Deliver Superior AI-Powered Defense Solutions Faster with Deci's Platform

Introduction

Deep learning is applied in various defense applications that leverage its capabilities in analyzing complex data and making intelligent decisions. Autonomous systems, including unmanned vehicles, drones, and robots, are just a few examples of the many defense applications of deep learning. Tasks such as object detection, object identification, and classification help generate actionable insights through streaming video and image analytics.

Implementing deep learning on edge devices holds immense potential for revolutionizing defense applications by enabling real-time intelligence, enhanced situational awareness, and decentralized decision-making capabilities. However, there are challenges to overcome in implementing deep learning on edge devices. Limited computational resources and power constraints pose challenges for running complex deep learning models in real-time on resource-constrained devices.

With Deci, you can boost your models' performance and maximize hardware utilization to deliver accurate and cost-efficient inference on cloud or edge devices. Below are several case studies of leading defense companies that, using Deci's platform, significantly improved their models' performance, resulting in new defense applications that can run reliably and cost-efficiently on various edge devices.

Case Study 1

Achieve Real Time Inference at the Edge & Maximize HW Utilization

1.58x Throughput Increase

25% Dev Cost Reduction

50% Savings on HW Expenditure

Overview

- Defense
- Image Denoising
- NVIDIA Jetson Orin AGX 32GB

The Challenge

A defense company developing electro-optics solutions for space, airborne, ground, and maritime applications was looking to improve the throughput of an image-denoising model for video stream analysis in order to deliver real-time insights at the edge.

The team was also looking for a way to free up GPU resources in order to support additional parallel tasks on the same edge device.

Before using Deci, the team spent over 6 months experimenting with various optimization techniques but failed to reach the desired results. They even considered replacing their inference hardware from the NVIDIA Jetson Orin AGX 32GB to a more expensive edge device in order to reach their target performance and enable parallel processing.

The Solution

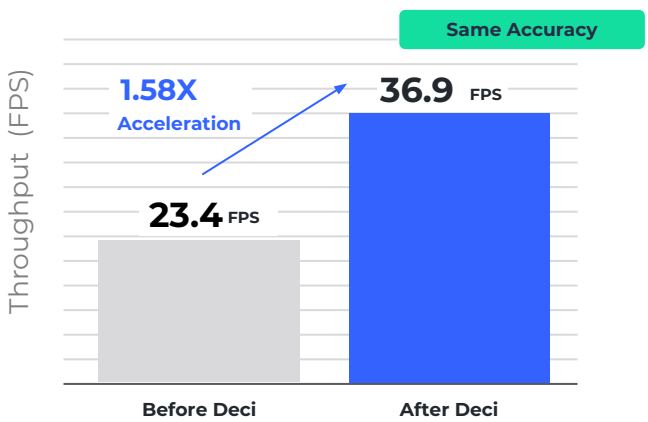
Using the Deci platform and its NAS-based engine, the team built a new architecture that delivered a 1.58x acceleration of throughput while maintaining the original model's accuracy. The team trained the new architecture on-premise using Deci's SuperGradients library and achieved the desired performance within 10 days. Once the team trained the model, they easily compiled and quantized it to TensorRT FP16 using Deci's platform.

The new efficient model enabled the team to not only achieve the desired performance but also utilize the same device for processing an additional task leading to significant savings on hardware expenditure.

Moreover, the team shortened their time to market by eliminating months of additional trial and error experimentations.

1.58X Higher Throughput

Measured on NVIDIA Jetson Orin AGX 32GB



Computer Vision & Deep Learning Team Lead at a large defense company

"Deci's AutoNAC engine quickly produced a new model architecture that surpassed our optimized model - a model we'd been working on for 18 months prior. The development time and cost saved is remarkable and provides a significant advantage for our business. Moving forward, we intend to utilize Deci for every new model we develop."

Case Study 2

Enabling a New Detection Solution to Run on Highly Constrained Compute Resources

3.08x
Throughput Increase

40%
Model Size Reduction

Overview

- Defense
- Object Detection
- NVIDIA Jetson Xavier NX - 10 Watts power mode

The Challenge

A defense company specializing in advanced land systems needed to process high-resolution images for object detection and tracking tasks running on an NVIDIA Jetson Xavier NX device. While their original model reached good accuracy, the team struggled to achieve the desired inference speed given the hardware's constrained compute resources. For the system to become operational, the team needed to run in a 10-watt power mode and achieve a throughput of 10 frames per second.

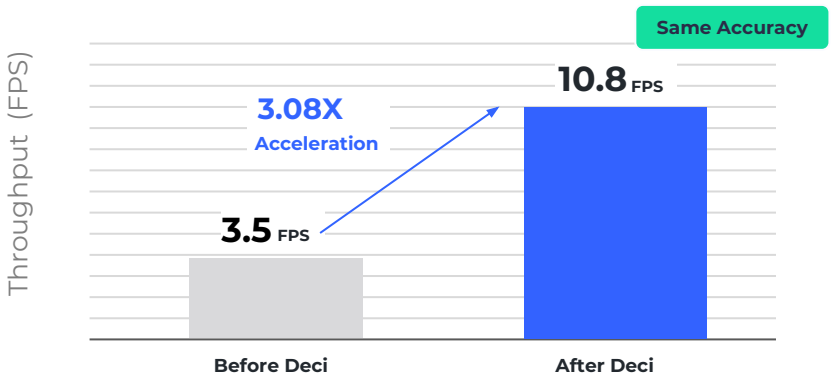
The Solution

Using Deci's platform and its NAS-based engine, the team generated a new model architecture and trained it on-premise using Deci's open-source training library called SuperGradients.

