



# Software für die automatische Zuordnung von Sensordaten

Valide Daten sind entscheidend für die erfolgreiche Entwicklung der Algorithmen von Assistenzsystemen für das autonome Fahren. Die eingesetzten Systeme benötigen dafür einen intensiven Abgleich mit den sogenannten Ground-Truth-Informationen. Dabei müssen bereits bei der Validierung der Bilderfassung pro Fahrzeug und Tag mehr als 100 TB Daten aufgezeichnet und ausgewertet werden. Bertrandt hat mit seinem Data Labeler ein Werkzeug entwickelt, um diese Datenmenge performant zu handhaben und automatisch zu analysieren.

## AUTOREN



**Dipl.-Phys. Stefan Schindler**  
ist Data Scientist im Bereich  
Machine Learning bei Bertrand  
in Rüsselsheim.



**Dipl.-Bioinf. Holger Bergmann**  
ist Data Scientist im Bereich  
Computer Vision bei Bertrand  
in Rüsselsheim.



**Torsten Bohnet, M. Sc.**  
ist Lead Engineer im Bereich  
Analytics bei Bertrand  
in Rüsselsheim.



**Dr. Dipl.-Phys. Yusuf Erdogan**  
ist Teamleiter Analytics bei  
Bertrand in Rüsselsheim.

## HERAUSFORDERUNGEN BEI DER ABSICHERUNG AUTONOMER FAHRFUNKTIONEN

Die Funktionsweise der Fahrerassistenzsysteme auf SAE-Level 3 und höher basiert auf dem Zusammenspiel zahlreicher Sensoren. Die Absicherung beziehungsweise der Test der Funktionalität wird durch den Abgleich der Messungen dieser Sensoren mit der zugrundeliegenden realen Information erreicht, der sogenannten Ground-Truth-Information. Der Ausdruck „Data is the new oil in the future of automated driving“ [1] greift hier auf besondere Weise: Daten sind sowohl der Schmierstoff, der die Stellglieder antreibt und schmiert, als auch das (schwarze) Gold im Sinne des wertvollen, aber auch durchaus kostspieligen Bestandteils sämtlicher Funktionen für autonomes oder unterstütztes Fahren, das es zu heben gilt.

Um die erforderliche Absicherungsgüte für die gesicherte Funktion von Fahrerassistenzsystemen zu erreichen, benötigt der Systementwickler eine große Menge an aufgezeichneten Umgebungsinformationen. Beispielsweise fallen beim Einsatz von bildgebender Sensorik in einer typischen Testphase mit 10.000 h Messfahrt bei einer Bildrate von 30 Bildern pro Sekunde mehrere hundert Millionen Kamerabilder an. Neben dem Soll-Ist-Abgleich der im Betrieb gewonnenen Werte müssen die im Fahrzeug eingesetzten Algorithmen für maschinelles Lernen mit großen Mengen an hochgenauen Realdaten trainiert werden. Diese Ground-Truth-Daten kommen dafür zum Einsatz, indem sie die Position und die Eigenschaften von auf den einzelnen Bildern vorkommenden Objekten sehr genau angeben und dadurch die Verifizierung der Messungen erlauben. Zu diesen Informationen kommen zahlreiche nicht-objektspezifische Informationen hinzu, die beispielsweise das Wetterverhältnis oder die Eingruppierung der Fahrscenarien beschreiben (globale Labels).

## STEIGENDE DATENMENGEN ERSCHWEREN MANUELLE ANNOTATION

Die Generierung dieser Informationen wurde bis jetzt aufgrund der hohen Anforderungen an die Datenqualität und -sicherheit häufig von geschultem Perso-

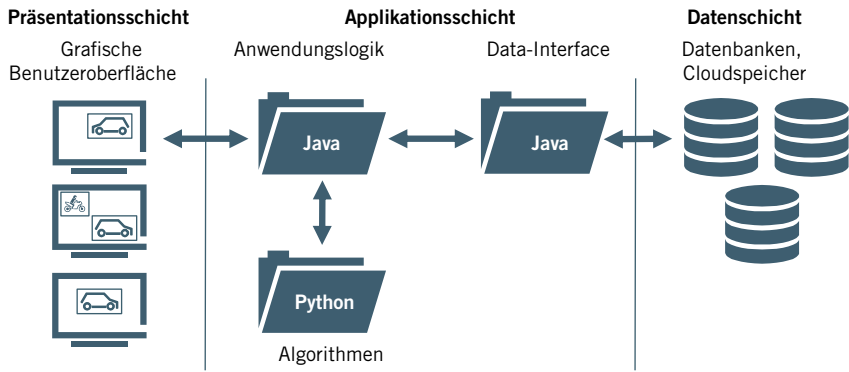


BILD 1 Schematische Darstellung der Schichtarchitektur des Labeling-Werkzeugs (© Bertrand)

