GPU TECHNOLOGY CONFERENCE

GTC 2016 ディープラーニング最新情報

エヌビディア合同会社 エンタープライズビジネス事業部
シニアマネージャー　井﨑 武士

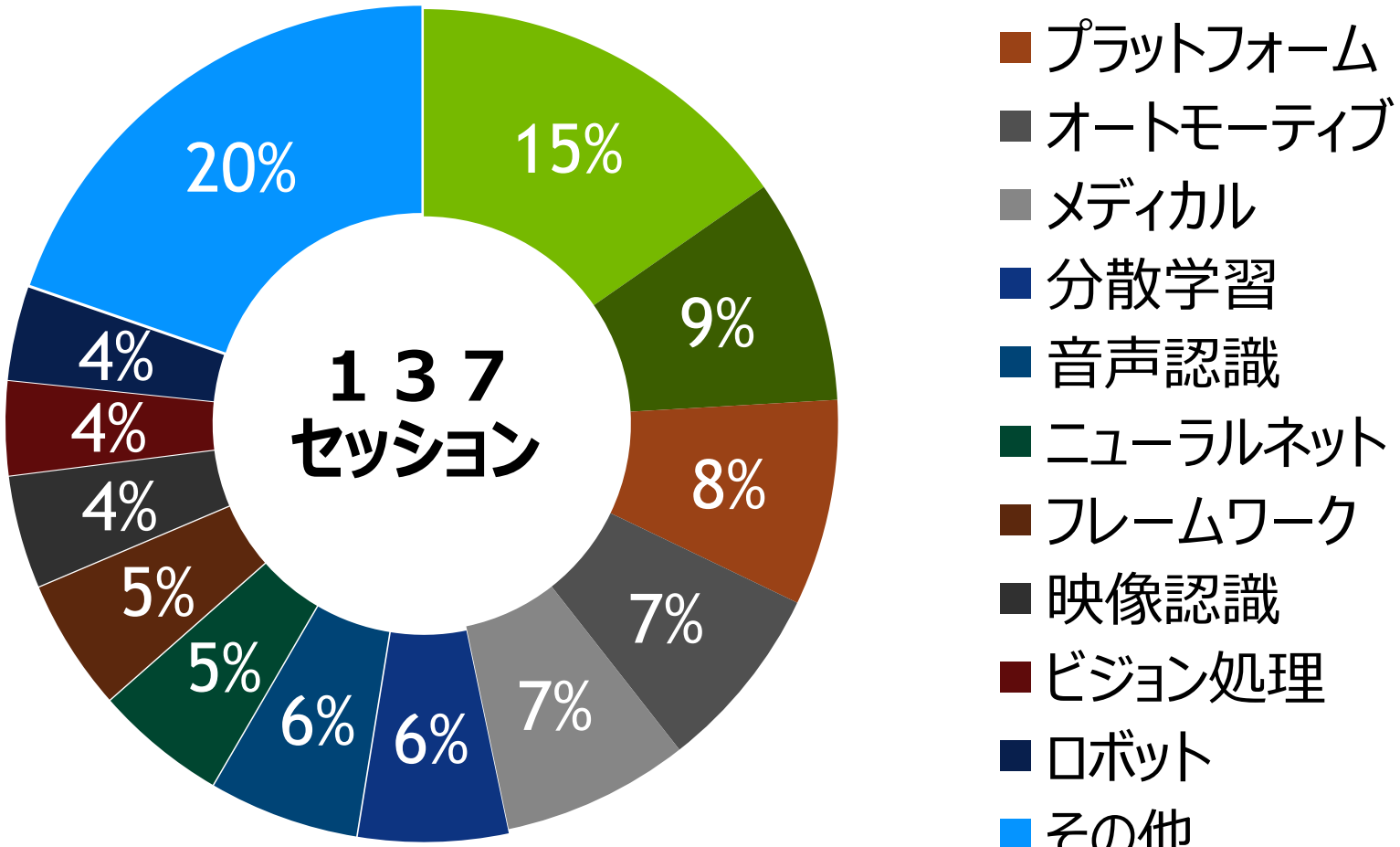
PRESENTED BY
NVIDIA.

DEEP LEARNING関連セッション

- 合計826セッション中166セッション

種類	件数
講演	116
パネル・ポスター	21
ハンズオン	17
ハングアウト	8
チュートリアル	4

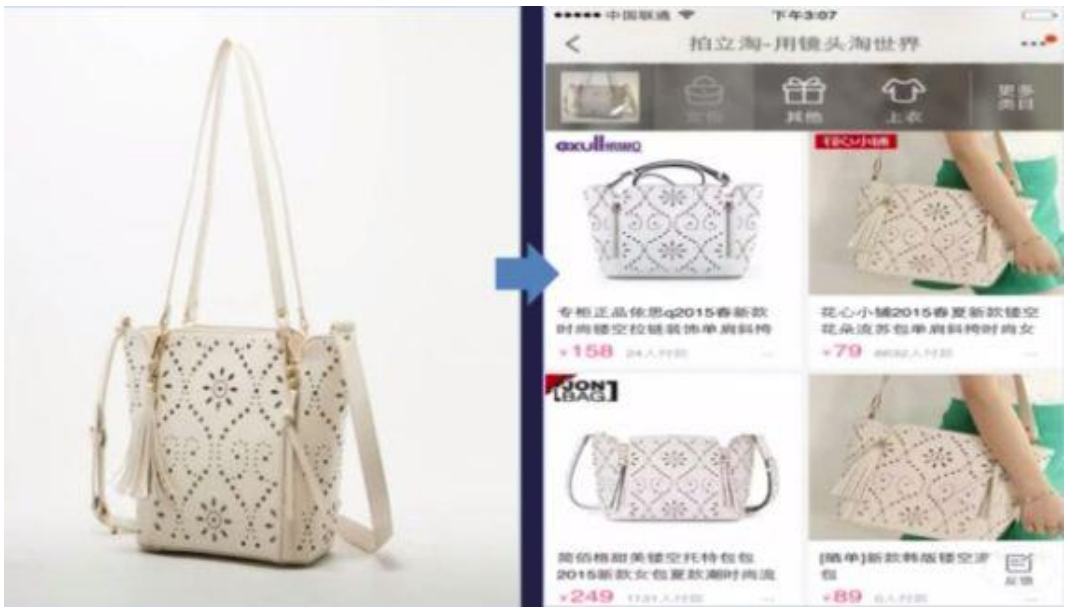
講演・パネル・ポスターセッション



DEEP LEARNING IN REAL-WORLD LARGE-SCALE IMAGE SEARCH AND RECOGNITION

Xian-Sheng Hua Senior Director/Researcher, Alibaba Group

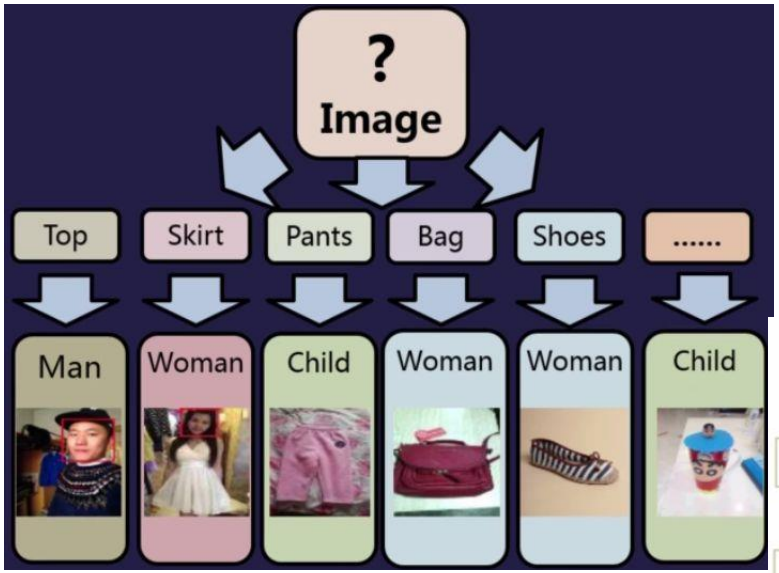
商品認識と検索


拍立淘-用镜头淘世界
其他
上衣
¥158
¥79
¥249
¥89

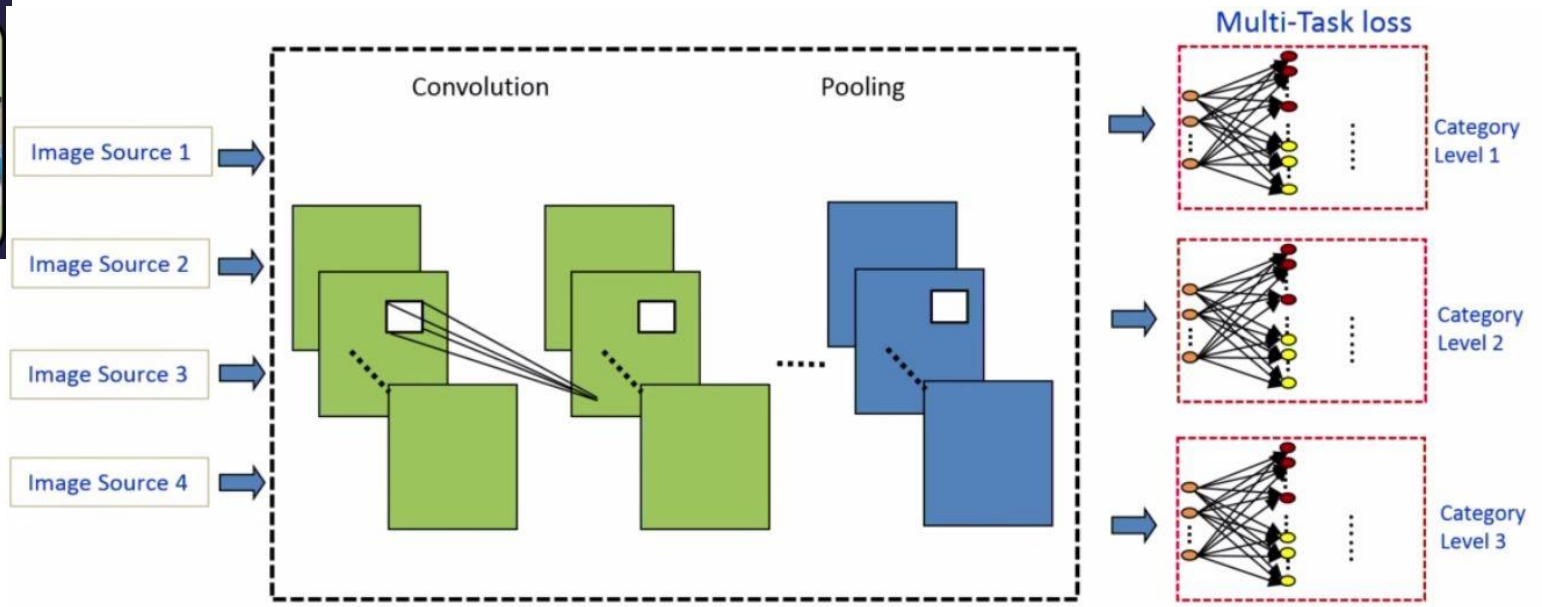

宝贝详情
图文详情
尺码参数
已经下架
拍立淘-用镜头淘世界
其他
上衣
PUMP FURY
¥459
¥151.20

特徴抽出は難しい

カテゴリ分類



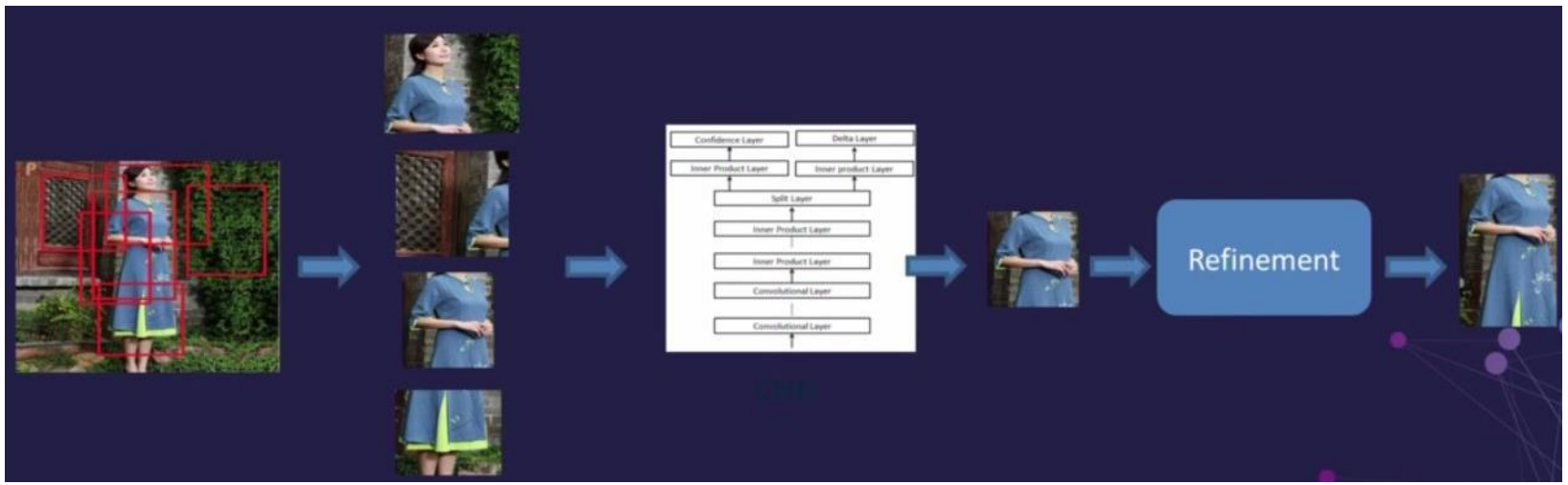
Level1：60+
Level2：1200+
Leaf：10000+



オブジェクト検出

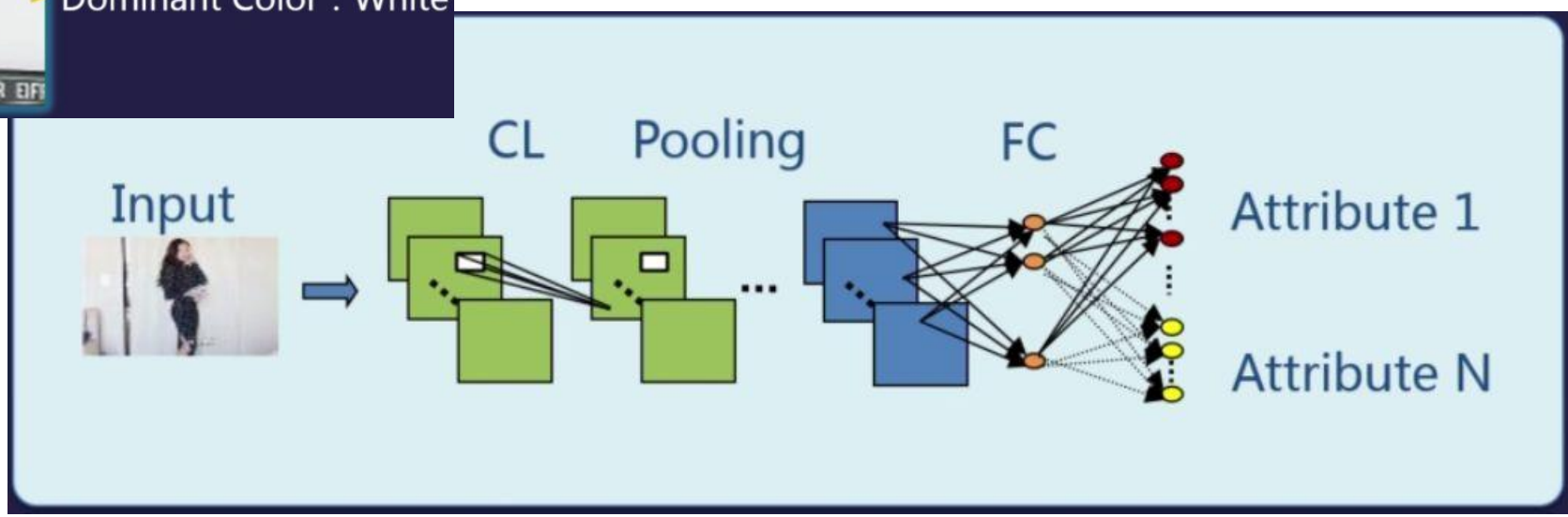
正確なバンディングボックス

背景の写りこみ
小さいオブジェクト



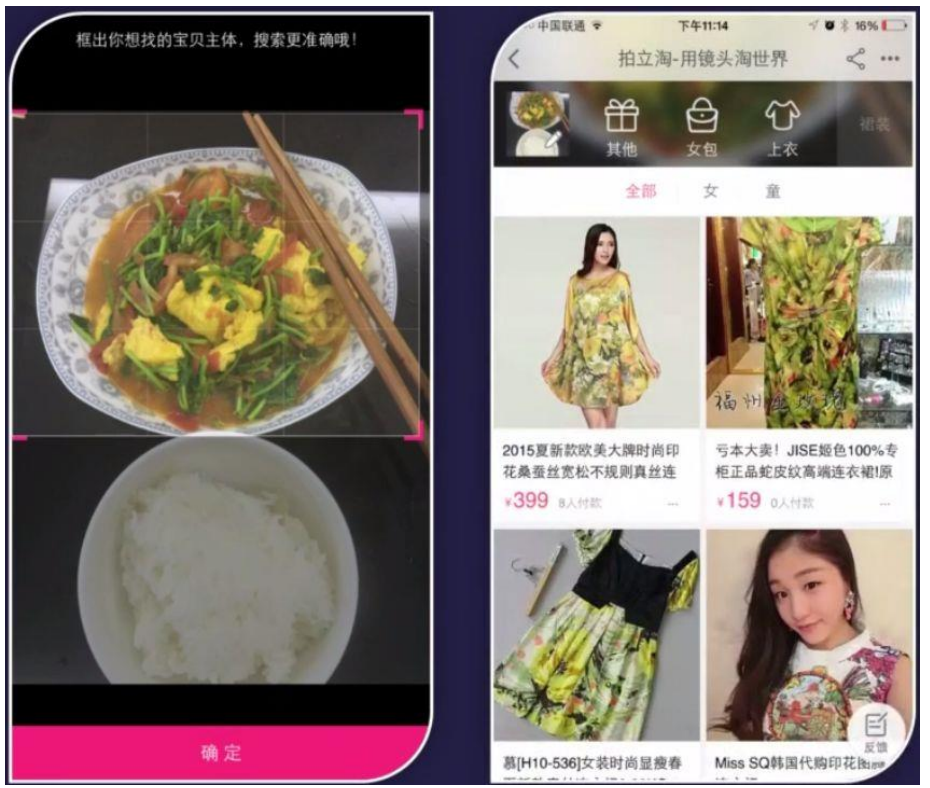
屬性分類


Collar Style : Round
Sleeve: Short
Pattern: Cartoon
Dominant Color : White

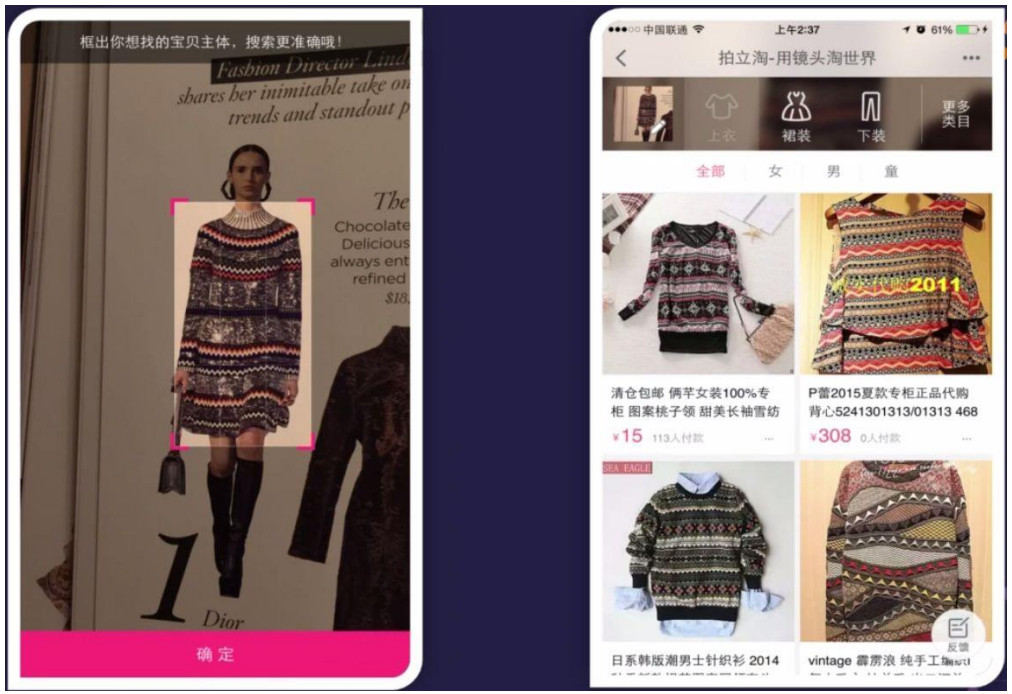

Input
CL
Pooling
FC
Attribute 1
Attribute N

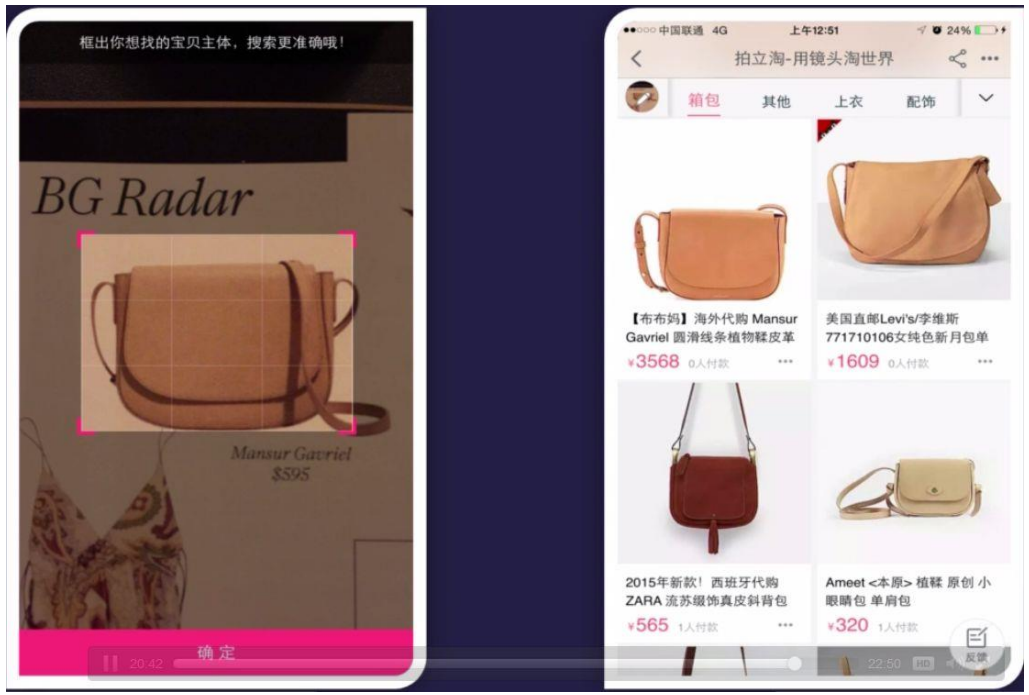
類似デザイン検索


框出你想找的宝贝主体，搜索更准确哦！
确 定
拍立淘-用镜头淘世界
新款预宽～包邮～
¥999 0人付款
欧美 夏季连衣裙露肩 荷叶边喇叭袖 A字裙宽松腰大码清
¥39.90 0人付款
2015夏季新款海滩波西米亚风雪纺长裙
¥55 0人付款
乔兰朵2015夏露肩荷叶边花朵短袖A字裙大摆裙漏肩大花
¥139 101人付款
童装2015新款女童中国风牡丹花布纱连衣裙子民族六一
¥60 0人付款
大花亮色连衣花裙子
¥78 1人付款
反馈
顶部


