

NVIDIA Doubles Performance for Deep Learning Training

New Releases of DIGITS, cuDNN to Deliver 2x Faster Neural Network Training; cuDNN to Enable More Sophisticated Models

NVIDIA today announced at the International Machine Learning Conference updates to its GPU-accelerated deep learning software that will double deep learning training performance.

The new software will empower data scientists and researchers to supercharge their [deep learning](#) projects and product development work by creating more accurate neural networks through faster model training and more sophisticated model design.

The NVIDIA® [DIGITS™ Deep Learning GPU Training System](#) version 2 (DIGITS 2) and NVIDIA CUDA® [Deep Neural Network library](#) version 3 (cuDNN 3) provide significant performance enhancements and new capabilities.

For data scientists, DIGITS 2 now delivers automatic scaling of neural network training across multiple high-performance GPUs. This can double the speed of deep neural network training for image classification compared to a single GPU.

For deep learning researchers, cuDNN 3 features optimized data storage in GPU memory for the training of larger, more sophisticated neural networks. cuDNN 3 also provides higher performance than cuDNN 2, enabling researchers to train neural networks up to two times faster on a single GPU.

The new cuDNN 3 library is expected to be integrated into forthcoming versions of the deep learning frameworks Caffe, Minerva, Theano and Torch, which are widely used to train deep neural networks.

"High-performance GPUs are the foundational technology powering deep learning research and product development at universities and major web-service companies," said Ian Buck, vice president of Accelerated Computing at NVIDIA. "We're working closely with data scientists, framework developers and the deep learning community to apply the most powerful GPU technologies and push the bounds of what's possible."

DIGITS 2 - Up to 2x Faster Training with Automatic Multi-GPU Scaling

DIGITS 2 is the first all-in-one graphical system that guides users through the process of designing, training and validating deep neural networks for image classification.

The new automatic multi-GPU scaling capability in DIGITS 2 maximizes the available GPU resources by automatically distributing the deep learning training workload across all of the GPUs in the system. Using DIGITS 2, NVIDIA engineers trained the well-known AlexNet neural network model more than two times faster on four [NVIDIA Maxwell™ architecture](#)-based GPUs, compared to a single GPU.¹ Initial results from early customers are demonstrating better results.

"Training one of our deep nets for auto-tagging on a single NVIDIA GeForce GTX TITAN X takes about sixteen days, but using the new automatic multi-GPU scaling on four TITAN X GPUs the training completes in just five days," said Simon Osindero, A.I. architect at Yahoo's Flickr. "This is a major advantage and allows us to see results faster, as well letting us more extensively explore the space of models to achieve higher accuracy."

cuDNN 3 - Train Larger, More Sophisticated Models Faster

cuDNN is a GPU-accelerated library of mathematical routines for deep neural networks that developers integrate into higher-level machine learning frameworks.

cuDNN 3 adds support for 16-bit floating point data storage in GPU memory, doubling the amount of data that can be stored and optimizing memory bandwidth. With this capability, cuDNN 3 enables researchers to train larger and more sophisticated neural networks.

"We believe FP16 GPU storage support in NVIDIA's libraries will enable us to scale our models even further, since it will increase effective memory capacity of our hardware and improve efficiency as we scale training of a single model to many GPUs," said Bryan Catanzaro, senior researcher at Baidu Research. "This will lead to further improvements in the accuracy of our models."

cuDNN 3 also delivers significant performance speedups compared to cuDNN 2 for training neural networks on a single GPU. It enabled NVIDIA engineers to train the AlexNet model two times faster on a single NVIDIA GeForce® GTX™ TITAN X GPU.²

Availability

The DIGITS 2 Preview release is available today as a free download for NVIDIA registered developers. To learn more or download, visit the [DIGITS website](#).

The cuDNN 3 library is expected to be available in major deep learning frameworks in the coming months. To learn more visit the [cuDNN website](#).

To Keep Current on NVIDIA:

- Keep up with the [NVIDIA Blog](#), and follow us on [Facebook](#), [Google+](#), [Twitter](#), [LinkedIn](#) and [Instagram](#).
- View NVIDIA videos on [YouTube](#) and images on [Flickr](#).
- Use the [Pulse news reader](#) to subscribe to the NVIDIA Daily News feed.

About NVIDIA

Since 1993, [NVIDIA](#) (NASDAQ: [NVDA](#)) has pioneered the art and science of [visual computing](#). The company's technologies are transforming a world of displays into a world of interactive discovery -- for everyone from gamers to scientists, and consumers to enterprise customers. More information at <http://nvidianews.nvidia.com/> and <http://blogs.nvidia.com/>.

(1) DIGITS 2 performance vs. previous version on an NVIDIA DIGITS DevBox system with NVIDIA GeForce GTX TITAN X GPUs.

(2) cuDNN 3 performance vs. previous version on Ubuntu 14.04 LTS with NVIDIA GeForce TITAN X GPU and Intel Core i7-4930K @ 3.40GHz.

Certain statements in this press release including, but not limited to, statements as to: the benefits, impact, performance and availability of the NVIDIA DIGITS Deep Learning GPU Training System version 2 and NVIDIA CUDA Deep Neural Network library version 3 are forward-looking statements that are subject to risks and uncertainties that could cause results to be materially different than expectations. Important factors that could cause actual results to differ materially include: global economic conditions; our reliance on third parties to manufacture, assemble, package and test our products; the impact of technological development and competition; development of new products and technologies or enhancements to our existing product and technologies; market acceptance of our products or our partners' products; design, manufacturing or software defects; changes in consumer preferences or demands; changes in industry standards and interfaces; unexpected loss of performance of our products or technologies when integrated into systems; as well as other factors detailed from time to time in the reports NVIDIA files with the Securities and Exchange Commission, or SEC, including its Form 10-Q for the quarterly period ended April 26, 2015. Copies of reports filed with the SEC are posted on the company's website and are available from NVIDIA without charge. These forward-looking statements are not guarantees of future performance and speak only as of the date hereof, and, except as required by law, NVIDIA disclaims any obligation to update these forward-looking statements to reflect future events or circumstances.

© 2015 NVIDIA Corporation. All rights reserved. NVIDIA, the NVIDIA logo, NVIDIA DIGITS, CUDA, Maxwell, GeForce and GTX are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. Other company and product names may be trademarks of the respective companies with which they are associated. Features, pricing, availability and specifications are subject to change without notice.

About NVIDIA

Since 1993, [NVIDIA](#) (NASDAQ : NVDA) has pioneered the art and science of [visual computing](#). The company's technologies are transforming a world of displays into a world of interactive discovery — for everyone from gamers to scientists, and consumers to enterprise customers. More information at <http://nvidianews.nvidia.com/> and <http://blogs.nvidia.com/>.

© 2014 NVIDIA Corporation. All rights reserved. NVIDIA and the NVIDIA logo are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. Other company and product names may be trademarks of the respective companies with which they are associated. Features, pricing, availability, and specifications are subject to change without notice.

Media Contacts

George Millington
+1 408 562 7226
gmillington@nvidia.com