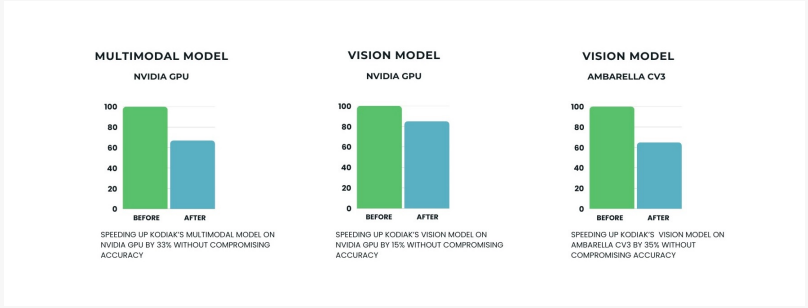
Embedl Turbocharges Kodiak's GenAI Edge Deployment

Autonomous trucking leader Kodiak Robotics uses Embedl's edge AI tools to effortlessly deploy generative models across hardware platforms.

Embedl Turbocharges Kodiak's GenAI Edge Deployment



(Figure 1) The subsequent optimization with the Embedl SDK resulted in dramatically faster inference speeds.

CA, UNITED STATES, January 8, 2025 /EINPresswire.com/ --

A HARD PROBLEM

Autonomous vehicles (AVs) have captured the imagination of the public for years, but it is only very recently that we have seen fully autonomous vehicles on the roads. Most of them are robotaxis in restricted areas in certain cities and with a remote backup assistant able to intervene if necessary. Heavy truck autonomy is also emerging now, and you can argue that those solve a more real and important problem - lack of drivers and faster delivery of goods. Kodiak Robotics stands out as the only company currently operating fully autonomous trucks. Their vision is to automate dangerous, dirty and dull truck driving.

EDGE AI MODELS RUN THE SHOW

At the heart of autonomous driving lies artificial intelligence, in the form of neural network models running on computers on board the vehicle (edge AI), enabling real-time decision-making. Kodiak Robotics exemplifies this by deploying advanced neural networks that process vast amounts of data directly on the vehicle.

Autonomous trucks must be capable of making split-second decisions while simultaneously

planning ahead, anticipating conditions and obstacles hundreds of meters beyond their current position. Kodiak's system integrates multimodal sensor fusion models, combining camera and lidar point cloud data from multiple sources to create a detailed understanding of the vehicle's surroundings.

Kodiak is also unique in its ability to use Generative AI to handle complex driving scenarios. GenAI allows Kodiak to generalize its autonomy system to drive in novel environments, such as terrain with a pile of thick snow in a dense forest. Additionally, Kodiak uses vision-language models (VLMs) to interpret out-of-distribution situations that it has never seen before. The result is a fleet of trucks capable of navigating highways, offroad, and gravel roads with unmatched precision, safety, and reliability.

FROM GPUS TO EMBEDDED HARDWARE ACCELERATORS

Developing and testing an AI system is just the beginning. It's one thing for the neural networks to work in a server machine with a large GPU and massive compute power. But those same neural nets must be significantly altered and improved to perform well on an AV's on-vehicle hardware, with embedded hardware accelerators. This is where Embedl comes in, with our state-of-the-art and battle-proven software product that automates this process and optimizes the performance on the target edge device.

HARDWARE ABSTRACTION

Embedl's software provides an innovative hardware abstraction technology that enables AI developers to easily manage target inference hardware during the design, training and optimization of the neural network models. The target inference hardware is the computer where the real-time decisions are made by the edge AI models, e.g. in the vehicle. The automatic model optimization is hardware aware and makes use of the very specific characteristics of the selected hardware. While this technical solution allows the optimization algorithms to fully exploit the optimal solution for that target hardware, it also allows companies like Kodiak to support multiple target hardware platforms in parallel. This greatly simplifies development for companies like Kodiak, which today uses both NVIDIA and Ambarella hardware to run their software stack.

"Having the ability to deeply inspect the cognitive blocks of our AI models, perform hardware-aware optimization, benchmark various layers, and deploy models through seamless hardware abstraction is truly game-changing." said Shubham Shrivastava, Head of Machine Learning at Kodiak.

FACTS AND FIGURES

Using the Embedl SDK, engineers at Kodiak successfully ported vision and multimodal generative AI (GenAI) models from NVIDIA to Ambarella platforms within weeks. The subsequent optimization with the Embedl SDK resulted in dramatically faster inference speeds on both hardware platforms.
(Figure 1)

CHECK IT OUT
If you are interested in improving your deep learning edge AI development, please contact ola@embedl.com.

Ola Tiverman
Embedl
+46 73 352 82 09
email us here
Visit us on social media:
LinkedIn
YouTube
Facebook
Instagram

This press release can be viewed online at: https://www.einpresswire.com/article/774806326

EIN Presswire's priority is source transparency. We do not allow opaque clients, and our editors try to be careful about weeding out false and misleading content. As a user, if you see something we have missed, please do bring it to our attention. Your help is welcome. EIN Presswire, Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.
© 1995-2025 Newsmatics Inc. All Right Reserved.