框出你想找的宝贝主体，搜索更准确哦！
确 定
拍立淘-用镜头淘世界
其他
女包
上衣
全部
女
童
2015夏新款欧美大牌时尚印花桑蚕丝宽松不规则真丝连
¥399 8人付款
亏本大卖！JISE姬色100%专柜正品鳄皮纹高端连衣裙!原
¥159 0人付款
慕[H10-536]女装时尚显瘦春
Miss SQ韩国代购印花
反馈

同一商品、類似品検索


框出你想找的宝贝主体，搜索更准确哦！
确 定
拍立淘-用镜头淘世界
上衣
裙装
下装
更多类目
全部
女
男
童
清仓包邮 俩芊女装100%专柜 图案桃子领 甜美长袖雪纺
¥15 113人付款
P蕾2015夏款专柜正品代购 背心5241301313/01313 468
¥308 0人付款
日系韩版潮男士针织衫 2014
vintage 霹雳浪 纯手工


框出你想找的宝贝主体，搜索更准确哦！
BG Radar
Mansur Gavriel
$595
确 定
拍立淘-用镜头淘世界
箱包
其他
上衣
配饰
【布布妈】海外代购 Mansur Gavriel 圆滑线条植物鞣皮革
¥3568 0人付款
美国直邮Levi's/李维斯 771710106女纯色新月包单
¥1609 0人付款
2015年新款！西班牙代购 ZARA 流苏缀饰真皮斜背包
¥565 1人付款
Ameet <本原> 植鞣 原创 小眼睛包 单肩包
¥320 1人付款

DEEPFONT: FONT RECOGNITION AND SIMILARITY BASED ON DEEP LEARNING

Hailin Jin *Principal Scientist, Adobe*

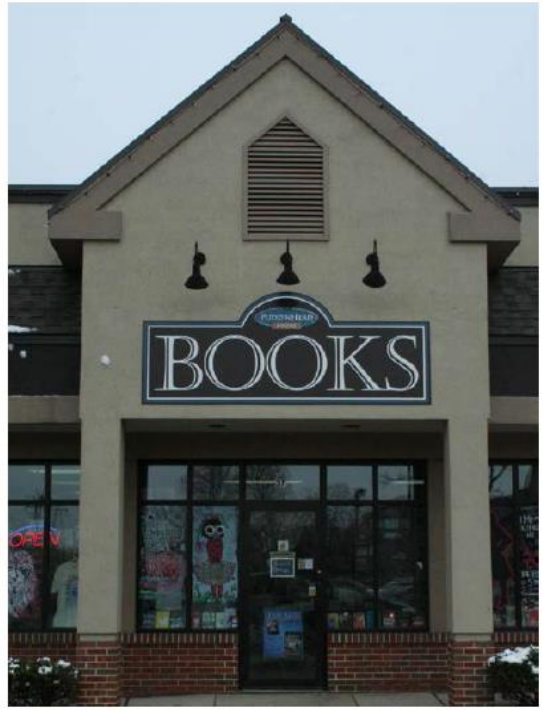
Deep Font:

フォントの認識

フォントの類似性

問題点

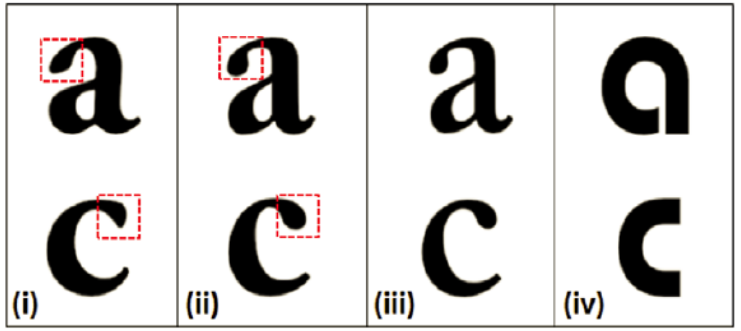
- 効果的に使用されているか
- デザインが独創的か
- テキスト/イメージ編集での活用



THE QUICK BROWN FOX	JAF Mashine Extra Light
THE QUICK BROWN FOX	JAF Mashine Rounded Extra Light
THE QUICK BROWN FOX	Restore Light
THE QUICK BROWN FOX	JAF Mashine Light
THE QUICK BROWN FOX	Pirulen Light
THE QUICK BROWN FOX	Ethnocentric ExtraLight
THE QUICK BROWN FOX	JAF Mashine Rounded Light
THE QUICK BROWN FOX	Aviano Sans Light
THE QUICK BROWN FOX	Aviano Future Regular
THE QUICK BROWN FOX	Good Times Light

課題

フォントの種類は莫大
　分かっているだけで10万フォント
認識の難しさ
　微妙なデザインの違い



実際の社会における学習データを集めるのが極めて難しい
　学習データとテストデータが異なる
　人工の学習データを作る必要がある

Deep Font

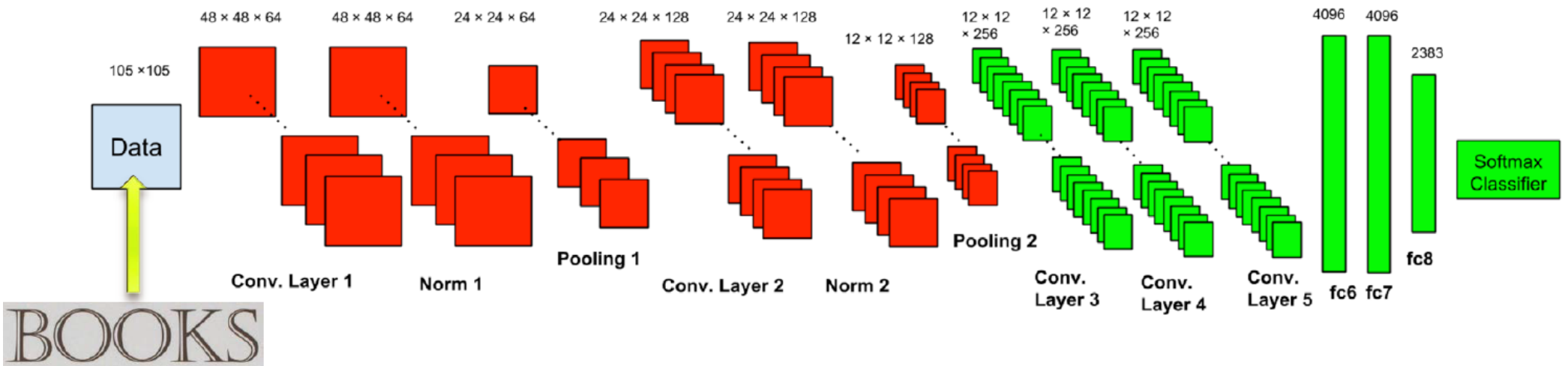
Deep Learning-　CNN

　大量のデータを処理するのに効果的

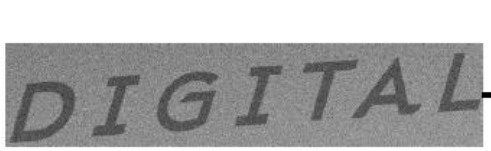
　きめ細かい認識に効果的

OCRの必要が無い

End to End学習



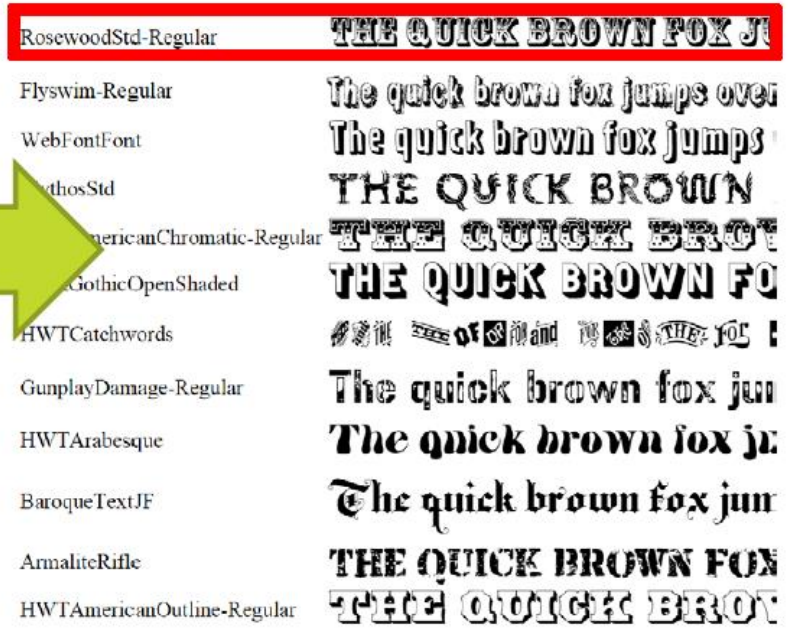
DeepFontのシステム



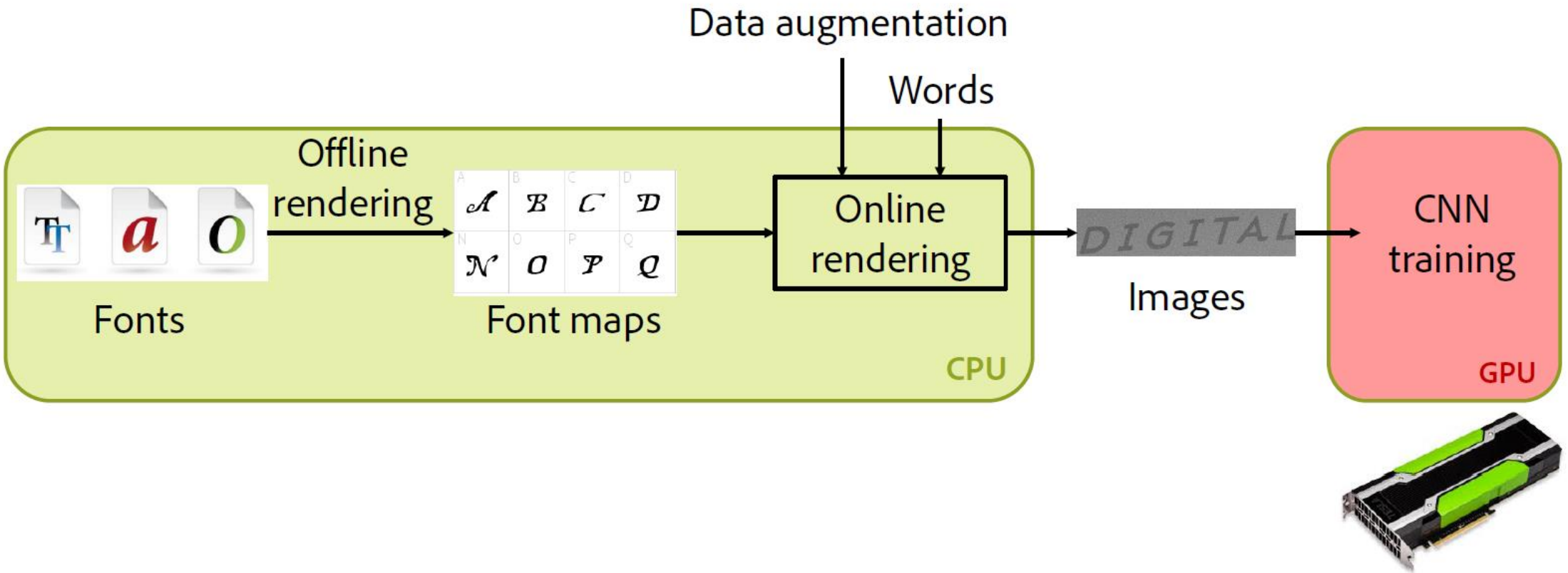
Localization Network

Recognition Network

#	weight	predicted font name	sample
1	0.927	ParkwayHotel	DIGITAL
2	0.018	Adrianna-DemiboldItalic	DIGITAL
3	0.015	Cantarell-BoldOblique	DIGITAL
4	0.014	LeHavreBoldItalic	DIGITAL
5	0.005	AdriannaCondensed-DemiBoldItalic	DIGITAL



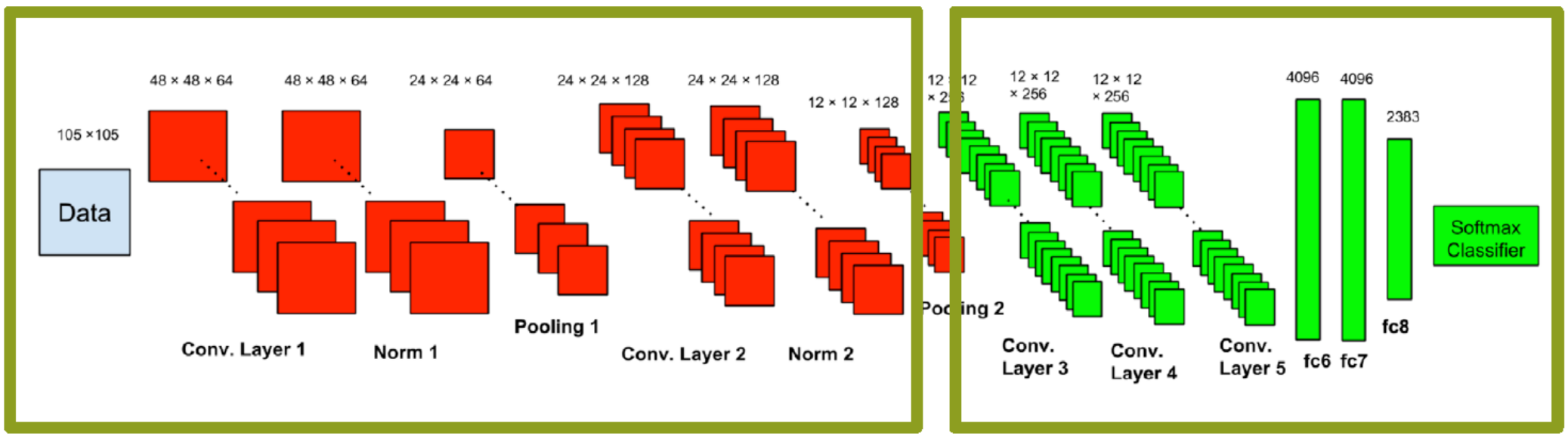
DeepFontの学習



データオーグメンテーション

ノイズ
ぼんやりとさせる
変形
影
文字の空き具合
アスペクト比

ネットワーク・デコンポジション



Unsupervised cross-domain variable layers

Supervised domain-specific

結果

Font Forumでの4383の実在のイメージでテスト

Model	Augmentation	Decomposition	Top-1 accuracy	Top-5 accuracy
LFE (CVPR'04)	Y	N/A	42.6%	60.3%
DeepFont	N	N	42.5%	49.2%
DeepFont	Y	N	66.7%	79.2%
DeepFont	Y	Y	**71.4%**	**81.8%**

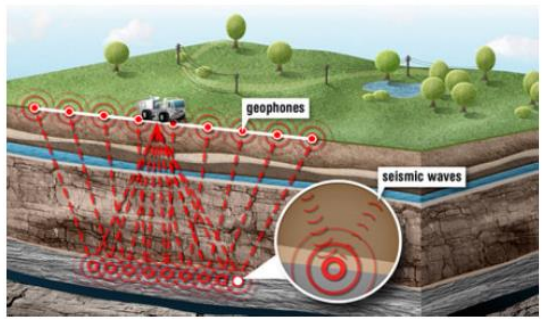
AUTOMATED GEOPHYSICAL FEATURE DETECTION WITH DEEP LEARNING

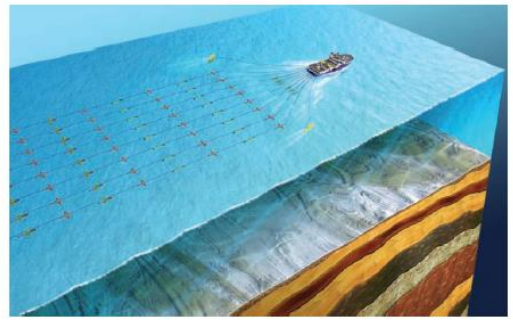
Chiyuan Zhang PhD Student, MIT

弾性波探査

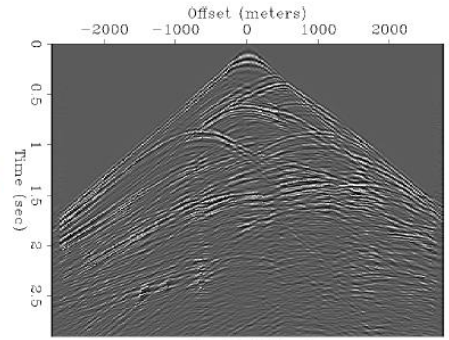
探査段階：弾性波データは石油・ガス産業で非常に重要。深層にある石油を見つけるために使用され石油・ガス探査における様々なフェーズで初期および発掘時に現場の特徴づけに使用される

Data acquisition, on/off shore.

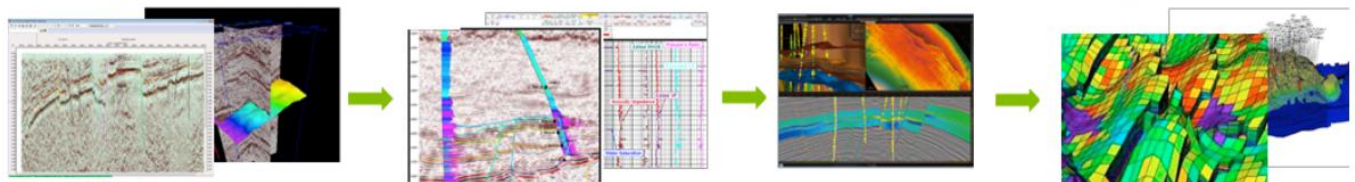




Seismic traces
waveforms (time series) indexed
by shot id and receiver id



Data processing: iterations could take multiple months with human experts.



