



NVIDIA Deep Learning 說明資料

GPUコンピューティング

GPUとは

Graphics Processing Unit

PCやモバイル機器等に搭載されるグラフィックス
描画用半導体のこと

コンピュータゲームで3D CGを高速に描画するため、

大量の行列計算を高速に行う需要から発達し、
現在はCADや画像処理・画像認識の分野でも
不可欠な半導体となっている

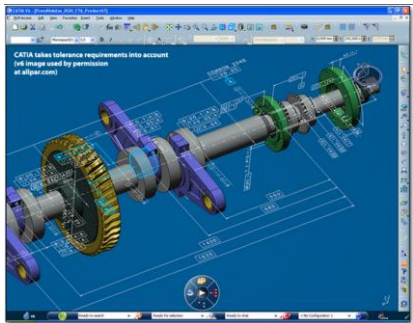
CATIA (CAD)

FINAL FANTASY (ゲーム)

スマートフォンでの顔認識



GPU



GPUの汎用化へ

GPUコンピューティングとは

CGや画像処理だけでなく、
汎用目的にGPUを活用する手法

NVIDIAが**2004年に**
《Strategic Investment》を
始め、**GPUコンピューティング用の**
プログラミング言語《CUDA》を
開発して、2007年から提供を開始

現在、CUDAは大学・研究機関や
ソフトベンダーに広く普及している

NVIDIA Introduction and HPC Evolution of GPUs



- Public, based in Santa Clara, CA | ~\$4B revenue | ~5,500 employees
- Founded in 1999 with primary business in semiconductor industry
 - Products for graphics in workstations, notebooks, mobile devices, etc.
 - Began R&D of GPUs for HPC in 2004, released first Tesla and CUDA in 2007
- Development of GPUs as a co-processing accelerator for x86 CPUs

HPC Evolution of GPUs

- **2004: Began strategic investments in GPU as HPC co-processor**
- 2006: G80 first GPU with built-in compute features, 128 cores; CUDA SDK Beta
- **2007: Tesla 8-series based on G80, 128 cores – CUDA 1.0, 1.1**
- **2008: Tesla 10-series based on GT 200, 240 cores – CUDA 2.0, 2.3**
- **2009: Tesla 20-series, code named “Fermi” up to 512 cores – CUDA SDK 3.0**

3 Generations of
Tesla in 3 Years

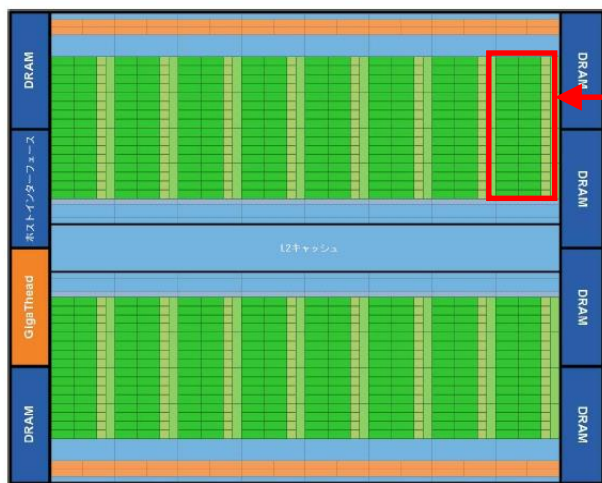
GPUのアーキテクチャ

HPC市場でCPUではなくGPUを選択する理由

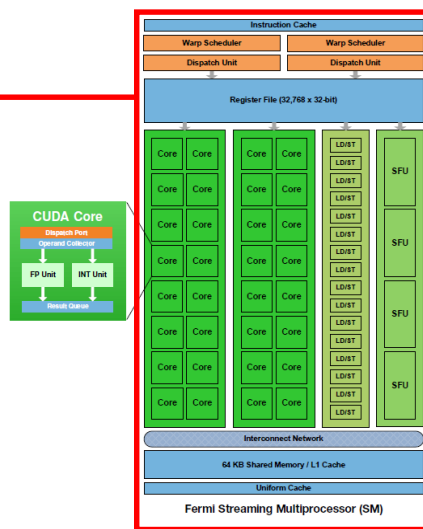
GPUコンピューティング = **Massively Parallel Computing (超並列計算)**

