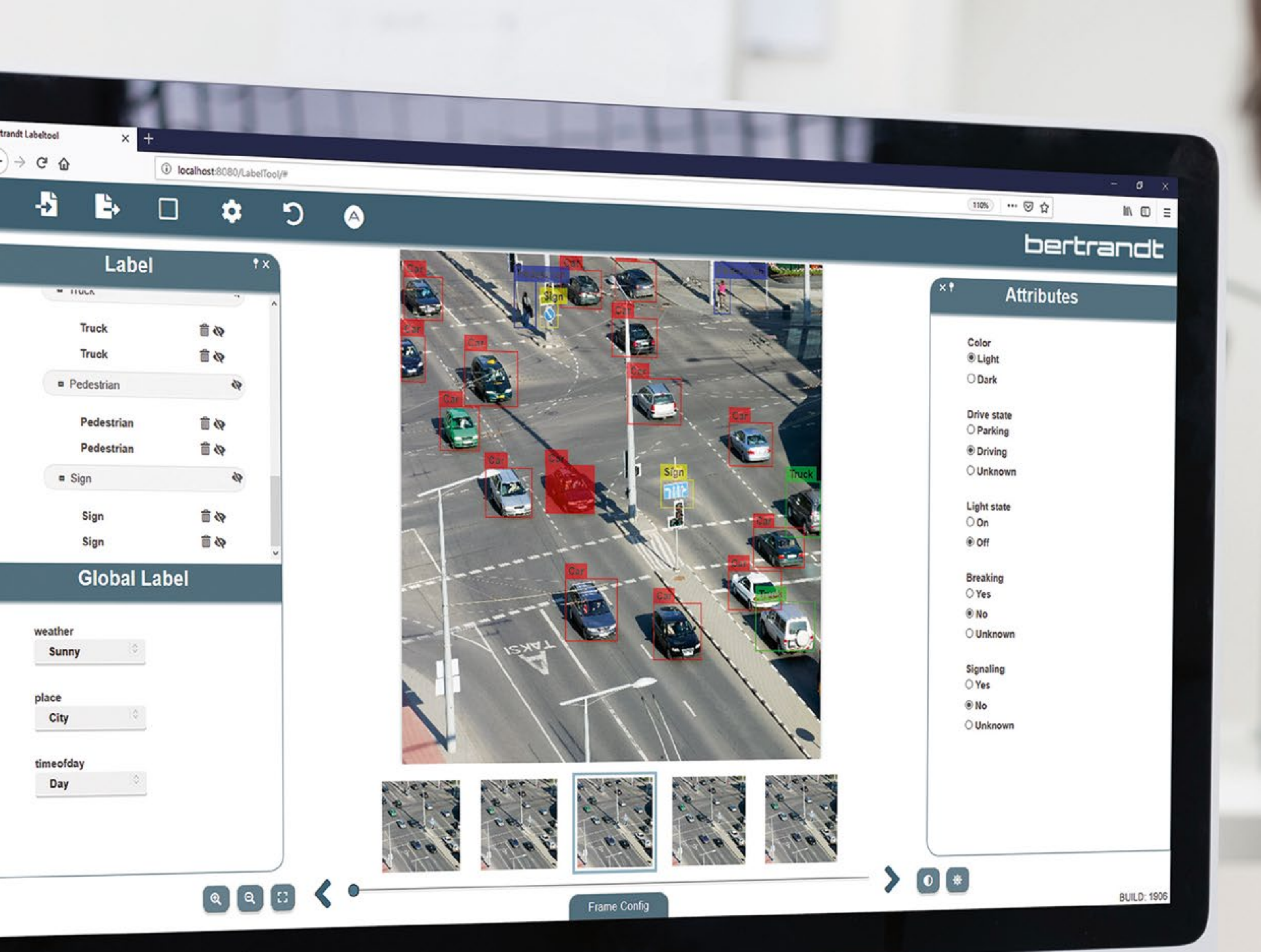

Software for the Automated Annotation of Sensor Data

Valid data is crucial for the successful development of algorithms for assistance systems for autonomous driving. The operating systems need a thorough annotation with the so-called ground truth information. During the validation of image capture, more than 100 TB of data have to be recorded and analyzed per vehicle per day. With the Data Labeler, Bertrandt has developed a tool in order to realize high-performance handling and automated analysis of these data.