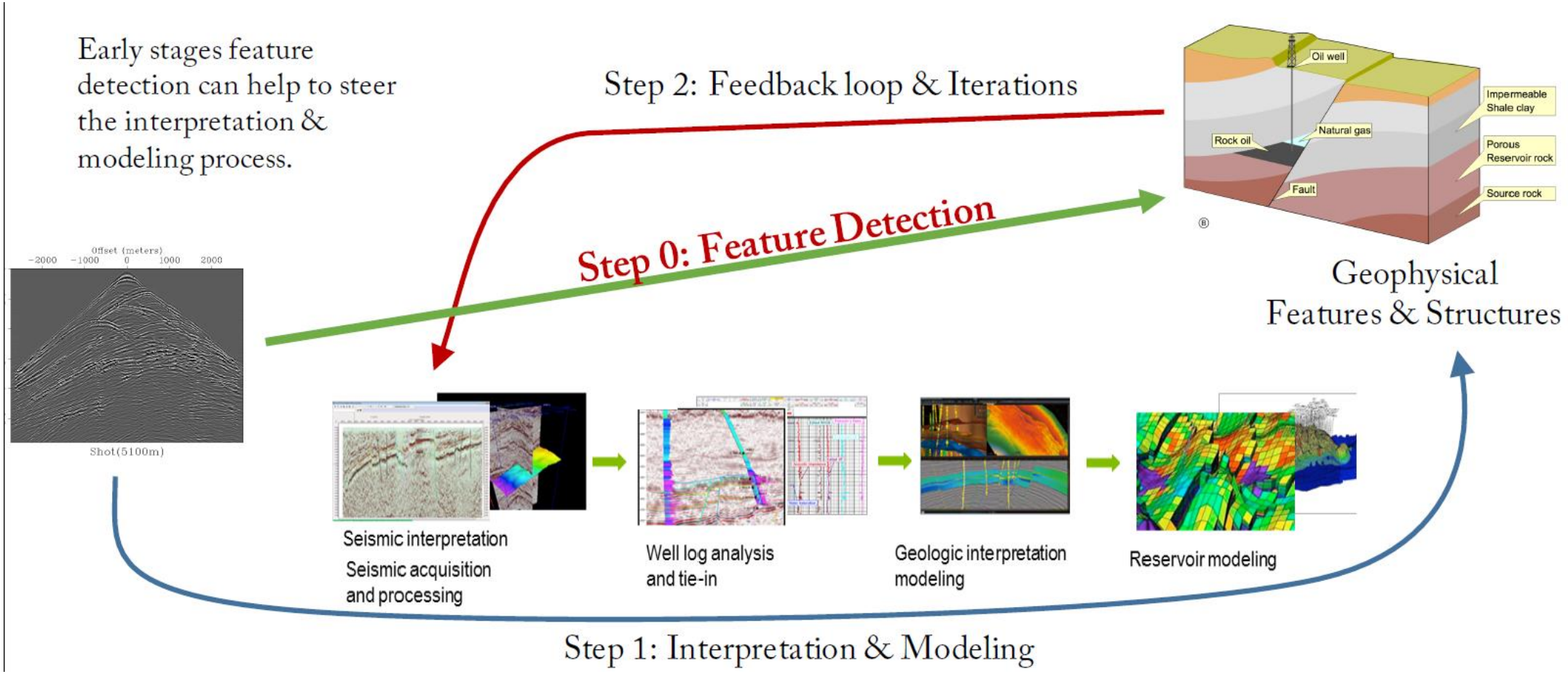
Seismic interpretation
Seismic acquisition
and processing

Well log analysis
and tie-in

Geologic interpretation
modeling

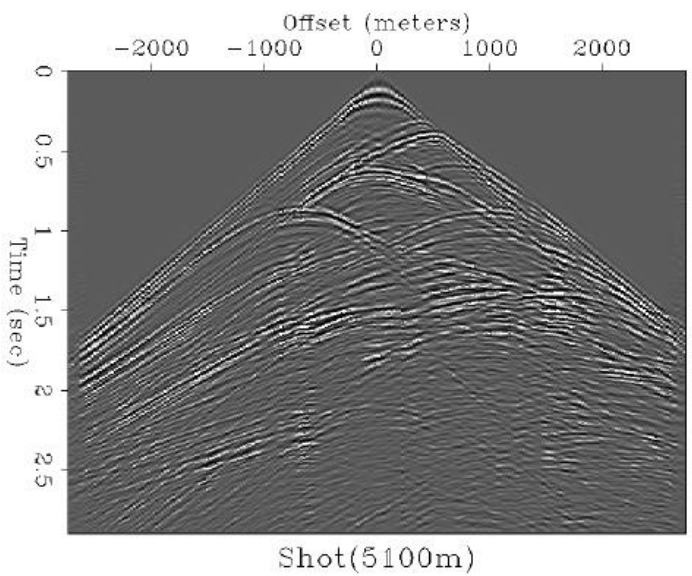
Reservoir modeling

地球物理学的特徴検出の自動化

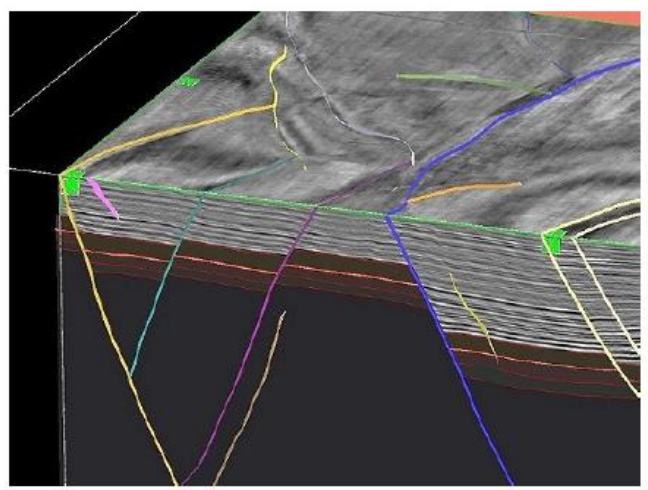


地球物理学的特徴検出の自動化

Seismic Survey



Machine Learning

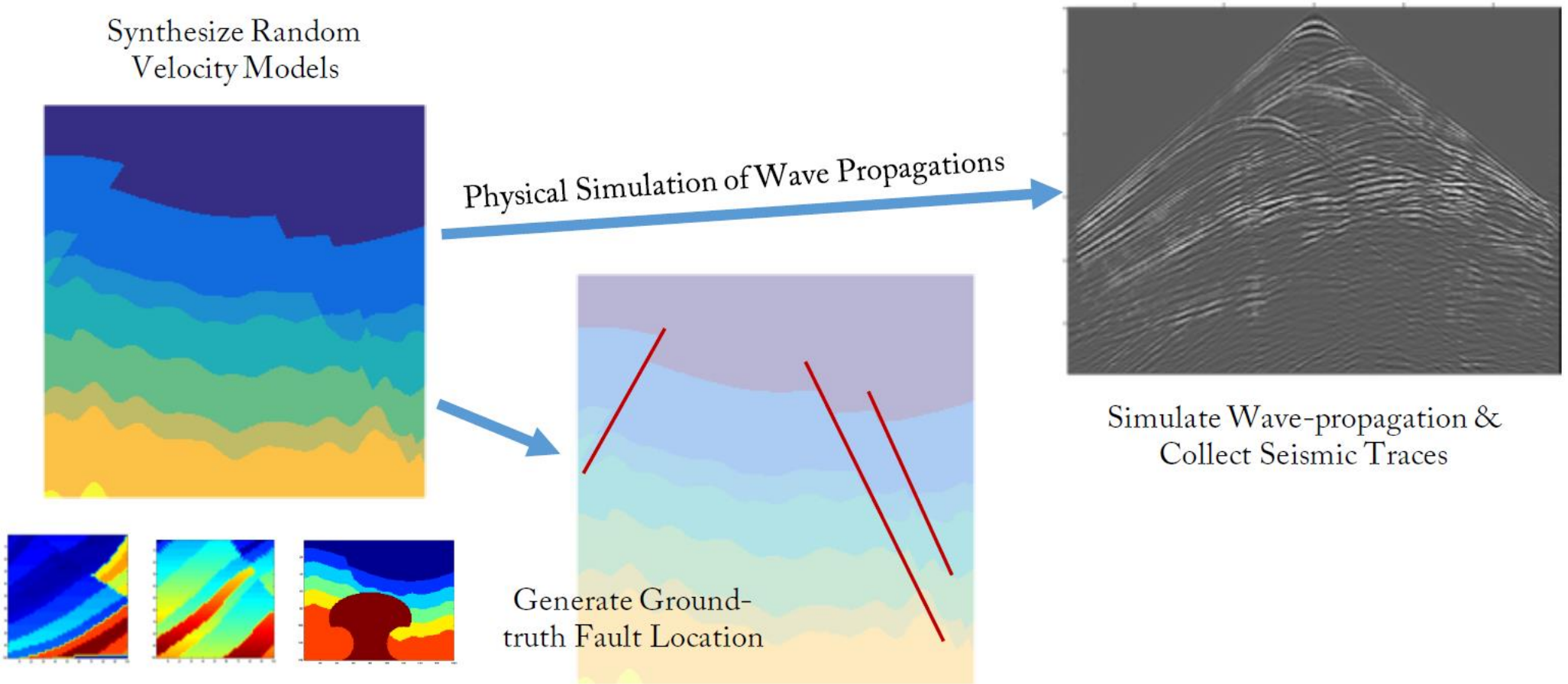


From raw seismic traces, discover (**classification**) and locate (**structured prediction**) faults in the underground structure, before running migration / interpretation.

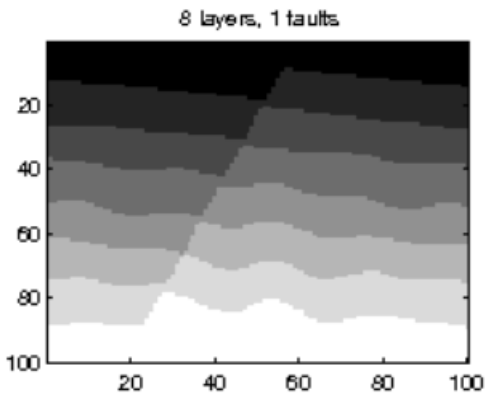
機械学習による断層の検知

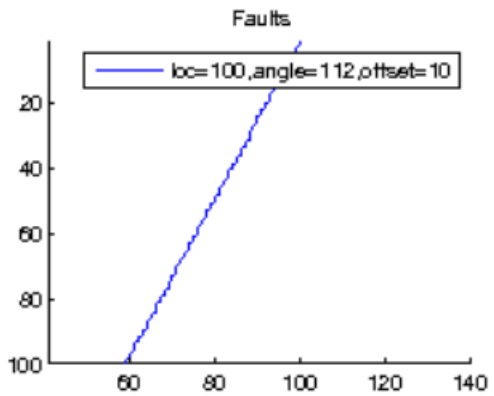
Challenge	Solution
❶ Unlike simple classification, the output space is structured.	Wasserstein-loss based structured output learning.
❷ The mapping from traces to location of faults is a very complex nonlinear function.	Using deep neural networks for modeling.
❸ DNNs need a lot of training data.	Generate random synthesized training data (geological/geophysical model design + physical simulation + generative probabilistic modeling)
❹ Computational issue.	Julia + GPU computation with NVidia CUDA.

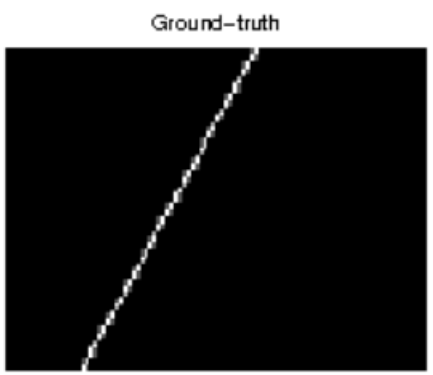
学習データの合成

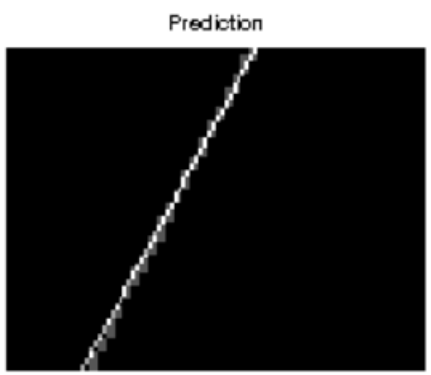

Synthesize Random
Velocity Models
Physical Simulation of Wave Propagations
Simulate Wave-propagation &
Collect Seismic Traces
Generate Ground-
truth Fault Location

結果：プロット　単一断層







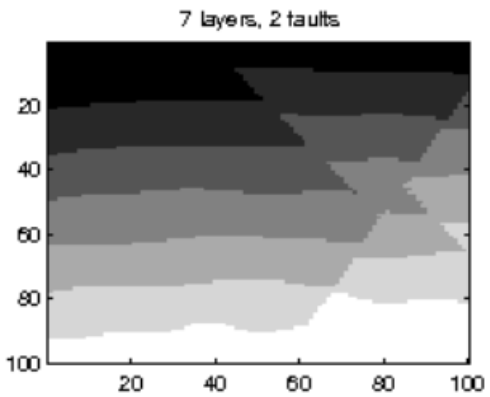


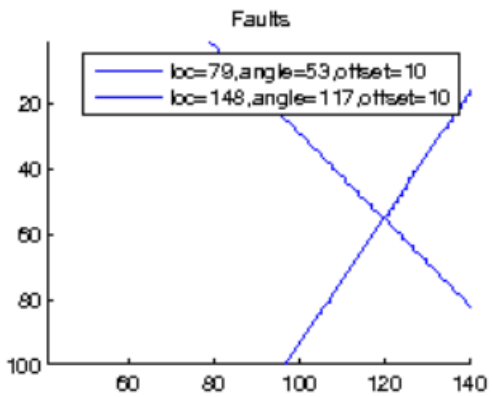
Test case: 10k models, 510k traces, SGD 250k iterations. No noise, **1 fault**, no salt body, **downsample 64**.
DNN arch: 4 layers,1024 neurons

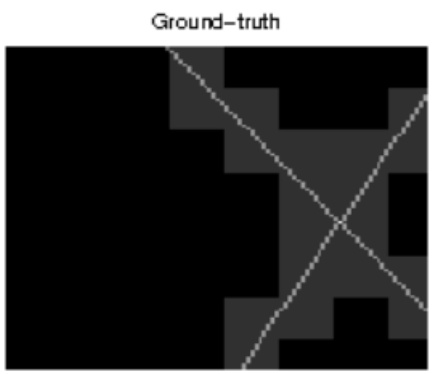
Prediction accuracy:

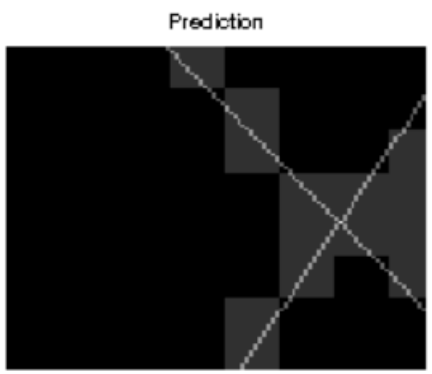
- Area under Curve (AUC): **77**%
- Intersection over Union (IOU): **71**%

結果：プロット、複数断層







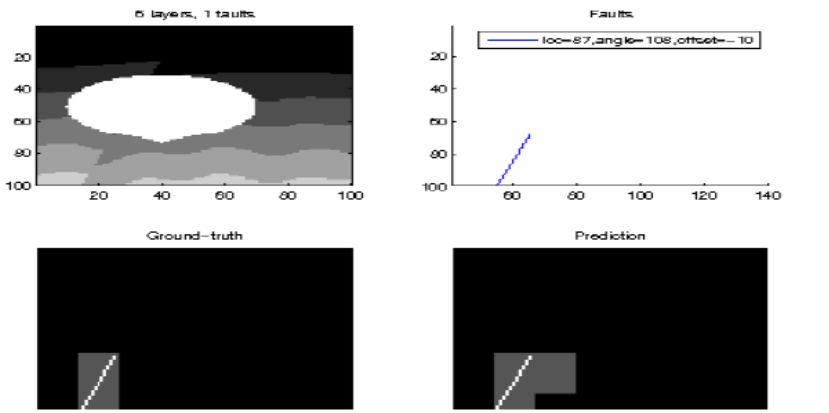


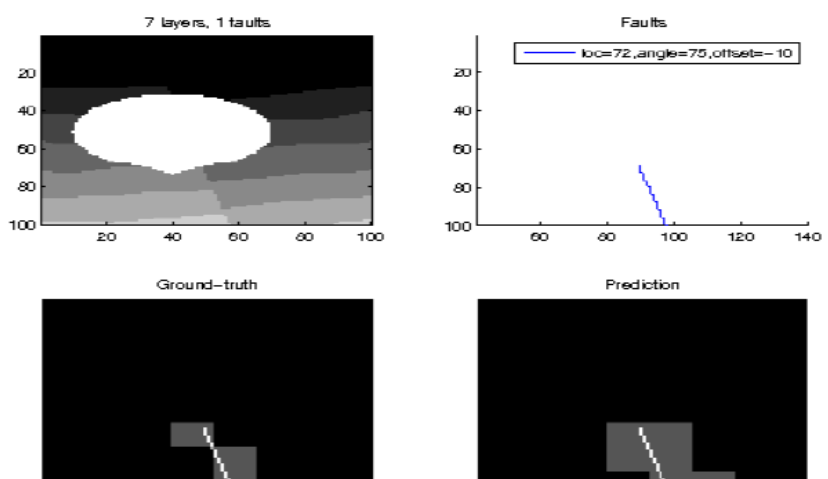
Test case: 10k models, 510k traces, SGD 250k iterations. No noise, **2 faults**, no salt body, **downsample 8**. DNN arch: 4 layer, 768 neurons

Prediction accuracy:

- Area under Curve (AUC): **86%**
- Intersection over Union (IOU): **75%**

結果：プロット、岩塩

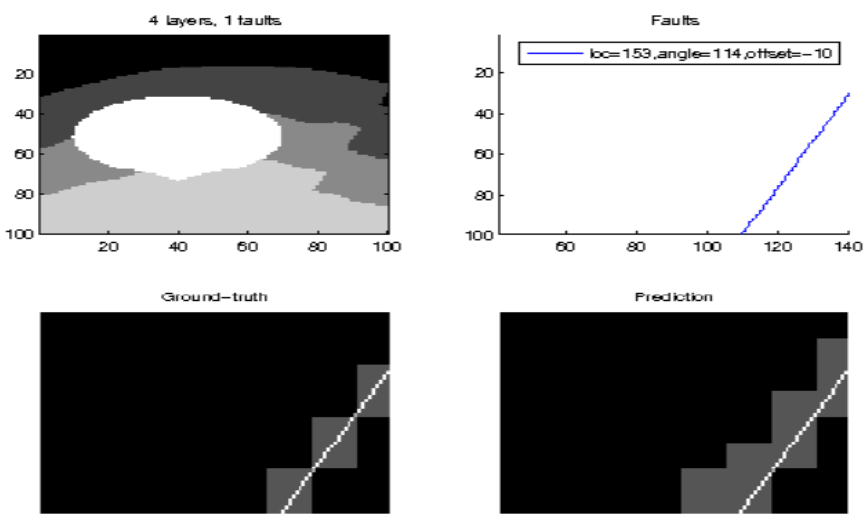




Test case: 10k models, 510k traces,
SGD 250k iterations. No noise, **1 fault**,
Salt body, **downsample 8**. DNN arch: 2, 256

Prediction accuracy:

