

# Fahrzeuge visuell klassifizieren

**Ein erneuter Paradigmenwechsel?** | Die Grundlage von Verkehrstelematik-Lösungen ist die Erfassung und Charakterisierung des Fahrzeugaufkommens mit Sensoren. Aktuell wird zwar vorwiegend auf Laserscanner gesetzt, aber Vision-basierte Ansätze können mittlerweile mit Laserscannern mithalten. Zwei Forschungsprojekte untersuchten das neue Potenzial von Vision-Lösungen.

KLAUS ZAHN, ANDREAS CADUFF, PETER BUCHER

Seit Mitte der 1990er-Jahre werden weltweit vorwiegend Laserscanner für Anwendungen im Strassen- und Schienenverkehr eingesetzt. Die vor 20 Jahren existierenden Videokamerasysteme für Klassifikations- und volumetrische Aufgaben wurden innert kürzester Zeit verdrängt, da die Distanzinformation der Laserscanner prinzipiell eine sehr einfache, robuste und nicht-rechenintensive Analyse garantiert. Neben der Nummernschilderkennung wurden Kamerasysteme lediglich für Nischenanwendungen, wie z.B. die Überwachung von Kreuzungen eingesetzt.[1] Allerdings stösst man mit Laserscannern mittlerweile an technische Grenzen, insbesondere bedingt durch die endliche Scanrate, welche vor allem bei schnellen Fahrzeugen nur eine sehr limitierte räumliche Auflösung erlaubt. Dieser Effekt wird bei widrigen Witterungseinflüssen noch verstärkt.

Durch die immensen wissenschaftlich-technologischen Fortschritte der letzten Jahre in Computer-Vision sowohl bei der Hardwareentwicklung (Kamera und eingebettete Rechnerplattformen) als auch bei den Algorithmen für maschinelles Lernen, u. a. für die Klassifikation mit Deep-Learning-Methoden, zeichnet sich ab, dass die Videokamerasysteme den Markt zurückerobern, weil sie prinzipiell durch die Fusion von Daten mehrerer Kameras neben der Grauwert- bzw. Farbinformation zusätzlich noch Tiefeninformation liefern – bei voller Videorate und Auflösung.

Im Rahmen der beiden von Innosuisse geförderten Projekte «Kontaktlose Achserfassung für die Verkehrsüberwachung» und «Visionbasierte

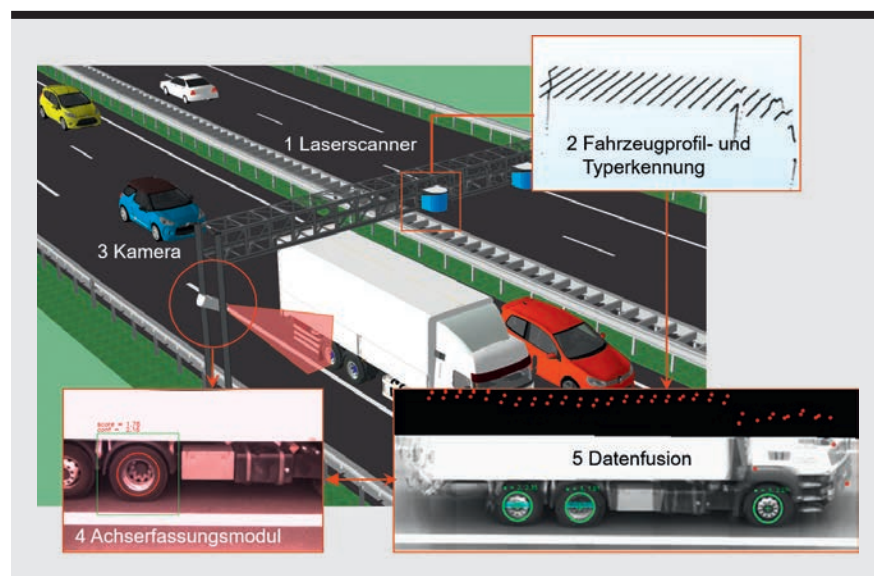
Fahrzeugklassifikation» haben das Kompetenzzentrum Intelligente Sensoren und Netzwerke, CC ISN, und der Umsetzungspartner Sick System Engineering verschiedene Verfahren für die visionbasierte Erfassung, Zählung und Klassifikation von Fahrzeugteilen und ganzen Fahrzeugen erforscht und implementiert.