GPUアーキテクチャ

1GPUチップに
14~16SMがある



1SMに
32コアがある



共通L2キャッシュを囲むように配置されているのが、FermiのSM、16個です。
各SMは長方形をしており、オレンジの部分がスケジューラーとディスパッチャー、
緑の部分が実行ユニット、薄い水色の部分がレジスタファイルとL1キャッシュです。

ポイント

- ✓ 多数の演算コア
- ✓ 演算コアと高速な通信ができるメモリ(GPUメモリ)

GPU



CPU



1つのGPUに500個前後の演算コアと超高速なメモリが階層型に接続されている

GPUコンピューティングの普及

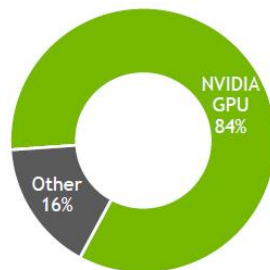
市場で拡大するNVIDIAの存在感

世界のデータセンターの**84%**でNVIDIAのGPUが採用される

- CUDAが動くGPUは**6億個**
- CUDAは**330万**ダウンロード
- CUDAコースは**800**大学
- CUDA関連論文は**6万編**





Data Center Accelerator Share



Source: Intersect360Research, Feb 2015

Broad Developer Reach


600M CUDA GPUs


3.3M CUDA downloads


800 CUDA courses


60K CUDA papers