- Area under Curve (AUC): **96**%
- Intersection over Union (IOU): **74**%



DEEP LEARNING ALGORITHMS FOR RECOGNIZING THE FEATURES OF FACIAL AGEING

Konstantin Kiselev Data Scientist, Youth Laboratories

美容：肌年齢測定から肌ケアへ

若く保つためのケア方法への探求

美容師 皮膚科医 その他の医師	部分的な意見 バイアス 一貫性が無い 時間＋お金
自己評価(鏡)	バイアス
周囲の人	部分的な意見 バイアス 一貫性が無い

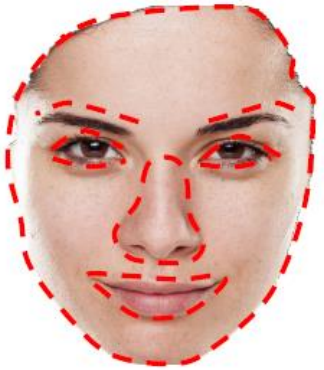
しわによる判断

従来のプロセス

Photo preprocessing

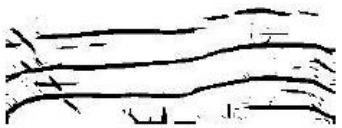
Face detection

Facial zone Alignment



Wrinkles area detection

Make wrinkles map

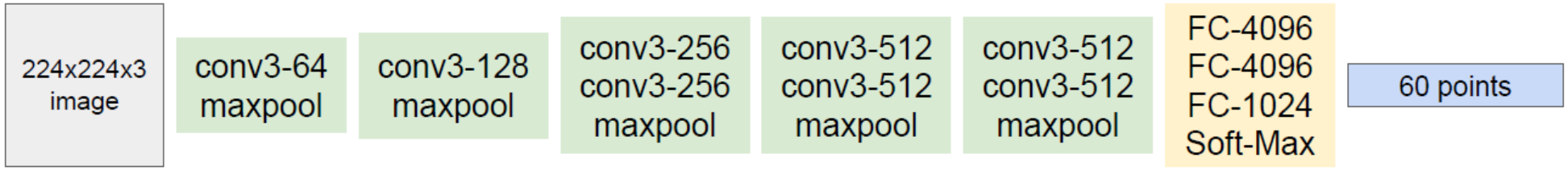


Each zone is processed separately, wrinkle-related features are detected

Calculate wrinkle score

Deep Learningによるアプローチ

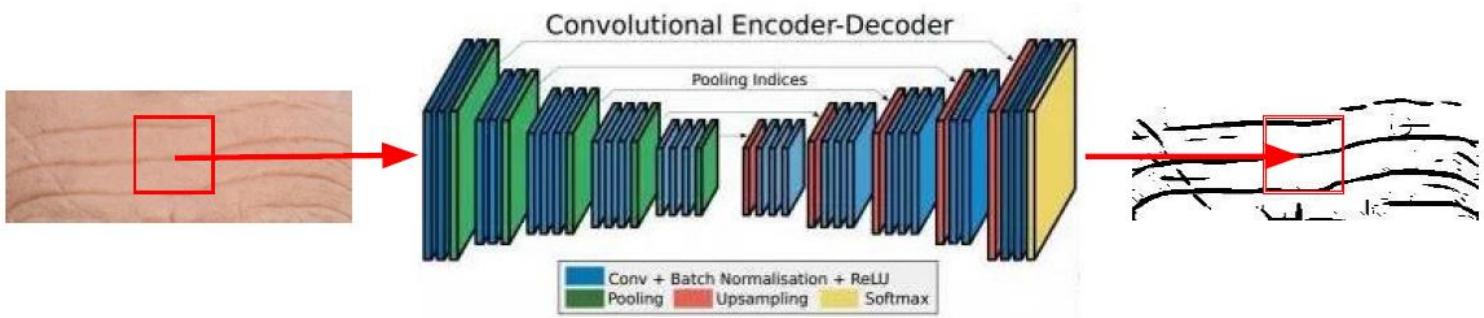
1. VGG-11 を用いて顔の領域を検出



2. しわスコアの算出

- SegNetを用いてしわマップを作成
- VGG-16を用いてRYNKLスコアを推定

*



データセットの集め方

AI判定による第1回目の国際ビューティコンテスト開催
（2015年12月1日～2016年1月18日）

約3000に上る画像（解像度２K以上）＋　情報（体重、身長、年齢、性別、人種、国）

第2回目のコンテストを2016年5月1日～開催予定

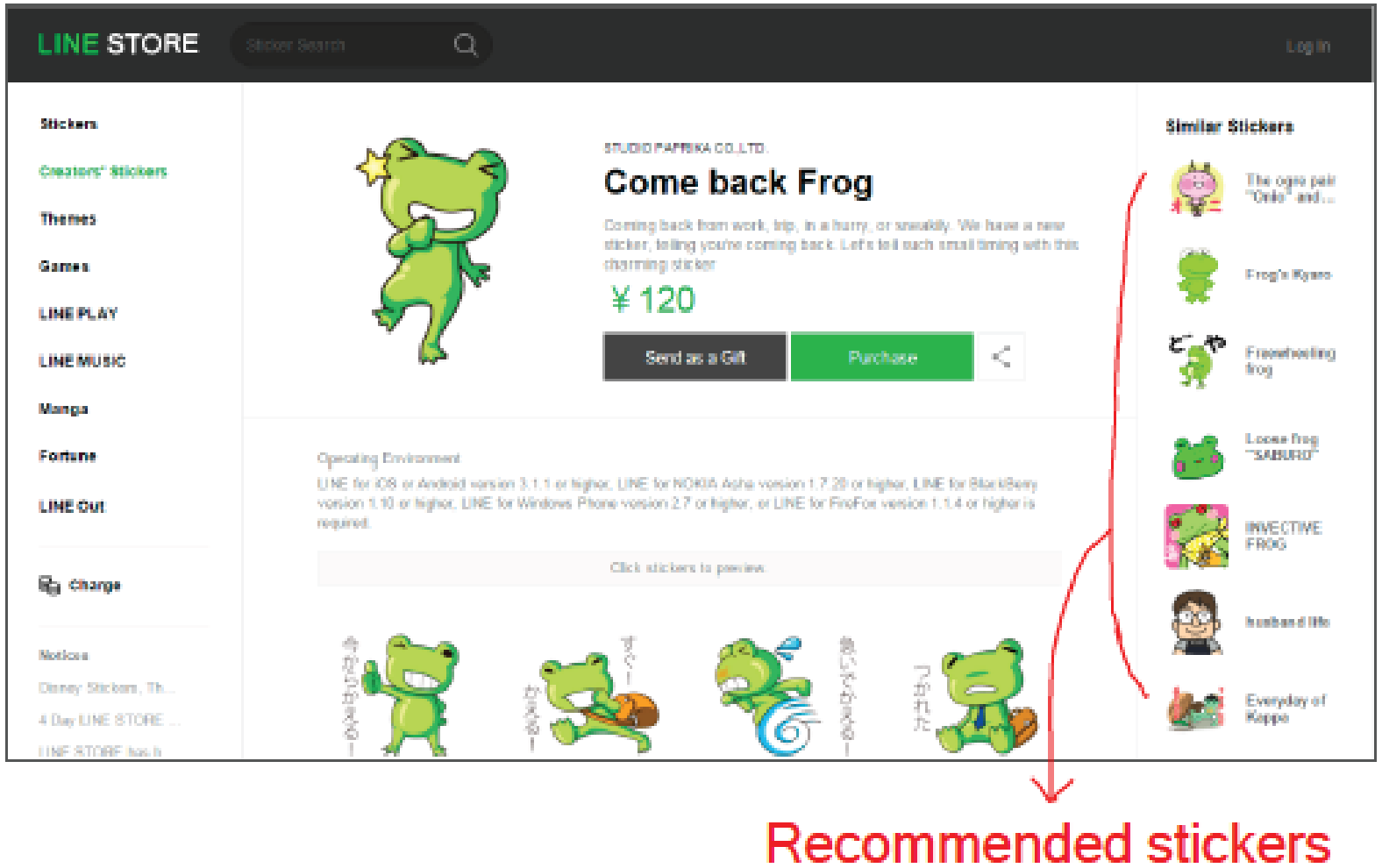
結果

平均二乗誤差：　従来手法　0.39、　Deep Learning　0.32

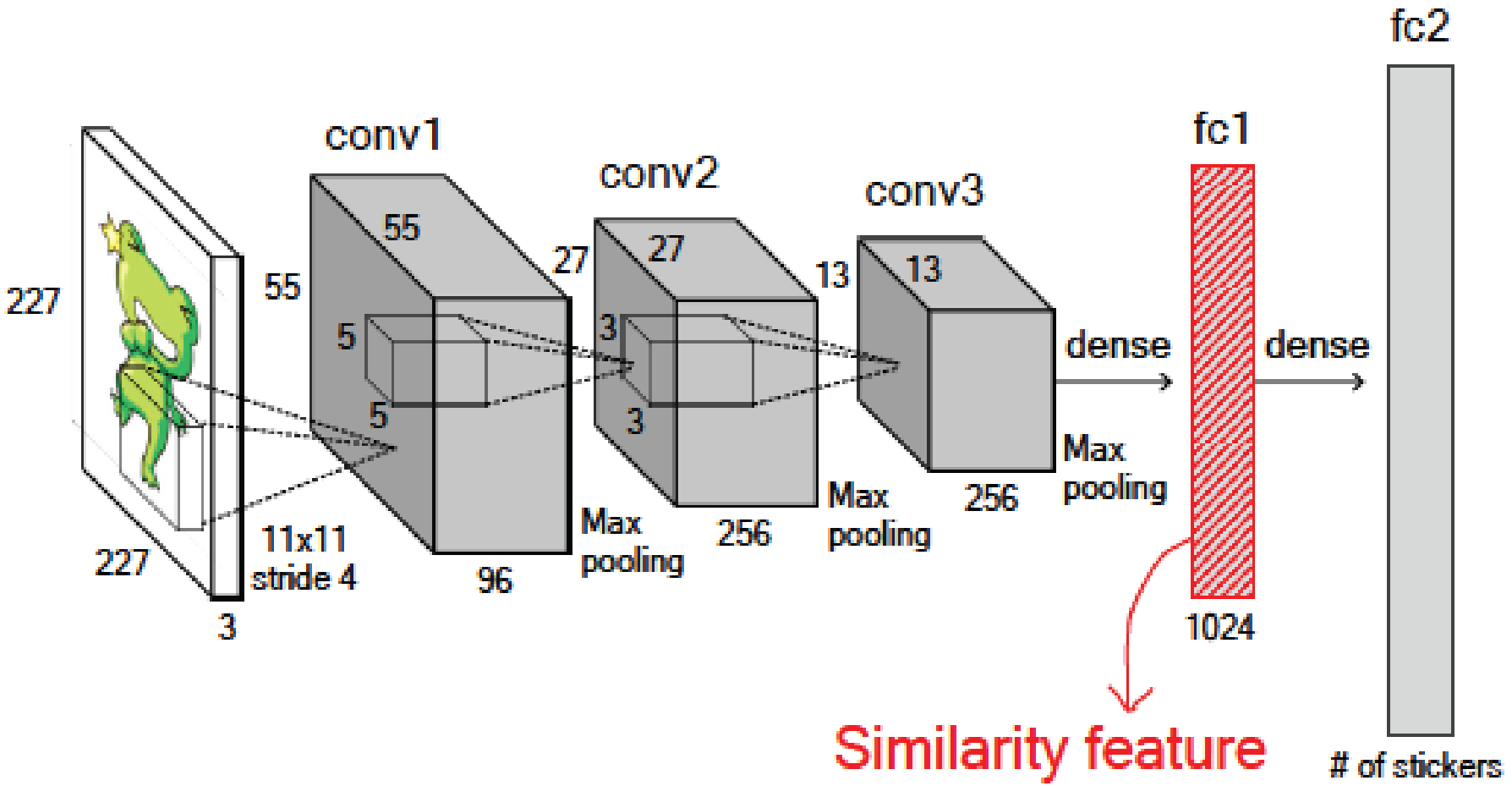
IMAGE-BASED STICKER RECOMMENDATION USING DEEP LEARNING

Jiwon Kim Senior Research Engineer, Naver Labs

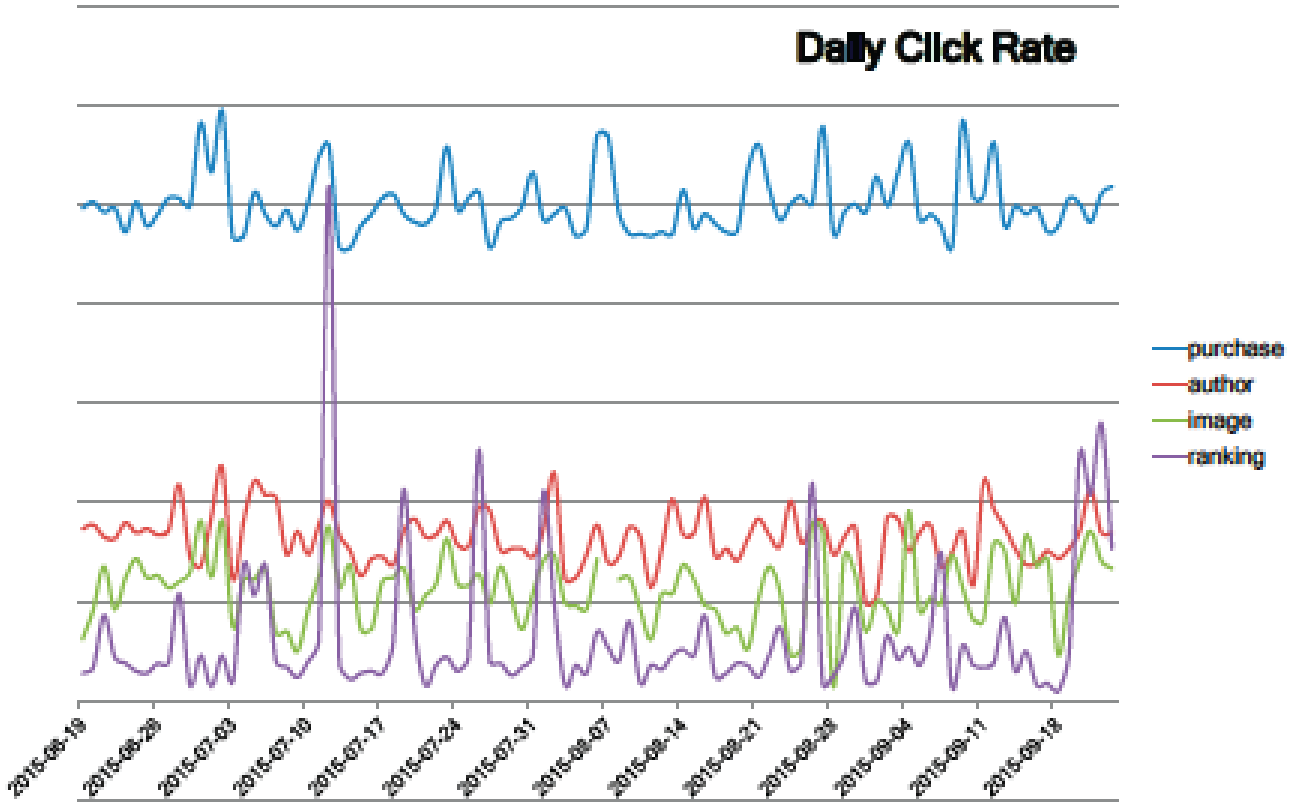
Lineスタンプのレコメンデーション



ネットワーク構成


fc2
conv1
conv2
conv3
fc1
227
55
55
27
27
13
13
5
5
3
3
dense
dense
Max
pooling
Max
pooling
Max
pooling
227
11x11
stride 4
3
96
256
256
1024
Similarity feature
# of stickers

クリック数評価



Purchase	Author	Image	Ranking
1.00	0.32	0.21	0.07

Total click rate, normalized with respect to purchase-based recommendation scheme

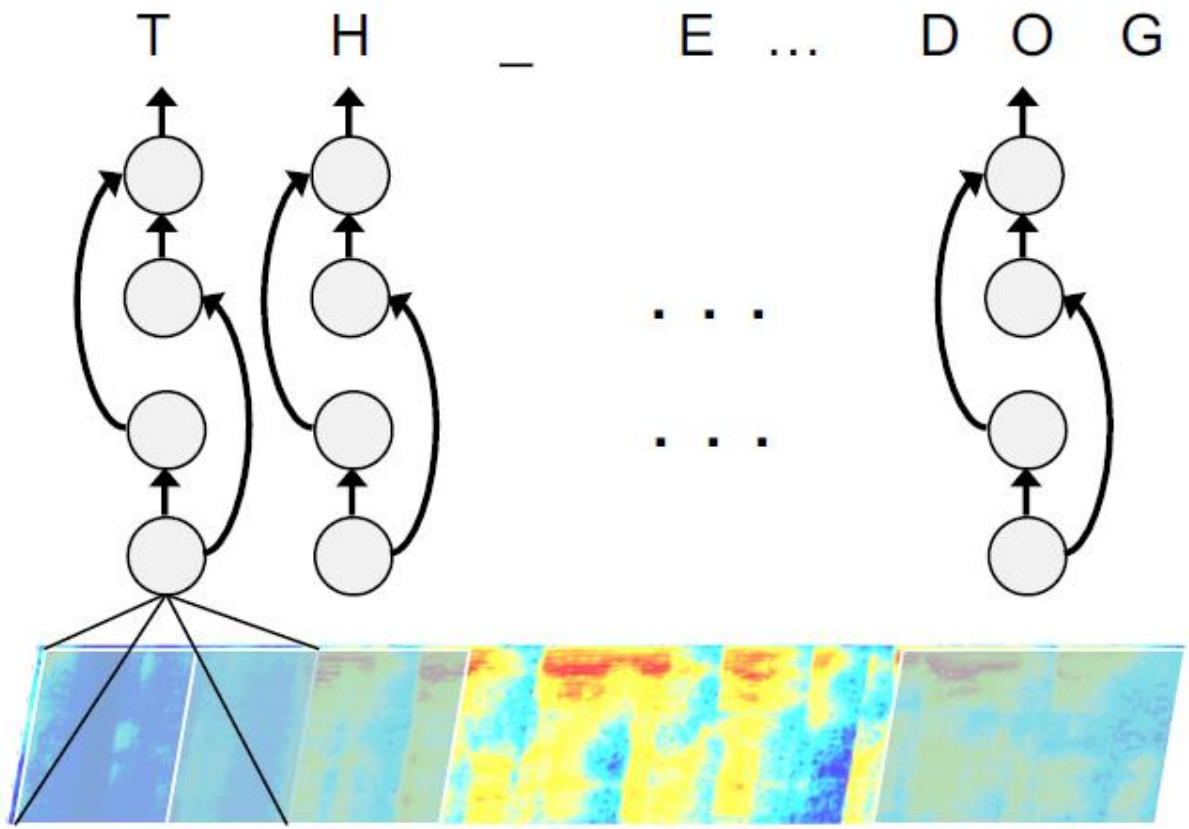
TRAINING AND DEPLOYING DEEP NEURAL NETWORKS FOR SPEECH RECOGNITION

Bryan Catanzaro Senior Researcher, Baidu Research

Deep Speech　音声認識

End to End Learning

音声から直接文字を推論するDNN



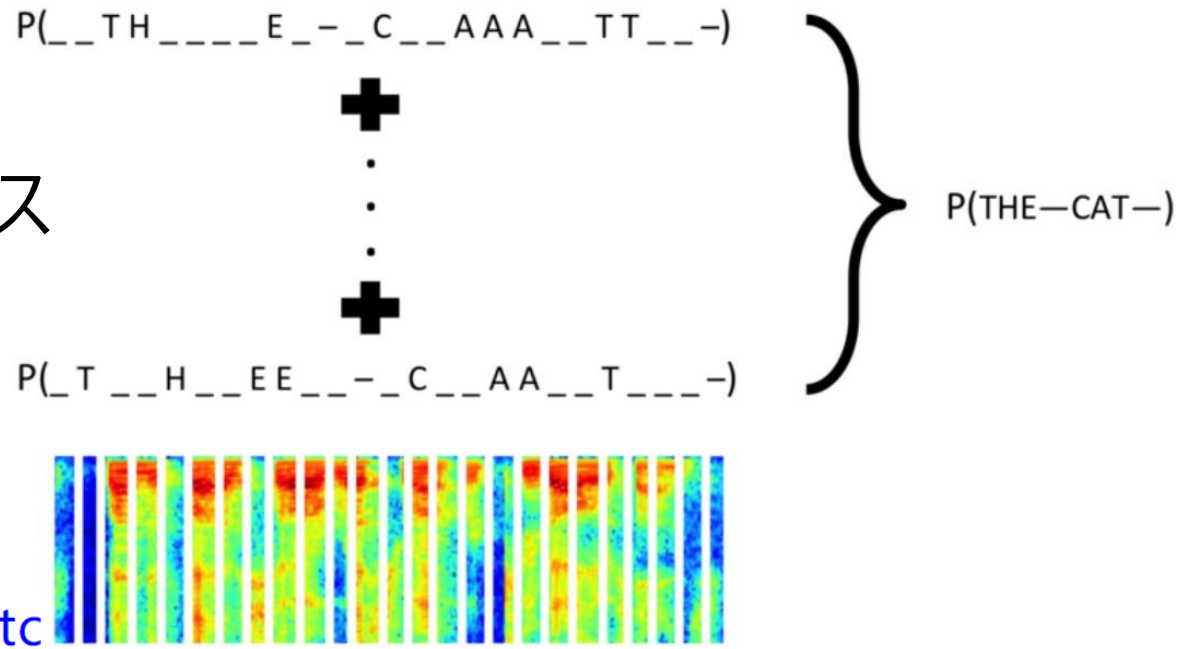
Warp-CTC

BaiduのOpen source化されたCTC実装

CPUとGPUの並列化に効果的

他の実装に比べ100~400倍高速

Apacheライセンス、Cインターフェイス



https://github.com/baidu-research/warp-ctc

Deep Speech2

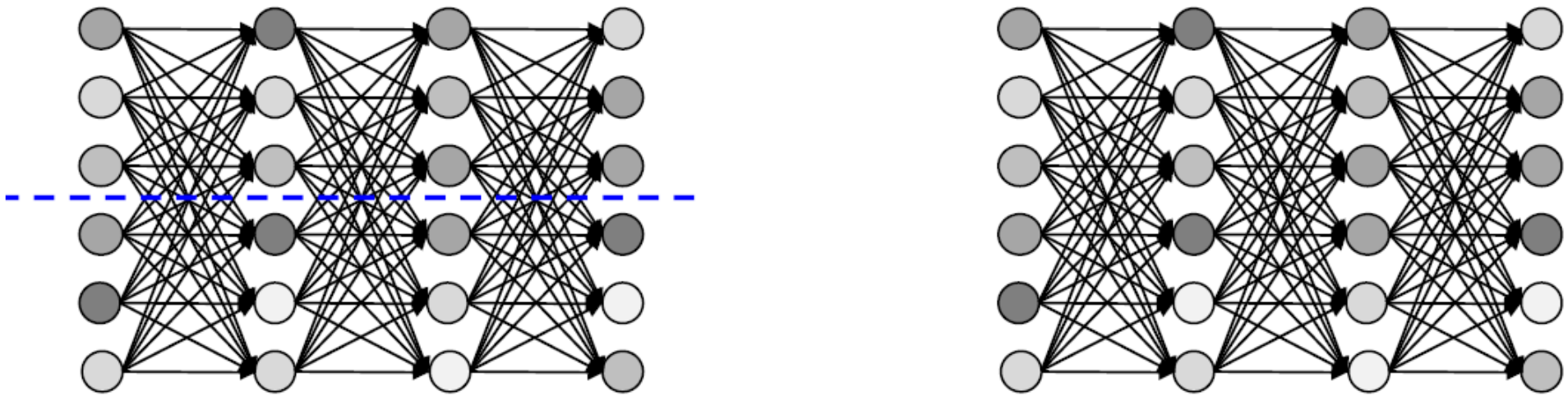
トレーニングデータ：1年半の蓄積データ（英語と北京語)

Fraction of Data	Hours	Regular Dev	Noisy Dev
1%	120	29.23	50.97
10%	1200	13.80	22.99
20%	2400	11.65	20.41
50%	6000	9.51	15.90
100%	12000	8.46	13.59

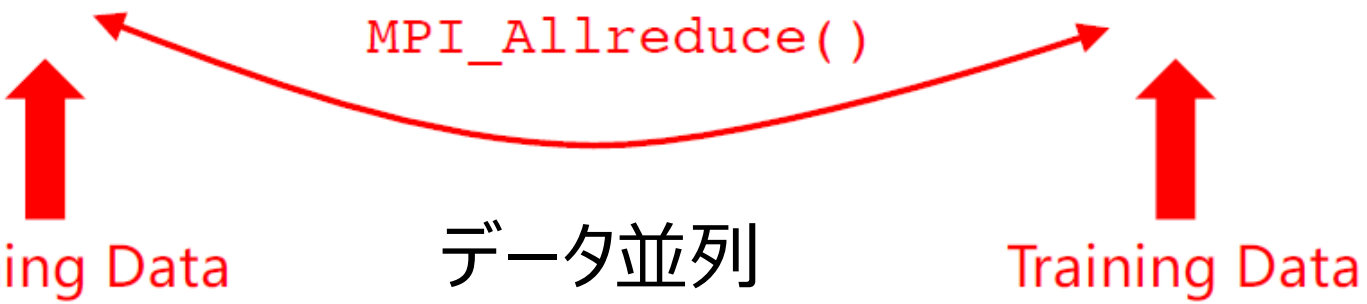
Batch Norm

Architecture	Hidden Units	Train		Dev	
		Baseline	BatchNorm	Baseline	BatchNorm
1 RNN, 5 total	2400	10.55	11.99	13.55	14.40
3 RNN, 5 total	1880	9.55	8.29	11.61	10.56
5 RNN, 7 total	1510	8.59	7.61	10.77	9.78
7 RNN, 9 total	1280	8.76	7.68	10.83	9.52

