deci.

Develop Superior ADAS Applications with Accelerated AI Inference Performance

Introduction

Advanced Driver Assistance Systems (ADAS) technology powered by deep learning models is transforming the mobility and transportation industry. However, AI developers are still facing an uphill battle when trying to transition from the lab to real-world deployments. A common barrier to deployment is the inability to achieve high accuracy and real-time inference performance on edge devices. Both factors are absolutely mission-critical for ensuring not only the application's usability but also safety for the users.

Leading automotive manufacturers and ADAS providers use Deci to boost their models' performance and deploy multiple models on edge devices. Below are three case studies demonstrating how such companies achieved the desired performance and delivered better products to their customers using Deci.

Case Study 1
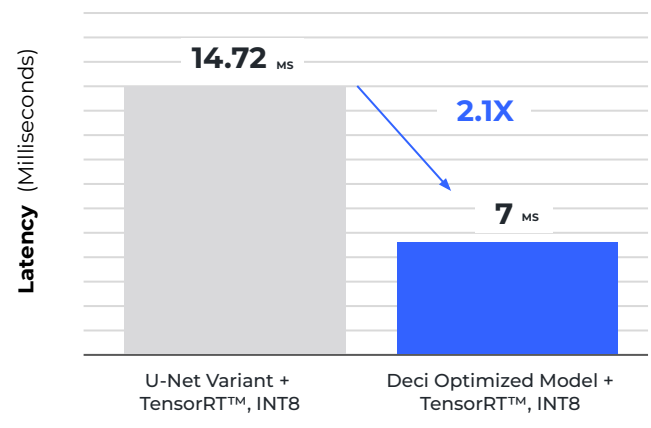### Enabling Real-Time Semantic Segmentation for an Automotive Application

2.1x
Latency
Acceleration

3x
Model Size
Reduction

3x
Memory Footprint
Reduction

An automotive company running a U-Net based segmentation model on a NVIDIA Jetson Xavier NX struggled to achieve the target latency in production.

Using Deci's AutoNAC engine, a faster and smaller model was generated. Latency was reduced by 2.1x, model size was reduced by 3x, and memory footprint was reduced by 3x – all while maintaining the original accuracy.

2.1X Lower Latency (Same Accuracy)

Measured on NVIDIA Jetson Xavier NX

Latency (Milliseconds)

14.72 MS

2.1X

7 MS

U-Net Variant +
TensorRT™, INT8

Deci Optimized Model +
TensorRT™, INT8

Case Study 2
Real-time Depth Estimation for an ADAS solution

2.5
Latency
Acceleration

50%
Reduction in Dev
Cost

A large company delivering solutions for ADAS/AD needed to improve the latency performance of a semantic segmentation model used for real-time depth estimation for an automotive use case. The original model was based on MobileNetV2.

The company's team used the Deci AutoNAC engine to build a custom model architecture to meet their inference performance targets while maintaining their original accuracy on the NVIDIA RTX 2080Ti hardware. As a result, the company can now deliver real-time inference performance for its automotive application.

2.5X Lower Latency (Same Accuracy)

Measured on NVIDIA RTX 2080Ti

Latency (Milliseconds)

9.2 MS

2.5x

3.7 MS

Case Study 3
Enabling Real-Time Classification at the Edge

+4%
Accuracy
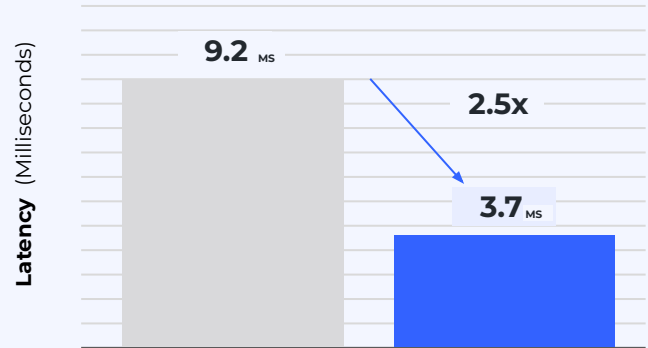Increase

2x
Latency
Acceleration

3.2x
Model Size
Reduction

An automotive company wanted to improve the latency of a model powering their road condition estimation system. The poor latency of their baseline image classification model was impacting real-time performance on their target hardware, a NVIDIA Xavier AGX.

The team used Deci's AutoNAC engine to build a customized model, resulting in a 2.1x acceleration in latency. Accuracy was improved by +4%, throughput was improved by 2x, and the model size was reduced by 3.2x.
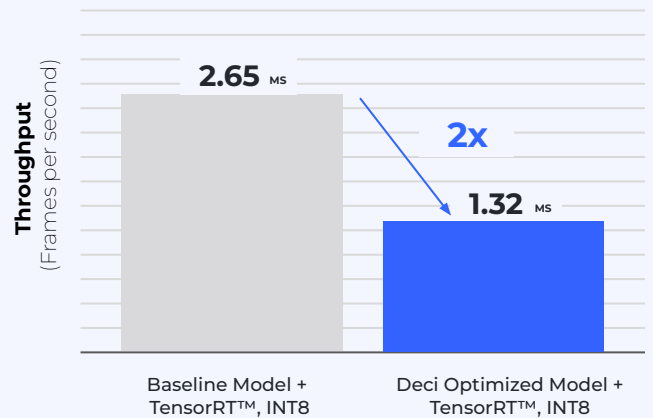
2X Lower Latency (Better Accuracy)

Measured on NVIDIA Jetson Xavier AGX

Throughput
(Frames per second)

2.65 MS

2x

1.32 MS

Baseline Model +
TensorRT™, INT8

Deci Optimized Model +
TensorRT™, INT8

Build Better Models Faster with Deci's Deep Learning Development Platform

The Deci platform is used by data scientists and machine learning engineers to build, optimize, and deploy highly accurate and efficient models. Teams can easily develop production-grade models and gain unparalleled accuracy and speed tailored for any performance targets and hardware environment.

