



エヌビディアが加速するディープラーニング ～進化するニューラルネットワークとその開発方法について～

ディープラーニング ソリューションアーキテクト 兼 CUDAエンジニア

村上 真奈

AGENDA

ディープラーニングの現在

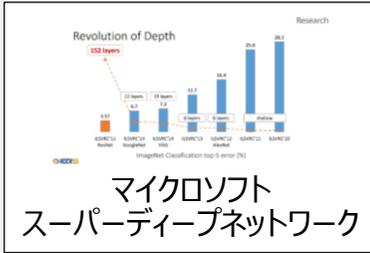
ディープラーニングSDK

NVIDIA DIGITS

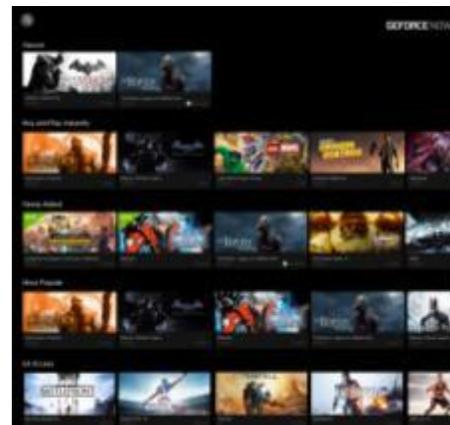
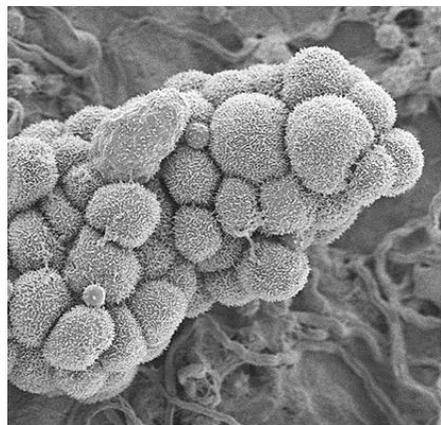
ディープラーニング・フレームワーク / ディープラーニングの基礎

ディープラーニングの現在

人工知能にとって驚くべき一年



様々な分野でディープラーニングを応用



インターネットとクラウド

画像分類
音声認識
言語翻訳
言語処理
感情分析
推薦

医学と生物学

癌細胞の検出
糖尿病のリスク付け
創薬

メディアとエンターテイメント

字幕
ビデオ検索
リアルタイム翻訳

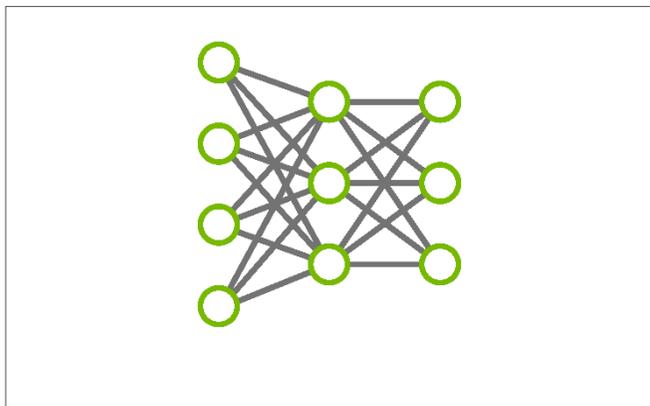
セキュリティと防衛

顔検出
ビデオ監視
衛星画像

機械の自動化

歩行者検出
白線のトラッキング
信号機の認識

ディープラーニングを加速する3つの要因



DNN



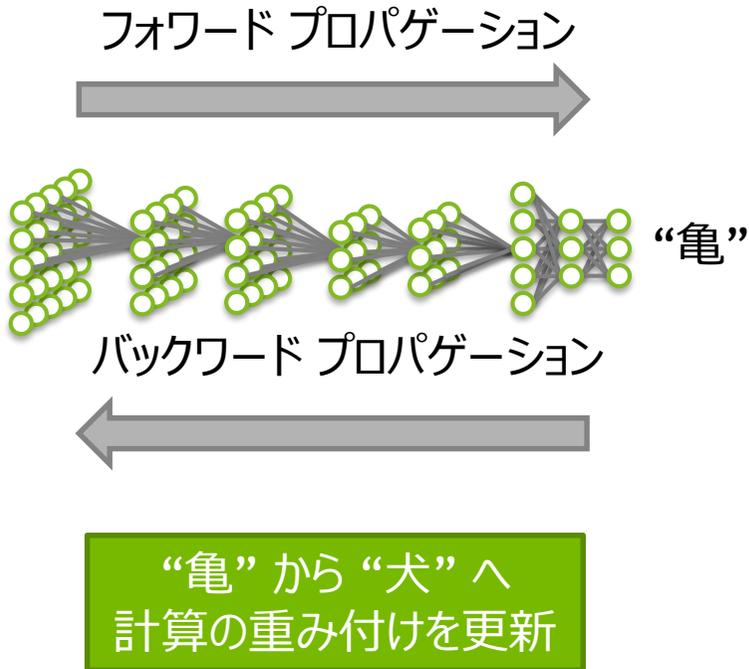
ビッグデータ



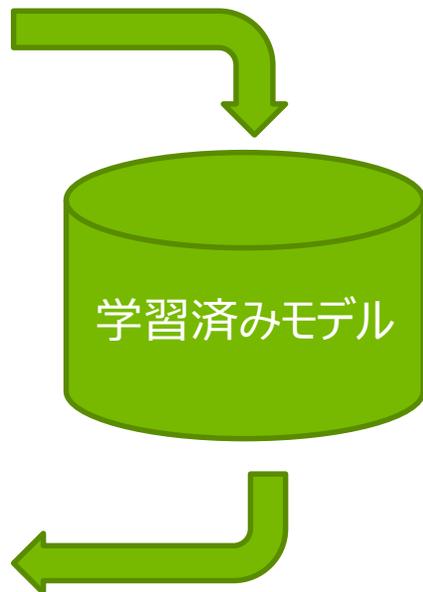
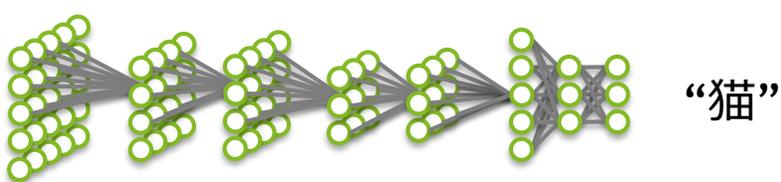
GPU

“The GPU is the workhorse of modern A.I.”

機械学習ソフトウェア



推論
(インファレンス)



ディープラーニングSDK

ディープラーニング SDK

ディープラーニングを加速するディープラーニングライブラリ



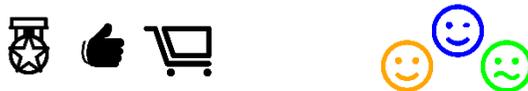
IMAGENET
画像分類
物体検出

コンピュータビジョン



音声認識
言語翻訳

ボイス&オーディオ

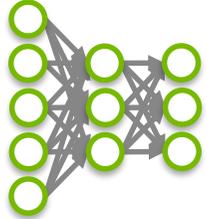


推薦エンジン
感情分析

自然言語処理

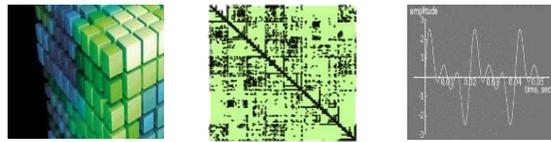


ディープラーニングフレームワーク



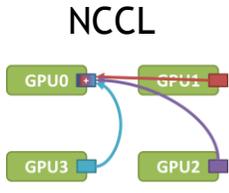
cuDNN

ディープラーニング



cuBLAS
cuSPARSE
cuFFT

CUDA 数学ライブラリ



NCCL

GPU0 GPU1 GPU2 GPU3

マルチGPU間通信



GIE

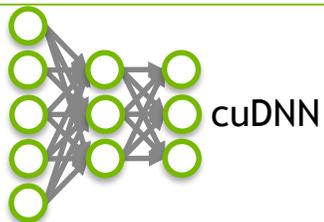
インファレンス

ディープラーニング SDK

ディープラーニングを加速するディープラーニングライブラリ

GPUで高速化されたディープラーニングのアプリケーションを設計、開発する為の強力な開発ツールおよびライブラリ

- ▶ NVIDIA GPUの為の高速なディープラーニング環境を保障
- ▶ 最新のディープラーニング・アルゴリズムをGPUに最適化した状態で提供
- ▶ ディープラーニング・アプリケーションを簡単に構築



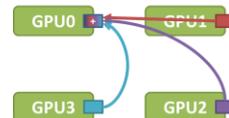
ディープラーニング

cuBLAS cuSPARSE cuFFT



CUDA 数学ライブラリ

NCCL



マルチGPU間通信

GIE



インファレンス

NVIDIA CUDNN

最新はバージョンは**5.1RC**

GPU上でディープニューラルネットワークの計算を高速に行う為のプリミティブ群

- ▶ ディープニューラルネットワークの学習の高いパフォーマンスを発揮
- ▶ Caffe、CNTK、Tensorflow、Theano、Torch、Chainerなどディープラーニングフレームワークを高速化
- ▶ バージョンアップ毎にパフォーマンスが向上



“NVIDIAはcuDNNのリリースの度により多くの操作ができるような機能拡張をおこなっており、同時に計算速度も向上させている”