World's Fastest Supercomputers

Titan (U.S.): 18,000 GPUs

CSCS (Europe): 5,200 GPUs

Tsubame (Japan): 4,000 GPUs

ディープレイヤーニングの市場

AlphaGo

プロ棋士を破った世界初の囲碁のコンピュータプログラム

モンテカルロ木探索とディープラーニングを組み合わせたシステム

3週間、50GPUを用いて3億4千万回のトレーニング・ステップでDNNをトレーニング

対局: 40の検索スレッド、1202個のCPUと176個のGPUを使用して計算

バリューネットワークで基盤の状況を予測し、ポリシーネットワークで手の評価を行う

全5局の対局、囲碁のヨーロッパチャンピオンと世界チャンピオンに勝利

<http://www.nature.com/nature/journal/v529/n7587/full/nature16961.html>

<http://deepmind.com/alpha-go.html>

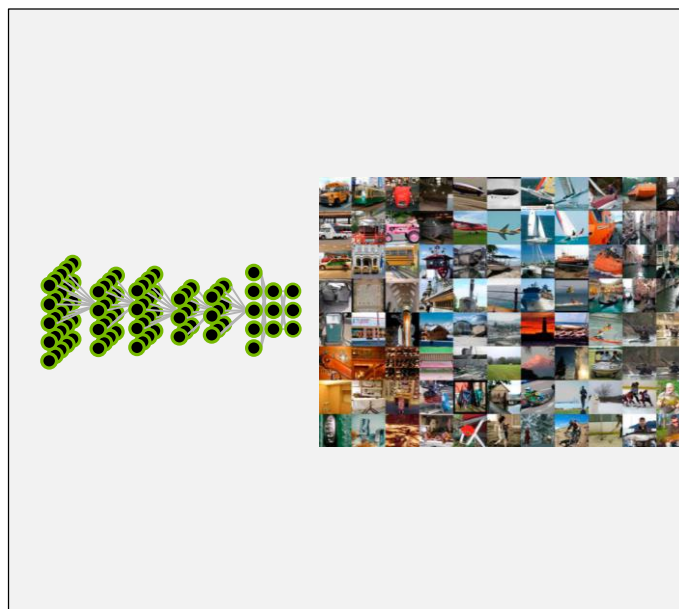


ディープラーニング

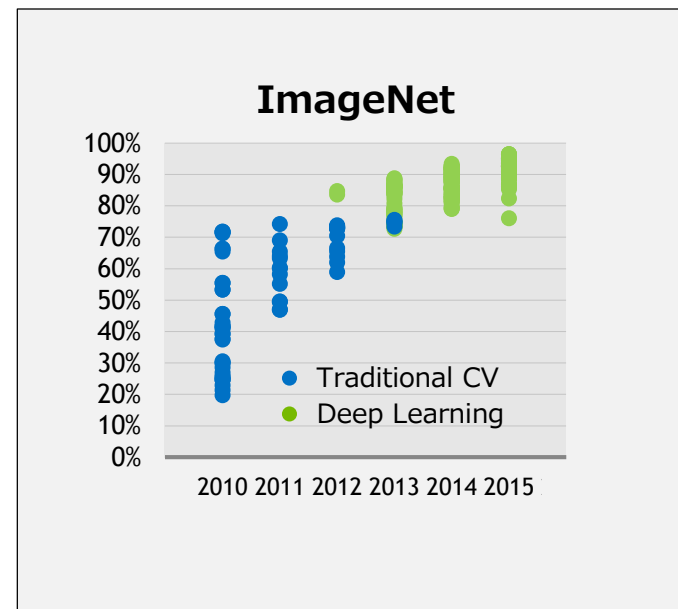
新しいコンピューティングモデル



従来からのコンピュータービジョン
専門家 + 時間

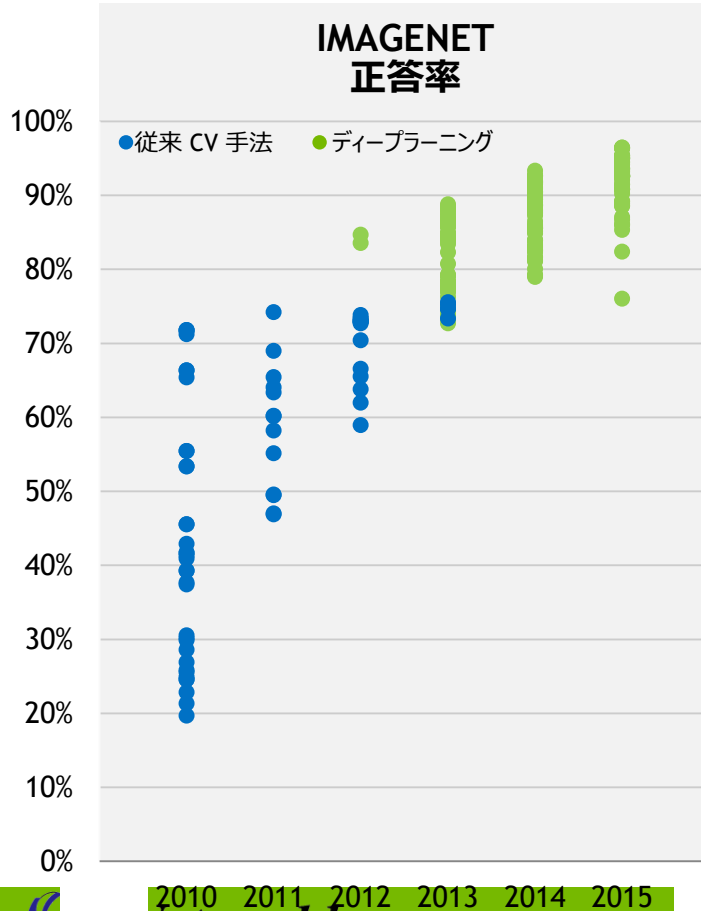