AUTHORS

Dipl.-Phys. Stefan Schindler
is Data Scientist in the Field of Machine Learning at Bertrandt in Rüsselsheim (Germany).

Dipl.-Bioinf. Holger Bergmann
is Data Scientist in the Field of Computer Vision at Bertrandt in Rüsselsheim (Germany).

Torsten Bohnet, M. Sc.
is Lead Engineer in the Field of Analytics at Bertrandt in Rüsselsheim (Germany).

Dr. Dipl.-Phys. Yusuf Erdogan
is Team Leader Analytics at Bertrandt in Rüsselsheim (Germany).

CHALLENGES IN THE VALIDATION OF AUTONOMOUS DRIVING FUNCTIONS

The functioning of driver assistance systems on SAE level 3 and higher is based on interaction between a multitude of sensors. This functionality is validated by comparing the measurements of these sensors with the underlying true information, also referred to as ground truth information. The proverb “Data is the new oil in the future of automated driving” [1] does apply here in a special way; data is as well the lubricant that keeps all elements going as it is the “black gold” in sense of precious yet also costly component of the functions to come for future autonomous or assisted driving that is to be explored.

Achieving the required validation quality requires a huge amount of data. This typically amounts to several hundred million camera images (typically from a total of 10,000 h of measurement test driving at a data rate of 30 frames/s). In addition to this comparison between actual values and target values, the machine learning algorithms that are increasingly being used in vehicles must be trained by using large quantities of highly accurate data. Ground truth data are therefore also used in this case, as they very precisely indicate the position and the properties of the objects shown in the individual images. This information is also supplemented by the addition of numerous, non-object-specific data, which, for example, describe the weather conditions or the classification of the driving scenarios (global labels).

INCREASING QUANTITIES OF DATA COMPLICATE MANUAL ANNOTATION

Until now, the high requirements regarding data quality, safety, and security meant that these data often had to be generated by trained employees who manually annotated the raw sensor data. As the quantities of data continue to increase and algorithms work better and better, in combination with a reduction in the cost of computing power, this process will, in the future, no longer be time-efficient nor attractive with regard to cost. For that reason, Bertrandt is developing a user-friendly, freely configurable, and high-performance cloud-

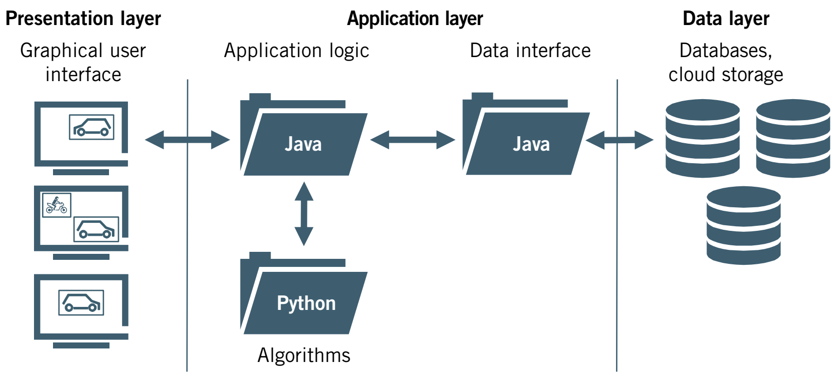

FIGURE 1 Schematic representation of the layer architecture of the labeling tool (© Bertrandt)

based software called Data Labeler to automate the annotation of camera data to the greatest possible extent.