—UC バークレー、Caffe リードデベロッパー、Evan Shelhamer

developer.nvidia.com/cudnn

NVIDIA CUDNN

ディープラーニング計算の為に高速なプリミティブ群

- 畳込み以外にもプーリング、ソフトマックス、活性化、batch normalization、テンソルのトランスポーズなどディープラーニングの順伝播・逆伝播でよく使われるレイヤータイプ、計算手法をサポート
- リカレント・ニューラルネットワーク(LSTM/GRU/RNN)をサポート。
- マルチスレッド化が容易なコンテキストベースのAPI
- 以下のOSとGPUの組み合わせで仕様が可能
 - Windows(7/10),Linux(x64,power8/ARM),MacOS
 - Pascal/Maxwell/Kepler世代のGPU, Tegra K1およびTegra X1

cuDNN 5 のアップデート

Pascal世代GPU, リカレントニューラルネットワーク, パフォーマンスの向上

Torch上でLSTM リカレントニューラルネットワークの
計算を6倍高速化

パフォーマンスの向上:

- VGG、GoogleNet、ResNetsのような3x3 畳み込み層を持つネットワークの高速化
- 3次元畳み込み
- Pascal世代のGPUの為のFP16ルーティンの追加
- Bilinear Spatial Transformer / ドロップアウト追加

5.9x

char-rnn
RNN レイヤーの速度向上

2.8x

DeepSpeech 2
RNNレイヤーの速度向上

Performance relative to torch-rnn
(<https://github.com/jcjohnson/torch-rnn>)
DeepSpeech2: <http://arxiv.org/abs/1512.02595>
Char-rnn: <https://github.com/karpathy/char-rnn>

cuDNN 5.1

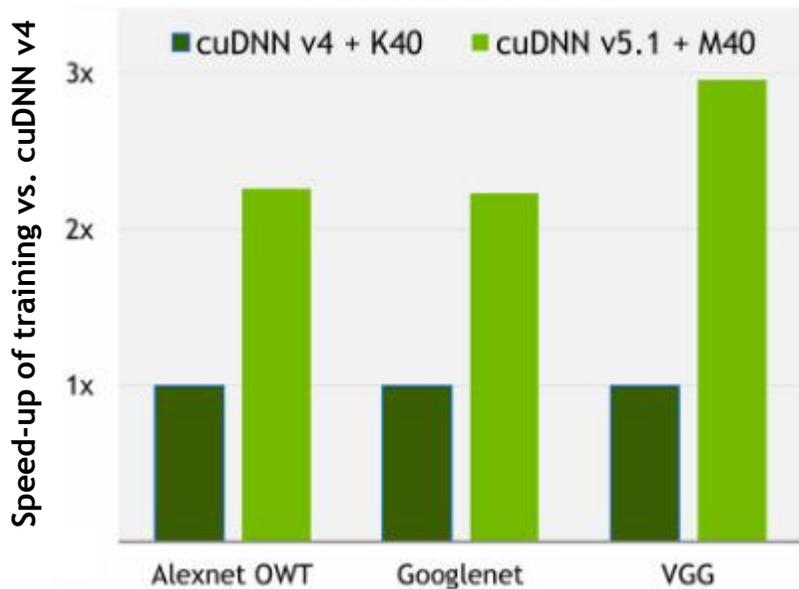
ディープラーニング計算の為に高速なプリミティブ群

- 3x3と5x5の畳込みの順伝播・および逆伝播計算の為に新しいWinogradアルゴリズムが追加。
 - CUDNN_CONVOLUTION_FWD_ALGO_WINOGRAD_NONFUSED
 - CUDNN_CONVOLUTION_FWD_ALGO_WINOGRAD_NONFUSED
- FP16の順伝播、逆伝播計算のパフォーマンスの向上
- CUDA8.0のサポート

cuDNN 5.1

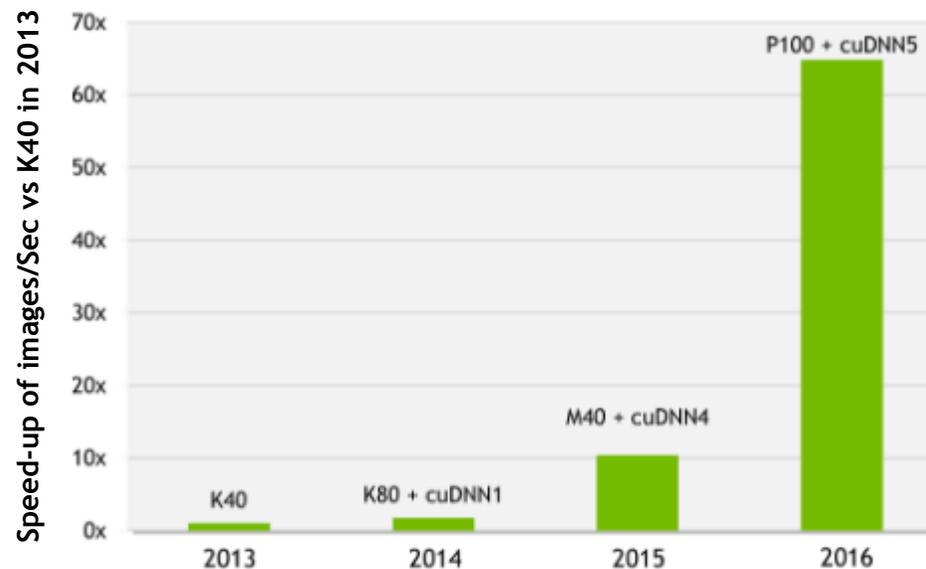
パフォーマンス

3x3のコンボリューション層をもつ
ネットワークの学習が2.7倍高速に



cuDNN 4 + K40 vs. cuDNN 5.1 RC + M40 on Torch and Intel Xeon Haswell
Single-socket 16-core E5-2698 v3 @2.3GHz 3.6GHz Turbo

この3年間で60倍学習が高速に



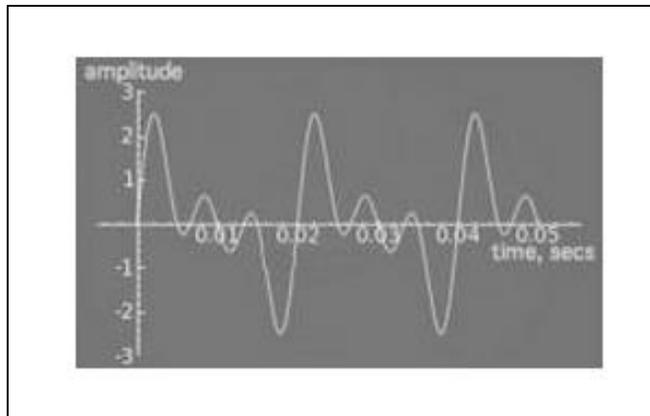
Alexnet training throughput on:

CPU: 1xE5-2680v3 12 Co 2.5GHz 128GB System Memory, Ubuntu 14.04
M40 bar: 8xM40 GPUs in a node. P100: 8xP100 NVLink-enabled

DL学習を加速するCUDAライブラリ

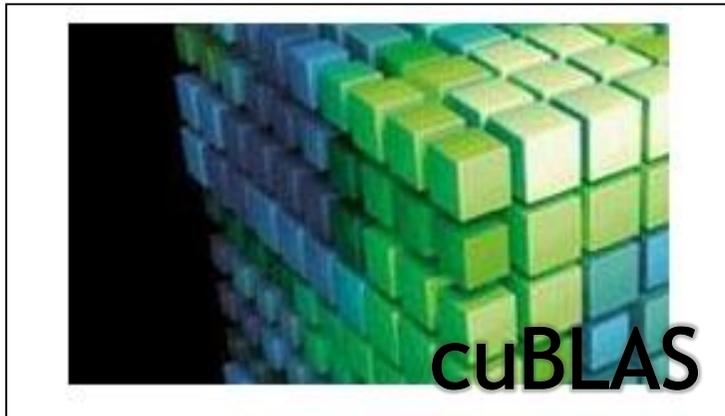
cuFFT

高速フーリエ変換用ライブラリ



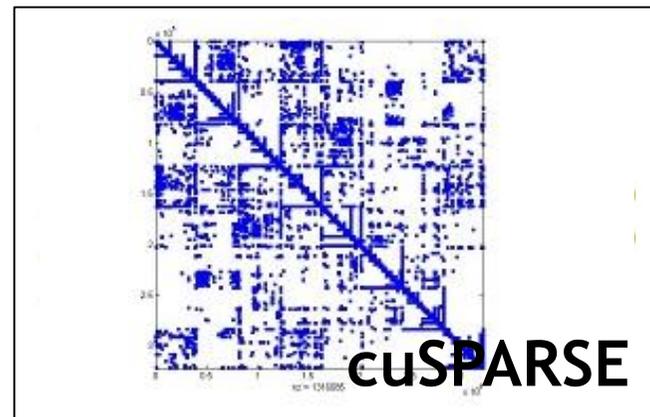
cuBLAS

密行列演算ライブラリ



cuSPARSE

疎行列演算ライブラリ



CUDA8: nvGraph

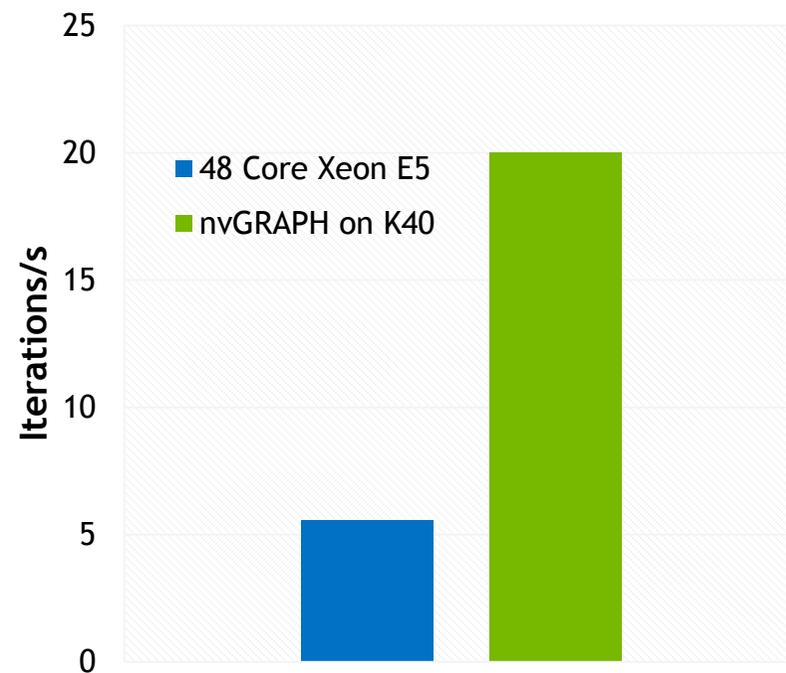
グラフ解析の高速化のためのライブラリ

シングルGPU (24GB M40)で25億エッジのグラフ探索

多岐に渡るアプリケーションの高速化が可能

ページランク	最短経路探索	ワイド経路探索
検索	ロボティックス 経路探索	IP ルーティング
推薦エンジン	電力ネットワーク プランニング	チップデザイン / 半導体
ソーシャル広告の配置	物流 & サプライチェーン プランニング	交通状態に敏感な ルーティング