並列処理

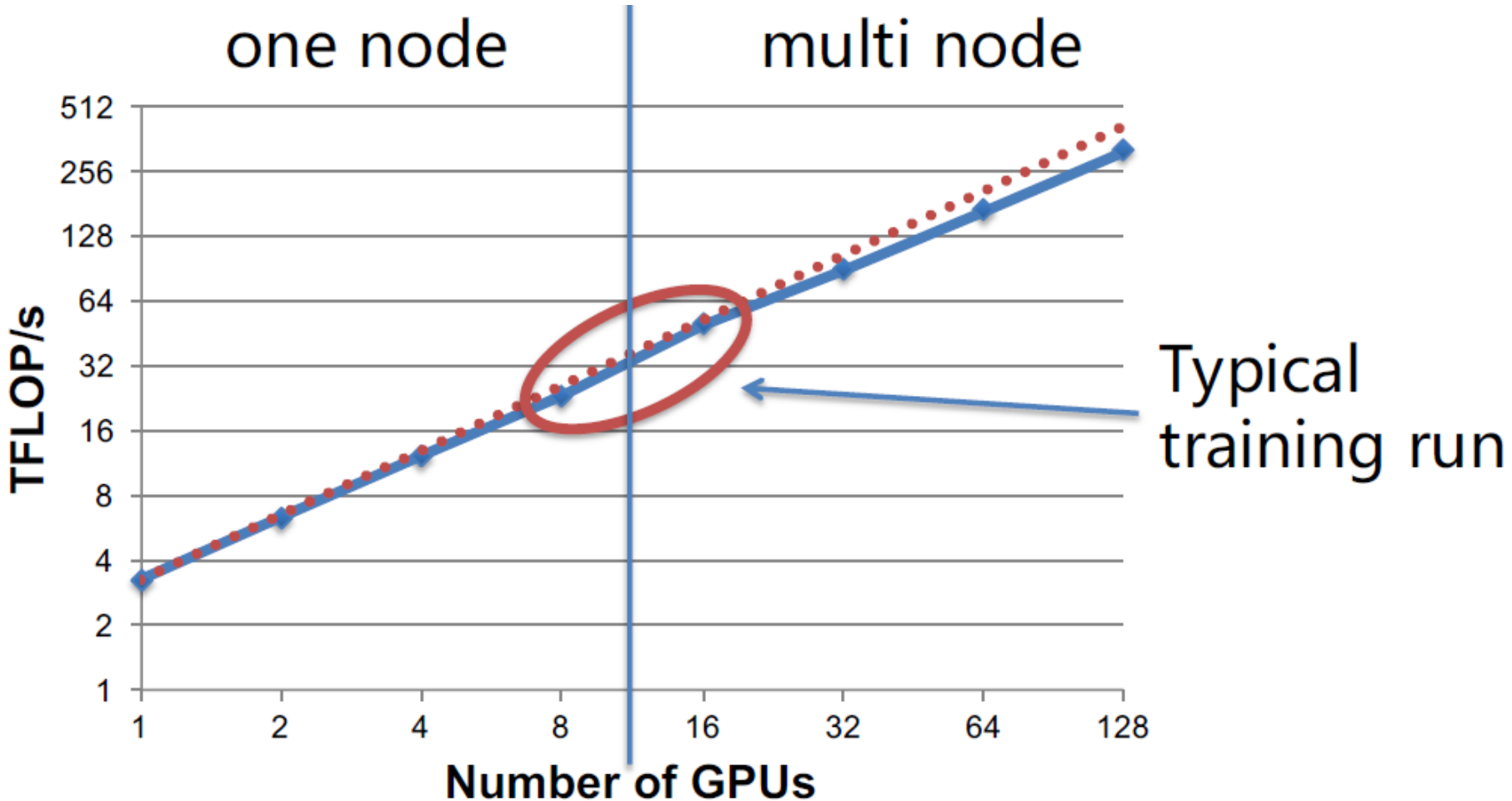


モデル並列



データ並列

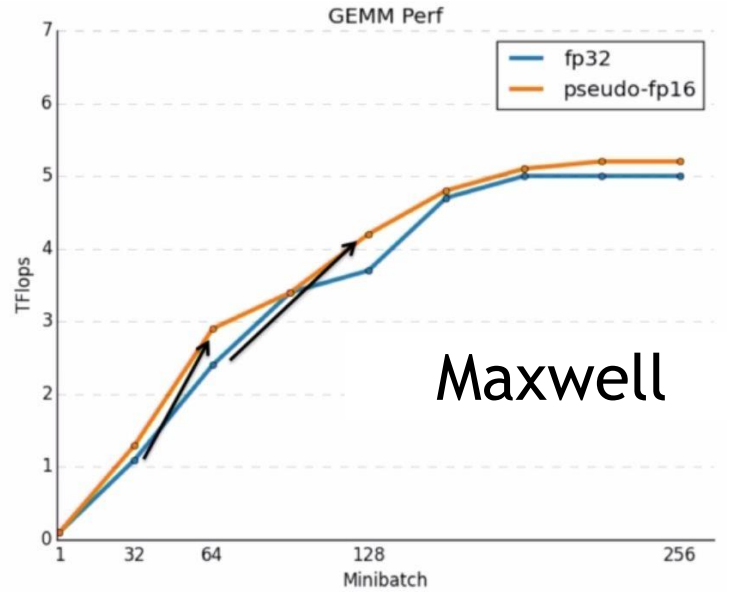
RNNトレーニング性能



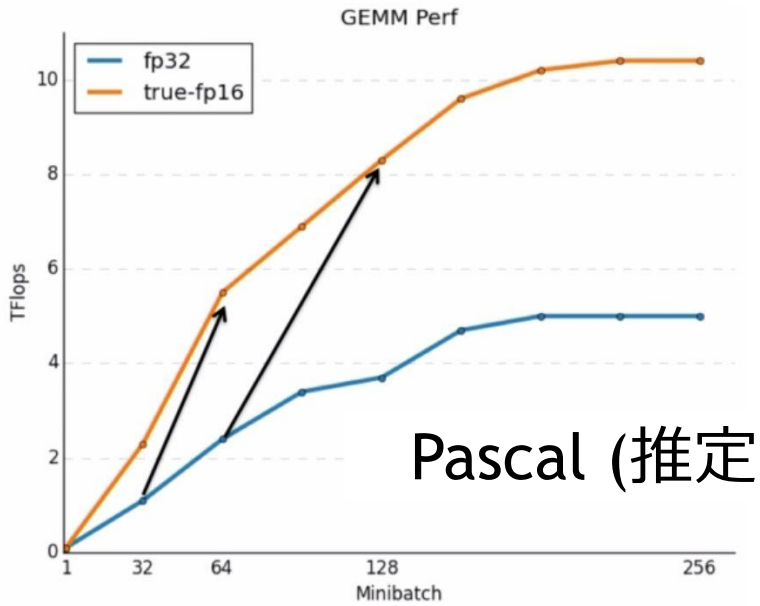
All reduce / FP16

独自のAllreduceを実装

GPU	OpenMPI all-reduce	Our all-reduce	Performance Gain
4	55359.1	2587.4	21.4
8	48881.6	2470.9	19.8
16	21562.6	1393.7	15.5
32	8191.8	1339.6	6.1
64	1395.2	611.0	2.3
128	1602.1	422.6	3.8



Maxwell



Pascal (推定)

DEEP CONVOLUTIONAL NEURAL NETWORKS FOR SPOKEN DIALECT CLASSIFICATION OF SPECTROGRAM IMAGES USING DIGITS

Nigel Cannings Chief Technical Officer, Intelligent Voice Limited

CNNを用いた方言分類

NIST LRE Competition

6言語、20方言

アラビア語(エジプト、イラク、レバノン、
マグレビ、標準語)

中国語(広東、北京、上海、台湾)

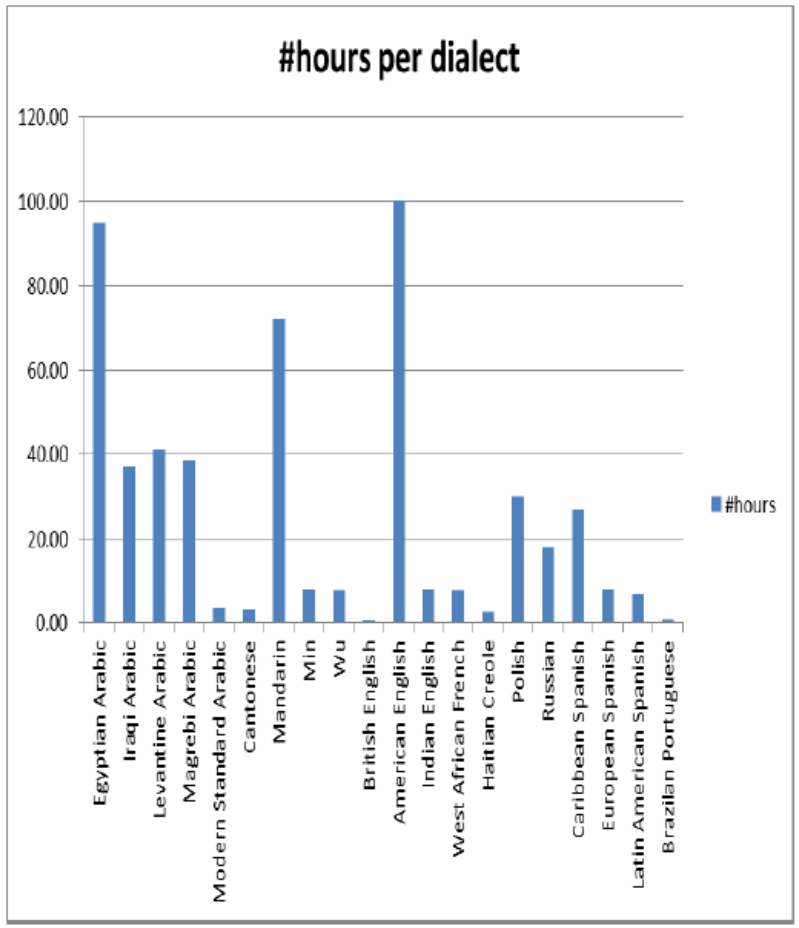
英語(英国、米国、インド)

フランス語(西アフリカ、ハイチ)

リベリア語 （カリブスペイン、ヨーロッパスペイン
ラテンアメリカスペイン、ブラジルポルトガル)

スラブ語(ポーランド、ロシア)

500時間以上のスピーチデータ

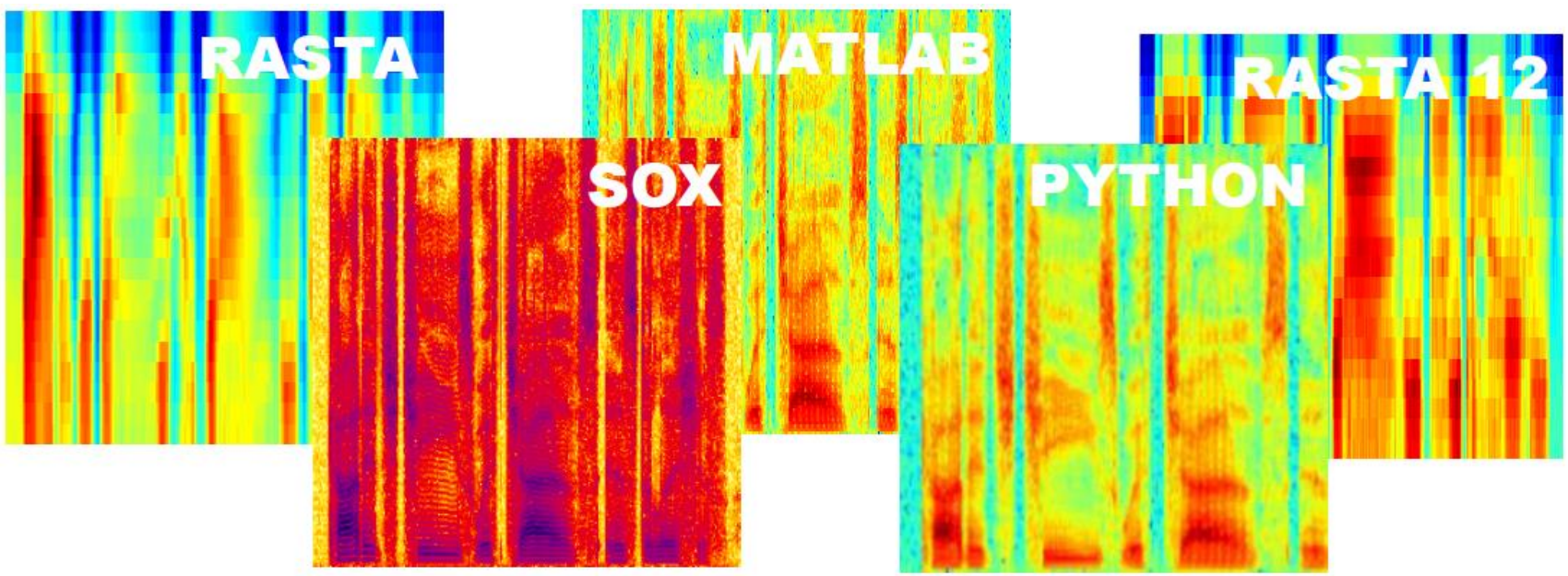


スペクトログラム＋CNN

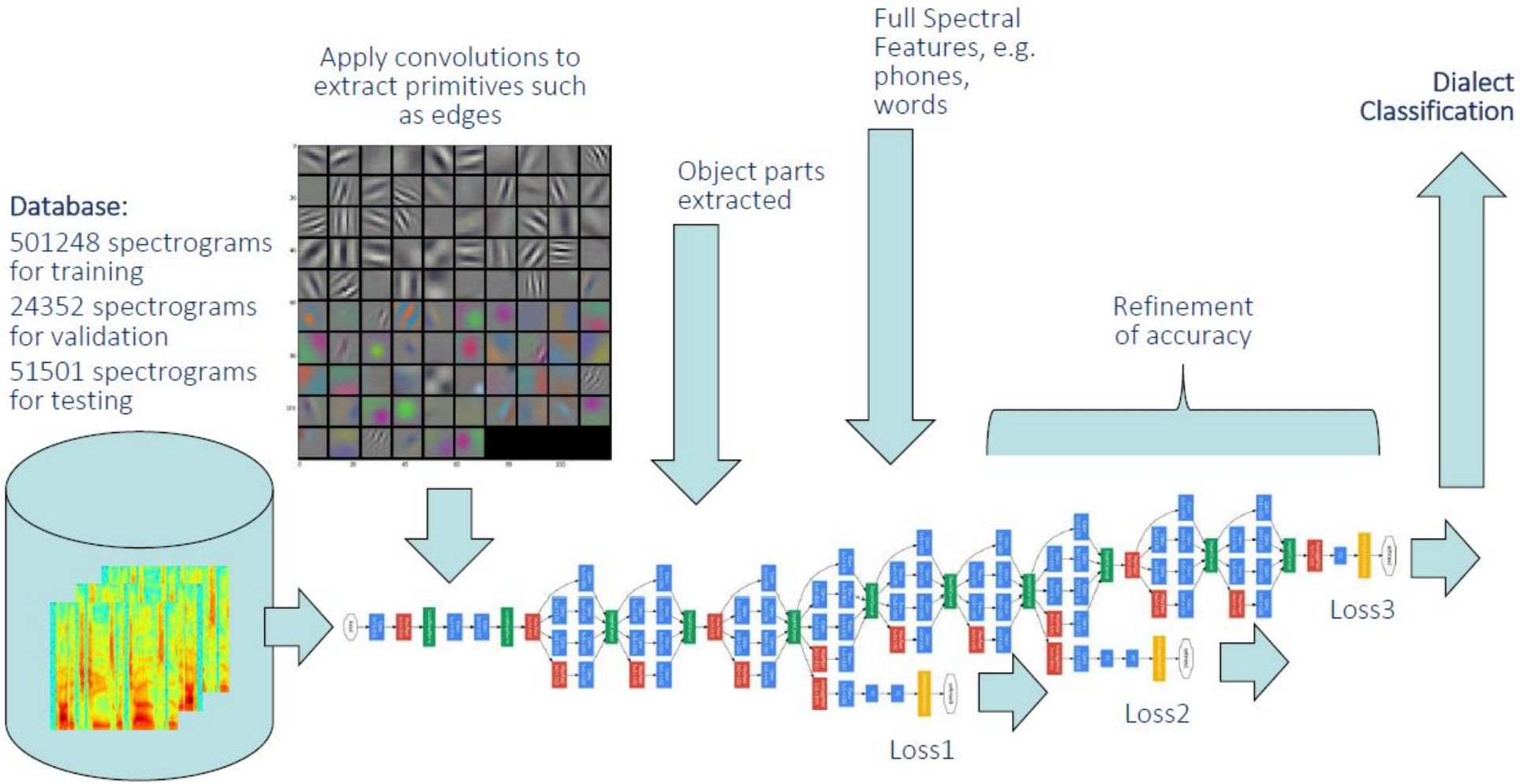
環境：NVIDIA DIGITS　GoogLeNet

会話データを256x256のスペクトログラムに変換

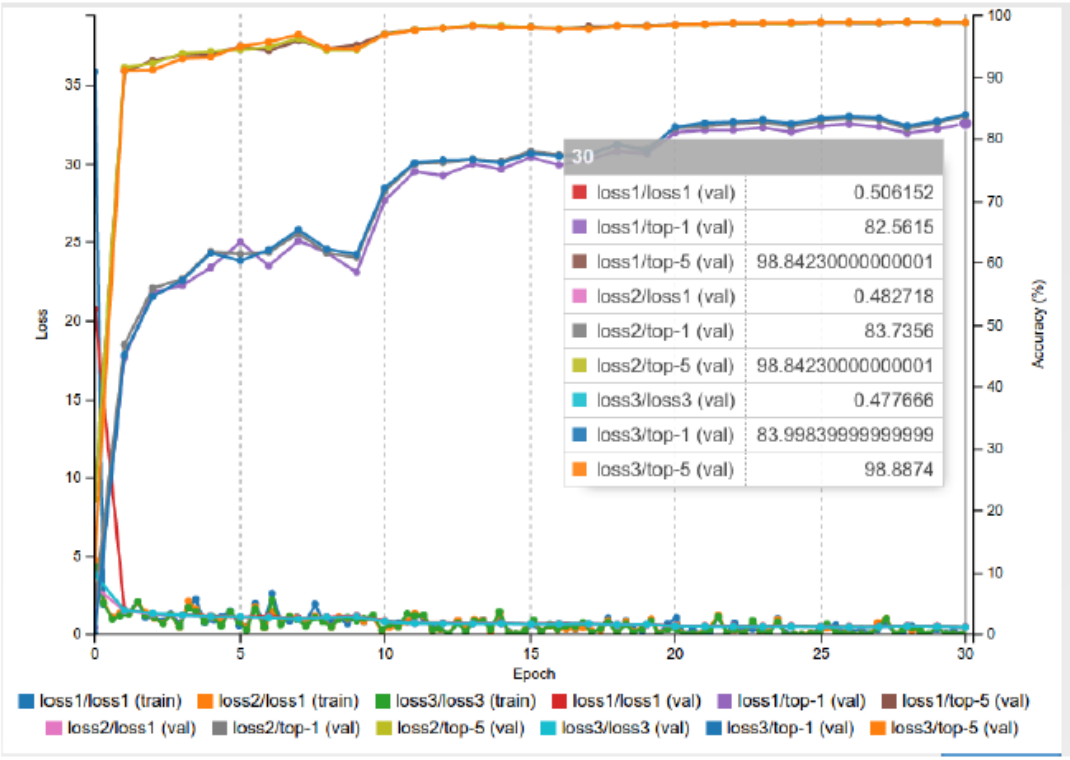
異なるスペクトル表現やコーディングを試行

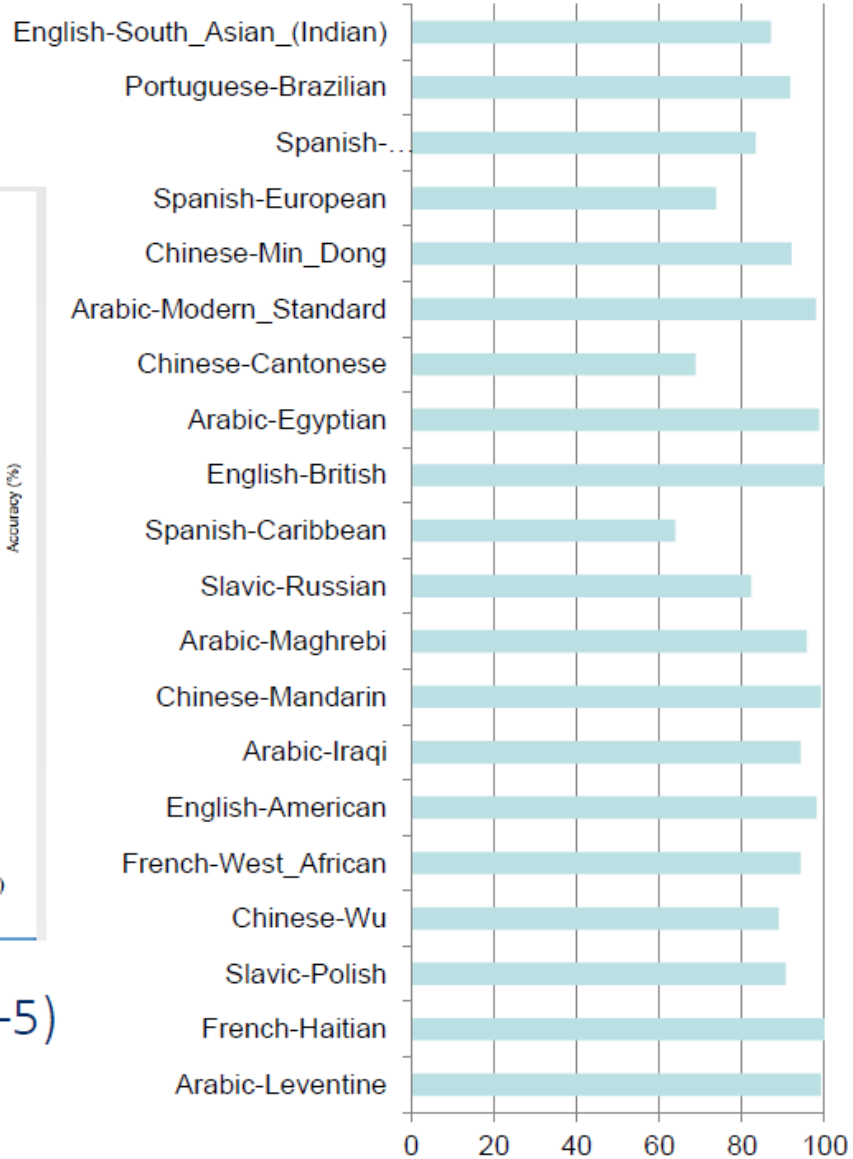


GoogLeNetでの処理



結果





- Accuracy – 83.99 (Top-1), 98.89% (Top-5)