nal durch manuelles Annotieren der Sensor-Rohdaten durchgeführt. Durch die größer werdenden Datenmengen wird diese Vorgehensweise wegen des steigenden Zeitaufwands stetig komplexer und damit ab einem bestimmten Zeitpunkt zu arbeits- und kostenintensiv – zumal sich durch die Entwicklungsarbeit der daran forschenden Teams die eingesetzten Algorithmen zum Abgleich ständig verbessern sowie immer günstiger werdende große Rechenleistung zur Verfügung steht. Bertrand hat deswegen für die Annotation von Kameradaten eine frei konfigurierbare und performante Software namens Data Labeler auf Cloudbasis entwickelt, um diesen Vorgang so weit wie möglich zu automatisieren.

**FLEXIBLE SOFTWARE-ARCHITEKTUR FÜR SKALIERBARE LÖSUNGEN**

Durch die Vielfalt der möglichen zu annotierenden Daten ergeben sich mit Blick in die Zukunft besondere Anforderungen an die Architektur und Algorithmen des Werkzeugs, BILD 1. Der Annotationsprozess ist vom Auftraggeber, Zielsystem und Anwendungsgebiet abhängig und sehr spezifisch. Konfigurierbarkeit und Flexibilität waren daher Kernanforderungen bei der Entwicklung des Produkts. Die Architektur des Werkzeugs verknüpft klassische und bewährte Ansätze mit neuartigen Aspekten. Einzelnen Kompo-

nenten des Gesamtsystems sind eindeutigen Aufgaben zugeordnet und diese in Schichten aufgeteilt. Die gesamte Architektur wurde modular aufgebaut, um die Flexibilität sowie Erweiterbar- und Wartbarkeit des robusten Werkzeugs zu gewährleisten. Die Architektur ist so konzipiert, dass die Software mit unterschiedlichsten Dateiformaten und Speicherlösungen wie Massenspeicherung in der Cloud umgehen und in Zukunft hardwareunabhängig weiterentwickelt werden kann.

Der Workflow ist eine Java-Implementierung, die auf Python basierende Algorithmen für das assistierte und automatisierte Annotieren steuert. Dem Nutzer steht eine Weboberfläche mit intuitiven Bedienkonzepten zur Verfügung. In dieser Oberfläche werden alle Funktionen, wie das manuelle Annotieren mit Boundingboxen, Segmentierungen, Punkten, Linien und Polygonen oder die Validierung der durch die Algorithmen erzeugten Labels, bereitgestellt.

**FUNKTIONEN FÜR ASSISTIERTES LABELING**

Zur Unterstützung des Annotationsprozesses kann der Anwender verschiedene Algorithmen aus dem Bereich Computer Vision (CV) verwenden. Besonders Objektverfolgung und Objekterkennung lassen sich nutzen, um das (teil-)automatisierte Annotieren zu ermöglichen. Bei der Objektverfolgung handelt es sich um die kontinuierliche, aber mit bestimmten Abbruchkriterien versehene Lokalisierung eines oder mehrerer Zielobjekte in einer Videosequenz, BILD 2. Für Objekte, die einer geradlinigen Trajektorie folgen, gibt es eine lineare Interpolationsfunktion, die zwischen zwei bereits annotierten Bildern die fehlenden Annotationen für ein oder mehrere Objekte hinzufügt. Um jedoch ein Objekt tatsächlich zu lokalisieren, bildet die Software ein initiales Modell des Objekts auf dem ersten Bild der Videosequenz. Dieses Modell besteht aus den Erscheinungs- und den Positionsparametern des gewählten Objekts und wird mithilfe einer gesetzten Boundingbox definiert. Der Inhalt einer Boundingbox dient somit als Trainingsbereich. Eine kontinuierliche Lokalisierung ist gewährleistet, indem das Modell von Videobild zu Videobild aktualisiert wird, BILD 3.

