Sie spielt allgemein eine entscheidende Rolle in der Bildverarbeitung und ermöglicht es einem Computer, Objekte in Bildern nicht nur zu erkennen, sondern darüber hinaus pixelgenau zu lokalisieren. Die Erkennung und exakte Abgrenzung von Objekten im Bild wird durch die Klassifizierung jedes einzelnen Pixels erreicht, d.h. jedem Pixel wird eine definierte Objektklasse zugewiesen *(Bild 6)*.

Das im letzten Abschnitt dargestellte Netzwerk (CNN) verarbeitet typischerweise nur kleine Bilder mit einer fest vorgegeben Größe (z.B. 224 x 224 Pixel, größere Bilder müssen zunächst immer auf diese Größe skaliert werden). Um ein, wie im vorliegenden Fall, 4.000 x 1.000 Pixel großes Bild zu verarbeiten, müsste für je-



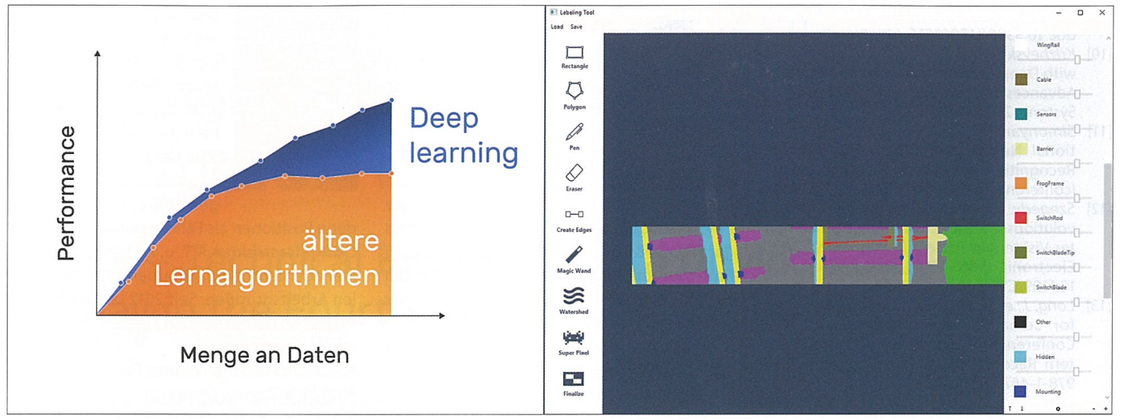
I Bild 7: Prinzipieller Aufbau eines speziellen tieffaltenden neuralen Netzwerks (FCN) zur semantischen Bildsegmentierung



I Bild 8: Typische Ergebnisse einer semantischen Bildsegmentierung, oben: 3D-Scan-Eingangsbilder, unten: farblich markiert die unterschiedlichen erkannten Objekte (Schotter – schwarz, Gleis – gelb, Schrauben – blau, Schwellen – rosa, Pflanzen – grün)

den der 4 Mio Pixel eine kleine Bildregion ausgeschnitten und dem Neuronalen Netzwerk zur Klassifikation zugeführt werden – eine Aufgabe, die selbst mit High-End-GPUs (Spitzen-Grafikprozessoren) zu völlig inakzeptablen Rechenzeiten führen würde.

Der für das Weichenstopf-Assistenzsystem verwendete Segmentierungsansatz basiert auf dem sogenannten Fully Convolutional Network FCN (spezielles tieffaltendes neurales Netzwerk) [13], einem beliebten Algorithmus für semantische Segmentierung. Die Idee hinter einem FCN-Netzwerk besteht darin, ein klassisches CNN, das üblicherweise zur Bildklassifikation benutzt wird, so zu erweitern, das der fully connected layer am Ende durch einen 1x1xn convolution layer ersetzt sowie um einen sogenannten convolution transpose layer (Faltung-Umsetzungs-Schicht) erweitert wird *(Bild 7)*. Im Detail verwendet dieses Netzwerkmodell verschiedene Convolution- und Pooling-Schichten, um ein Bild zu analysieren und auf einen Bruchteil seiner ursprünglichen Größe zu komprimieren (meist 1/32). Dieses komprimierte Zwischenergebnis enthält eine Klassenvorhersage auf dieser Auflösungsstufe. Schließlich verwendet es einen convolution transpose layer, um das Bild wieder auf seine ursprünglichen Abmessungen hochzuskalieren. Durch Erweitern eines CNN zu einem FCN wird also der Klassifizierungsteil durch einen Faltungsteil ersetzt. Beim Anlegen eines größeren Bildes liefert nun das Netz nicht nur einmal Wahrscheinlichkeiten für jede trainierte Klasse, sondern ein ganzes array (Feld) solcher Wahrscheinlichkeiten. Ein Element dieses arrays steht dann für die erkannte Klasse eines Teilbereichs des Eingangsbilds. Nach der Hochskalierung dieses arrays auf die ursprüngliche Größe des Eingabebilds mit Hilfe des convolution transpose layers erhält man schließlich eine heatmap (Aktivierungskarte) der