nvGRAPH: 4倍の速度向上



ページランク on Wikipedia 84 M link dataset

developer.nvidia.com/nvgraph

NCCL(NVIDIA Collective Collection Library)

ディープラーニング SDK

マルチGPU集合通信ライブラリ

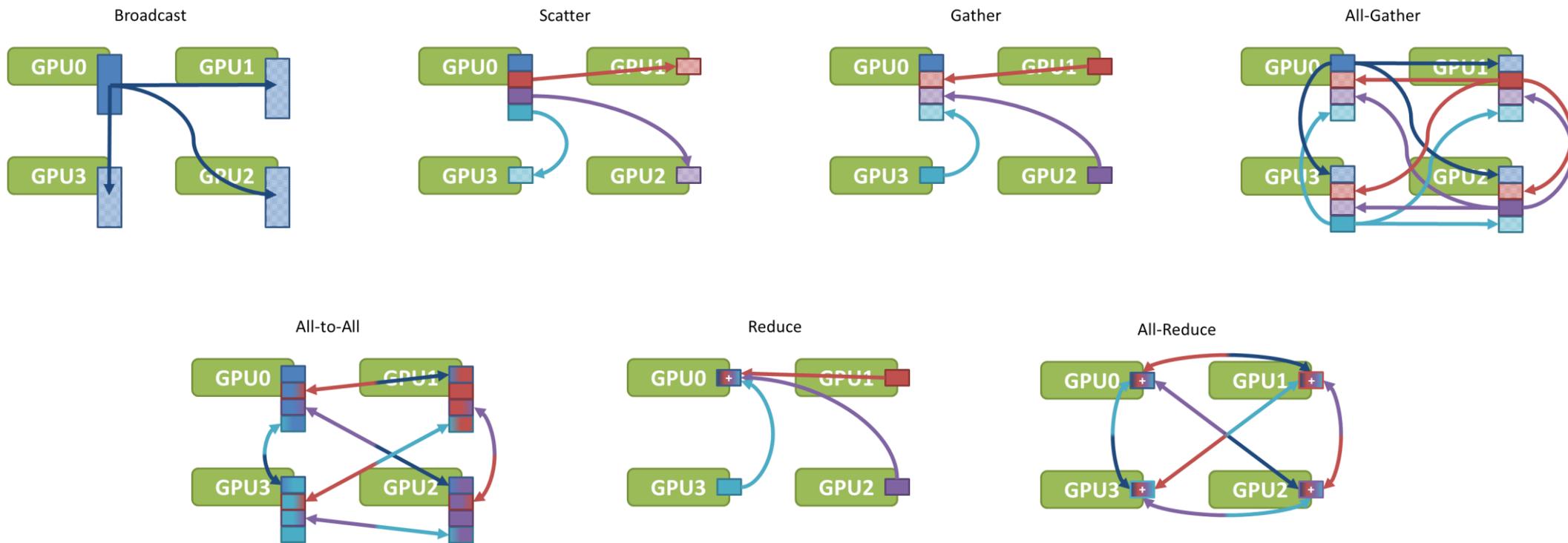
- 最新リリースはv1.2.3
- <https://github.com/NVIDIA/nccl>

all-gather, reduce, broadcast など標準的な集合通信の処理をバンド幅が出るように最適化

シングルプロセスおよびマルチプロセスで使用する事が可能

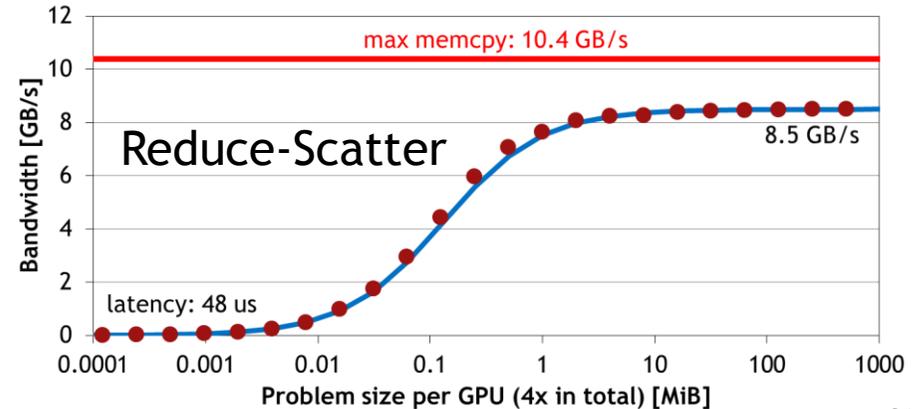
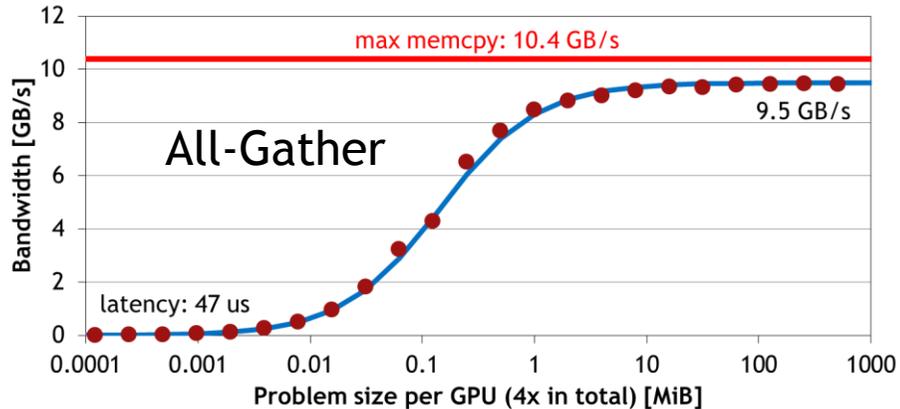
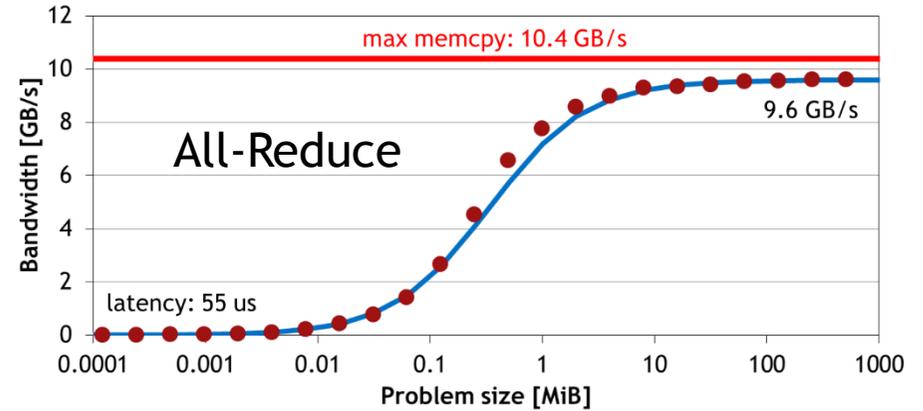
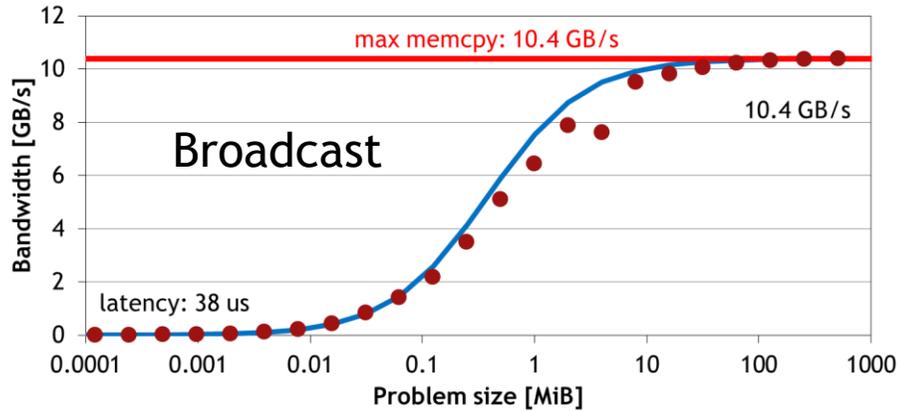
NCCL(NVIDIA Collective Collection Library)

NCCLの集合通信処理



NCCL パフォーマンス

Bandwidth at different problem sizes (4 Maxwell GPUs)