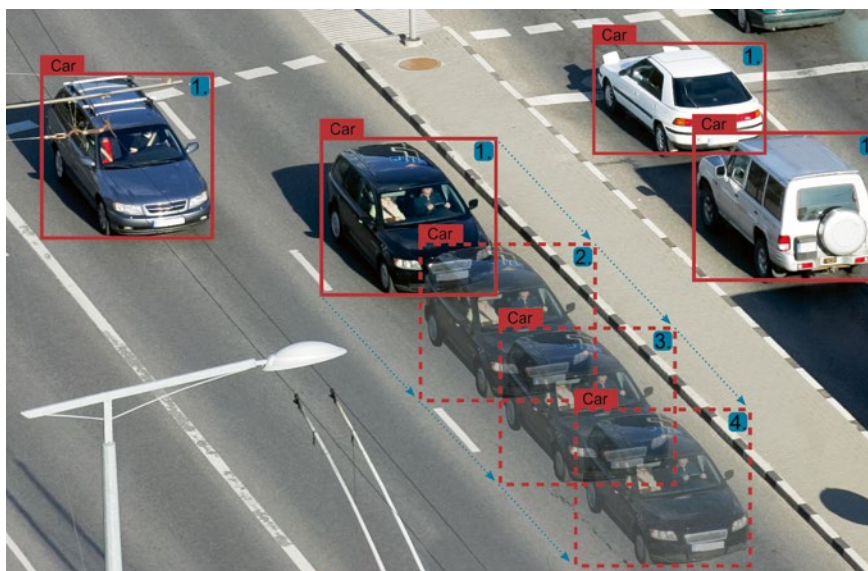
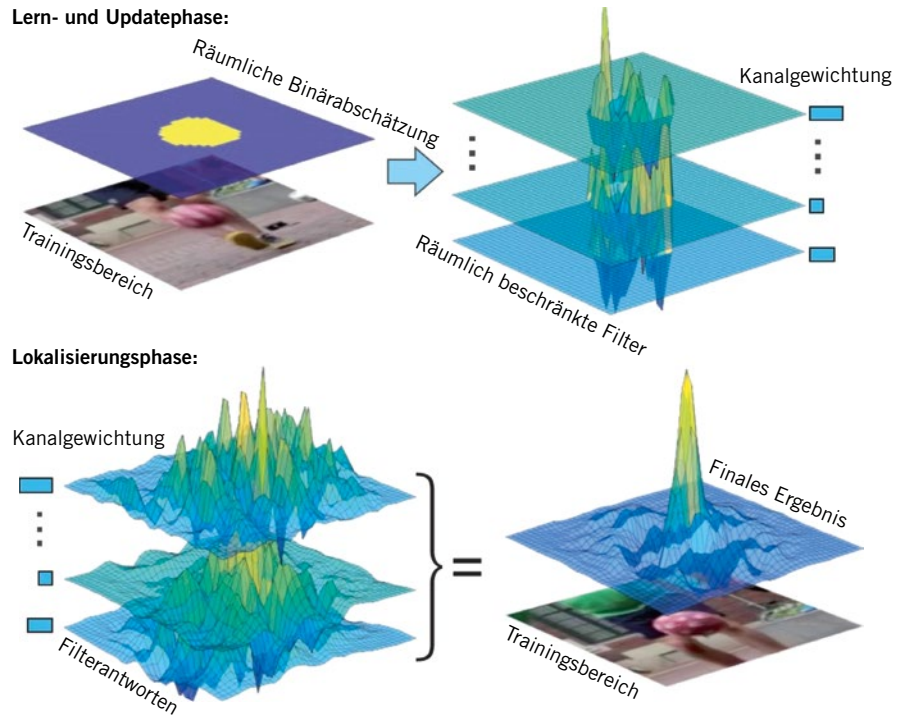


BILD 2 Annotationen in einer Videosequenz können automatisch über die nächsten Frames verfolgt werden; der Inhalt einer gesetzten Boundingbox dient dabei der Erstellung eines Erscheinungs- und Bewegungsmodells; die automatisch gesetzten Annotationen (gestrichelte Boundingboxen) stellen schematisch die Verfolgung eines Objekts bis zum vierten Bild dar (© Bertrand)