I Bild 9: Vergleich zwischen klassischen Lern-Algorithmen und Deep-Learning-Algorithmen (links); Software-Tool für das Generieren von Lerndaten, die durch eine Vielzahl an Funktionen die zeitintensive Arbeit des Image-Labeling unterstützen (rechts)

örtlichen Verteilung der detektierten Klassen, da jedes Pixel in diesem Ergebnisbild einem klassifizierten feature vector entspricht. Ergebnisse einer semantischen Bildsegmentierung zeigt *Bild 8*.

4.2 Lerndaten und Image Labeling (Bildkennzeichnung)

Eine generelle Herausforderung bei Deep Learning ist die enorme Menge an Lerndaten, die notwendig ist, um entsprechend gute Ergebnisse zu produzieren *(Bild 9, links)*. CNNs sowie FCNs sind überwachte Lernansätze, d.h. es müssen Bilder zum Trainieren der Netzwerke zur Verfügung gestellt werden, die die Grundwahrheit darstellen. Insbesondere bei der semantischen Segmentierung ist der diesbezügliche Aufwand beträchtlich, da hier eine pixelgenaue Markierung der Bilder notwendig ist, d.h. jedes Objekt im Bild muss zunächst exakt markiert und mit genauen Grenzkonturen hinterlegt werden (image labeling). Unterstützung dabei bieten spezielle Software-Tools mit umfangreichen Zeichen-Funktionalitäten und halb automatischen Markierungs-Funktionen *(Bild 9, rechts)*.

Die Experimente der Autoren haben gezeigt, dass sich neben komplett segmentierten Bildern auch kleine Bildausschnitte (patches) ausgezeichnet zum Training eignen, da die FCN-Architektur Eingangsdaten in unterschiedlichen Bildgrößen erlaubt. Damit lässt sich aus einer einzigen Tiefenbildaufnahme eine Vielzahl an Trainingsbildern generieren.

5 Zusammenfassung und Ausblick

In diesem Beitrag wurde ein Bildverarbeitungssystem für ein teilautonomes Weichenstopf-Assistenzsystem vorgestellt, das auf der Basis von Deep-Learning-Algorithmen in der Lage ist, vollautomatisch relevante Infrastruktur im Arbeitsbereich einer Stopfmaschine zu lokalisieren und zu klassifizieren. Dabei kommt ein 3D-Laserscanner zum Einsatz, der eine exakte und robuste Erfassung der Maschinenumgebung ermöglicht, unabhängig von der jeweiligen Lichtsituation (Tag, Nacht, Nebel etc.). Es wurden kurz die Grundlagen neuronaler Netzwerke im Anwendungsfeld der Bildverarbeitung umrissen und die Funktionsweise moderner Convolutional Neural Networks für die semantische Bildsegmentierung dargestellt. Erste Versuche im Echtbetrieb bestätigen die ausgezeichnete Eignung des beschriebenen Verfahrens.

Abschließend kann gesagt werden, dass Deep Learning die praktische Realisierung sehr anspruchsvoller Bildverarbeitungsaufgaben ermöglicht, ohne die ein robuster Outdoor-Einsatz unter sehr harschen Bedingungen nicht möglich wäre. Die eingesetzten Algorithmen werden dabei laufend verbessert – so wird aktuell an einer neuen Generation von Netzwerkarchitekturen und Segmentierungsansätzen gearbeitet, die eine weitere Verbesserung der Detektionseigenschaften von sehr kleinen Objekten verspricht (etwa Kabel etc.) und die mögliche Anzahl der zu klassifizierenden Objekte noch einmal wesentlich erhöht.

#452_A3

(Bildnachweis: 1 bis 9, Plasser & Theurer)
(Tabellennachweis: 1, Verfasser)