## Proof of Concept: Kontaktlose Achserfassung

Im ersten Projekt sollte die grundsätzliche Eignung der Kamertechnologie im «harten» Umfeld der Verkehrstelematik unter Beweis gestellt werden. Die Kernanforderung an das System war die kontaktlose Zählung und Positionsbestimmung der Achsen mit einer Genauigkeit von über 98% (auf Fahrzeugebene) bzw. einem Positionsfehler unter 10 cm. Dies bei Fahrzeugen im fließenden Verkehr bis zu einer

Geschwindigkeit von 140 km/h, wofür mit aktueller Lasertechnologie keine zufriedenstellende Lösung möglich war. Hierfür wurde das in **Bild 1** gezeigte Konzept umgesetzt. Es besteht aus einem Standardsystem, dem «Traffic Information Collector» (TIC), welcher um eine Achserfassungskamera erweitert wurde:

- Das TIC-System besteht aus Laserscannern (ein Gerät pro Fahrbahn) in Überkopfperspektive, welche das Fahrzeugprofil erfassen und dessen Typ bestimmen.
- Mit einer Kamera am Fahrbahnrand werden die Achsen detektiert. Hierzu werden die Kamerabilder an ein Achserfassungsmodul weitergeleitet.
- Da die Kamerabilder jeweils nur einen Teil des Fahrzeuges abdecken, werden sie mit den Laserdaten zu einem Gesamtbild des Fahrzeuges zusammengefügt, das mit den Laser-



**Bild 1** Prinzip der Achserfassung durch Fusion der Laser- (1-2) und Videodaten (3-4).

Bilder: HSLU

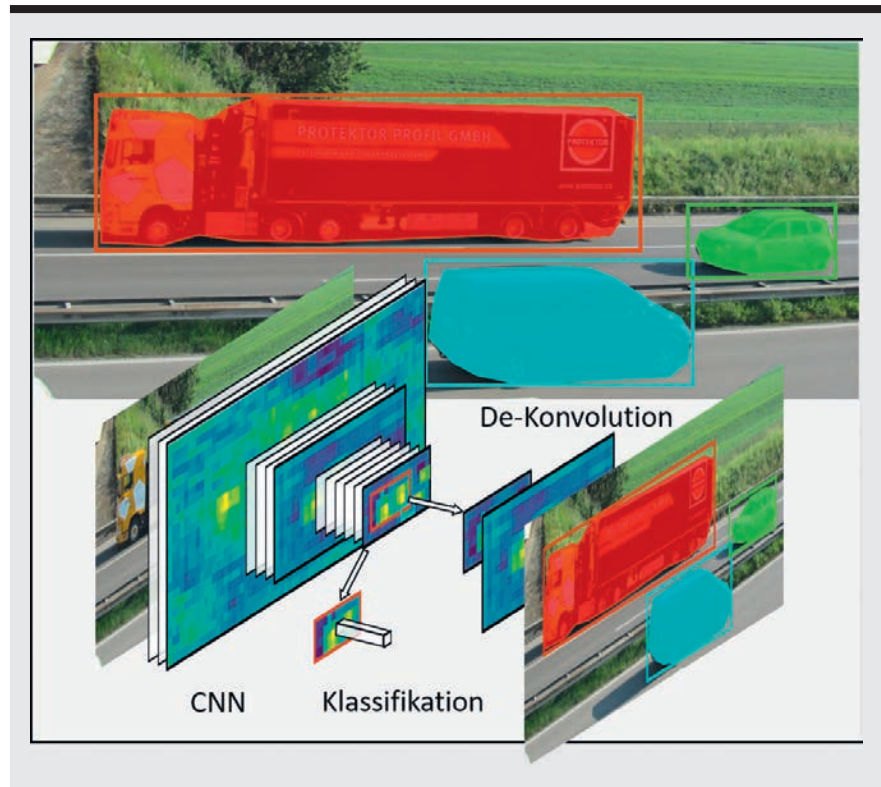
daten fusioniert wird (In dieser Konfiguration erfolgt die Achserfassung nur auf der rechten Spur. Für die linke Spur wäre eine weitere Kamera notwendig).