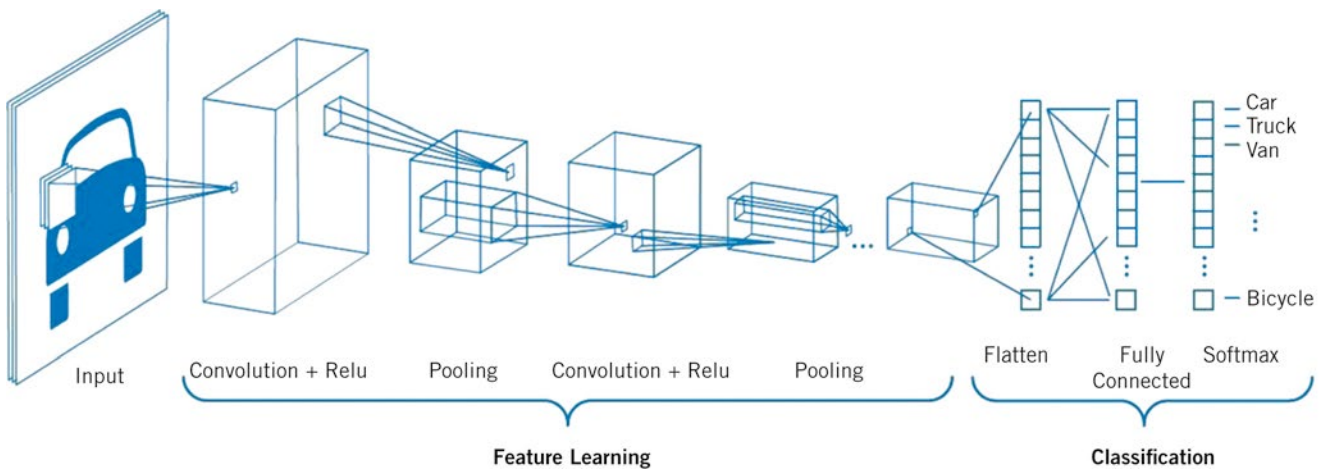
## KORRELATIONSFILTER FÜR OBJEKTERKENNUNG

Zur Erkennung eines Objekts im jeweiligen Folgebild nutzt die Software dessen Erscheinungsparameter, sogenannte Korrelationsfilter, die es ermöglichen, die Ähnlichkeit eines Pixels aus dem Trainingsbereich mit einem Pixel auf dem Folgebild im Suchbereich abzuschätzen [2, 3]. Der Suchbereich kann anhand der Positionsparameter des Modells so eingeschränkt werden, dass der Lokalisierungsalgorithmus eine effiziente Laufzeitkomplexität beibehält statt den gesamten Frame durchsuchen zu müssen. Um bei der Aktualisierung des Modells möglichst wenig Hintergrundverunreinigung zu erhalten, identifiziert das Programm mithilfe einer räumlichen Binärabschätzung für einen Korrelationsfilter diejenigen Pixel, die beim Lernen der Erscheinungsparameter ignoriert werden sollten, da sie nicht das Zielobjekt, sondern den Hintergrund darstellen [2].

Die konsistente Verfolgung von annotierten Objekten und deren relativer



**BILD 3** Erstellung der räumlichen Binärabschätzung, um die Erscheinungsparameter des Objekts zu erlernen (oben); die Gewichtung der einzelnen Kanäle dient während der Lokalisierung des Objekts zur Reduzierung des Rauschens (unten) [2] © Bertrandt)



**BILD 4** Schematische Darstellung des Aufbaus eines Convolutional Neural Nets (CNN) © Bertrandt)

räumlicher Anordnung zueinander erlaubt es, eine zusätzliche Situationslogik zu implementieren, die den Annotationsprozess effizienter gestaltet und die Genauigkeit erhöht. Die unterstützenden Algorithmen erhöhen somit die

Effizienz des Anwenders. Für ihre Verwendung ist allerdings eine Nutzerinteraktion notwendig, wie die Auswahl eines Startpunkts. Daher spricht man an dieser Stelle von einem teilautomatischen beziehungsweise assistierten Vorgang.

## LERNENDE ALGORITHMEN ÜBERNEHMEN DIE FLEISSARBEIT

Um das Labeln der Daten noch stärker zu automatisieren, greift das Werkzeug auf etablierte Algorithmen für maschi-

nelles Lernen zurück. Im Bereich der Computer Vision hat sich die Klasse der Convolutional Neural Nets (CNN) in der Objekterkennung durchgesetzt. Diese Netze bestehen aus teilweise hunderten von Schichten, gefüllt mit Zahlen, in denen die Informationen in verschiedenen Abstraktionsstufen abgespeichert sind (sogenannte verborgene Schichten). In den unteren Schichten speichert das Netz sehr rudimentäre Informationen wie Ecken, Kanten oder Farbverläufe. Diese sind für die meisten Objekte gleich und können daher wiederverwendet werden. Erst in den letzten Schichten bilden sich Merkmale aus zusammengesetzten Features aus, **BILD 4**.