MINING AUDIO INFORMATION ON WEB VIDEOS AND RECORDINGS

Benjamin Elizalde PhD Student, Carnegie Mellon University

ビデオから都市を特定

オーディオで特定

10種類の典型的な都市の音

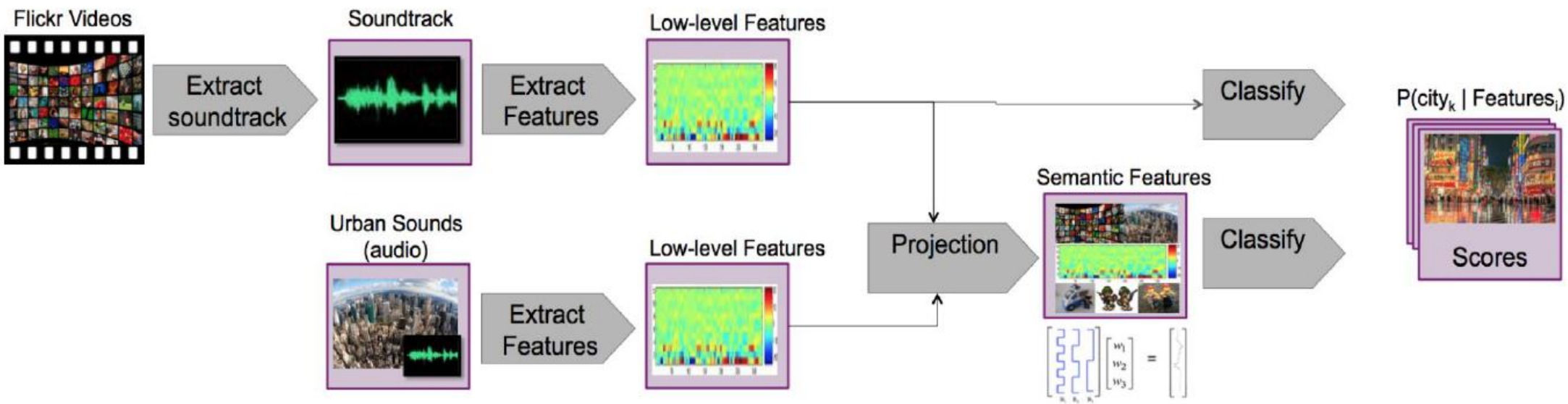
空調機、クラクション、子供の遊び声、犬の泣き声、アイドリング、
銃声、手持ち削岩機、サイレン、ドリル、ストリートミュージック

18都市

バンコク、バルセロナ、北京、ベルリン、シカゴ、ヒューストン、ロンドン
ロサンゼルス、モスクワ、ニューヨーク、パリ、プラハ、リオ、ローマ、
サンフランシスコ、ソウル、シドニー、東京

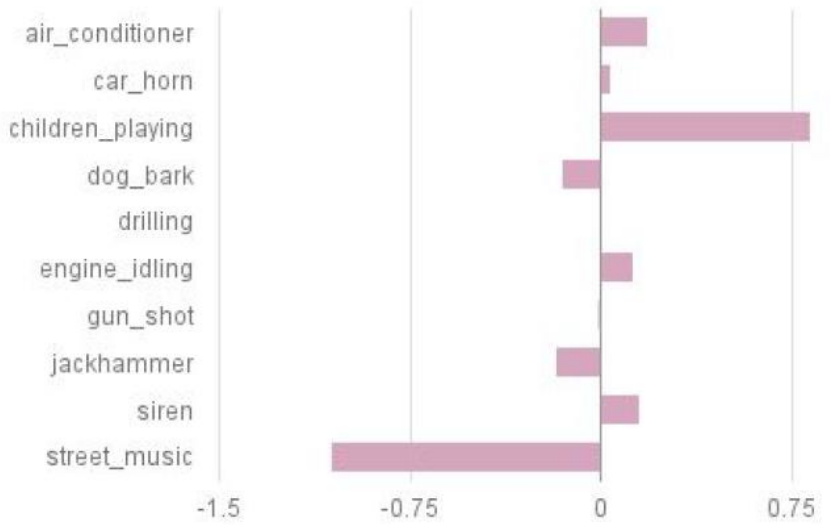
$$City - \widehat{soundtrack} \approx carW_1 + sirenW_2 + \ldots + drillingW_{10}$$

都市の認識フロー


Flickr Videos
Extract soundtrack
Soundtrack
Extract Features
Low-level Features
Classify
P(city$_k$ | Features$_i$)
Scores
Semantic Features
Urban Sounds (audio)
Extract Features
Low-level Features
Projection
Classify

認識例

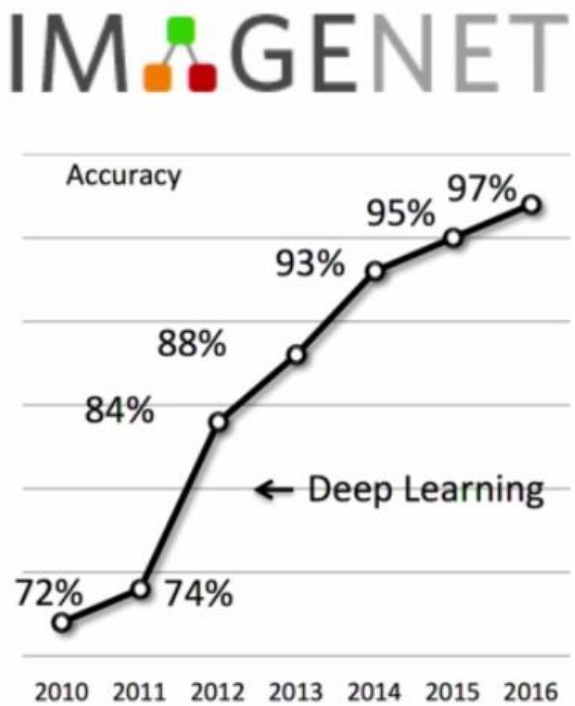
Children Playing and Siren in Rome



3D DEEP LEARNING

Jianxiong Xiao Assistant Professor, Princeton University

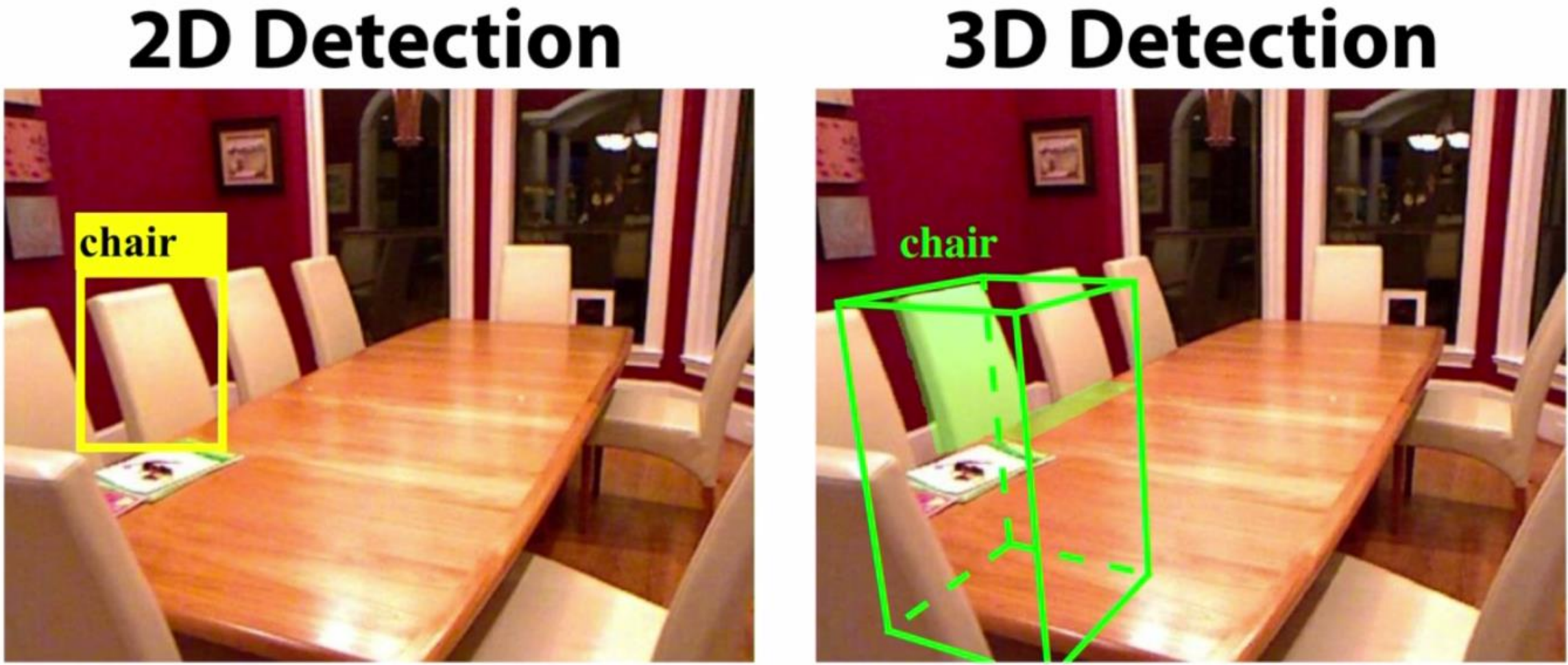
ロボットのための3次元 Deep Learning 認識



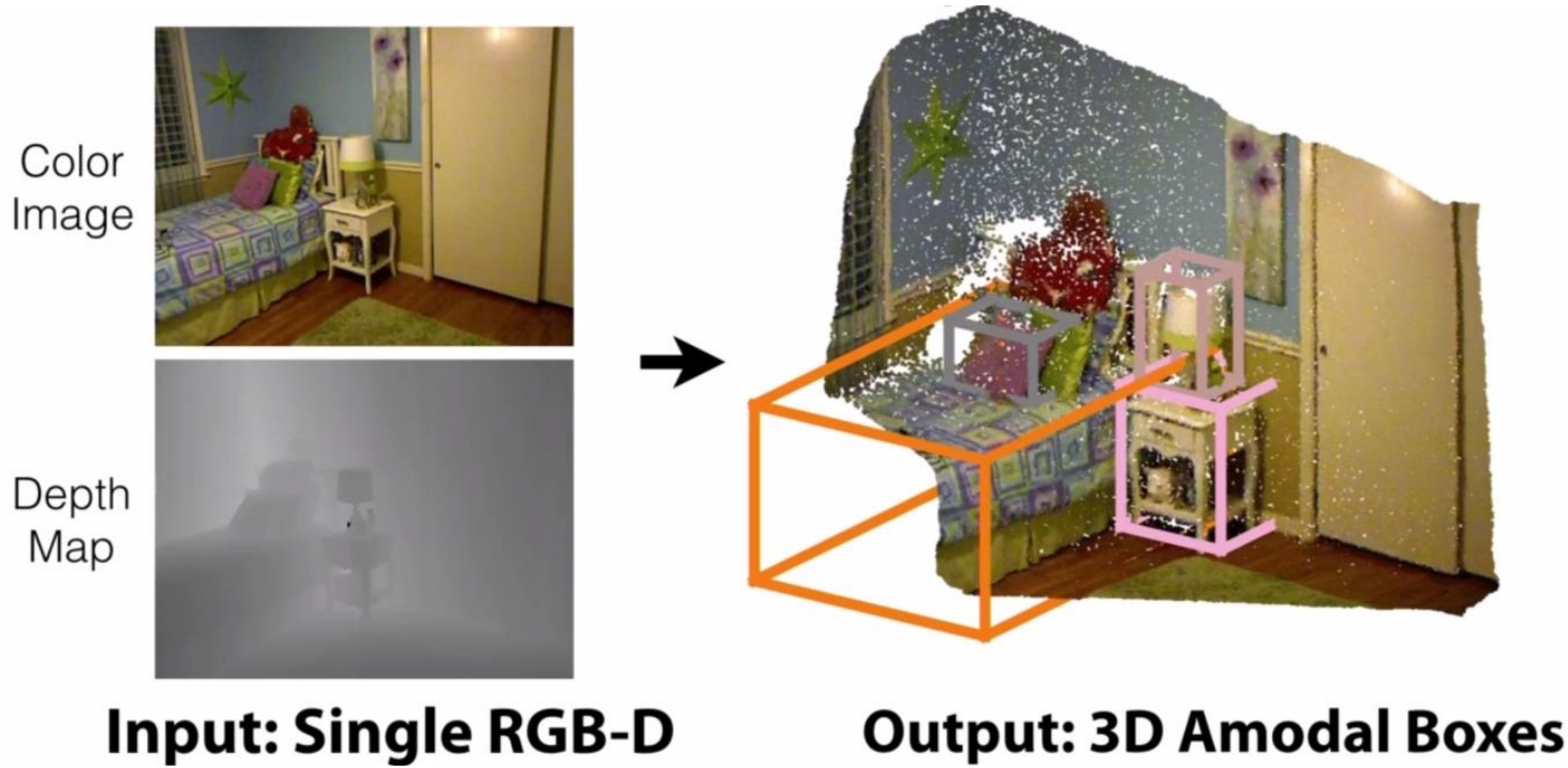
Computer Vision:
A Huge Breakthrough!

Robot Perception:
Still Doesn't Work!

３次元での認識



３次元アモーダル物体検出



2 次元物体検出

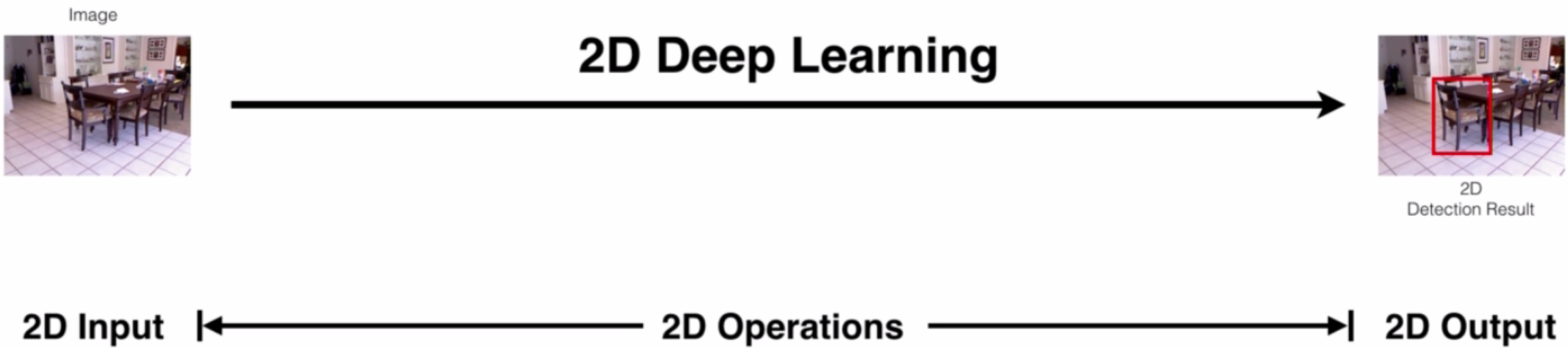
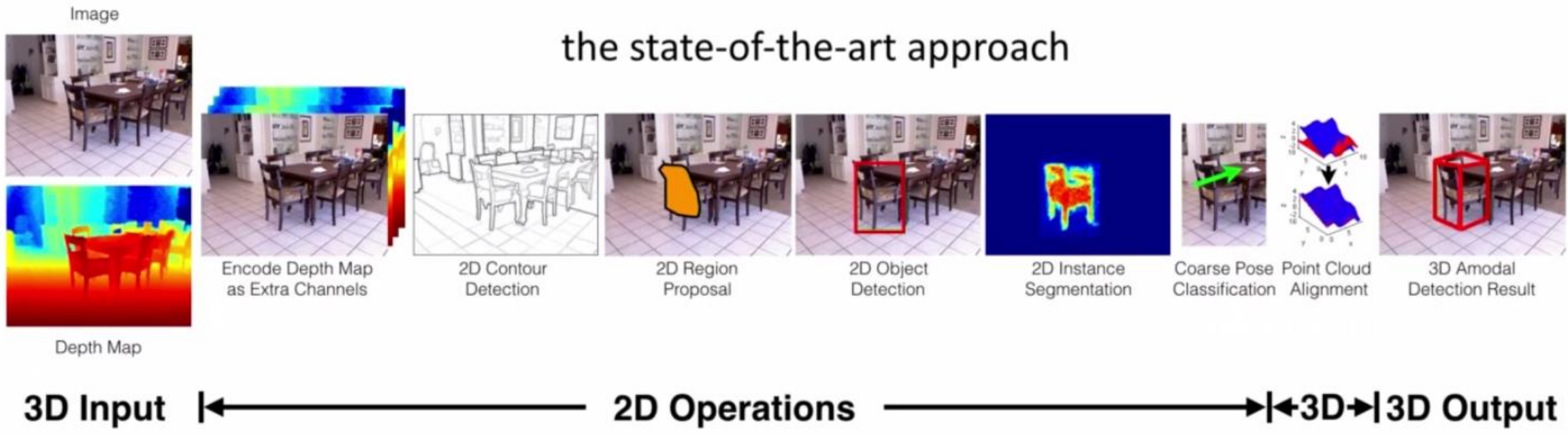
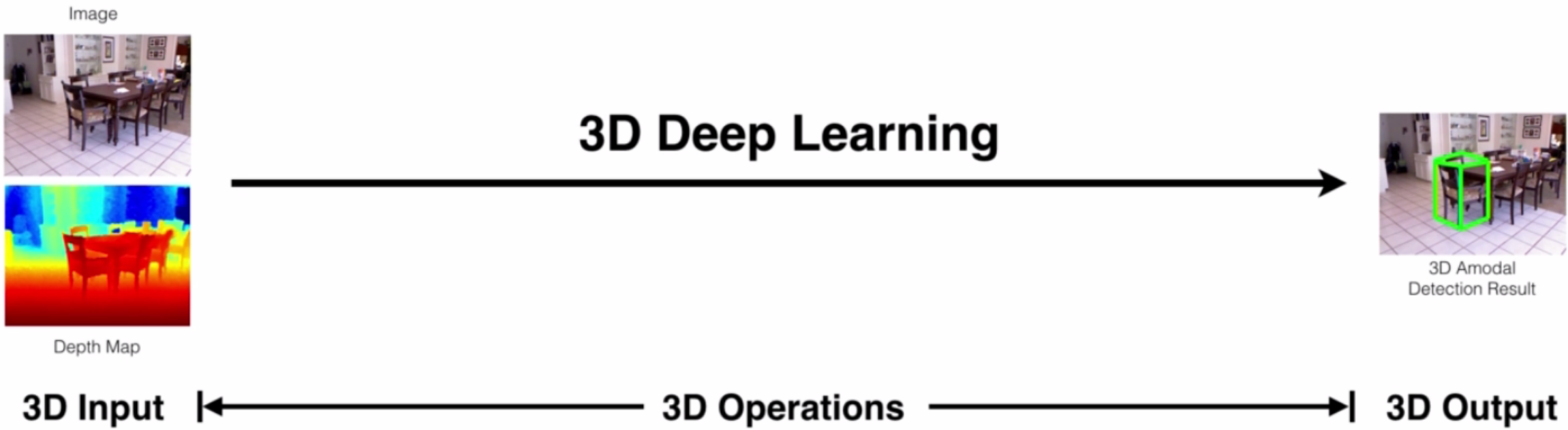