Literatur

[1] Homepage Plasser & Theuer: https://www.plassertheurer.com/en/machines- systems/tamping.html, abgerufen April 2019.
[2] *Auer, F.; et al.*: Smart Tamping – Anwendungsmöglichkeiten des Weichenstopf-Assistenzsystems. ZEVrail 142 (2018) Sonderheft Digitalisierung im Schienenverkehr Juli 2018, S.38-41.
[3] *Bürger, M.; et. al.*: Hilfe beim Stopfen – Entwicklung eines Weichenstopf-Assistenzsystems. Eisenbahn-Ingenieur 2017, Nr. 6, S. 16-18.
[4] *Bürger, M.; Zauner, G.*: Automatischer Maschinenbetrieb bei Gleisbauarbeiten zur Optimierung des Rad/Schiene-Kontakts. ZEVrail 143 (2019) Tagungsband SFT Graz 2019, S. 212-217.
[5] *Gibert, X; et al.*: Deep Multi-task Learning for Railway Track Inspection. IEEE Transactions on Intelligent Transportation Systems, Volume 18, Issue 1, Jan. 2017, doi: 10.1109/TITS.2016.2568758.
[6] *Singh, A. K.; et al.*: Vision based rail track extraction and monitoring through drone imagery. KICS – ICT Express, Dec. 2017, doi: 10.1016/j.icte.2017.11.010.
[7] *Chen, J.; et al.*: High-speed railway catenary components detection using the cascaded convolutional neural networks. 2017 IEEE International Conference on Imaging Systems and Techniques (IST), Electronic ISBN: 978-1-5386-1620-8.
[8] *Faghih-Roohi, S.; et al.*: Deep convolutional neural networks for detection of rail surface defects. 2016 International Joint Conference on Neural Networks (IJCNN), Electronic ISBN: 978-1-5090-0620-5.
[9] *Ye, T.; et al.*: Automatic Railway Traffic Object Detection System Using Feature Fusion Refine Neural Network under Shunting

Mode. Sensors (Basel) 2018 Jun 12, 18(6), doi: 10.3390/s18061916.
[10] *Krizhevsky, A.; et al.:* ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems 25 (NIPS 2012).
[11] *Simonyan, K.; et al.:* Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv technical report, 2014 (conference paper presented at ICLR 2015).
[12] *Szegedy, Ch.; et al.:* Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Electronic ISBN: 978-1-4673-6964-0, doi: 10.1109/CVPR.2015.7298594.
[13] *Long, J.; et al.:* Fully Convolutional Networks for Semantic Segmentation. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Electronic ISBN: 978-1-4673-6964-0.

Dr. techn. Gerald Zauner (45). Studium der Automatisierungstechnik an der Fachhochschule Oberösterreich (FH OÖ), Fakultät für Technik und Angewandte Naturwissenschaften. Dissertation an der TU Wien am Institut für Angewandte Physik (2005). Leiter des Research Center Wels der FH OÖ F&E GmbH (2005 bis 2015). Seit 2015 Professur für Signal Processing an der FH OÖ Campus Wels. Seit 2015 beschäftigt bei Plasser & Theurer (Forschung & Versuch).
Anschrift: Plasser & Theurer, Pummererstraße 5, 4021 Linz/Donau, Österreich.
E-Mail: gerald.zauner@plassertheurer.com

Dipl.-Ing. Dr. techn. Florian Auer (42). Studium des Bauingenieurwesens an der Technischen Universität Graz, Vertiefung Verkehrstechnik. Dissertation "Zur Verschleißreduktion von Gleisen in engen Bögen". 2002 bis 2012 verschiedene Positionen als Fahrweg-Ingenieur in der ÖBB-Infrastruktur AG, u.a. Leiter Life-Cycle-Management Team. Mitarbeit in internationalen Arbeitsgruppen. Seit 2012 beschäftigt bei Plasser & Theurer, seit 2017 Leiter Technologie und Innovation.
Anschrift: Plasser & Theurer, Pummererstraße 5, 4021 Linz/Donau, Österreich.
E-Mail: florian.auer@plassertheurer.com

Dipl.-Ing. Martin Bürger (47). Studium der Informatik an der Johannes Kepler Universität Linz. Bis 2007 bei Siemens in der Softwareentwicklung, unter anderem als Architekt, Tester, Projektleiter und Prozess-Berater, seit 2007 in der agilen Softwareentwicklung als Scrum Master und Product Owner tätig. Abteilungsleiter der Softwareentwicklung bis 2010. 2009 Lektor an der Fachhochschule Wels für Automatisierungstechnik. Bis 2012 Software Qualitätsbeauftragter bei Magna Powertrain Engineering Center. Seit 2012 bei Plasser & Theurer in Forschung & Entwicklung, seit 2018 Abteilungsleiter der Abteilung Digital Track Systems.
Anschrift: Plasser & Theurer, Pummererstraße 5, 4021 Linz/Donau, Österreich.
E-Mail: martin.buerger@plassertheurer.com

Dipl.-Ing. Tobias Müller (26). Studium der Automatisierungstechnik an der Fachhochschule Oberösterreich FH OÖ. Seit 2018 beschäftigt bei Plasser & Theurer in der Abteilung „Digital Track Systems" mit Schwerpunkt Deep Learning in der Bildsegmentierung.
Anschrift: Plasser & Theurer, Pummererstraße 5, 4021 Linz/Donau, Österreich.
E-Mail: tobias.mueller@plassertheurer.com