

← back

30. April 2021 - Providentia Editors

How artificial intelligence recognizes objects and looks into the future

What vehicles are on the road, where they are likely to go, and when they will probably change lanes: Artificial intelligence sifts through gigantic amounts of data in fractions of a second, providing new insights into traffic and laying the foundation for value-added services.

An intelligent car knows which lane it should drive in to keep traffic moving. It speeds up and slows down when appropriate. It can even recognize that a certain constellation at an intersection is very likely to cause an accident and suggest what to do. **Christoph Schöller**, a scientist at fortiss, is looking into the future with the help of artificial intelligence. He calls it "motion prediction." His goal is clear: "Being able to reliably predict collisions is like the holy grail of our research."

"Being able to reliably predict collisions is like the holy grail of our research," says Christoph Schöller from fortiss.



AI: Solving problems independently

This has not yet been achieved, then. A great deal of preliminary work was and is necessary for Schöller to be able to press on with his research and to complete his PhD thesis by the end of the year. Recognizing objects (I.), fusing data (II.), and, finally, predicting motion (III.): These are the three steps that scientists from the consortium leader TU Munich and partners such as fortiss have been researching for four and a half years in the Providentia++ project. Wherever the system learns and independently finds answers and solves problems, artificial intelligence (AI) is helping not only to develop a precise digital twin of traffic, but also to create the basis for apps that can make concrete recommendations.

1. Neural networks for object recognition

In order to predict traffic motion, vehicles must be digitally recorded and displayed in a digital twin of the traffic. To do this, sensors such as cameras, radars, and lidars observe traffic.

Detecting the vehicles is particularly challenging: In order to identify them as reliably as possible, neural networks require so-called annotated data. This is raw data in which the positions of the vehicles have already been manually readjusted and optimized. Since this process takes a lot of time, the researchers use existing public data sets that already have very precise annotations. And as many of them as possible: "You can never have enough annotated data. It's hard to get hold of, and generating it yourself takes a lot of effort," says **Walter Zimmer**, a scientist at the TU Munich who uses existing data sets (MS COCO and Visdrone) to identify road users. But these data sets each have their own individual drawbacks: MS COCO is based on 118,000 annotated images, but does not detect vans. Visdrone offers more than twice as many annotated images (265,000), with the added bonus of being recorded from the air by a drone. "That serves us well, since we also look down on vehicles from above with our sensors on the overhead signs," Zimmer says. The downside: sometimes the neural network doesn't get the differentiation between vans and passenger cars right, and jumps back and forth between these classes. That's why the computer scientist Zimmer has developed his own program that semi-autonomously annotates data on the Providentia++ test track. He released the first version of the program at the Intelligent Vehicles Symposium in Paris in 2019. "The program automates many things," he says. It can also be used for 3D detection, which is now the basis for the Provid-21 data set.

Also used for 3D detection is the KITTI data set, which Zimmer explains is "a comparatively small data set on which you can train more quickly." The scientists also make use of the nuScenes and Waymo data sets, which capture a greater variety of objects.



"You can never have enough annotated data. It's hard to get hold of, and generating it yourself takes a lot of effort," says TU Munich scientist Walter Zimmer.

2. Classical algorithms in fusion

Next, the "insights" from the detections are brought together in a high-level fusion. This involves the use of "classical algorithms," which do not learn from data but can use the information from AI image recognition. The advantage: they generalize new and unknown situations better, and their function can be proven, unlike that of neural networks. According to Christoph Schöller, however, this is not artificial intelligence. "Statistical methods are used to fuse data," the fortiss scientist explains. "The system does not learn in a data-driven way, so it cannot draw conclusions from history." See also: [Many sensors, one digital twin: How does it work?](#)

3. Neural networks in motion prediction

Artificial intelligence becomes all the more important when it comes to predicting vehicles' movements, known as trajectories. In order to extract the most important features from a large amount of data, such as that on the A9 highway, one typical characteristic of neural networks helps: "If you train a neural network on five trajectories, it remembers how each vehicle moves according to them," says Christoph Schöller. "But if more are added, it becomes more and more difficult. When the network can't remember the huge amount of data, it is forced, so to speak, to recognize generalizable patterns." The neural network's necessity becomes a virtue for Schöller, for these generalized patterns can be used to deduce which trajectories vehicles are most likely to follow. "We extrapolate into the future," says Schöller, who is currently already preparing for a particular challenge – traffic at the intersections that will be equipped with elaborate sensor technology in the test field expansion. "On the highway, movements are fairly uniform, speeds largely constant, and lane changes traceable. Maneuvering at an intersection with bicyclists, pedestrians, and multiple options for turning is far more complex," says Schöller, who has already developed a model for this that incorporates the movements of all road users ("agents") and the road layout into the prediction. A great advantage for this model is provided by the external infrastructure in Providentia++: Its bird's-eye view allows it to see further than before and to detect vehicles that are obscured by others. This enables more reliable prediction.

Value-added services are only feasible with AI

All that is really missing now, as a final step, is an app that passes on the aforementioned services to connected vehicles. The conditions are good: "Test vehicles from Valeo and Elektrobit are already getting the digital twins on board in quasi-real time of 40 ms," explains the scientist Zimmer. Special software has been programmed for this purpose, which now "just" needs to be further developed into an app of its own. Once again, the advantage of such an app lies in the external infrastructure, because far more computing power is available for the digital twins. The complex calculations of the trajectories have already been made before they are transmitted on board a vehicle.