Image
2D Deep Learning
2D
Detection Result
2D Input
2D Operations
2D Output

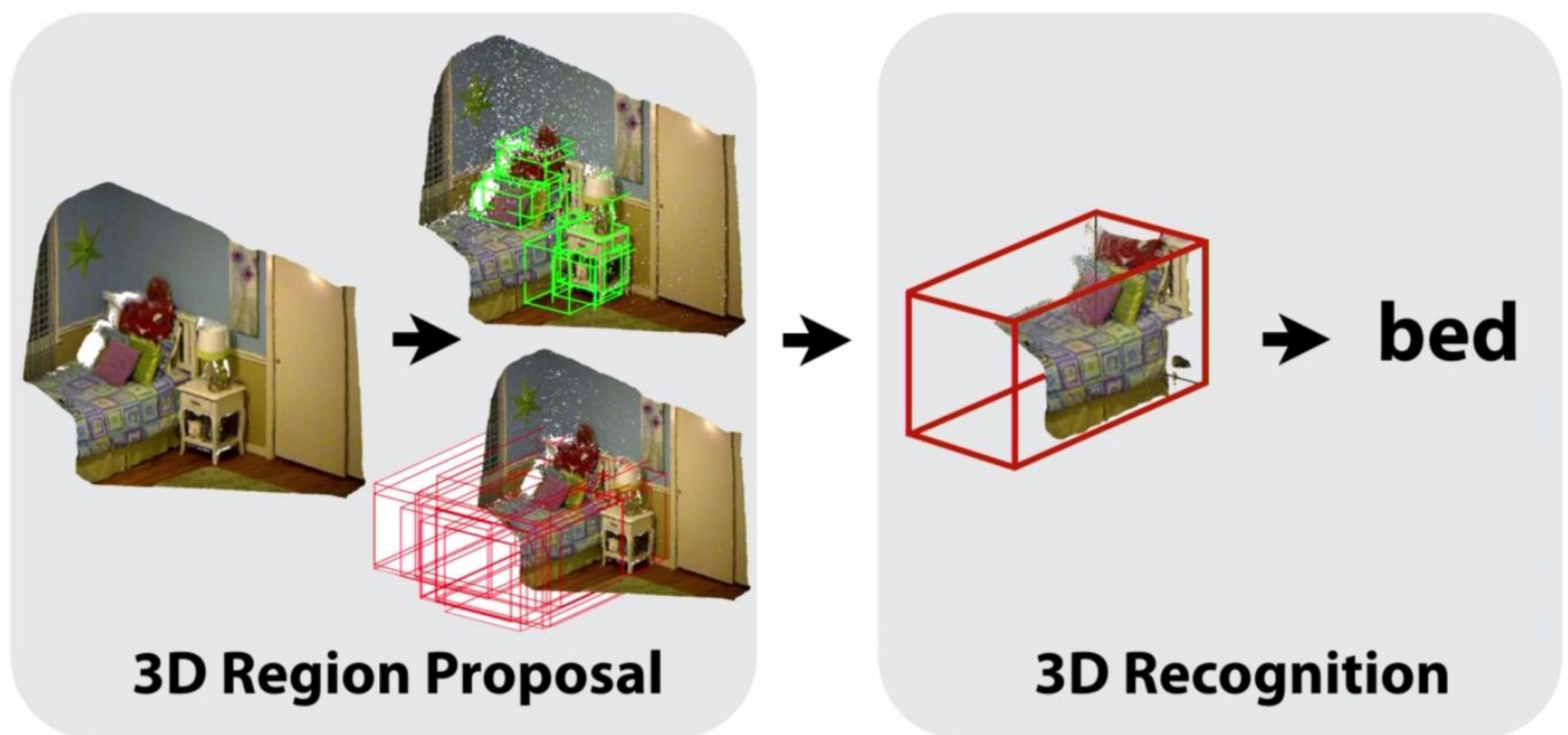
３次元アモーダル物体検出



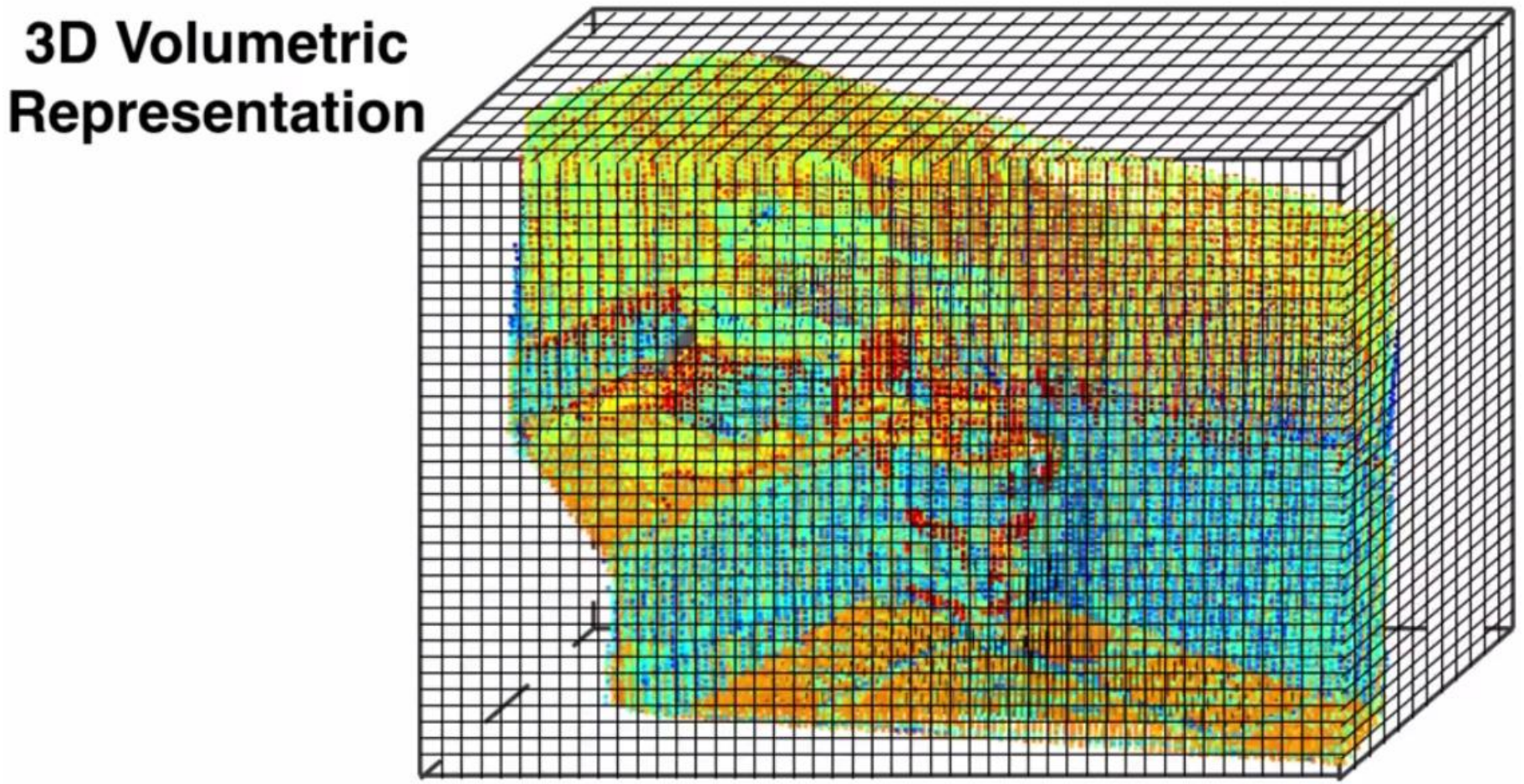
３次元アモーダル物体検出



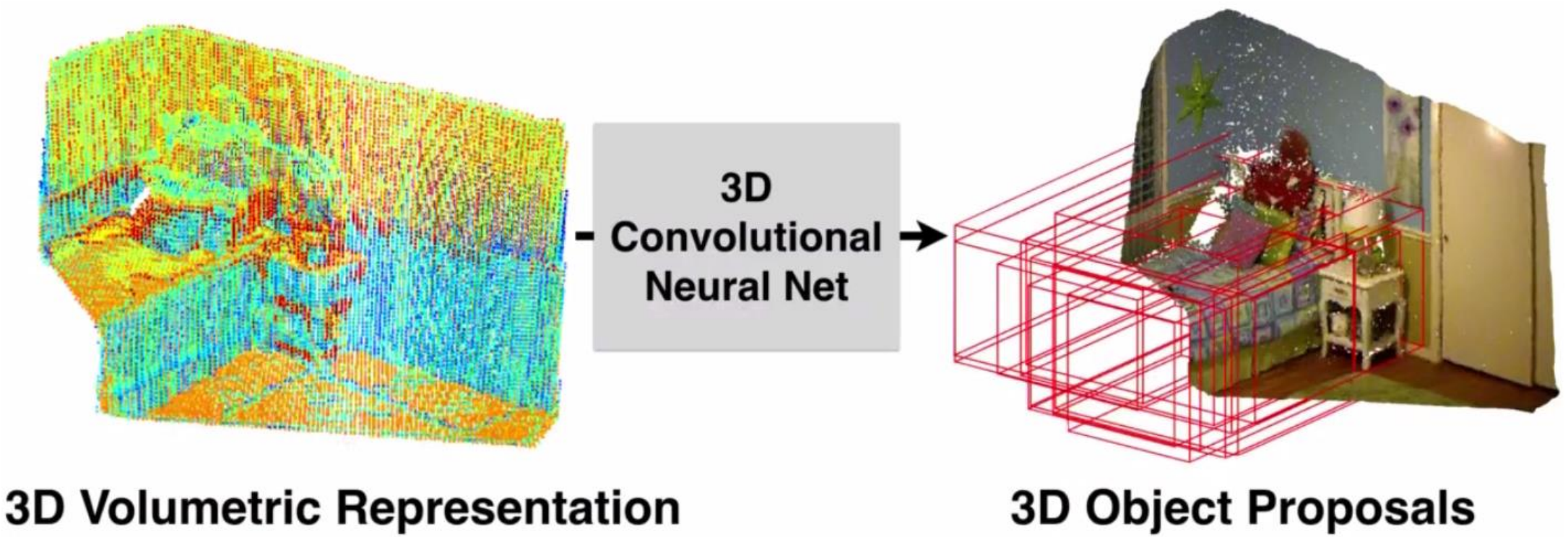
3次元 Deep Learning



３次元情報の符号化

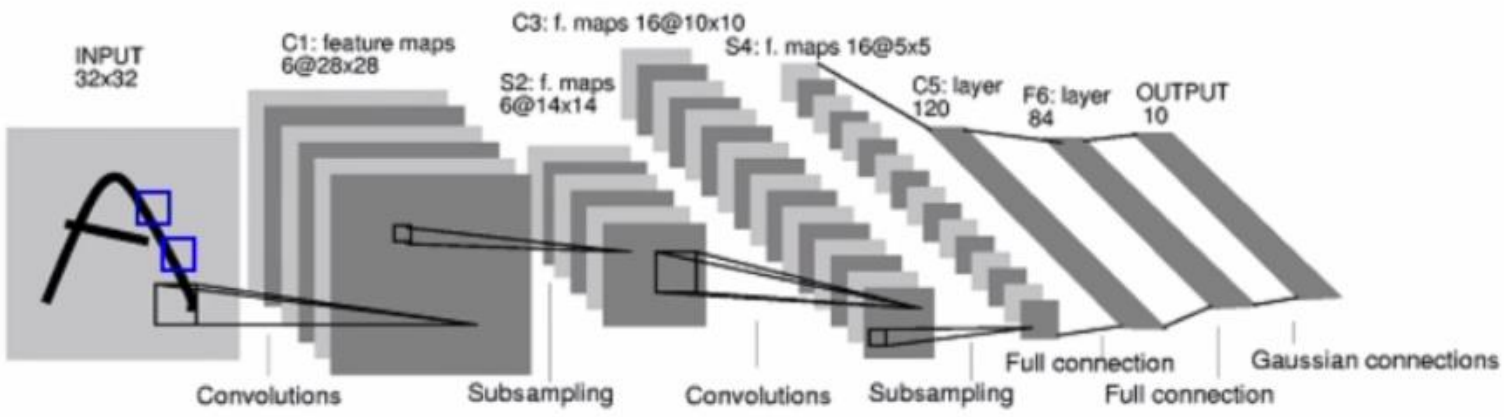


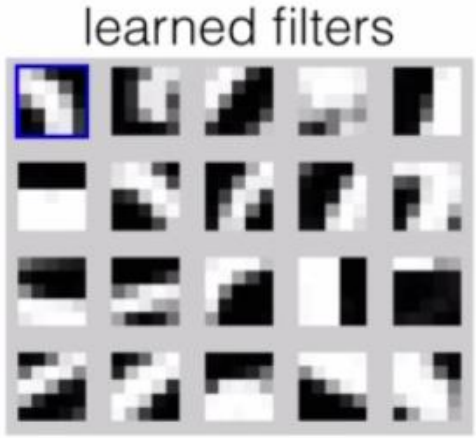
3 次元物体提案 ネットワーク



２次元コンボリューショナル・ニューラル・ネットワーク

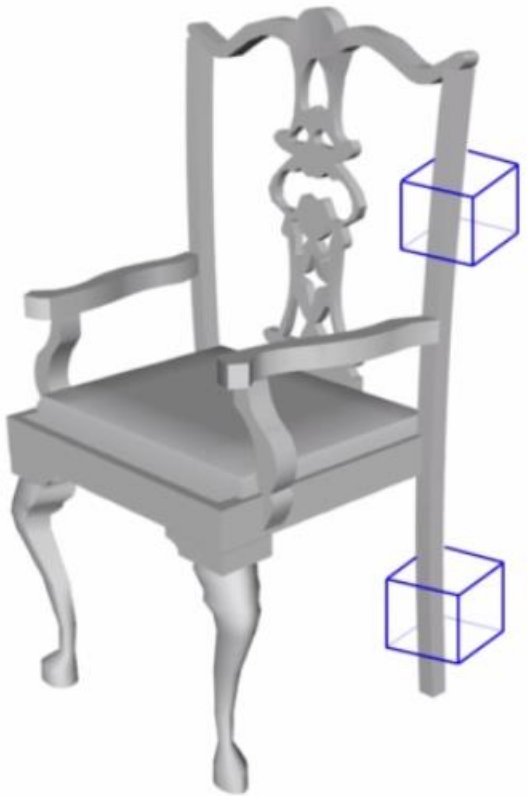
Idea: Because of the spatial nature of 2D images, filter weights can be re-used at different spatial locations of the image.



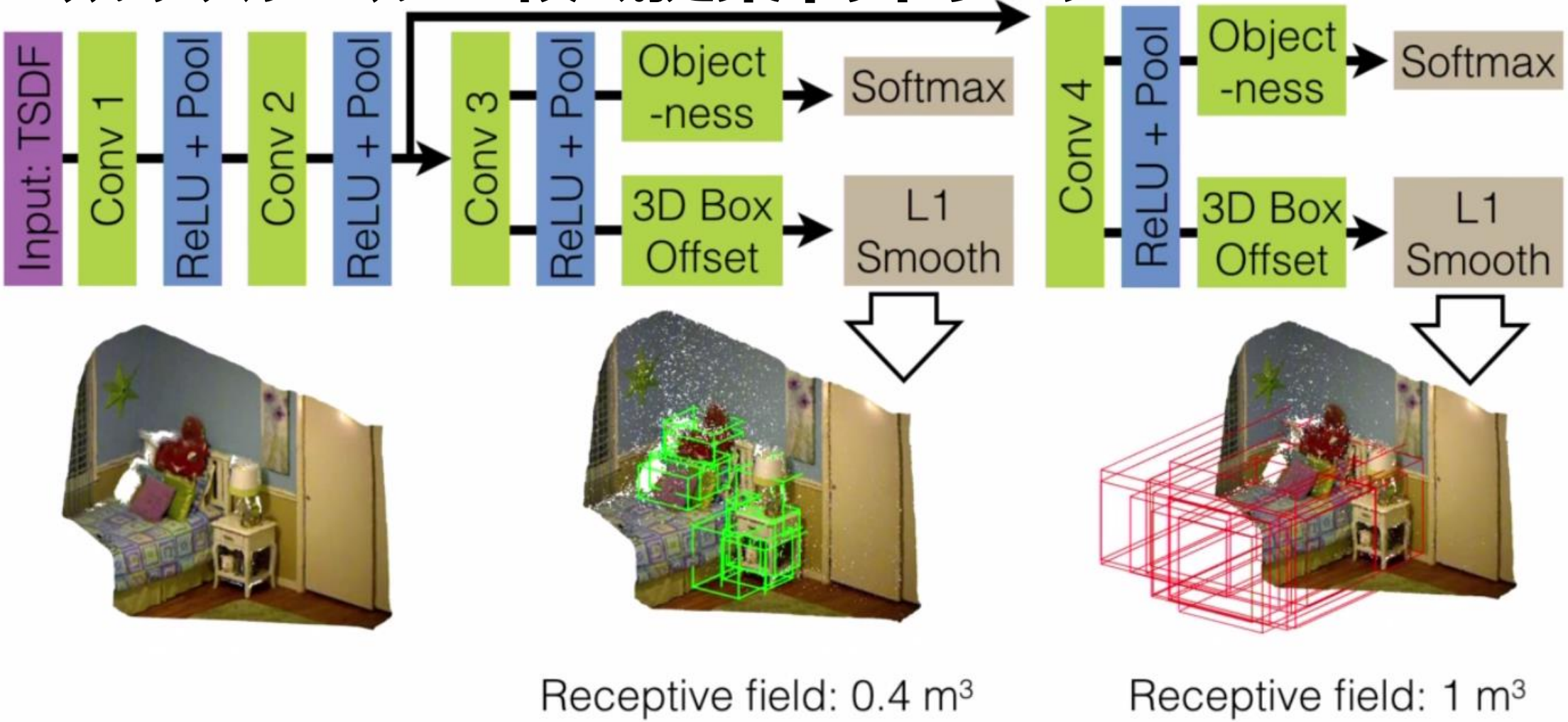


３次元コンボリューショナル・ニューラル・ネットワーク

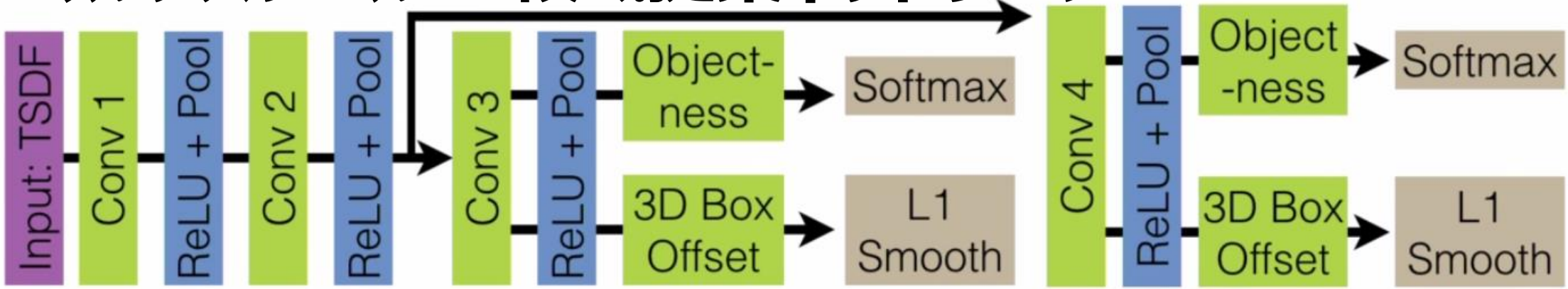
Idea: Because of the spatial nature of 3D shapes, filter weights can be re-used at different spatial locations of the 3D volume.



マルチスケール3D領域提案ネットワーク

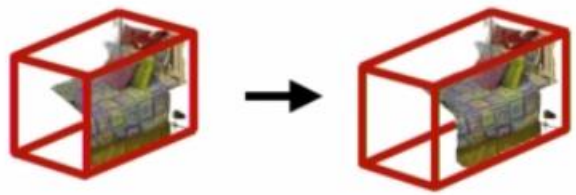


マルチスケール3D領域提案ネットワーク

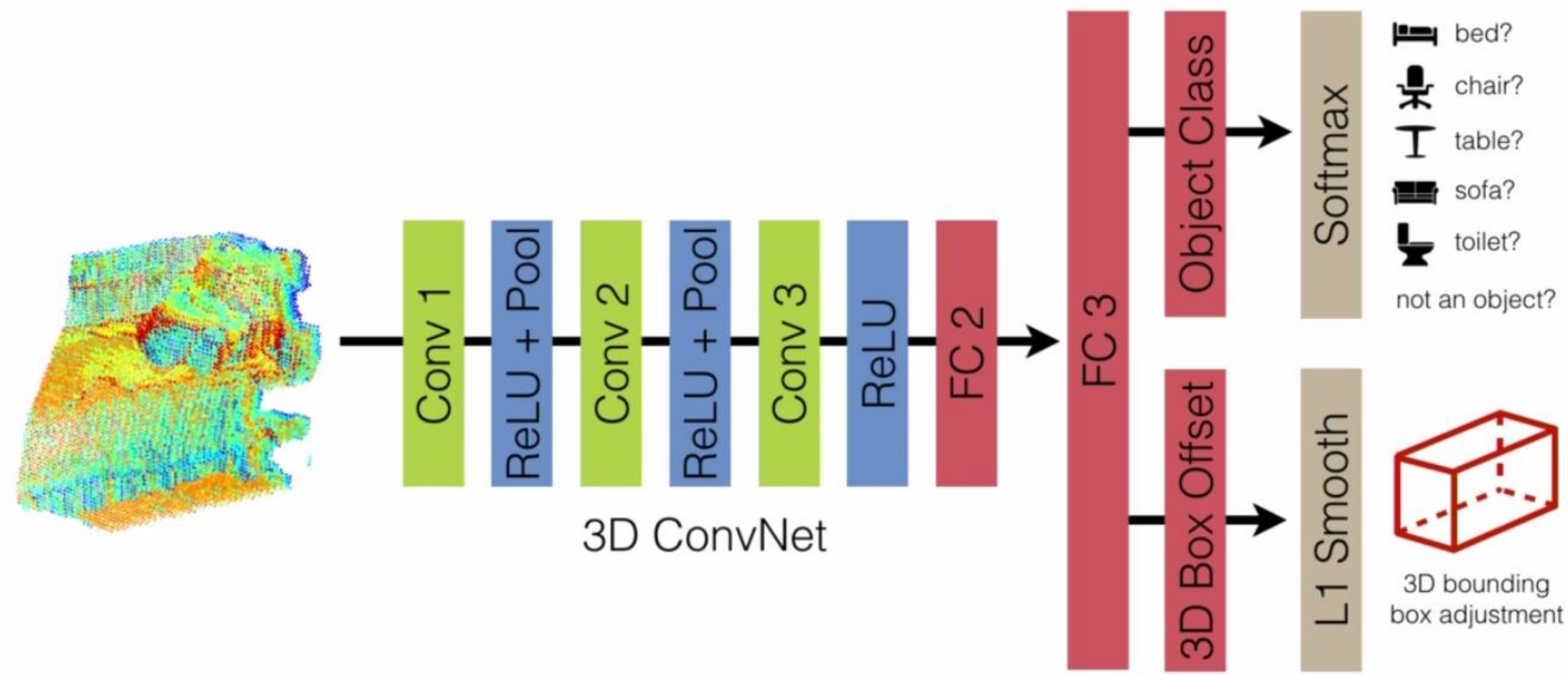


- **Objectness**: Is there an object in this box?