Die wesentlichen Herausforderungen, welche es im Rahmen des Projektes zu lösen galt, waren eine ausreichende Bildqualität und die Echtzeitverarbeitung. Aufgrund der hohen Fahrzeuggeschwindigkeiten musste die Verschlusszeit der Kamera vor allem in der Nacht ausreichend kurz gehalten werden, um Bewegungsunschärfe zu vermeiden. Hier konnte von der Entwicklung leistungsfähiger Kameras mit Infrarotbeleuchtungen, wie sie u.a. auch für die Nummernschilderkennung entwickelt wurden, profitiert werden. Wesentlich für eine konstant hohe Bildqualität war aber zusätzlich eine optimale Einstellung der Kamera- und Beleuchtungsparameter in Echtzeit. Das Achserfassungsmodul berechnete diese aus den aktuellen Umgebungslichtbedingungen und dem Bildinhalt und steuerte dann Kamera und Beleuchtung entsprechend. Hier konnte davon profitiert werden, dass offene http-Schnittstellen zur Steuerung der Parameter bei IP-Kameras mittlerweile Stand-der-Technik sind. Eine weitere Herausforderung war die Geschwindigkeit der verwendeten Algorithmen, die das Ergebnis trotz beschränkter Prozessorressourcen auf lüfterlosen Industrierechnern in Echtzeit liefern mussten. Die Erforschung, Implementierung und Optimierung der notwendigen Algorithmen stellte dabei den wesentlichen Forschungsinhalt des Projektes dar.

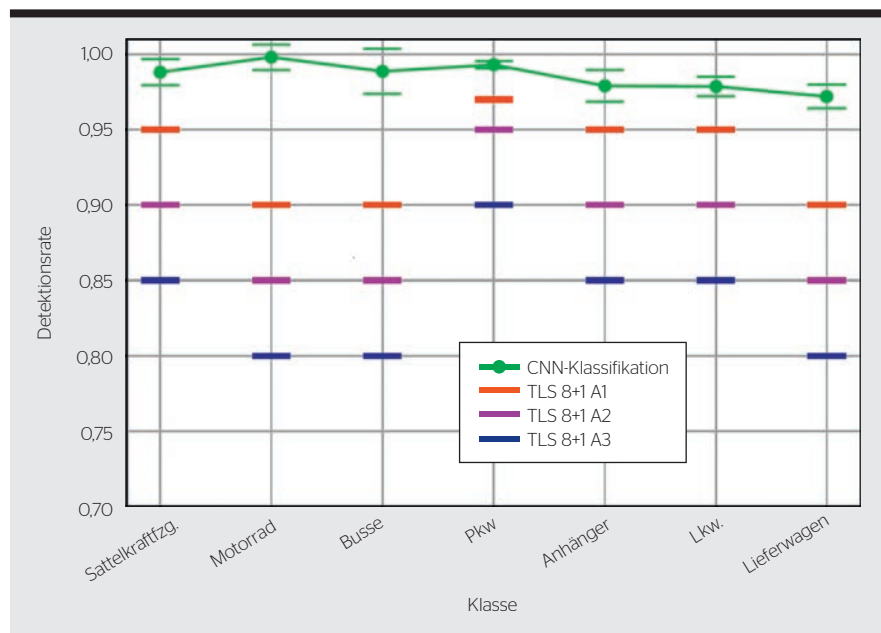
In einem Feldtest konnte gezeigt werden, dass die geforderte Genauigkeit von 98% unter allen Bedingungen zu jeder Tageszeit erfüllt wird. Dabei bezieht sich die Genauigkeit auf die Fahrzeuge (d.h. für 98% aller Fahrzeuge wurde die Achszahl korrekt ermittelt), sodass die algorithmische Genauigkeit der Achszählung noch deutlich höher ist.

