intech
an Infosys company

AI on the Edge in Industry

Efficient and Secure Processes Thanks to AI-Supported Data Processing On-Site

www.in-tech.com

Utilizing artificial intelligence directly on local devices ("Edge Devices") offers numerous advantages for industrial applications. By processing data on-site instead of in the cloud, latency and data privacy risks are eliminated. in-tech uses this technology to make industrial processes more efficient while simultaneously increasing data security. The reduced data traffic also facilitates integration into existing infrastructures and optimizes process control.

High-Precision Traffic Sign Recognition Thanks to AI on the Edge

Modern neural networks have long surpassed classical image recognition methods. Today, powerful Edge AI systems enable high-precision image recognition even without cloud connectivity.

Our showcase demonstrates how an autonomous ADAS system recognizes traffic signs and adheres to the visible speed limits. Three key technologies are employed:

1 An **Edge AI system** with a connected camera for traffic sign recognition and transmission of the recognized speed to an MBSE model.

2 A mini-computer with the **MBSE model** to ensure the correct response to the recognized speed limit.

3 A gaming/simulation computer that uses the **CARLA simulation environment** to depict the traffic situation and adherence to speed through a cruise control system.

Industry-Grade Hardware

To be suitable for industrial use and deployable in production systems, the showcase hardware meets the following criteria:

- Form Factor: Compact design for integration into machines and systems.
- Computing Power: Sufficient memory and GPU capacity for multimodal models and real-time applications.
- Software SDK: Compatibility with AI stacks like NVIDIA CUDA.

A NVIDIA Jetson Orin NX is used, providing the necessary performance features and enabling real-time AI model processing.

Powerful System and Software Architecture

The architecture is based on CUDA, a parallel computing platform from NVIDIA. This allows graphics processors to be used for computationally intensive tasks like AI models. OpenCV is used for image processing to break down video streams into individual frames and pass them to the AI. AI recognition is performed with specialized Python libraries like Torch, enabling both machine learning and the application of pre-trained neural networks.

This architecture allows for **rapid prototyping of AI applications**. Customer requirements can be implemented and demonstrated in a Minimal Viable Product (MVP). Subsequently, the MVP application can be ported to more cost-effective hardware for series deployment. This is achieved through model optimization techniques such as quantification, distilling, pruning, or OBD to make the AI model more efficient.

Specifically Adaptable AI Model for Industrial Application

The model used in the showcase is based on YOLOv5, a pre-trained model for object recognition. Transfer learning was used to adapt the model to traffic signs for the specific task. This enables efficient fine-tuning with minimal data effort and shortened training times.

An important platform for pre-trained models is Hugging Face, which offers a wide selection of models for text and image processing. Through transfer learning, these models can be adapted for specific industrial use cases.

Future Perspectives: Multimodal AI and Semantic Image Processing

Classical AI-based image recognition uses fixed trained categories and is already established in many application areas. The recent development of multimodal Large Language Models (LLMs) now allows the combination of image recognition with linguistic description. This enables semantic image understanding and complex applications that combine visual and linguistic information.

A prominent example is Gemma 3 from Meta, characterized by on-edge capabilities and flexible model sizes. Through Retrieval-Augmented Generation (RAG), Gemma 3 can be expanded with domain-specific knowledge to interpret precise technical terms and process flows in specialized industrial areas.

Potentials for Industry: Efficient Processes with Edge AI

Advancements in Edge AI open up new possibilities for machinery and plant engineering. In addition to person recognition, semantic image recognition now allows material-dependent configuration tasks and quality controls at a new level. With Edge AI, industrial processes can be made smarter, safer, and more efficient, while minimizing data privacy and latency issues.

in-tech supports companies in profitably integrating these technologies into their products and solutions.

Contact

If you have any questions, please do not hesitate to contact us. Please contact us via **smart-industry@in-tech.com**

in-tech contact:

Klaus Wiltschi
Director Smart Industry Germany
E-Mail: klaus.wiltschi@in-tech.com