FLEXIBLE SOFTWARE ARCHITECTURE FOR SCALABLE SOLUTIONS

The wide variety of possible data to be analyzed will present special challenges for the architecture and algorithms of the tools in the future, **FIGURE 1**. The annotation process depends on the customer, the target system, and the application field and is very specific. Configurability and flexibility were therefore core requirements for the product. For the architecture of the labeling tool, classical and proven approaches have been linked with modern and innovative aspects. Individual components of the overall system are assigned to clear tasks and divided into layers. The complete architecture is modular in design, thus guaranteeing maximum flexibility. As a result, robustness, expandability, ease of maintenance, and security for the future are optimized. The architecture is designed in such a way that the software can deal with many different file formats and innovative storage solutions such as mass storage in the cloud and can be further developed in the future independent of the hardware.

The workflow is a Java implementation which controls Python-based algorithms for assisted and automated annotation. The user is provided with a web user interface with intuitive operating concepts. This interface provides all functions such as bounding boxes, segmentations, points, lines, and polygons as well as the validation of the labels generated by the algorithms.

FUNCTIONS FOR ASSISTED LABELING

To support the annotation process, the user can employ state-of-art computer vision algorithms. In particular, object tracking and object detection are used to enable (semi-) automated annotation. Object tracking is the localization, with certain termination criteria, of one or more target objects in a video sequence, **FIGURE 2**. For objects that are following a straight-line trajectory, the user can employ the linear interpolation function, which adds the missing annotations for one or more objects between the two already annotated frames. However, in order to actually localize an object, an initial model of the object is created in the first frame of the video sequence. This model consists of the appearance and the position parameters of the selected object and is designed by defining a bounding box. The contents of a bounding box can therefore also be described as a training area. Continuous localization is guaranteed by updating the model from video frame to video frame, **FIGURE 3**.

CORRELATION FILTER FOR OBJECT RECOGNITION

For the recognition of an object in each subsequent frame, the software uses the object's appearance parameters, so-called correlation filters, which make it possible to estimate the similarity of the pixel from the training area to a pixel in the subsequent frame in the search area [2, 3]. The search area can be restricted on the basis of the position parameters in such a way that the localization algorithm maintains an efficient runtime complexity, which means that it is not necessary to search the entire frame. In order to minimize background contamination during model updating, a spatial binary estimation for a correlation filter is used to identify those pixels that should be

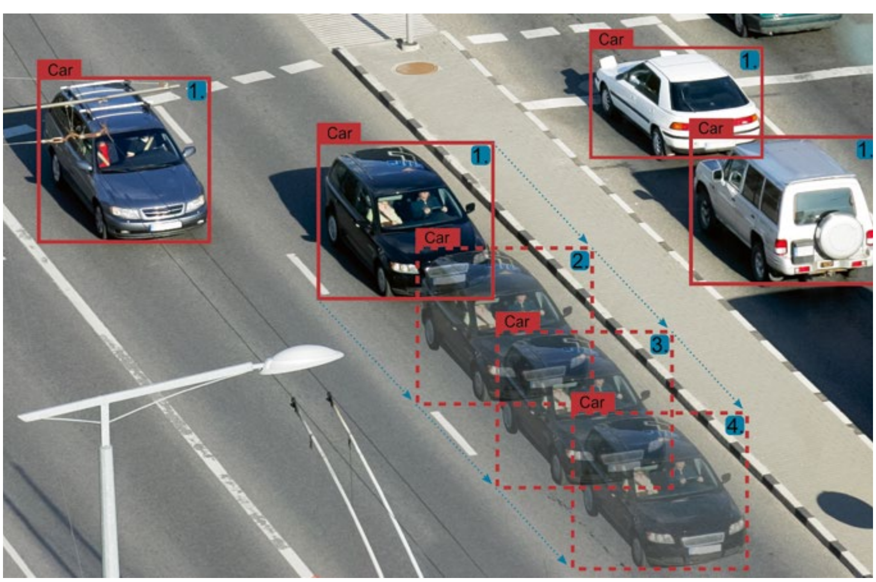

FIGURE 2 Annotations in a video sequence can be automatically tracked through the next frames; the contents of a bounding box that has been defined are used to create an appearance and motion model; the automatically set annotations (dotted bounding boxes) schematically represent an object being tracked to the fourth frame (© Bertrandt)