### Vision-basierte Fahrzeugklassifikation

Nachdem sich im Rahmen der Achserfassung gezeigt hatte, dass mittlerweile visionbasierte Ansätze für die Verkehrstelematik das Potenzial haben, die Genauigkeit von Laserscannern zu übertreffen, wurde ein weiteres Projekt



**Bild 2** Prinzip eines Mask R-CNN zur Segmentierung der Fahrzeuge im Bild.



**Bild 3** Erste Ergebnisse der Klassifikation mit einem CNN basierend auf etwa 60 000 Fahrzeugen. Die grünen Punkte stellen die erreichten Detektionsraten (inkl. statistischem Fehler) dar. Die farbigen Linien zeigen die jeweils geforderten Mindestraten der TLS 8+1, in Rot die höchste Anforderung gemäss TLS 8+1 A1.

initiiert, welches die Klassifikation der Fahrzeuge zum Ziel hat. Hierbei sollte nun aber aus wirtschaftlichen Gründen ein reines Kamerasystem – d.h. ohne Fusion mit dem Laserscanner – verwendet werden. Als Grundlage für die

Klasseneinteilung wurde, wegen ihrer hohen Relevanz als Standard, die TLS 8+1 (Technische Lieferbedingungen für Streckenstationen [2]) genutzt. Im Projekt sollte untersucht werden, ob visionbasierte Methoden alleine in der

Lage sind, die Norm mit der höchsten Genauigkeit – die TLS 8+1 A1 – zu erreichen. Diese Anforderung wird aktuell von keinem Laserscanner (oder Vision System) erreicht.

Die Tatsache, dass die Fahrzeugklassifikation als reines Kamerasystem funktionieren soll, stellte eine deutliche Steigerung der Komplexität der Aufgabenstellung dar. Für die Achserfassung erfolgte die Detektion des Fahrzeuges noch durch den Laserscanner, welcher die Achserfassung durch die Kamera triggert (**Bild 1**). Nun sollte diese Aufgabe allein durch die Kamera erfüllt werden. Daher musste für das Projekt neben der eigentlichen Fahrzeugklassifikation auch deren Detektion – die sogenannte Segmentierung im Bild – gelöst werden.

Grundlage für die Lösung von beiden Aufgaben stellen CNNs (Convolutional Neural Networks) dar, welche in den letzten Jahren im Rahmen der Objektklassifikation im Zusammenhang mit Deep-Learning-Methoden intensiv untersucht wurden. Dabei funktioniert ein CNN für die Klassifikation so, dass für ein Eingabebild entschieden wird, welcher vortrainierten Klasse das Bild angehört. Im Rahmen der TLS 8+1 müssen die Fahrzeuge z. B. in die Klassen Motorrad, Pkw, Lieferwagen, Lkw, Pkw mit Anhänger, Lkw mit Anhänger, Sattelkraftfahrzeuge und Busse eingeteilt werden. Allerdings wird hier vorausgesetzt, dass das zu klassifizierende Bild nur eine Klasse (d.h. Fahrzeug) enthält, mit anderen Worten, dass die Segmentierung der Fahrzeuge schon erfolgt ist. Zur Lösung

dieses Problems wurde eine Weiterentwicklung von CNNs, das sogenannte Mask R-CNN, verwendet.