GPU インファレンス エンジン

ディープラーニングにおける最高の推論パフォーマンス

高性能なフレームワークで、
GPU上の推論エンジンを、容易に開発する。

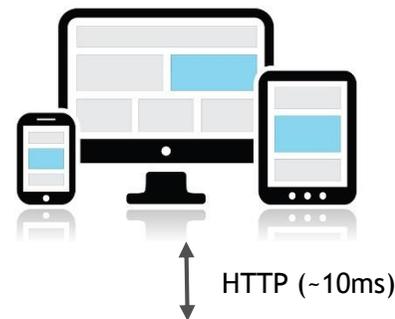
ディープラーニングによる推論を用いた、
製品展開のためのソリューション

トレーニング済みネットワーク、ターゲットGPUに対して、
推論エンジンの性能を最適化

32ビット、16ビット 演算を用いた推論エンジン

Hyperscale, ADAS, Embedded 向け

developer.nvidia.com/gpu-inference-engine



データセンターにおける GPU インファレンス エンジン

画像分類

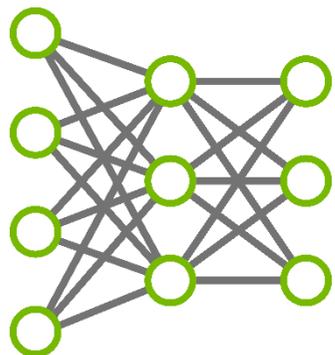
オブジェクト
検知

音声認識

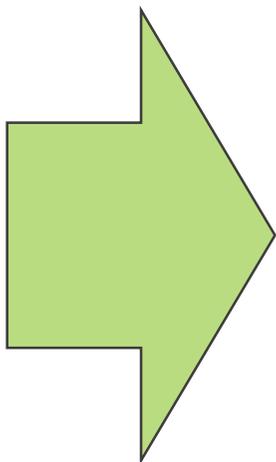


GPU インファレンス エンジン

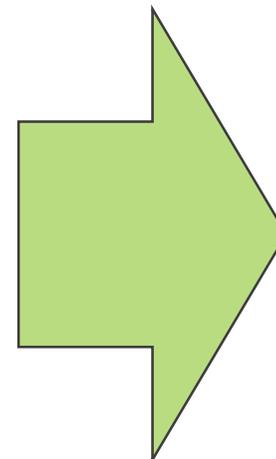
ニューラルネットワークの最適化



トレーニング済みの
ニューラルネットワーク



- ネットワークレイヤの融合
- Concatenationレイヤの除去
- カーネル特殊化
- ターゲットプラットフォームに対する
オートチューニング
- 最適なテンソルレイアウトの選択
- バッチサイズのチューニング



最適化された
推論エンジン

GPU インファレンス エンジン

GoogleNet パフォーマンス

BATCH=1	M4	TX1	TX1 FP16
GIE	3.7 ms	13.9 ms	16.5ms (N=2)
Caffe	15 ms	33 ms	n/a

BATCH=16	M4	TX1	TX1 FP16
GIE	39 ms	164 ms	99 ms
Caffe	67 ms	255 ms	n/a

Jetson TX1 HALF2 column uses fp16

DIGITS

The background is a solid light green color. On the right side, there are several overlapping, semi-transparent geometric shapes, primarily triangles and polygons, in a slightly darker shade of green. These shapes are arranged in a way that creates a sense of depth and complexity, resembling a network or a digital structure.

NVIDIA DIGITS

GPUで高速化されたディープラーニングトレーニング・システム

学習データの作成

The screenshot shows the 'Image Classification Dataset' page. At the top, it displays 'voc_cropped@256x256' with a link icon. Below this, there are two main sections: 'Job Information' and 'Parse Folder (train/val)'. The 'Job Information' section includes fields for 'Job Directory' (a path), 'Image Type' (Color), 'Image Dimensions' (256x256), and 'Resize Mode' (half_crop). The 'Parse Folder (train/val)' section includes 'Folder' (a URL), 'Number of categories' (20), 'Training images' (26759), and 'Validation images' (8917 (25.0%)). At the bottom, there is a 'Create DB (train)' section with 'Input file' (train.txt) and 'DB Entries' (26759). A small bar chart is visible at the bottom left.

モデルの作成

The screenshot shows the 'New Model' page. It features a 'Select Dataset' dropdown menu with options like 'PASCAL VOC', 'ILSVRC 2012', and 'MNIST Dataset'. Below this is the 'Solver Options' section, which includes 'Training epochs' (30), 'Validation interval (in epochs)' (1), 'Batch size' (100), and 'Base Learning Rate' (0.01). There is a 'Custom Network' section with a code editor containing a JSON-like configuration for a convolutional layer. At the bottom, there is a 'Model Name' field with 'ImageNet' and a 'Create' button.

学習過程の可視化

The screenshot shows the 'Image Classification Model' page. It displays 'Solver' options (solver.prototxt, Network (train/val), Network (deploy)) and 'Dataset' information (voc_cropped@256x256, Done Wed Mar 11, 05:16:57 PM). Below this is a line graph showing 'Loss (train)' (red), 'Loss (val)' (orange), and 'Accuracy (val)' (blue) over 10 epochs. The graph shows training loss decreasing and validation accuracy increasing over time.

モデルのテスト

The screenshot shows the 'Test Image' page. It displays a test image of a handwritten digit '8'. To the right, there is a 'Predictions' table with the following data:

Predictions
8
3
0
6
4

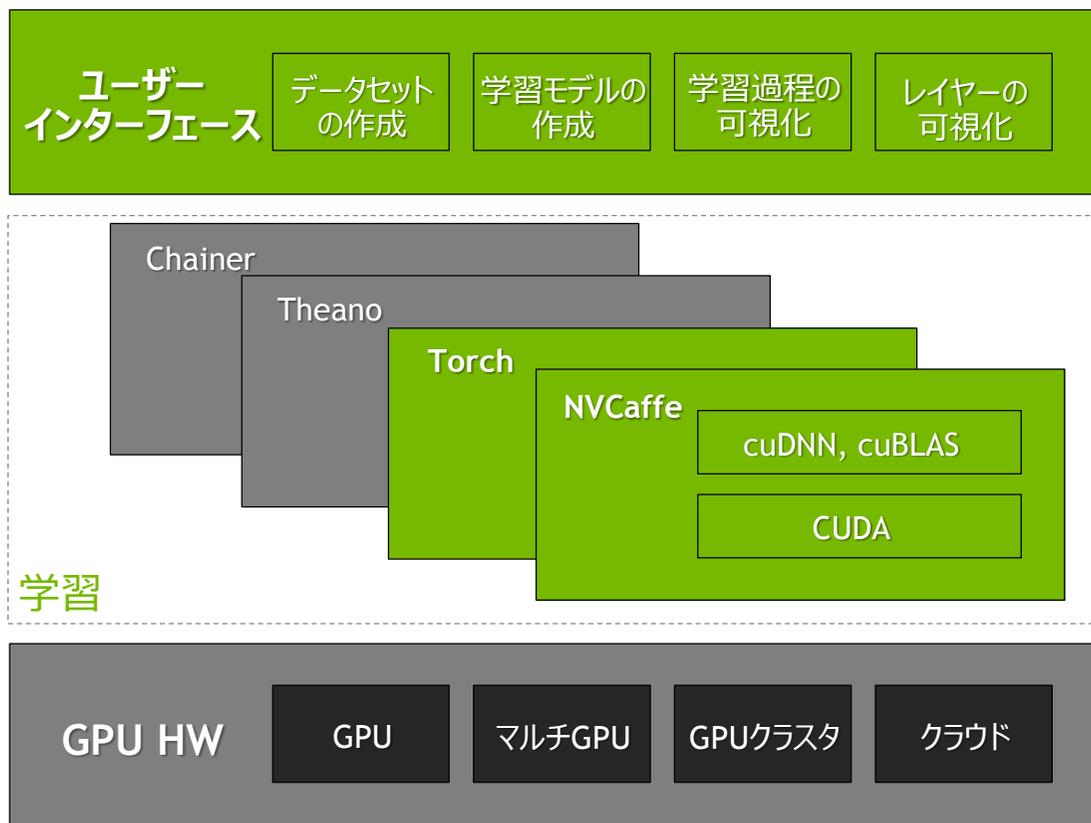
Below the predictions, there is a 'Layer Activations' section. It shows the 'conv1' layer activation as a heatmap of the digit '8'. Below that, it shows the 'pool1' layer activation as three smaller heatmaps of the digit '8'.