The new models delivered a 3.08x increase in the throughput from 3.5 FPS to 10.8 FPS, and the model size was reduced by 40% from 40.04MB to 24.14 MB.
As a result, the team obtained smooth object detection and tracking, allowing the company to deliver a new and superior product to the market.

3.08X Higher Throughput

Measured on NVIDIA Jetson Xavier NX 10 Watts



Case Study 3

Ship Superior Product to Market within a Few Weeks

6.5x
Throughput Increase

14 Days
From data to production ready model

Overview

- Security
- Object Detection
- A wide range of Intel CPUs

The Challenge

A company developing an AI-powered video analytics technology for security applications was looking to improve one of its existing products by providing real-time alerts to its users. While the solution was already in production, running on a wide range of Intel CPUs, the team struggled to improve the model's speed without compromising accuracy.

Before using Deci, the team experimented with various inference optimization techniques, including pruning and INT8 quantization was unable to achieve an optimal trade-off between accuracy and throughput.

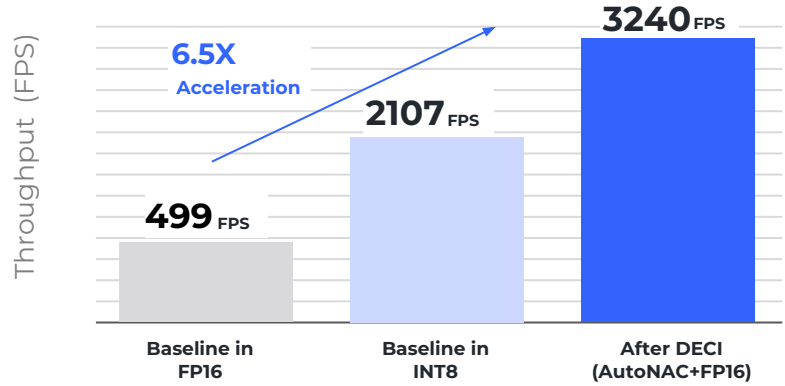
The Solution

Using the Deci platform, the company's team generated a new model architecture that delivered a 6.5x acceleration of throughput over the original SOTA architecture while maintaining the original model's accuracy. The team trained, compiled, and quantized the new model on-premise, all within 14 days.

With this improved performance, the company was able to introduce a superior product and significantly increase its market share.

6.5X Higher Throughput

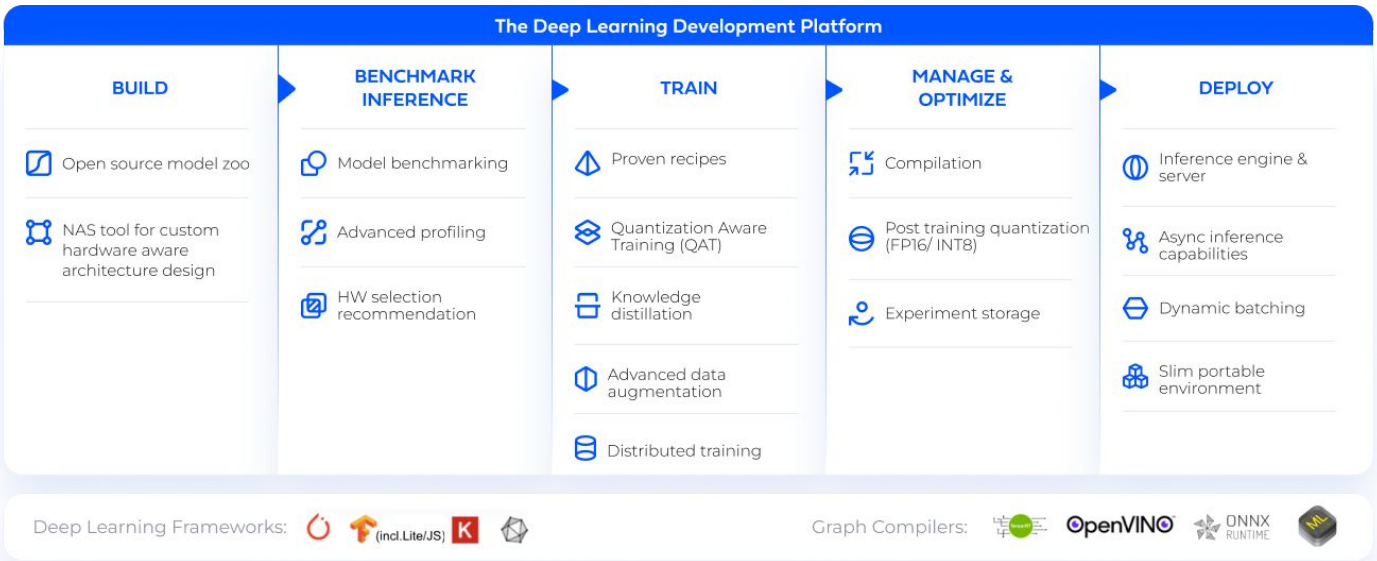
Measured on Intel CPU



deci

Deci's Deep Learning Platform

Powered by Neural Architecture Search



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.

80%
Shorter development process on average

30%
Lower development costs on average

5x
Inference Acceleration on average

5x
Inference & training cost reduction on average

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

The Deci platform is used by data scientists and machine learning engineers to build, optimize, and deploy highly accurate and efficient models to production. 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 market. Leading enterprises are using Deci to boost their deep learning models' inference performance, shorten development cycles, enable new use cases on edge devices, and reduce computing costs.

BOOK A DEMO

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

deci