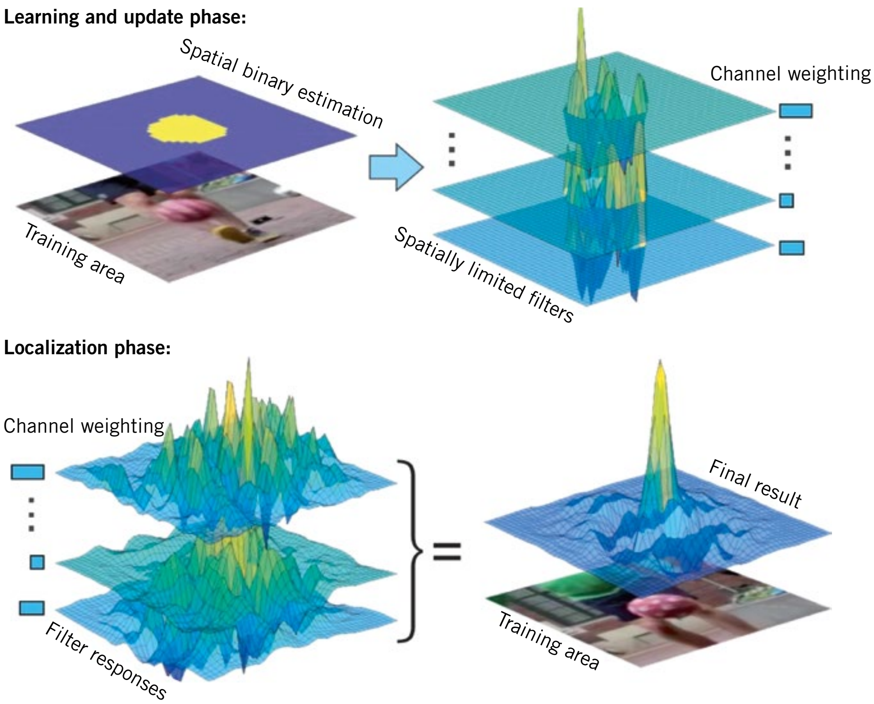

FIGURE 3 Creation of the spatial binary estimation for learning the appearance parameters of the object (top); the weighting of the individual channels is used to reduce noise during the localization of the object (bottom) [2] (© Bertrandt)

ignored when the appearance parameters are being learned, as they do not represent the target object but only the background [2].

The possibility of consistently tracking annotated objects and their relative spatial arrangement to each other makes it possible to implement additional situation logic in order to make the annotation process more efficient and to increase accuracy. The supporting algorithms therefore increase the user's efficiency. However, user interaction, for example the selection of a starting point, is necessary for their use. For that reason, one refers to this as a semi-automatic or assisted process.

LEARNING ALGORITHMS DO THE HARD WORK

In order to further automate the labeling of the data, the tool makes use of machine learning algorithms. In the field of computer vision, the class of Convolutional Neural Networks (CNN) has become established in object recognition. In some cases, these networks consist of hundreds of layers filled with numbers in which information is stored at different stages of abstraction (so-called hidden layers). In the lower levels, the network "stores" very rudimentary information such as corners, edges, or color gradients. These are the same for most objects and can therefore be reused. Only in the last layers is a classification formed from the assembled features, **FIGURE 4**.

The accuracy of the machine learning algorithms in the Bertrandt labeling tool is a priority for safe and reliable localization and the classification of objects such as traffic signs or road users in an image or image sequence. Particularly suitable