<http://developer.nvidia.com/digits>

NVIDIA DIGITS

GPUで高速化されたディープラーニングトレーニング・システム

- ▶ 手元のPCからWebブラウザでアクセス可能なディープラーニングトレーニングシステム



- 画像分類と物体検出の為のDNNのデザインと可視化の機能を提供
- NVIDIAが最適化したフレームワークで高速に学習が可能
- ハイパーパラメータ・チューニングを強力にサポート
- DL学習のジョブを簡単にスケジューリング、リアルタイムにaccuracyとlossを監視
- 様々な学習データフォーマット、フレームワークに対応

DIGITS 4

画像分類だけでなく、様々なDLワークフローに対応

- DIGITS4 RCは**7月公開予定**
- 物体検出の学習に対応
 - 顔や人やその他のオブジェクトを検出するニューラルネットワークを学習する機能を提供
- ハイパーパラメータチューニング機能の強化
 - バッチサイズおよび学習率の複数值指定に対応
 - モデルの認識精度を向上させる為のチューニングが容易に

The screenshot displays the DIGITS 4 web interface. At the top, there are navigation links for 'DIGITS', 'kaspy_walkthrough', 'Test One', 'Login', and 'Info-'. The main content area is divided into several sections:

- Source image:** Shows an aerial view of a construction site with several vehicles. A green button labeled 'Infer Model Done' is visible to the right.
- Inference visualization:** Shows the same aerial view with red bounding boxes around the vehicles, indicating object detection results.
- Statistics:** A histogram showing the distribution of the 'data' activation values. The x-axis is labeled 'Value' and the y-axis is labeled '38'. The histogram is blue and shows a peak around 16.5.
- Description:** A section for 'data' with the following statistics:
 - Data shape: [3 1024 1536]
 - Mean: -11.9555
 - Std deviation: 48.7568
- transformed_data:** A section for 'transformed_data' with the following statistics:
 - Data shape: [3 1024 1536]
 - Mean: -138.956
 - Std deviation: 48.7568

物体検出

DIGITS 4の新しい機能

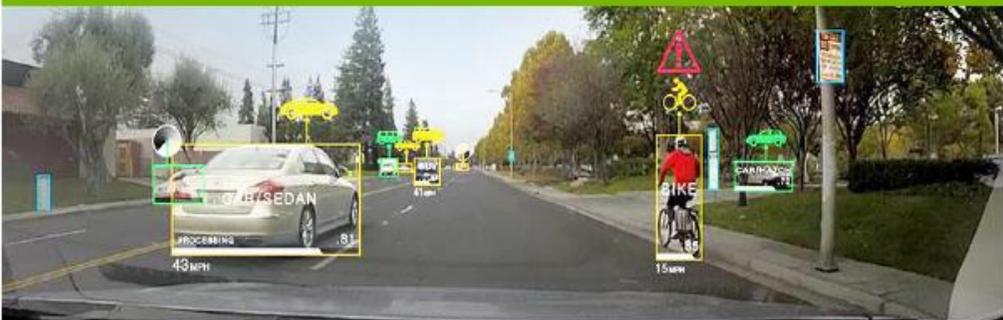
リモートセンシング



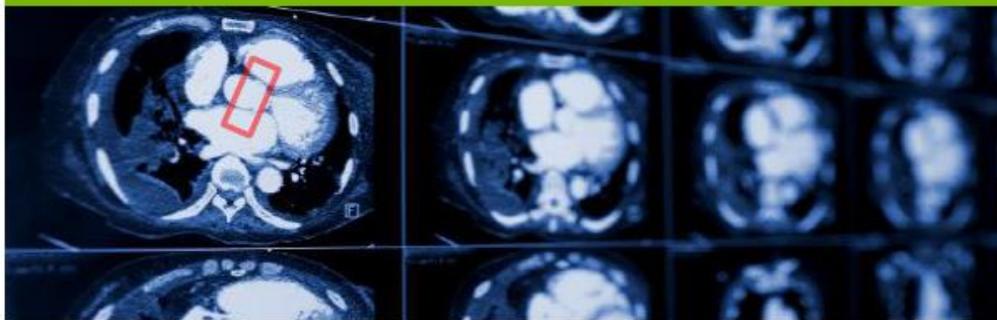
インテリジェント・ビデオアナリシス



ADASシステム



医療診断



ハイパーパラメータチューニング機能強化

DIGITS 4の新しい機能

- [指定学習率の数]x[指定バッチサイズ数]のジョブが一度に投入可能
 - ハイパーパラメータ・チューニング時に必要な手間を軽減

Base Learning Rate ? multiples allowed

Batch size ? multiples allowed



Home

0/2 GPUs available

Name	Start Time	Status	Training Loss	Progress	
G&B learning_rate:0.2 batch_size:2	02:36:15 PM	Initialized		0%	Abort
G&B learning_rate:0.01 batch_size:2	02:36:14 PM	Running		0%	Abort
G&B learning_rate:0.2 batch_size:1	02:36:14 PM	Running		0%	Abort
G&B learning_rate:0.015 batch_size:2	02:36:14 PM	Initialized		0%	Abort
G&B learning_rate:0.015 batch_size:1	02:36:13 PM	Running		8%	Abort
G&B learning_rate:0.01 batch_size:1	02:36:13 PM	Running		12%	Abort

Showing 1 to 6 of 6 entries

一度に複数のジョブ投入が出来る

DIGITS ダウンロード

developer.nvidia.com/digits

github.com/NVIDIA/DIGITS

The screenshot shows the NVIDIA DIGITS website homepage. At the top, there is a navigation bar with the NVIDIA logo and links for 'CUDA ZONE', 'Getting Started', 'Downloads', 'Training', and 'Ecosystem'. Below the navigation bar, the breadcrumb trail reads 'Home > CUDA ZONE > Tools & Ecosystem > GPU Accelerated Applications > NVIDIA® DIGITS™ – Interactive Deep Learning GPU Training System'. The main heading is 'NVIDIA® DIGITS™ – Interactive Deep Learning GPU Training System'. Below the heading, there is a paragraph describing the system: 'The NVIDIA Deep Learning GPU Training System (DIGITS) puts the power of deep learning in the hands of data scientists and researchers. Quickly design the best deep neural network (DNN) for your data using real-time network behavior visualization. Best of all, DIGITS is a complete system so you don't have to write any code. Get started with DIGITS in under an hour.' Below this paragraph is a quote from Simon Osindero, AI Architect at Flickr: 'DIGITS makes it way easier to design the best network for the job. The DIGITS interface makes it super easy to track key diagnostics during training. The field will definitely benefit from having tools like this for configuration and introspection.' At the bottom of the page, there are three screenshots illustrating the DIGITS interface: 'DIGITS training multiple DNNs and building a new dataset in parallel.', 'DIGITS real-time monitoring of DNN model training', and 'DIGITS visualizing a test image'.

NVIDIA® DIGITS™ – Interactive Deep Learning GPU Training System

The NVIDIA Deep Learning GPU Training System (DIGITS) puts the power of **deep learning** in the hands of data scientists and researchers. Quickly design the best deep neural network (DNN) for your data using real-time network behavior visualization. Best of all, DIGITS is a complete system so you don't have to write any code. Get started with DIGITS in under an hour.

DIGITS makes it way easier to design the best network for the job. The DIGITS interface makes it super easy to track key diagnostics during training. The field will definitely benefit from having tools like this for configuration and introspection.

— Simon Osindero, AI Architect at Flickr

DIGITS training multiple DNNs and building a new dataset in parallel.

DIGITS real-time monitoring of DNN model training

DIGITS visualizing a test image

The screenshot shows the GitHub repository page for NVIDIA/DIGITS. At the top, there is a search bar and navigation links for 'Explore', 'Features', 'Enterprise', and 'Blog'. Below the search bar, the repository name 'NVIDIA / DIGITS' is displayed, along with 'Watch' and 'Star' buttons. The description reads 'Deep Learning GPU Training System <https://developer.nvidia.com/digits>'. Below the description, there are statistics: '61 commits', '3 branches', '3 releases', and '8 contributors'. The main content area shows a commit history table with columns for the commit message and the time ago. The commit history table is as follows:

Commit Message	Time Ago
Quick solution for classifying many images	11 hours ago
Change run_me.sh to runme.sh in documentation	6 days ago
Fix capitalization of DIGITS	11 days ago
Add .coverage to .gitignore	3 days ago
Move LICENSE.txt to LICENSE	20 days ago
Update README to reflect new NVIDIA/caffe version	11 days ago
Fix capitalization of DIGITS figlet	6 days ago
Improve scripts for use in virtualenv	4 days ago
Improve scripts for use in virtualenv	4 days ago
Another fix to walkthrough - save MNIST unencoded	3 days ago
Initial commit	28 days ago

ディープラーニング・フレームワーク

ディープラーニング・フレームワーク

GPUで高速化されているディープラーニング・フレームワークが多数存在

Caffe

DL4J
Deeplearning4j

Microsoft
CNTK

mxnet

Purine

torch

TensorFlow

theano

Chainer

K
KERAS

Mocha.jl
julia

MatConvNet

MINERVA

OpenDeep

Pylearn2

ディープラーニング・フレームワーク

<https://developer.nvidia.com/deep-learning-frameworks>

ディープラーニング・フレームワーク

	Caffe	Torch7	Theano	TensorFlow	Chainer
インターフェース	C++/Python/Matlab	Lua/C	Python	C/C++/Python	Python
cuDNN	5	5	5	5	5
ライセンス	BSD-2	BSD	BSD	Apache 2.0	MIT
マルチGPU (1ノード)	○	○		○	○
モデルの柔軟性	△	◎	◎	○	◎
CNN	○	○	○	○	○
RNN	○ #2033	○	○	○	○
RBM	×	○	○	○	○
備考	高速 Caffe Model Zoo	多数のアルゴリズムを サポート	自動微分	自動微分 TensorBoard	Define by Run CuPy

ディープラーニングの基礎

NVIDIA ディープラーニング学習コース

NVIDIA ディープラーニング・インスティテュート

ディープラーニングの為に自習型のクラス。ハンズオンラボ、講義資料、講義の録画を公開。

ハンズオンラボは日本語で受講可能

<https://developer.nvidia.com/deep-learning-courses>

1. ディープラーニング入門
2. 画像分類の為にDIGITS入門
3. Caffe入門
4. Theano入門
5. Torch入門



qwikLABS:AWSを用いた自主学習ラボ

手元のPCにGPUが搭載されていなくてもOK

- <https://nvidia.qwiklab.com>に多数のラボを公開
 - 「ディープラーニング入門」は無料。日本語で受講可能。無料のアカウントを作る必要あり。



The screenshot shows a webpage titled "Introduction to Deep Learning Quest" with a "Domains" tab and a "登録" (Register) button. A vertical progress indicator on the left shows the current position in the quest. The main content is a table of courses:

Course Name	Level	Cost	Additional Info
ディープラーニング入門	Beginner	無料	🌐
ディープラーニングトレーニングシステムDIGITS入門	Beginner	15クレジット	🌐
ディープラーニングフレームワークCaffe入門	Intermediate	15クレジット	🌐
ディープラーニングフレームワークTheano入門	Expert	15クレジット	🌐
ディープラーニングフレームワークTorch7入門	Expert	15クレジット	🌐

At the bottom, there is a badge for "Introduction to Deep Learning Quest" with a green hexagon icon.

qwikLABS:AWSを用いた自主学習ラボ

qwikLABSを日本語で受講するには?