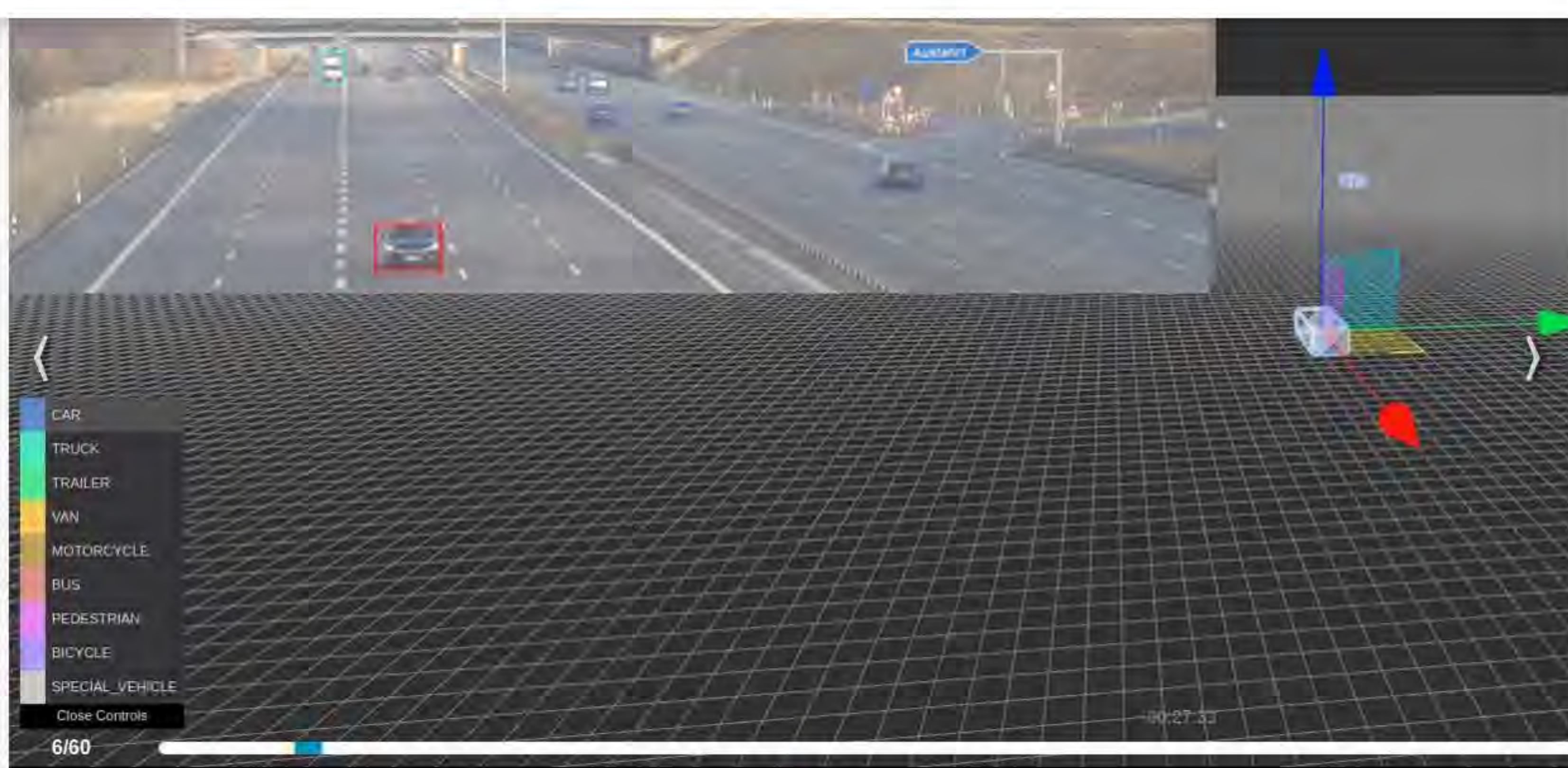
Further information

[AI Self-Driving Cars Divulgement: Practical Advances in Artificial Intelligence and Machine Learning](#), Lance Eliot, 7-2020

[FloMo: Tractable Motion Prediction with Normalizing Flows](#), Christoph Schöller and Alois Knoll

[What the Aravantinos Velocity Model Can Teach Us About Pedestrian Motion Prediction](#), Christoph Schöller, Vincent Aravantinos, Florian Lay, Alois Knoll

[YOLOv4: Optimal Speed and Accuracy of Object Detection](#), Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao



(1.) 3D Objektdetektion in LiDAR Daten (blau: PKW, orange: LKWs, grün: Fahrbahn, pink: Schilderbrücke)

(2.) proAnno-Labeling-Tool, mit dem der Providentia++ 3D Datensatz PROVID-21 erstellt wird.

(3.) 3D BAT: Bounding-Box-Annotation-Toolbox für das Labeling von LiDAR-Punktwolken in 3D.

Quelle: <https://ieeexplore.ieee.org/abstract/document/8814071>