



**QUALITY  
MATCH**

# **VERIFICATION & IMPROVEMENT OF DATASET QUALITY FOR COMPUTER VISION & MACHINE LEARNING APPLICATIONS**

**➤ A PROJECT BETWEEN BOSCH AND QUALITY MATCH**

“GARBAGE IN - GARBAGE OUT” is a saying in Data Science. The better the quality of data sets, the better the accuracy of machine learning models - ultimately leading to better products. As enterprises across all verticals, including automotive, rush to apply artificial intelligence (AI) to their business practices, they stumble upon one major obstacle: efficient data labeling at scale.

EDITOR: TIMO LINNEMANN

A team of data scientists, data engineers, and AI experts in the field of Driver Assistance Systems from Bosch have partnered up with Quality Match – a startup from Heidelberg optimizing data sets for machine learning applications. The teams aimed to improve label accuracy in semantic segmentation – an **annotation method** that prepares data for model training, in this case for autonomous driving. The goal was to find labeling errors and fixing these errors in order to create models with higher accuracy.

Bosch owns large annotated data sets that are used for training **Deep Neural Networks**, mainly in the field of Video-based perception software for Driver Assistance Systems and Autonomous Driving Technology that is continuously expanded. In order to develop the data sets as fast as possible, Bosch works together with a number of third party annotation service providers.

The challenge for Bosch was to improve one of their autonomous driving training sets by identifying and removing faulty annotated images, in the most efficient way possible: A method for quality checking at scale – this is where they partnered up with Quality Match.

The company's USP lies in how it breaks down complex quality control tasks into a series of trivial checks that can successfully be performed even by an untrained **clickworker / annotator**. With this approach, the startup not only finds hidden labeling errors, but unveils ambiguities in label taxonomy – **edge-cases** that cannot be explicitly determined by the annotator, and therefore by your ML models. This helps to create better annotation schemes.

## ↳ THE PROJECT

As the quality of data sets decisively influences the quality of the trained deep learning models, Bosch has



### ANNOTATION

The process of making an object recognizable to machines by linking each pixel in an image to an object class, e.g. a person or a car.



### DEEP NEURAL NETWORKS

A machine learning method that has recently become a standard tool for solving a variety of computer vision problems.



### CLICKWORKER

Originated in a NASA project, this term refers to freelancers doing independently small digital tasks – in this case annotating data.

### EXAMPLE EDGE-CASE

“Which of the following is a vehicle? A wheelchair, an airplane, a toy car, ... ?”

a bespoke interest in high annotation quality. Besides outsourcing quality control to third party providers, Bosch also does in-house quality checks. The challenge here lies in the ever-increasing speed of incoming annotated data that needs to be quality-checked. Quality Match was approached to help identify the most crucial and problematic object class in this case – the road.

First of all, the larger problem had to be broken down into smaller, highly intuitive annotation tasks, which any crowd workers could quickly solve in under 10 seconds. A decision tree based on the detection and classification of error sources was built. In this case: road **polygons**. The individual steps were then repeated many times by independent annotators in order to gain broader insights before an optimization process was started.

The team was able to optimize a 16-step decision tree to a 9-step one. Notably, here – the annotators tend not to differentiate between false negative (FN errors) and false positive (FP errors) but they were sensitive to rather any error type.

## ➤ FUTURE OUTLOOK

The POC is yet to be completed but on track to automate the detection and removal of images containing erroneous road polygons. The goal is to improve the model performance while reducing costs and time by 50% – instead of reannotating the dataset from scratch.

The next step will involve adjusting the defective image rather than removing it from the training set, before rolling out the workflow to other object classes like cars and pedestrians.



### POLYGONS

A precise way to annotate objects is by drawing a polygon around a certain object (instead of a box).

This project was made possible by Startup Autobahn:

STARTUP AUTOBAHN powered by Plug and Play is a neutral innovation platform moderating an in-depth and curated collaboration between core partners from industry and young tech companies. The program is stage agnostic and designed to accelerate startups by connecting them to the right business units of our Corporate Partners in order to lay the groundwork of a successful cooperation and eventually outstanding pilot projects, implementations or investments.

[www.startup-autobahn.com](http://www.startup-autobahn.com)



**WE CAN CHECK ANY SEMANTIC SEGMENTATION FOR ACCURACY USING THIS METHOD. IF YOU'D LIKE TO DISCOVER EDGE CASES AND ERRORS IN YOUR DATASET AS WELL, CONTACT US TO GET A DEMO.**

Quality Match GmbH  
Langer Anger 7-9  
69115 Heidelberg  
Germany

Daniel Kondermann  
[dk@quality-match.com](mailto:dk@quality-match.com)