- <https://nvidia.qwiklab.com>の右上のアカウント設定を選択
 - 「言語」→「日本語(Japanese)」を選択すると日本語で講座を受講可能に

The screenshot shows the user interface of the qwikLABS dashboard. At the top, there is a navigation bar with several tabs: 「受講者画面」, 「アカウント」 (highlighted with a green box), 「サインアウト」, 「ダッシュボード」, 「ラボカタログ」, 「よくある質問」, 「お問い合わせ」, and 「クレジットの購入」. Below the navigation bar, the account settings form is visible. It includes fields for 「*姓」 (Murakami), 「*会社名」 (NVIDIA Japan), and 「言語」. The 「言語」 dropdown menu is open, showing options: 「英語(English)」, 「日本語(Japanese)」 (highlighted in blue), 「中国語(Chinese)」, 「韓国語(Korean)」, and 「スペイン語(Spanish)」. A green arrow points from the 「アカウント」 tab to the account settings form. A green callout box with white text says 「日本語(Japanese)を選択」.

項目	内容
*姓	Murakami
*会社名	NVIDIA Japan
言語	日本語(Japanese)
	英語(English)
	日本語(Japanese)
	中国語(Chinese)
	韓国語(Korean)
	スペイン語(Spanish)

qwikLABS:AWSを用いた自主学習ラボ

手元のPCにGPUが搭載されていなくてもOK

- 「ディープラーニング入門」は無料。日本語で受講可能。



Self-Study Courses for Deep Learning

The NVIDIA Deep Learning Institute offers self-paced classes for deep learning that feature interactive hours Q&A with instructors. You'll learn everything you need to design, train, and integrate neural networks with widely used open-source frameworks and NVIDIA software. During the hands-on exercises, you'll

This is an introductory course, so previous experience with deep learning and GPU programming is not required.

Introduction to Deep Learning

Class: [Download](#) | [Stream](#) | [Slides](#)

Assignment: [Hands-on lab](#)

Q&A: [Download](#) | [Stream](#) | [Slides](#) | [Q&A Log](#)

<https://developer.nvidia.com/deep-learning-courses>



詳細情報

3112 ★★★★★

もしあなたがまだ登録していないならば、ディープラーニングラボ・シリーズにサインアップして下さい。

このインタラクティブクラスでは、GPUで加速することで、急速に発展するディープラーニングの技術をご紹介します。ディープラーニングの近年の進歩は、視覚、聴覚、触覚の段階的変化につながります。

<https://nvidia.qwiklab.com>

ラボの詳細	
トークンが必要です:	無料
レベル:	Beginner
時間	00時:45分
アクセス時間	00時:55分
セットアップ時間	00時:04分
タグ:	Deep Learning

ディープラーニング入門

AGENDA

ディープラーニングとは?

Caffe

Theano

Torch7

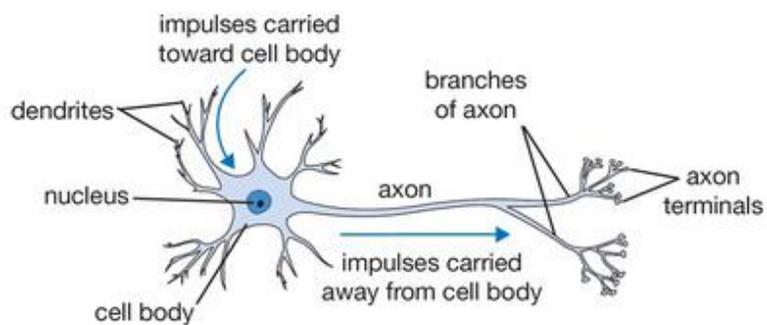
DIGITS

ディープラーニングとは？

人工ニューロン

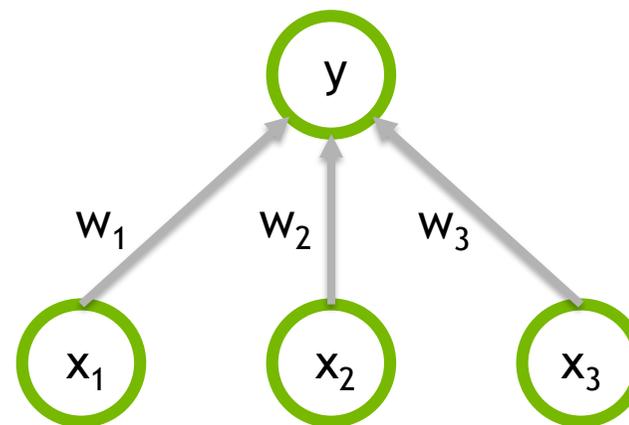
神経回路網をモデル化

神経回路網



スタンフォード大学cs231講義ノートより

人工ニューロン

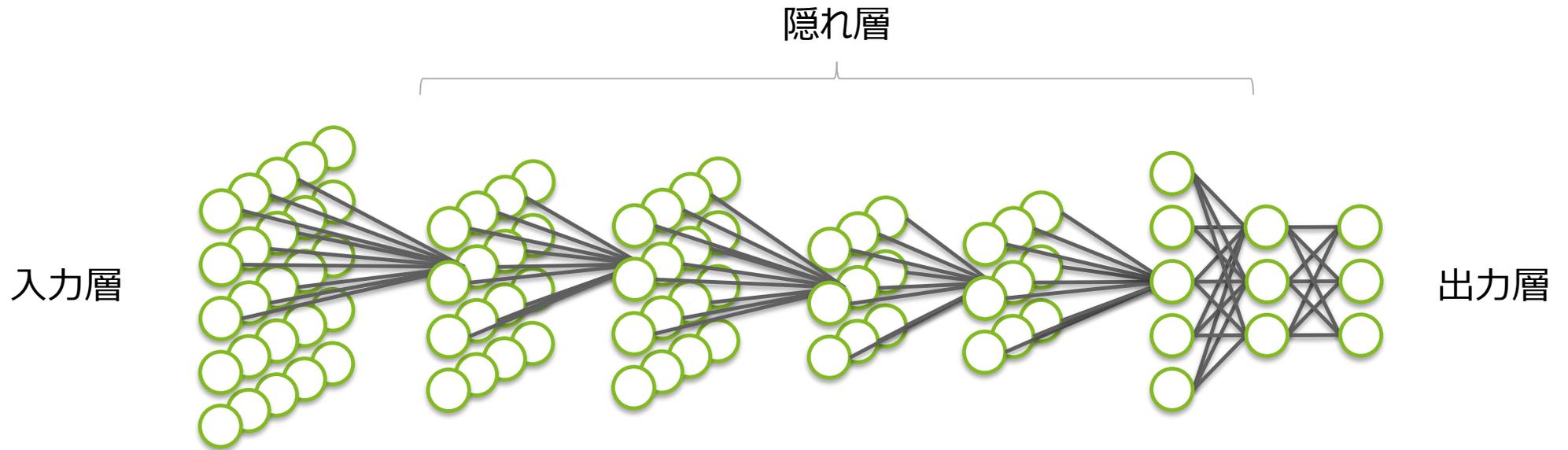


$$y = F(w_1x_1 + w_2x_2 + w_3x_3)$$

$$F(x) = \max(0, x)$$

人工ニューラルネットワーク

単純で訓練可能な数学ユニットの集合体
ニューラルネットワーク全体で複雑な機能を学習



十分なトレーニングデータを与えられた人工ニューラルネットワークは、入力データから判断を行う複雑な近似を行う事が出来る。

ディープラーニングの恩恵

- **ロバスト性**
 - 特徴量の設計を行う必要がない。 - 特徴は自動的に獲得される学習用データのバラツキの影響を押しさえ込みながら、自動的に学習していく
- **一般性**
 - 同じニューラルネットワークのアプローチを多くの異なるアプリケーションやデータに適用する事が出来る
- **スケーラブル**
 - より多くのデータで大規模並列化を行う事でパフォーマンスが向上する

従来型の機械学習を用いた認識

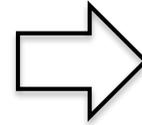
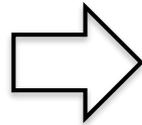
人間が設計した特徴量を使用

Rowデータ

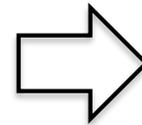
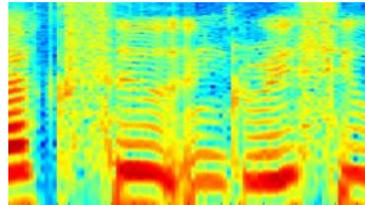
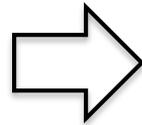
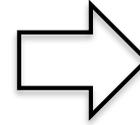
特徴抽出

分類・検出

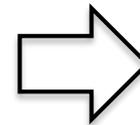
結果



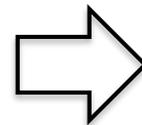
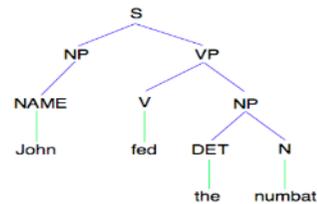
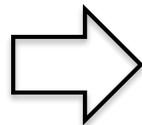
SVM,
shallow neural net,
...



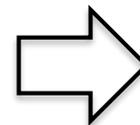
HMM,
shallow neural net,
...



Speaker ID,
speech transcription, ...



