

# **NVIDIA Announces Microsoft, Tencent, Baidu Adopting CV-CUDA for Computer Vision AI**

Forthcoming open beta optimizes pre- and post-processing for higher throughput at quarter the cost and energy.

Author: Michael Boone

Microsoft, Tencent and Baidu are adopting CV-CUDA for computer vision AI.

NVIDIA CEO Jensen Huang highlighted work in content understanding, visual search and deep learning Tuesday as he announced the beta release for NVIDIA's CV-CUDA — an open-source, GPU-accelerated library for computer vision at cloud scale.

"Eighty percent of internet traffic is video, user-generated video content is driving significant growth and consuming massive amounts of power," said Huang in his keynote at NVIDIA's GTC technology conference . "We should accelerate all video processing and reclaim the power."

CV-CUDA promises to help companies across the world build and scale end-to-end, AI-based computer vision and image processing pipelines on GPUs.

The majority of internet traffic is video and image data, driving incredible scale in applications such as content creation, visual search and recommendation, and mapping.

These applications use a specialized, recurring set of computer vision and image-processing algorithms to process image and video data before and after they're processed by neural networks.

While neural networks are normally GPU accelerated, the computer vision and image processing algorithms that support them are often CPU bottlenecks in today's AI applications.

CV-CUDA helps process 4x as many streams on a single GPU by transitioning the pre- and post-processing steps from CPU to GPU. In effect, it processes the same workloads at a quarter of the cloud-computing cost.

The CV-CUDA library provides developers more than 30 high-performance computer vision algorithms with native Python APIs and zero-copy integration with the PyTorch , TensorFlow2 , ONNX and TensorRT machine learning frameworks.

The result is higher throughput, reduced computing cost and a smaller carbon footprint for cloud AI businesses.

Adoption by industry leaders around the globe highlights the benefits and versatility of CV-CUDA for a growing number of large-scale visual applications. Companies with massive image processing workloads can save tens to hundreds of millions of dollars.

Microsoft is working to integrate CV-CUDA into Bing Visual Search , which lets users search the web using an image instead of text to find similar images, products and web pages.

In 2019, Microsoft shared at GTC how they're using NVIDIA technologies to help bring speech recognition, intelligent answers, text to speech technology and object detection together seamlessly and in real time.

Tencent has deployed CV-CUDA to accelerate its ad creation and content understanding pipelines, which process more than 300,000 videos per day.

The Shenzhen-based multimedia conglomerate has achieved a 20% reduction in energy and cost for image processing over their previous GPU-optimized pipelines.

And Beijing-based search giant Baidu is integrating CV-CUDA into FastDeploy , one of the open-source deployment toolkits of the PaddlePaddle Deep Learning Framework , which enables seamless computer vision acceleration to developers in the open-source community.

Applications for CV-CUDA are growing. More than 500 companies have reached out with over 100 use cases in just the first few months of the alpha release.

In content creation and e-commerce, images use pre- and post-processing operators to help recommender engines recognize, locate and curate content.

In mapping, video ingested from mapping survey vehicles requires preprocessing and post-processing operators to train neural networks in the cloud to identify infrastructure and road features.

In infrastructure applications for self-driving simulation and validation software , CV-CUDA enables GPU acceleration for algorithms that are already occurring in the vehicle, such as color conversion, distortion correction, convolution and bilateral filtering.

Looking to the future, generative AI is transforming the world of video content creation and curation, allowing creators to reach a global audience.

New York-based startup Runway has integrated CV-CUDA, alleviating a critical bottleneck in preprocessing high-resolution videos in their video object segmentation model.

Implementing CV-CUDA led to a 3.6x speedup, enabling Runway to optimize real-time, click-to-content responses across its suite of creation tools.

“For creators, every second it takes to bring an idea to life counts,” said Cristóbal Valenzuela, co-founder and CEO of Runway. “The difference CV-CUDA makes is incredibly meaningful for the millions of creators using our tools.”

To access CV-CUDA, visit the CV-CUDA GitHub .

Or learn more by checking out the GTC sessions featuring CV-CUDA. Registration is free .

Overcoming Pre- and Post-Processing Bottlenecks in AI-Based Imaging and Computer Vision Pipelines [S51182] ,

Building AI-Based HD Maps for Autonomous Vehicles [SE50001] ,

Connect With the Experts: GPU-Accelerated Data Processing with NVIDIA Libraries [CWES52014] ,

Advancing AI Applications with Custom GPU-Powered Plugins for NVIDIA DeepStream [S51612] .

Original URL: <https://blogs.nvidia.com/blog/2023/03/21/cv-cuda-ai-computer-vision/>