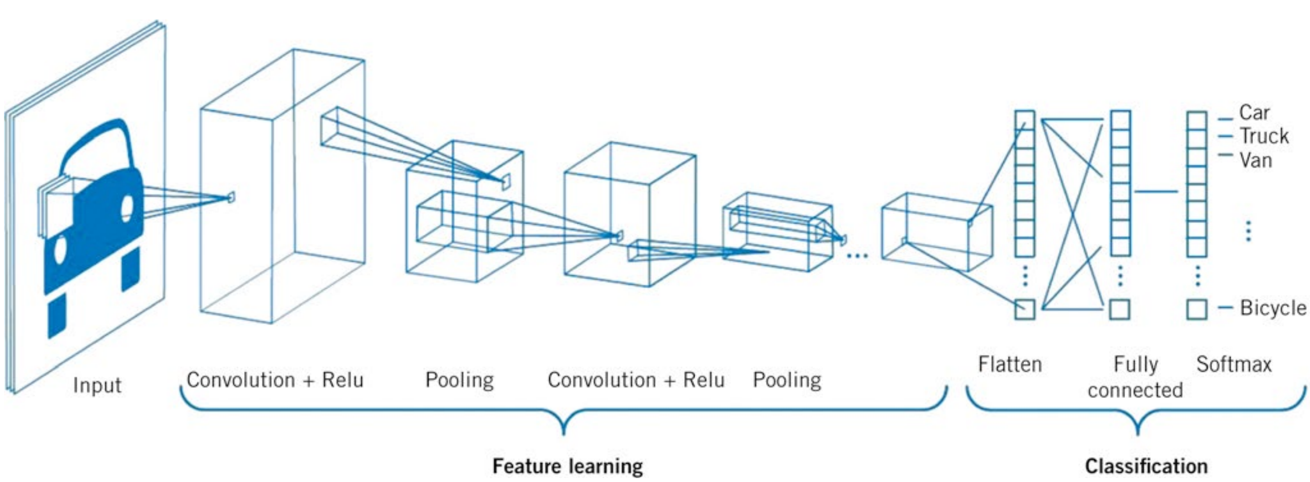

FIGURE 4 Schematic representation of the structure of a Convolutional Neural Network (CNN) (© Bertrandt)

for this application case are F-RCNNs (Faster Region Proposal CNNs) [4], supplemented by various backbone networks such as ResNet or Inception. For that reason, it is precisely this combination that is also used for the Bertrandt tool.

FASTER TRAINING OF NEURAL NETWORKS

As far as the training concept for the neural network is concerned, the Bertrandt labeling tool is designed in such a way that a completely new training session can be initiated or an already available model can be "trained up" (so-called transfer learning). Using already available, pretrained networks offers the advantage that significantly fewer annotated data are required in order to learn the features. The neural networks are trained in a cloud environment with hardware designed specifically for this purpose. Key features such as high computing power with dedicated, latest-generation graphics processors and connection to scalable cloud storage systems ensure high-performance and time-efficient learning. The training levels thus achieved can be retained for the individual customer. This ensures more efficient usage in the future by allowing new features to be added and trained instead of carrying out the complete learning process right from the start with old and known features.

Once the training has been completed, new data can be annotated fully automatically. It is no longer necessary to ensure that the data are available in a sequential order, as is still the case with semi-automatic tracking.

SUMMARY

Bertrandt has developed a tool for the annotation of sensor data. The main features of the Data Labeler are a new specific layer architecture, a user-friendly operating concept, scalability, and automation of the annotation process with algorithms as the core component. For Bertrandt, the subject of annotating sensor data is a sub-aspect of a comprehensive solution. There are many challenges, including the collection of data from test drives, high-performance data transfer, intelligent data management, and various data analyses. Each of these sub-aspects requires the development of sensible solutions in order to provide a consistent chain of action and therefore to achieve maximum time and cost efficiency.

REFERENCES:

[1] Krzanic, B.: Data is the new oil in the future of automated driving. Press release, 2016. Online: https://newsroom.intel.com/editorials/krzanich-the-future-of-automated-driving/#gs.ITTrSUIZ, access: February 6, 2019

[2] Lukežič, A.; Vojíř, T.; Čehovin, L.; Matas, J.; Kristan M.: Discriminative Correlation Filter with Channel and Spatial Reliability. In International Journal of Computer Vision, 2018 (modified)

[3] Bolme, D. S.; Beveridge, J. R.; A. Draper, B. A.; Lui, Y. M.: Visual Object Tracking using Adaptive Correlation Filters. In: Comp. Vis. Patt. Recognition. IEEE, 2010

[4] SRen, 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, access: February 6, 2019

optimize! softing

DIAGNOSTIC AND TEST SOLUTIONS BY SOFTING

AUTOMOTIVE

automotive.softing.com