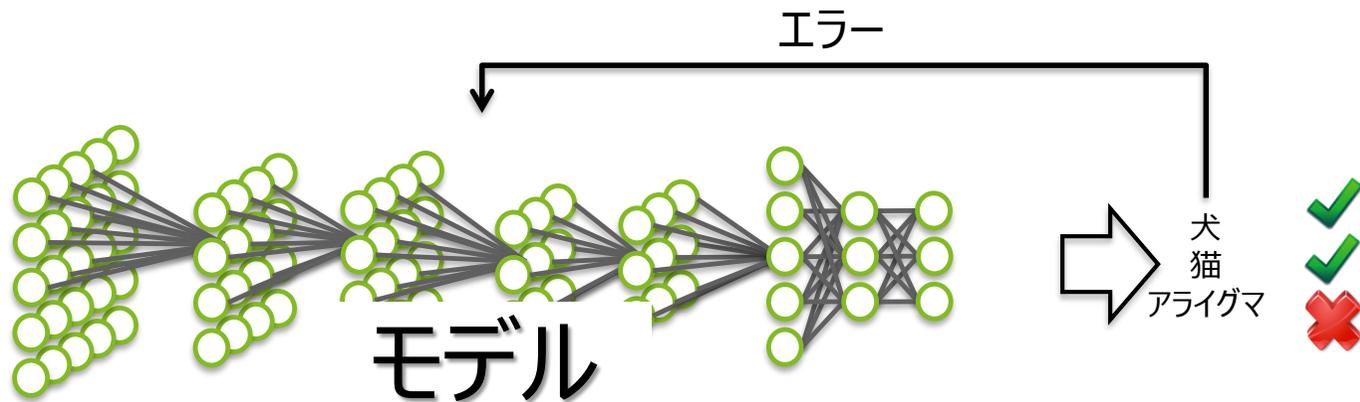
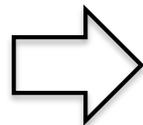
Clustering, HMM,
LDA, LSA
...



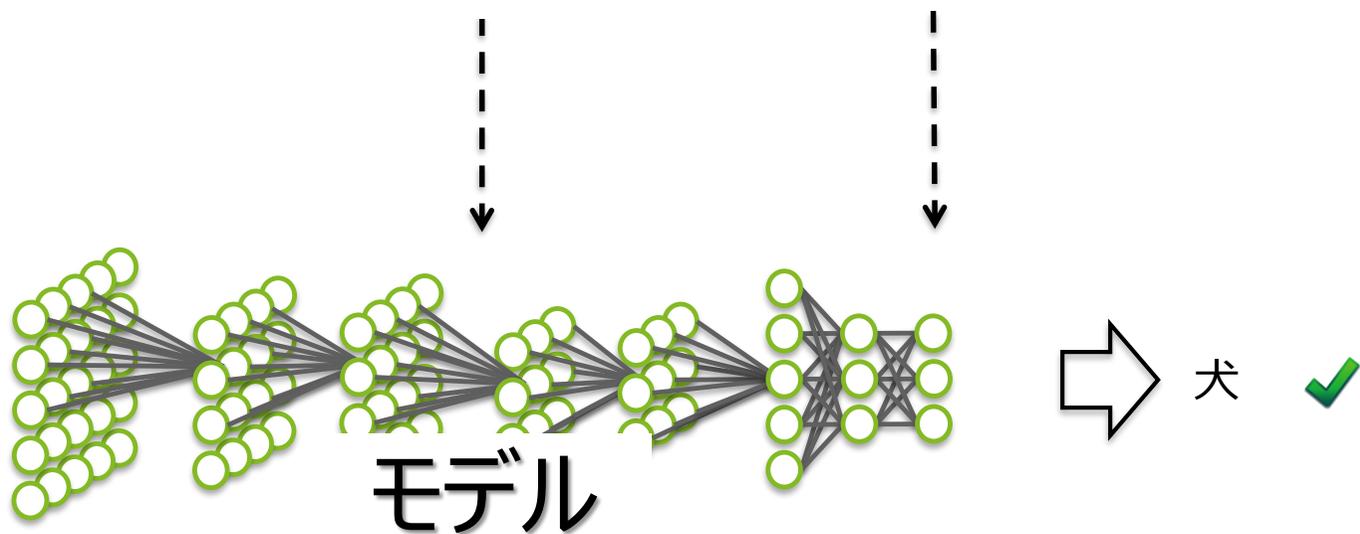
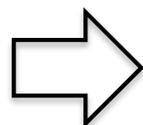
Topic classification,
machine translation,
sentiment analysis...

ディープラーニングのアプローチ

トレーニング:



認識:



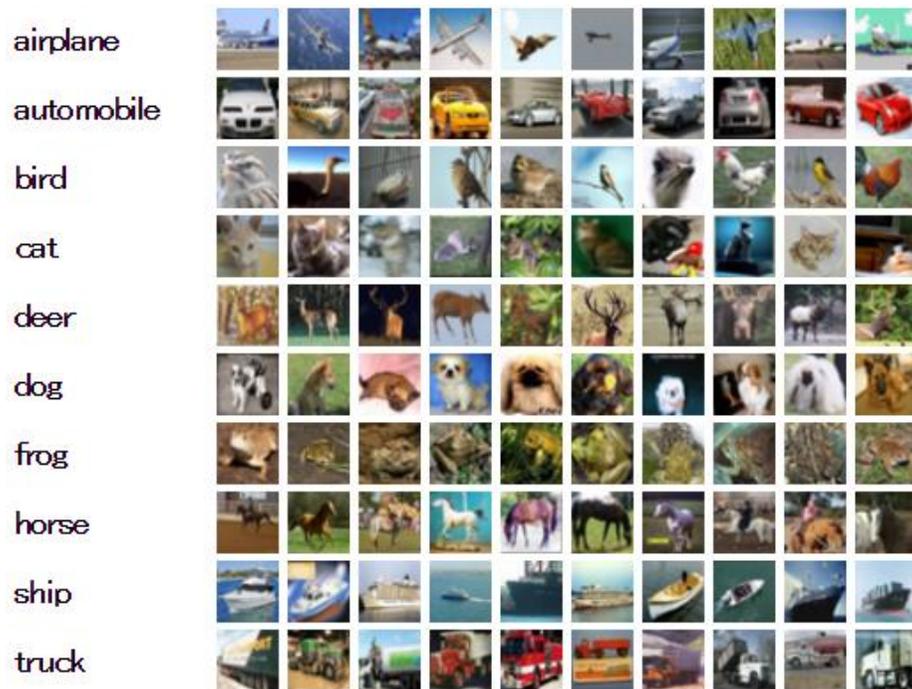
Caffe

The background of the slide is a vibrant green color. On the right side, there is a complex, abstract geometric pattern consisting of numerous overlapping, semi-transparent polygons and lines, creating a layered, crystalline effect. The overall aesthetic is clean and modern.

CAFFE

CIFAR10 を用いた学習

飛行機、自動車、鳥、猫、鹿、犬、蛙、馬、船、トラックの 10 個の分類
32 × 32 ピクセルの 6 万枚のカラー画像から成るデータセット



CAFFE

ネットワーク定義ファイル(例)

```
layer {  
  name: "conv1"  
  type: "Convolution"  
  bottom: "data"  
  top: "conv1"  
  param {  
    lr_mult: 1  
  }  
  param {  
    lr_mult: 2  
  }  
  convolution_param {  
    num_output: 32  
    pad: 2  
    kernel_size: 5  
    stride: 1  
  }  
}
```

使用するネットワークの定義を行う
レイヤーの数や、各レイヤーのタイプ
レイヤータイプ毎のパラメータなどを記述

CAFFE

ソルバー定義ファイル(例)

```
# トレーニングとテストに使用するネットワークの定義
net: "examples/cifar10/cifar10_quick_train_test.prototxt"
# テストイテレーション数
test_iter: 100
# 500トレーニングイテレーション毎にテスト
test_interval: 500
# 学習率・モーメントム
base_lr: 0.001
momentum: 0.9
weight_decay: 0.004
# 学習率のポリシー
lr_policy: "fixed"
# 100 イテレーション毎に表示
display: 100
# イテレーション最大数
max_iter: 4000
```

トレーニングの際に使用するパラメーターのリスト
DNNトレーニングアルゴリズムに渡される

CAFFE

学習の実行

以下のコマンドで学習を開始

```
./build/tools/caffe train --solver=examples/cifar10/cifar10_quick_solver.prototxt
```

Caffeコマンド

学習時は”train”を指定

ソルバー定義

Theano

The background of the slide is a solid light green color. On the right side, there is a decorative graphic consisting of several overlapping, semi-transparent green wireframe cubes or rectangular prisms, creating a 3D effect.