### Segmentierung mit Mask R-CNN

Mask R-CNNs [3] basieren prinzipiell auch auf einem CNN, bzw. enthalten dieses als eine wesentliche Architekturkomponente. Zusätzlich werden aus den Feature Maps des CNNs (d.h. den Ergebnissen der Konvolution) Kandidaten für Objektregionen ermittelt. Hierbei werden pro Bild sehr viele (mehrere Hundert oder Tausend) verschiedene Objektkandidaten ermittelt und hieraus mittels einer Klassifikation die relevanten selektiert. In **Bild 2** sind dies die farbigen Boxen um die drei Fahrzeuge. Damit können in einem Bild mehrere Objekte bzw. Klassen ermittelt werden. Zusätzlich wird für diese Objekte, im Rahmen einer sogenannten De-Konvolution, eine pixelgenaue Segmentierung ermittelt. Diese ist durch die farbig umrahmten Regionen im **Bild 2** dargestellt. Damit ist es nun möglich, alle Fahrzeuge – auch bei Überlappung – im Bild zu erfassen und mit zusätzlichen Methoden des «Trackings» von Bild zu Bild in der Sequenz zu verfolgen.

Prinzipiell findet im Mask R-CNN auch eine Klassifikation statt, sodass auch die detaillierte Fahrzeugkategorie direkt bestimmt werden könnte. Allerdings wird hierauf aus folgenden Gründen verzichtet:

- Die Anforderungen von Segmentierung und Klassifikation widersprechen sich. Mask R-CNNs sind sehr rechenintensiv, gerade bei grosser

Bildaufflösung. Die Segmentierung der Fahrzeuge muss für ein robustes Tracking aber bei hoher Bildrate (d.h. sehr schnell) erfolgen, wofür eine niedrige Bildauflösung vorzuziehen ist. Im Gegensatz hierzu benötigt die Klassifikation – vor allem bei subtilen Klassenunterschieden – eine hohe Bildauflösung. Zudem genügt es, die Klassifikation eines Fahrzeuges nur einmal im Verlaufe der Bildsequenz, während der es sichtbar ist, durchzuführen.

- Flexibilität bei der Klassifikation. Neben der Klassifikation gemäss TLS 8+1 gibt es noch weitere, auch deutlich detailliertere, Kategorisierungssysteme, welche projektspezifisch verwendet werden. Bei einer Änderung des Kategorisierungssystems müsste dann jedes Mal das vollständige Mask R-CNN neu trainiert werden, was einen deutlich höheren Aufwand darstellt als nur die Anpassung der Klassifikation mit einem Standard-CNN.

Daher wird das Mask R-CNN bei niedrigerer Auflösung im Wesentlichen nur für die Segmentierung der Fahrzeuge verwendet, wobei zusätzlich eine einfache Klasseneinteilung in «Pkw-ähnlich» und «Lkw-ähnlich» erfolgt. Nachdem das Fahrzeug den Sichtbereich der Kamera verlassen hat, erfolgt dessen Klassifikation auf Basis eines Bildes bei maximaler Auflösung.

### Klassifikation

Für die eigentliche Klassifikation wird dann ein Standard-CNN verwendet, welches mit einem (hochauflösenden)

## RÉSUMÉ

### Classification visuelle des véhicules

Un nouveau changement de paradigme ?

La saisie et la caractérisation du parc de véhicules à l'aide de capteurs sont à la base des solutions en matière de systèmes de transport intelligent (STI). Bien que pour ce faire les scanners laser soient actuellement privilégiés, les approches basées sur la vision sont désormais en mesure de les concurrencer.

Deux applications ont montré que la technologie actuelle des caméras permet aujourd'hui d'atteindre la précision des scanners laser en cas de problèmes de classification dans les applications STI. La combinaison d'un scanner laser avec une caméra et une fusion intelligente des données a permis d'augmenter la précision de la reconnaissance sans contact

des essieux des véhicules circulant dans des conditions de trafic fluides à un niveau sans précédent. Pour ce faire, des algorithmes d'apprentissage automatique « conventionnels » ont encore été utilisés, pour lesquels les classificateurs ont été conçus manuellement. Le plein potentiel de la technologie des caméras ne peut être exploité que si l'on dispose de suffisamment de données pour l'entraînement nécessaire aux méthodes de deep learning (apprentissage profond) telles que les réseaux neuronaux convolutifs (CNN). Ainsi, pour la première fois, il paraît possible de classer les véhicules très précisément, selon la TLS 8+1 A1, uniquement grâce à la technologie des caméras.