Die Genauigkeit der Algorithmen für maschinelles Lernen im Bertrand-Labeling-Werkzeug steht im Vordergrund, um Objekte wie Verkehrsschilder oder Verkehrsteilnehmer auf einem Bild oder einer Bildsequenz sicher zu lokalisieren und zu klassifizieren. Für diesen Anwendungsfall eignen sich insbesondere F-RCNNs (Faster Region proposal CNNs) [4], erweitert um verschiedene Backbone-Netze wie ResNet oder Inception. Daher wurde genau diese Kombination auch für das Bertrand-Werkzeug verwendet.

## NEURONALE NETZE SCHNELLER TRAINIEREN

Dem Labeling-Werkzeug liegt ein neuronales Netz zugrunde, dessen Trainingskonzept so ausgelegt ist, dass sich ein komplett neues Training anstoßen oder aber ein bereits vorhandenes Modell „auftrainieren“ lässt (sogenanntes Transferlernen). Die Nutzung bereits vorhandener, vortrainierter Netze bietet den Vorteil, dass wesentlich weniger annotierte Daten benötigt werden, um neue Merkmale (Features) zu erlernen. Das Trainieren der eingesetzten neuronalen Netze erfolgt in einer Cloud-Umgebung mit speziell dafür ausgelegter Hardware. Hauptmerkmale wie hohe Verarbeitungsleistung der Rechner mit dedizierten Grafikkarten der neuesten Generation und die Anbindung an skalierbare Cloud-Speicher erlauben ein performantes und zeiteffizientes Lernen. Auch das kundenspezifische Festhalten der so gewonnenen Trainingsstände ist möglich. Sie erlauben zukünftig eine effizientere Anwendung durch das Auftrainieren neuer Merkmale, anstatt

den kompletten Lernvorgang mit altbekannten Merkmalen immer wieder neu durchzuführen.

Nach abgeschlossenem Training können neue Daten vollautomatisch annotiert werden. Dabei muss der Anwender zudem nicht mehr darauf achten, dass die Daten in einer sequenziellen Reihenfolge vorliegen, wie es beim halbautomatischen Tracking noch der Fall ist.

## ZUSAMMENFASSUNG

Bertrand hat ein Werkzeug zum Annotieren von Sensordaten entwickelt. Der Data Labeler arbeitet mit einer spezifischen Schichtarchitektur, ist skalierbar und erledigt den Annotationsprozess weitgehend automatisch. Dabei ist das Thema Annotation von Sensordaten ein Teilaspekt eines gesamten Lösungsansatzes. Denn die Herausforderung fängt bei der Datenaufnahme in Erprobungsfahrten an und erstreckt sich über performante Datenübertragung und intelligentes Datenmanagement bis hin zu vielfältigen Analysen. Zu jedem dieser Teilaspekte benötigt man sinnvolle Lösungen, um eine konsistente Wirkkette abbilden zu können und so die maximale Zeit- und Kosteneffizienz zu erreichen.

## LITERATURHINWEISE

- [1] Krzanic, B.: Data is the new Oil in the Future of automated driving. Pressemitteilung, 2016. Online: <https://newsroom.intel.com/editorials/krzanic-the-future-of-automated-driving/#gs.ITTrSUIZ>, aufgerufen am 06.02.2019
- [2] Lukežič, A.; Vojir, T.; Čehovin, L.; Matas, J.; Kristan, M.: Discriminative Correlation Filter with Channel and Spatial Reliability. In: International Journal of Computer Vision, 2018 (modifiziert)
- [3] Bolme, D. S.; Beveridge, J. R.; Draper, B. A.; Lui Y. M.: Visual Object Tracking using Adaptive Correlation Filters. In: Comp. Vis. Patt. Recognition, IEEE, 2010
- [4] Ren, S.; He, K.; Girshick, R.; Sun, J.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. Online: <https://doi.org/10.1109/TPAMI.2016.2577031>, aufgerufen am 06.02.2019



### READ THE ENGLISH E-MAGAZINE

Test now for 30 days free of charge:  
[www.ATZelectronics-worldwide.com](http://www.ATZelectronics-worldwide.com)