THEANO

MNISTを用いた学習

0から9の数字の手書き文字画像データ

28x28ピクセル、7000枚のグレースケール画像からなるデータセット

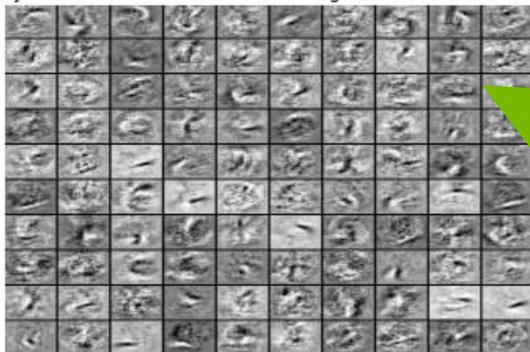


THEANO

MNISTを用いた学習:実行結果

Optimization complete. Best validation score of **98.397436** % obtained at iteration 6630, with test performance 98.387420 %

Layer 0 convolutional filters, training cost: 1.61258012821



可視化すると1万回反復してトレーニングしている間に
重みが変わっている事が分かる

最終的に分類精度は98%を超える

Layer 1 fully connected weights, training cost: 1.61258012821



Layer 2 fully connected weights, training cost: 1.61258012821



Torch7

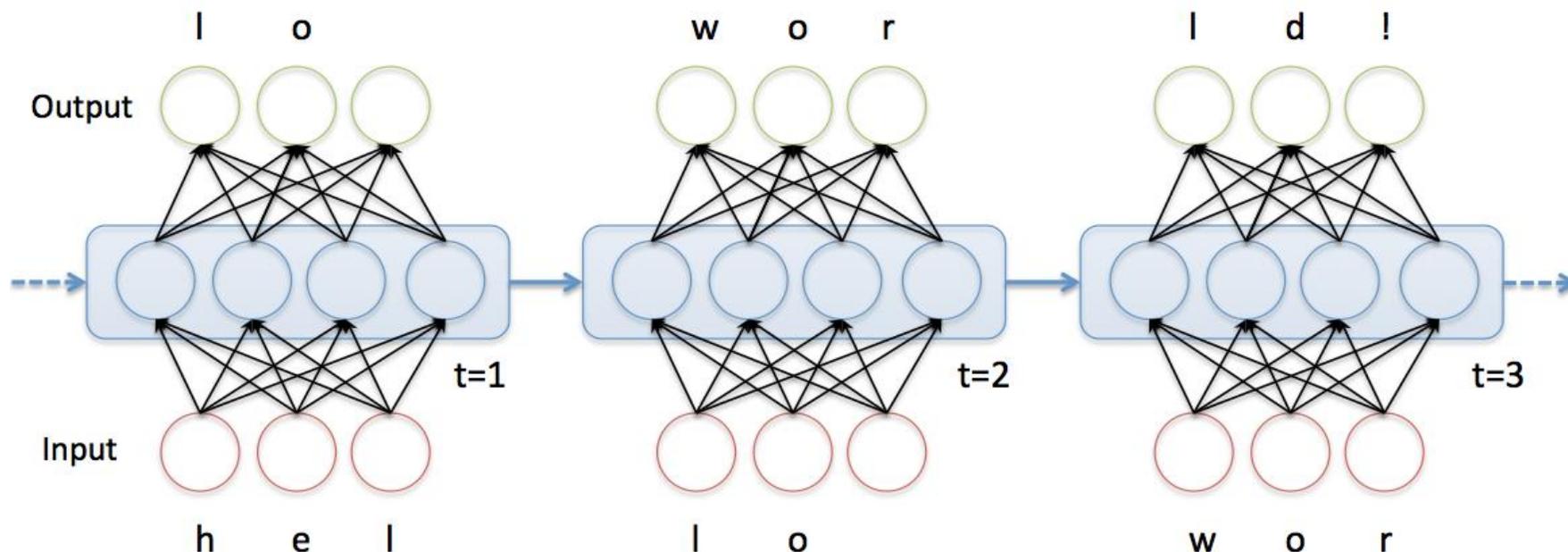
The background of the slide is a solid light green color. On the right side, there is a complex, abstract geometric pattern composed of overlapping, semi-transparent green polygons and lines, creating a mesh-like or crystalline structure that extends from the top right towards the bottom right.

TORCH7

リカレントニューラルネットワーク(RNN)を使ったテキスト生成

TorchベースのRNNを用いたテキスト生成

シェークスピア、トルストイ、オバマ大統領のスピーチ原稿作成者、Linux カーネルのソースコード作成者のいずれかのスタイルでテキストを生成



TORCH7

実行結果:Style=obama

Michelle, another defining place meant on the only time from the United States is a few years we're continued here to live up to balance on Ukraing; from Iraq.

Let me took off in, loans and the people of Hawaii will justify just like a two. One of what reform can't act as we can. Part of all the rights of somebody as so they described another accomplishine between 15-year believe.

So the quiet Republican orgence of broadband new majuri with our head, a leader American had train and making '

TORCH7

実行結果:Style=linux

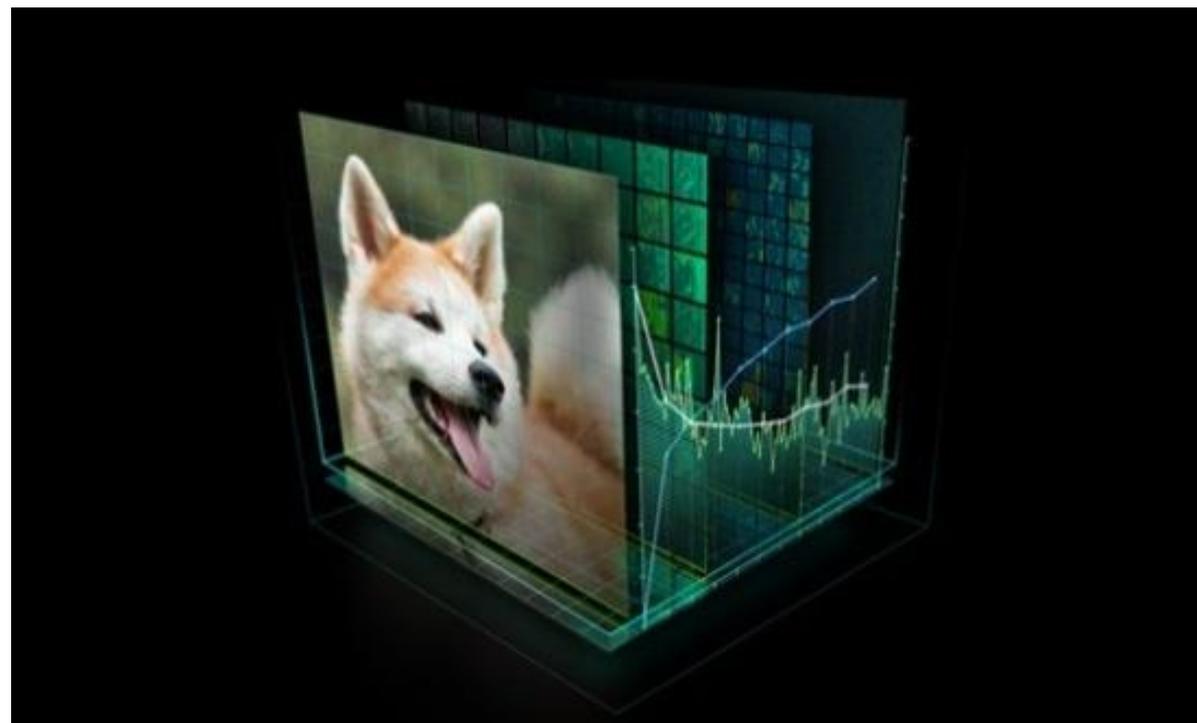
```
CPU_DEBUG + this_cpu->expect->state = PM_SUSPEND_MAX);
}
static inline void torture_ops_trace_trace_clock(struct ring_buffer_event *event)
{
    u64 rw = (struct cgroup_subsys_state *css && !LOCK_RMODED & fail);
    return ret;
}
static void perf_string_print(struct timespec __user *, bin_niveltist Clock_vidch,
                             struct pid_namespace *ns, unsigned int idx)
{
    unsigned long flags = a; if (handler_nid) return NULL;
    if (atomic_dec_and_test(&jiffies))
        put_futex_key := when allocate_st
```

NVIDIA ディープラーニング学習コース

NVIDIA ディープラーニング・インスティテュート

ディープラーニング・インスティテュートで今後公開予定のコース

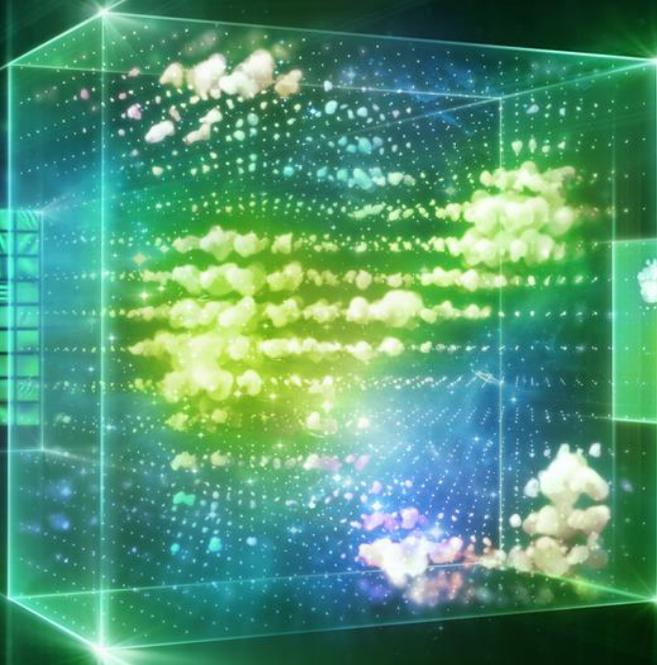
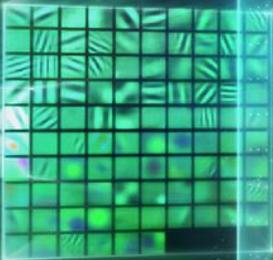
1. TensorFlow入門
2. Chainer入門
3. CNTK入門
4. DIGITS中級講座



参考

ディープラーニング・フレームワーク関連サイト

- ディープラーニング学習コース
 - <https://developer.nvidia.com/deep-learning-courses> (学習コース/講義資料・録画)
 - https://nvidia.qwiklab.com/lab_catalogue (ハンズオンラボ)
- Caffe
 - <http://caffe.berkeleyvision.org/> (公式)
- Torch
 - <http://torch.ch/> (公式)
- Theano
 - <http://deeplearning.net/software/theano/> (公式)
- TensorFlow
 - <http://tensorflow.org/> (公式)
- Chainer
 - <https://github.com/pfnet/chainer> (公式)
 - <https://github.com/hido/chainer-handson> (Jupyter Notebook)
- CNTK
 - <https://cntk.codeplex.com/> (公式)



THANK YOU