NO

Teilbild bestehend aus der Fahrzeugregion (farbige Boxen in **Bild 2**) «gefüttert» wird.

Wesentlich für das Erreichen einer hohen Genauigkeit bei der Klassifikation ist die Verfügbarkeit eines ausreichend grossen und repräsentativen Datensatzes für das Training des CNN. Hierfür konnte im Rahmen des Projektes auf Basis einer Vorselektion mit dem Laserscanner eine in dieser Form einzigartige Datenbank von über 270 000 Fahrzeugen aus unterschiedlichen Perspektiven und bei verschiedenen Witterungs- und Beleuchtungsbedingungen erstellt werden. Diese sind – manuell korrigiert – in über 20 Kategorien eingeteilt. Dabei fallen mehr als die Hälfte (150 000) in die «seltenen» d.h. Nicht-Pkw-Kategorien. Und diese stellen in der Regel die interessanten bzw. die schwierigen Fälle dar. Zusätzlich sind für alle Fahrzeuge sowohl die Objektregion als auch die Segmentierung (**Bild 2**) verfügbar. Im Vergleich hierzu umfasst der aktuell grösste, öffentlich verfügbare Datensatz zwar über 430 000 Fahrzeuge.[4] Allerdings entfallen von den sieben Fahrzeugklassen nur 110 000 in die «seltenen» Kategorien, und der überwiegende Anteil

stellt Pkw dar. Schliesslich ist die detaillierte Segmentierung nicht verfügbar.

Die Auswertung der Fahrzeugdatenbank, insbesondere das Training und die Evaluation der Neuronalen Netze ist aktuell noch in vollem Gange. In **Bild 3** sind erste Ergebnisse einer Klassifikation gemäss TLS 8+1 dargestellt [5], welche darauf hindeuten, dass mit einer visionbasierten Klassifikation die Norm TLS 8+1 A1 erfüllt werden kann.

### Fazit

Im Rahmen von zwei Anwendungen konnte gezeigt werden, dass aktuelle Kameratechnologie mittlerweile in der Lage ist, bei Klassifikationsproblemen in der Verkehrstelematik die Genauigkeit von Laserscannern zu erreichen. Durch die Kombination eines Laserscanners mit einer Kamera und einer intelligenten Datenfusion konnte die Genauigkeit der kontaktlosen Achserkennung von Fahrzeugen im fliessenden Verkehr auf ein bisher nicht erreichtes Mass gesteigert werden.[6] Dabei wurde noch auf «konventionelle» Algorithmen des maschinellen Lernens gesetzt, für welche ein manuelles Design der Klassifikatoren

erfolgte. Das volle Potenzial der Kameratechnologie kann erst ausgeschöpft werden, wenn ausreichend viele Daten zum Training von Deep-Learning-Methoden wie CNNs vorhanden sind. Hiermit erscheint zum ersten Mal die hochgenaue Fahrzeugklassifikation nach TLS 8+1 A1 rein mit Kameratechnologie machbar.

### Referenzen

- [1] [www.citilog.com](http://www.citilog.com)
- [2] [www.bast.de/DE/Verkehrstechnik/Publikationen/Regelwerke/Unterseiten/V5-tls-2012.html](http://www.bast.de/DE/Verkehrstechnik/Publikationen/Regelwerke/Unterseiten/V5-tls-2012.html)
- [3] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN, 2017.
- [4] Z. Luo, et al., «MIO-TCD: A New Benchmark Dataset for Vehicle Classification and Localization», IEEE Transactions on Image Processing, 27(10), 2018, 5129-5141.
- [5] A. Caduff, Anwendungsspezifische Optimierung eines Mask R-CNN, Vertiefungsmodul 1, HSLU, 2018.
- [6] [www.hslu.ch/de-ch/hochschule-luzern/ueber-uns/m Medien/m Medienmitteilungen/2017/10/18/achszahlen-genauer-bestimmen](http://www.hslu.ch/de-ch/hochschule-luzern/ueber-uns/m Medien/m Medienmitteilungen/2017/10/18/achszahlen-genauer-bestimmen)