ディープラーニングによる物体認識
DNN + データ + HPC



ディープラーニングが
人間を超える成果を達成

ディープラーニングによって加速する AI 競争



幅広いディープラーニングの応用例

応用分野

顔、場所、物体認識

音声認識および翻訳

自然言語処理

ビデオ検索および分析

インデキシングおよび検索

レコメンデーションエンジン

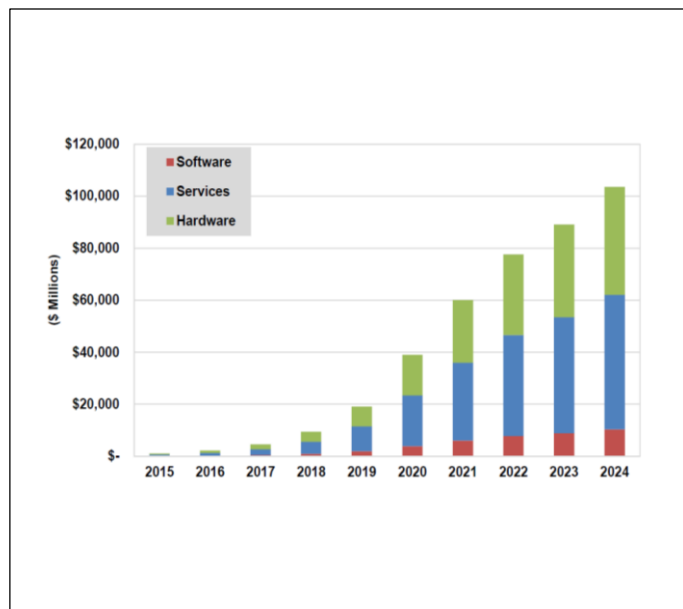
コンピュータビジョン自動運転



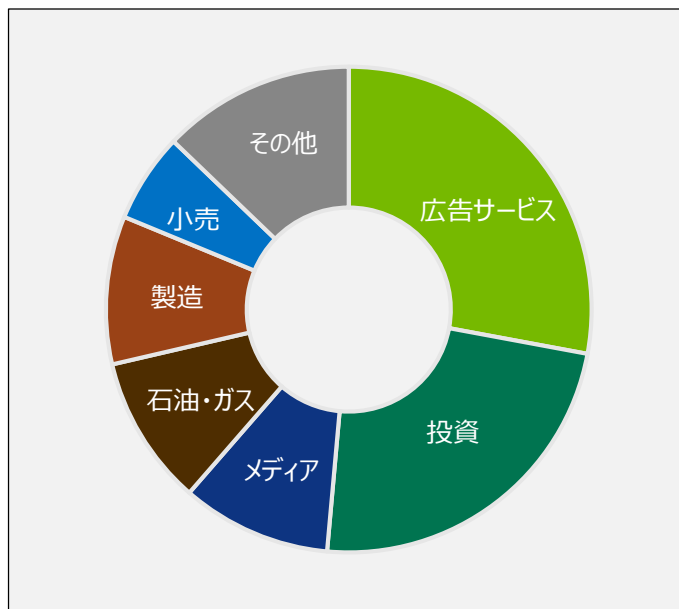
検索の50%は音声か画像になるだろう
Andrew Ng, Baidu

ディープラーニングが創出する市場価値

今後10年で50兆円の市場とされている



セグメント毎のディープラーニングの売上



産業別ディープラーニングソフトウェアの売上



IBM コグニティブビジネスは200兆円市場

ディープラーニングを加速する3つの要因

The GPU is the workhorse of modern A.I

GPUの処理能力の向上が必要不可欠



DNN



ビッグデータ



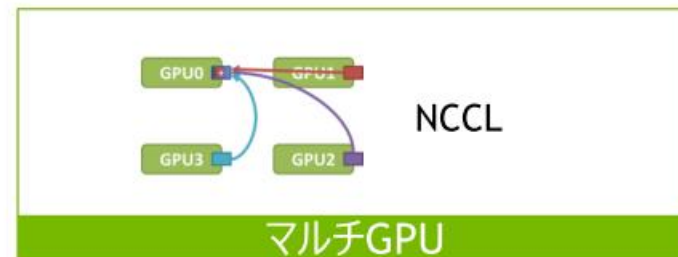
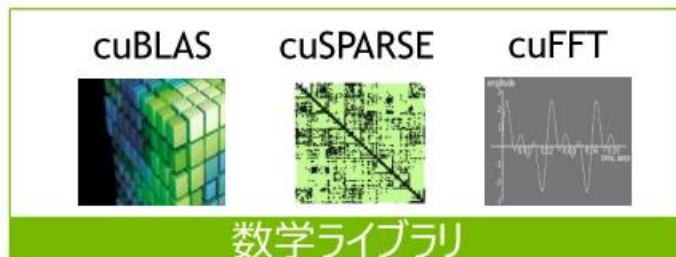
GPU