Deci is powered by AutoNAC (Automated Neural Architecture Construction), the most advanced and commercially scalable Neural Architecture Search engine in the market. Leading enterprises use Deci to boost their deep learning models' inference performance, shorten development cycles, enable new use cases on edge devices, and reduce computing costs.

Deci's platform recently won the Best Edge AI Developer Tool in 2023's Edge AI and Vision Product of the Year Awards.
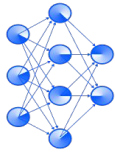

Trusted by Industry Leaders
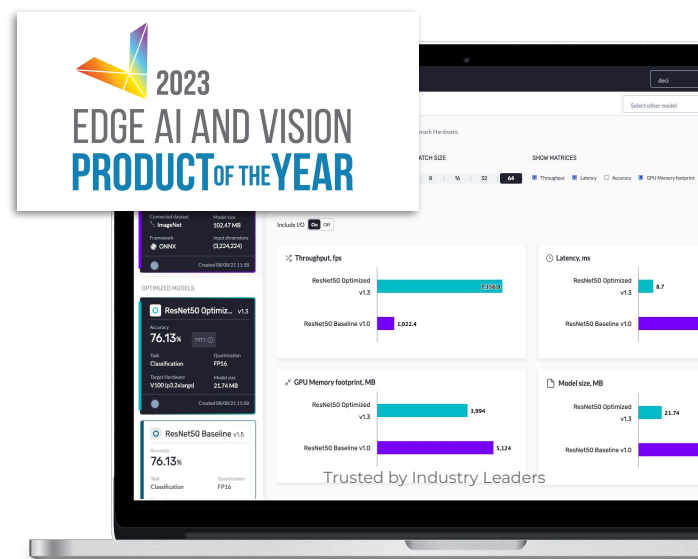
Deci's Deep Learning Platform
Powered by Neural Architecture Search

Deci Platform

Foundation or Custom Models
Choose an ultra-performant pre-trained model or generate a custom one.

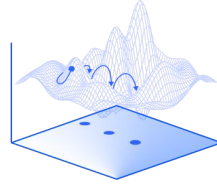`SaaS` **AutoNAC**
Neural Architecture Search Engine

`On Prem` **DataGradients™**
Dataset analyzer

Train or Fine-tune
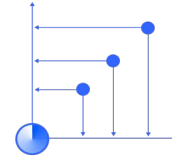Use Deci's library & custom recipe to train on-prem.

`On Prem` **SuperGradients™**
PyTorch Training library

Optimize & Run
Apply acceleration techniques. Run self-hosted inference anywhere.

`On Prem` **Infery**
Optimization & Inference Engine SDK

Training Frameworks

Runtime Frameworks   OpenVINO   ONNX RUNTIME

Main Capabilities Overview

Gain Superior Performance with Custom Architectures
Build accurate & efficient architectures tailored to your hardware and application's performance targets with Deci's proprietary Neural Architecture Search engine.

Maximize Accuracy with Advanced Training Techniques
Train models with SuperGradients. Leverage custom recipes and advanced training techniques (e.g., knowledge distillation, quantization-aware training) with one line of code.

Simplify Runtime Optimization
Easily compile and quantize your models (FP16/INT8) and evaluate different production settings with a click of a button.

Streamline Deployment with 3 Lines of Code
Deploy your models with Infery, Deci's simple-to-use, unified, model inference API. Streamline deployment and boost serving performance with parallelism and concurrent execution. Compatible with multiple frameworks and hardware types.

Easily Find the Best Hardware for the Job
Benchmark your models' inference performance across multiple hardware types with Deci's online hardware fleet. Get actionable insights and select the optimal hardware.

Automate Model Benchmarking
Easily measure and compare the performance of various models on your inference hardware.

For Everyone in the AI-Driven Organization

Tech Executives
Reduce time to market by 80% and lower development cost by 30%. Supercharge your AI teams with advanced tools.

Data Scientists
Deliver state-of-the-art models, faster than ever, without worrying about performance or model size. Focus on your core competency: solving business problems with AI.

ML Engineers
Easily optimize models for various hardware types with a few clicks. Seamlessly deploy and maximize application performance with advanced serving capabilities.

Product Leaders
Unlock new use cases and release new features to production faster without compromising on quality.

About Deci
Deci enables deep learning to live up to its true potential by using AI to build better AI. With the company's deep learning development platform, AI developers can build, optimize, and deploy faster and more accurate models for any environment including cloud, edge, and mobile, allowing them to revolutionize industries with innovative products. Founded in 2019, Deci's team of deep learning engineers and scientists are dedicated to eliminating production-related bottlenecks across the AI lifecycle.

Book a Demo

For more information, visit **deci.ai**

deci.