### Autoren

Prof. Dr. **Klaus Zahn** ist Forschungsgruppenleiter am CC Intelligent Sensors and Networks der Hochschule Luzern.  
→ HSLU, 6048 Horw  
→ klaus.zahn@hslu.ch

**Andreas Caduff** ist wissenschaftlicher Assistent am CC Intelligent Sensors and Networks der Hochschule Luzern.  
→ andreas.caduff@hslu.ch

**Peter Bucher** ist Leiter Application Development  
→ SICK System Engineering, 6374 Buochs  
→ peter.bucher@sick.ch



**Die einzigartige modulare Weiterbildung**

**Elektromagnetische  
Verträglichkeit EMV  
Elektrobiologie**

Elektrobiologie, EMV Grundlagen	28.08.2019
Niederfrequente Felder (NF)	11.09.2019
Hochfrequente Felder (HF)	25.09.2019
EMV Planung & Gesundheit	30.10.2019
Messtechnik richtig angewandt	14.11.2019

**SABE**® | Schweizerische Arbeitsgemeinschaft  
Biologische Elektrotechnik

Information und Anmeldung:  
[www.bildungszentrumbaubio.ch](http://www.bildungszentrumbaubio.ch)

## LANZ protected®

Steigleitungen 400 A 700 A 1250 A

Die neuen LANZ Steigleitungen mit **EMV-Schirmung auf 1 Mikrottesla in 1 m Abstand** und mit **Funktionserhalt im Brandfall E90** sind die besten feuersicheren Steigleitungen für Wohn- und Bürohochhäuser.

**LANZ – kompetent in Stromschienen. Rufen Sie an: 062 388 21 21. Verlangen Sie Auskunft und Offerte.**

..... STS2\_1

 **stromschienen  
lanz oensingen ag**

CH-4702 Oensingen [www.lanz-oens.com](http://www.lanz-oens.com) Tel. ++41/062 388 21 21  
Südringstrasse 2 [info@lanz-oens.com](mailto:info@lanz-oens.com) Fax ++41/062 388 24 24

**modino:**  
denkt weiter

## priamos

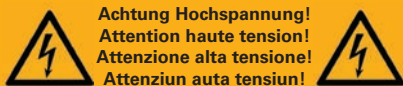
Das Installationsprogramm von modino.  
So individuell wie Sie.



- Grosse ergonomische Schaltflächen
- Optimale Steckdosennutzung
- Farblich auf die Wohnsituation abstimbar

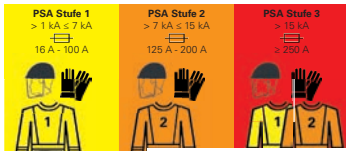
[www.modino.ch](http://www.modino.ch)

Privatkunden (B2C) können unsere Produkte über die Vertriebspartner im Detailhandel und E-Commerce beziehen.  
Für Firmenkunden (B2B) stehen Ihnen unsere Verkaufsberater gerne zur Verfügung. [www.maxhauri.ch](http://www.maxhauri.ch)



### Die 5 Sicherheitsregeln

- freischalten und allseitig trennen
- gegen Wiedereinschaltung sichern
- auf Spannungsfreiheit prüfen
- erden und kurzschliessen
- gegen benachbarte, unter Spannung stehende Teile schützen



Polizei 117 | Feuerwehr



**Achtung!**  
Bedienung der NH-Sicherungen  
nur durch instruiertes Personal!

**Wichtige Kleber für Ihre Sicherheit**

**Jetzt online bestellen!**

[www.electrosuisse.ch/aufkleber](http://www.electrosuisse.ch/aufkleber)