NVIDIA DeepLearning SDK

GPUで高速化されたディープラーニングの為の開発ツールおよびライブラリ

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、開発するための強力な開発ツールおよびライブラリ

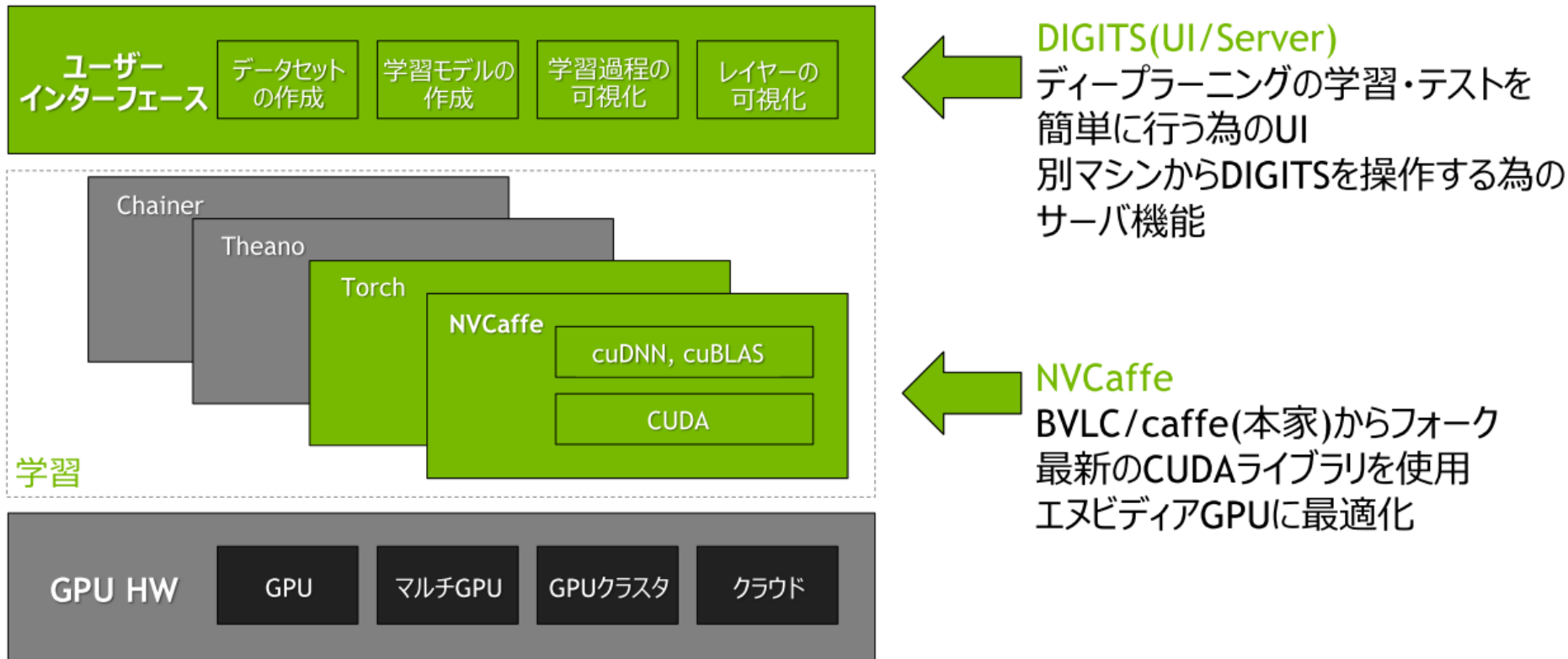
- NVIDIA GPUのための高速なディープラーニング環境を保障
- 最新のディープラーニング・アルゴリズムをGPUに最適化した状態で提供
- ディープラーニング・アプリケーションを簡単に構築



NVIDIA DIGITS

ディープラーニングSDK

WebベースのディープラーニングGPUトレーニングシステム



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

Job Information

Job Directory
/home/michael/.digits/jobs/20150311-171431-e0d8

Parse Folder (train/val)

Folder
http://sqf/data/images/voc_cropped/

Number of categories
20

Training images
26759

Validation images
8917 (25.0%)

Create DB (train)

Input file
train.txt

DB Entries
26759

モデルの作成

Select Dataset

PASCAL VOC
ILSVRC 2012
MNIST Dataset

Solver Options

Training epochs
30

Validation interval (in epochs)
1

Batch size
100

Base Learning Rate
0.01

Custom Network

```
layer {  
  name: "conv1"  
  type: "Convolution"  
  bottom: "data"  
  top: "conv1"  
  param {  
    k: mult: 1  
    decay_mult: 1  
  }  
}
```

Model Name
ImageNet

学習過程の可視化

Solver

Network (train/val)
train_val.prototxt

Network (deploy)
deploy.prototxt

Dataset

voc_cropped@256x256

Done Wed Mar 11, 05:16:57 PM

Image Size
256x256

Image Type
COLOR

Create DB (train)
26759 images

Create DB (val)
8917 images

Loss (train) Loss (val) Accuracy (val)

Epoch	Loss (train)	Loss (val)	Accuracy (val)
0.0	4.0	4.0	0
1.0	2.5	2.5	20
2.0	2.0	2.0	30
3.0	1.8	1.8	35
4.0	1.7	1.7	38
5.0	1.6	1.6	40
6.0	1.5	1.5	42
7.0	1.4	1.4	43
8.0	1.3	1.3	44
9.0	1.2	1.2	45
10.0	1.1	1.1	46

モデルのテスト

Predictions

8
3
0
6
4

Layer Activations

conv1

pool1



Networld

Thank you