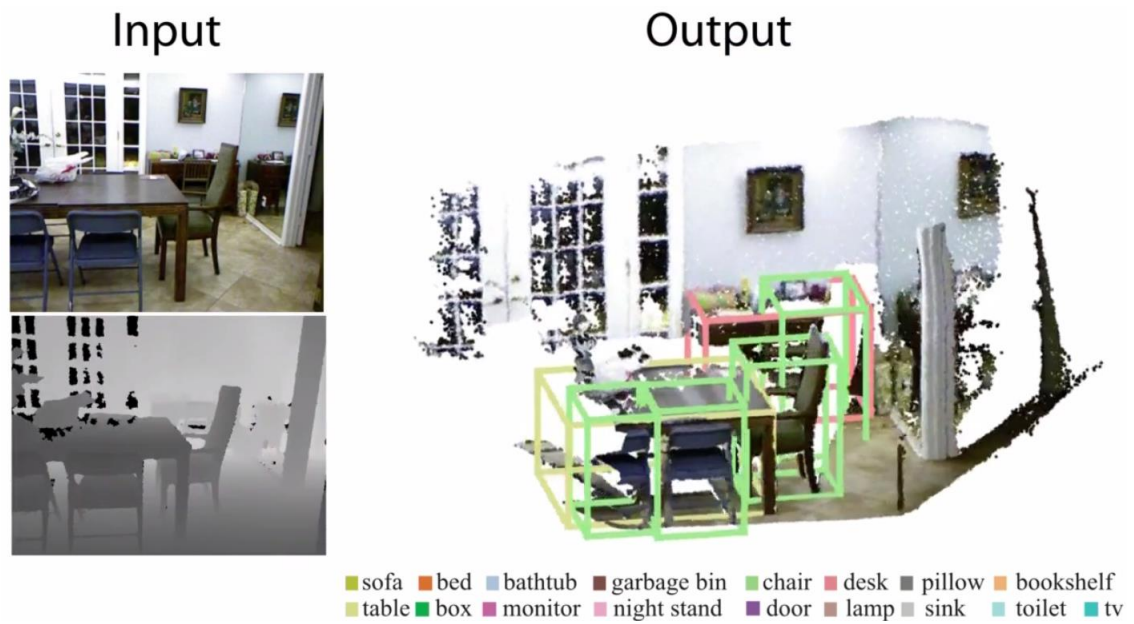
- **3D Box Offset**: How to adjust the 3D bounding box to better fit the object?

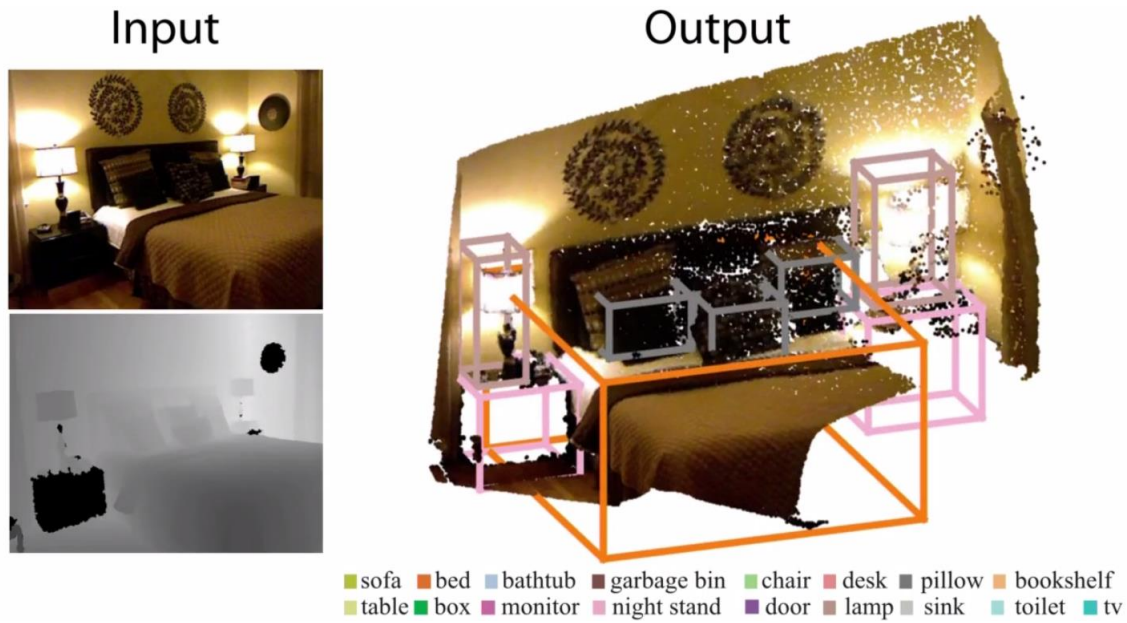


3 次元物体認識ネットワーク



3次元物体認識例

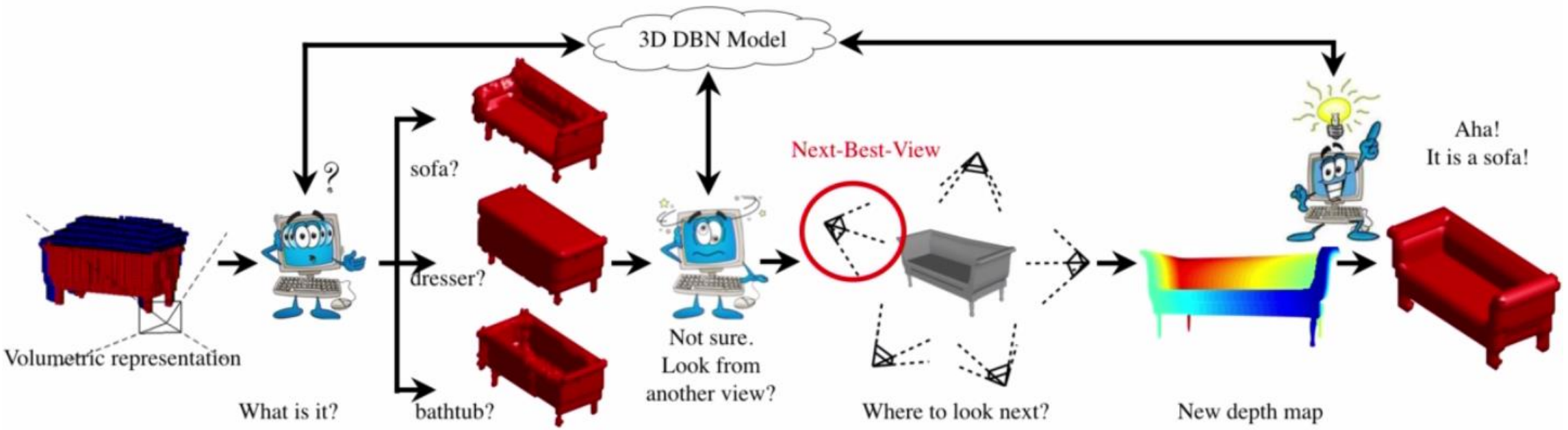

Input
Output
sofa bed bathtub garbage bin chair desk pillow bookshelf
table box monitor night stand door lamp sink toilet tv


Input
Output
sofa bed bathtub garbage bin chair desk pillow bookshelf
table box monitor night stand door lamp sink toilet tv

結果：性能比較

	Algorithm	Input						mAP
3D Non-Deep Learning	Sliding Shapes	Depth	33.5	29	34.5	33.8	67.3	39.6
2D Deep Learning	Depth-RCNN (segment)	Depth	71	18.2	49.6	30.4	63.4	46.5
	Depth-RCNN (segment)	RGB-D	74.7	18.6	50.3	28.6	69.7	48.4
	Depth-RCNN (CAD fit)	Depth	72.7	47.5	54.6	40.6	72.7	57.6
	Depth-RCNN (CAD fit)	RGB-D	73.4	44.2	57.2	33.4	84.5	58.5
3D Deep Learning	Ours	Depth	83.0	58.8	68.6	49.5	79.2	67.8
	Ours	RGB-D	**84.7**	**61.1**	**70.5**	**55.4**	**89.9**	**72.3**

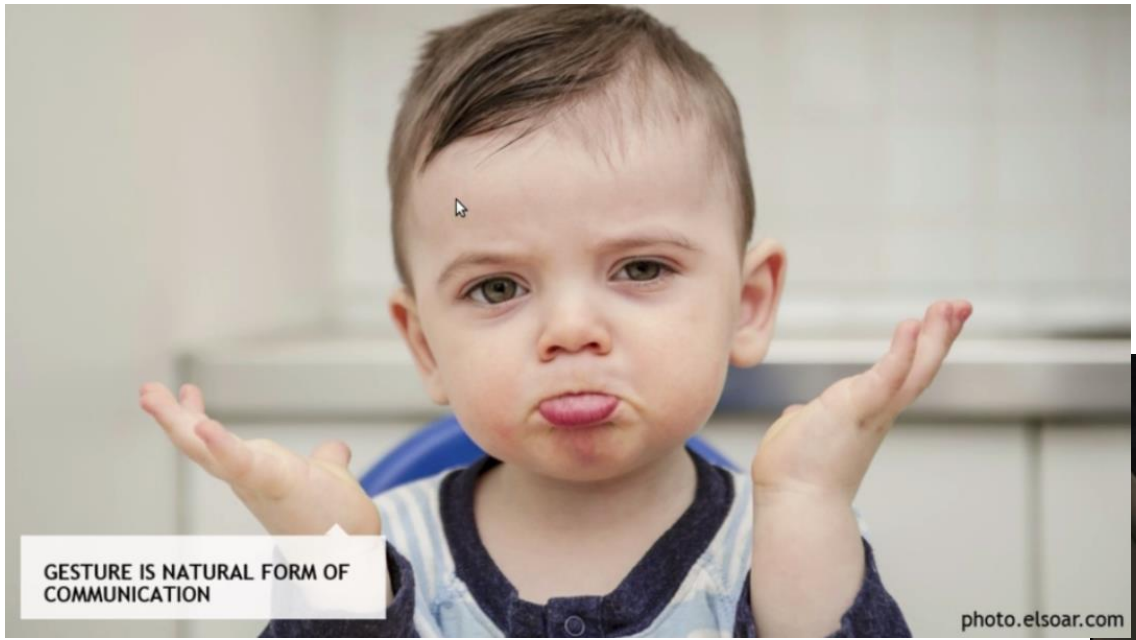
Deep View Planning



HAND GESTURE RECOGNITION WITH 3D CONVOLUTIONAL NEURAL NETWORKS

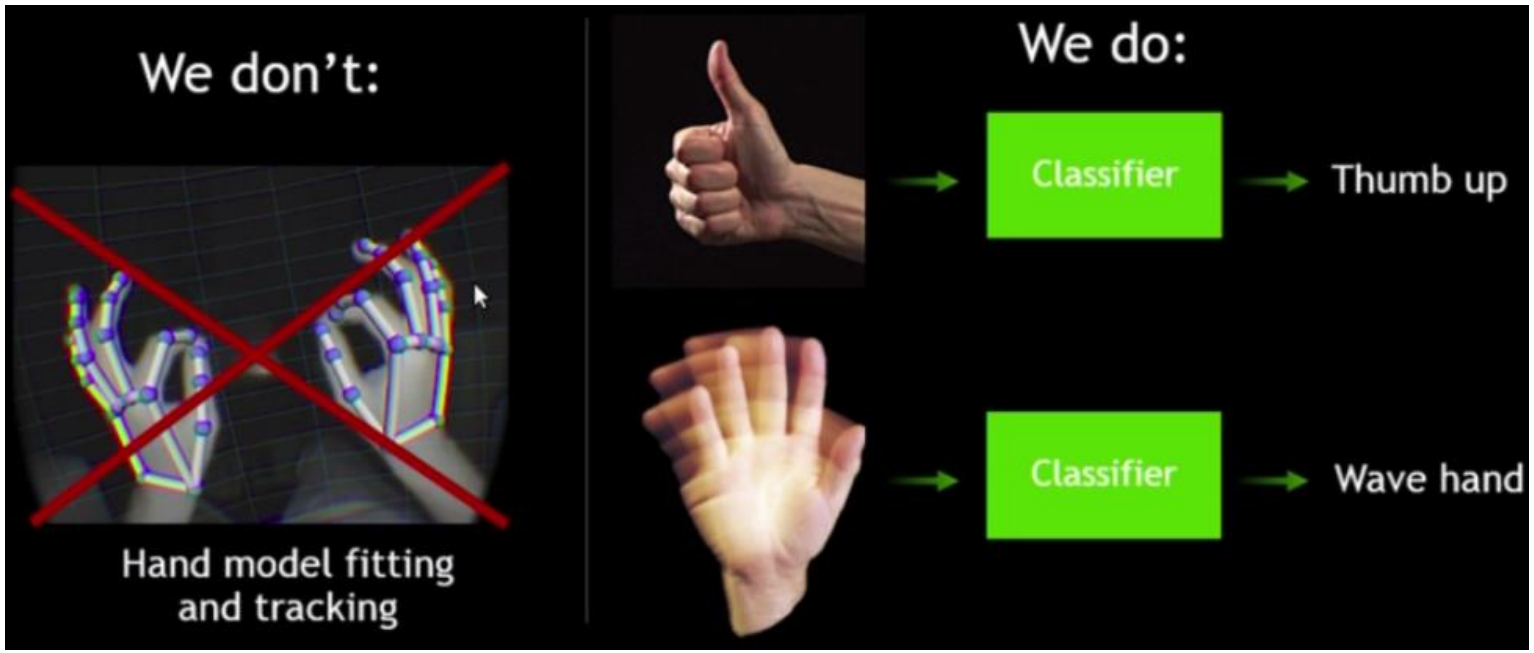
Pavlo Molchanov Research Scientist, NVIDIA

ジェスチャー認識


GESTURE IS NATURAL FORM OF COMMUNICATION
photo.elsoar.com


SAFE INTERFACES
@ bmw.com

本件のアプローチ方法


We don't:
Hand model fitting
and tracking
We do:
Classifier
Thumb up
Classifier
Wave hand


Kinectv1
SoftKinetic

最良の分類器の選択

VIVA CHALLENGE 2015 UCLA

19 classes, 8 subjects

Driver and passenger

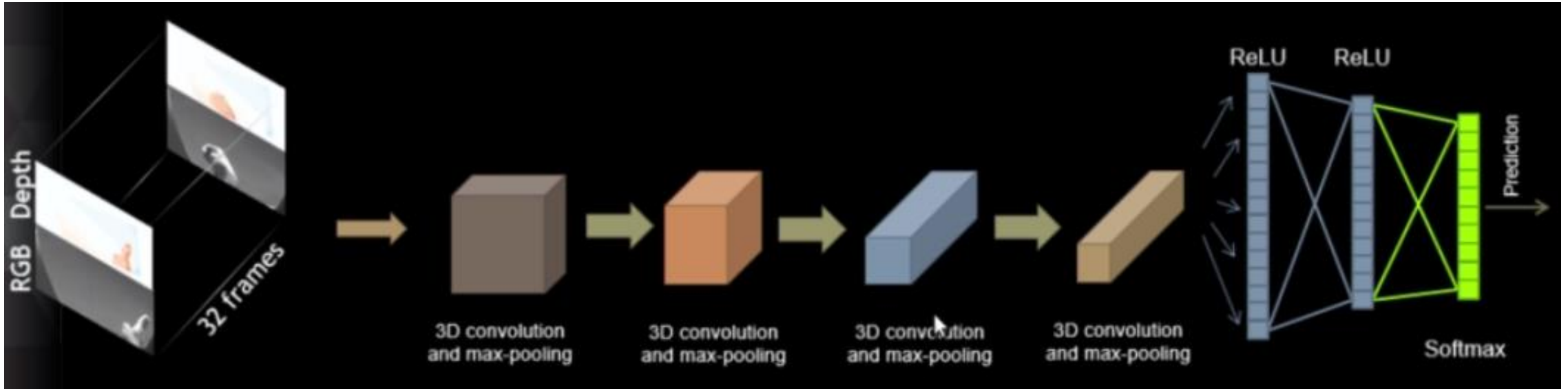
RGB + Depth from Microsoft Kinect

885 gestures in total

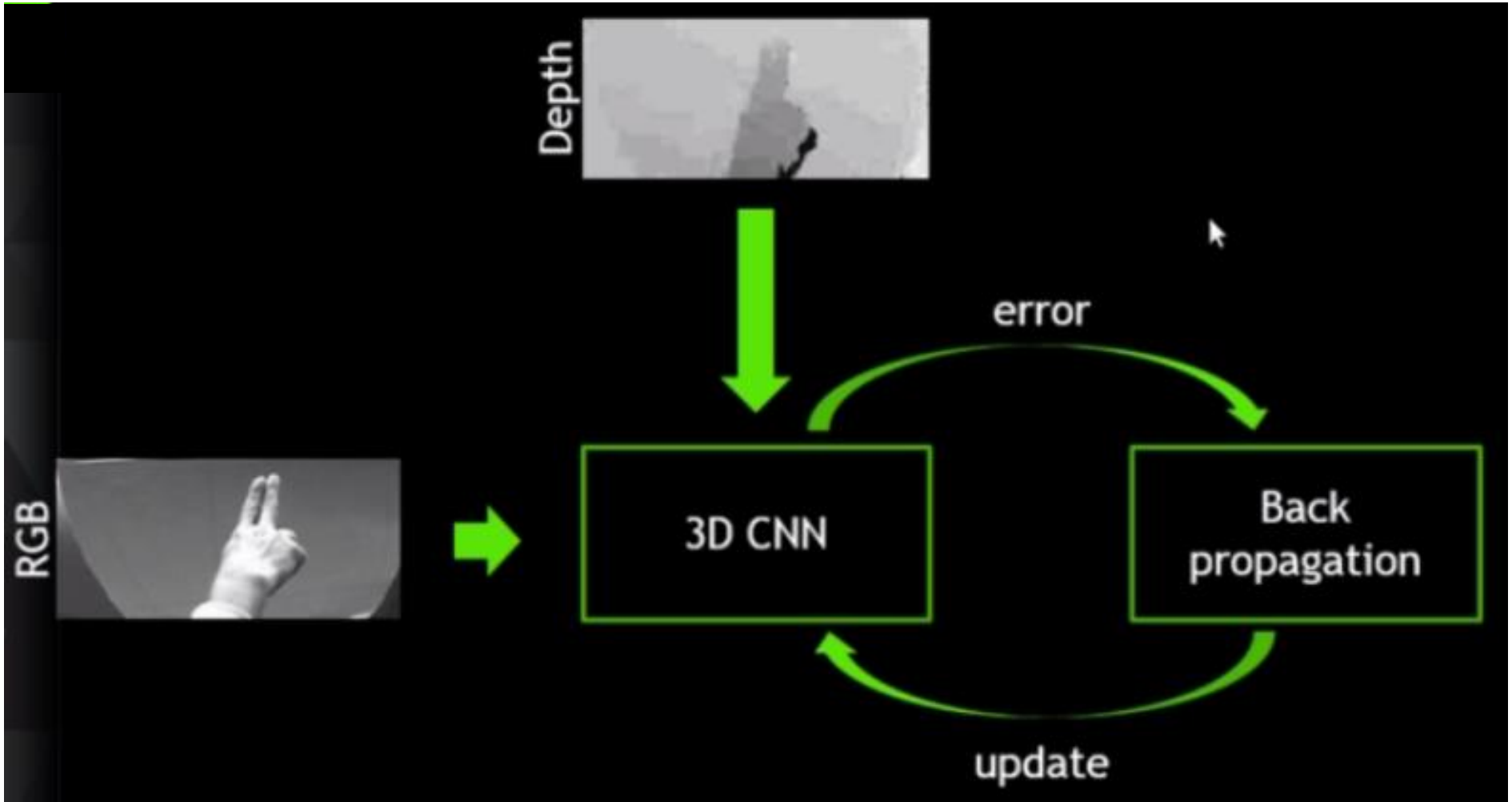
http://cvrr.ucsd.edu/vivachallenge/index.php/hands/hand-gestures/

最良の分類器の選択

3D Convolutional Neural Network



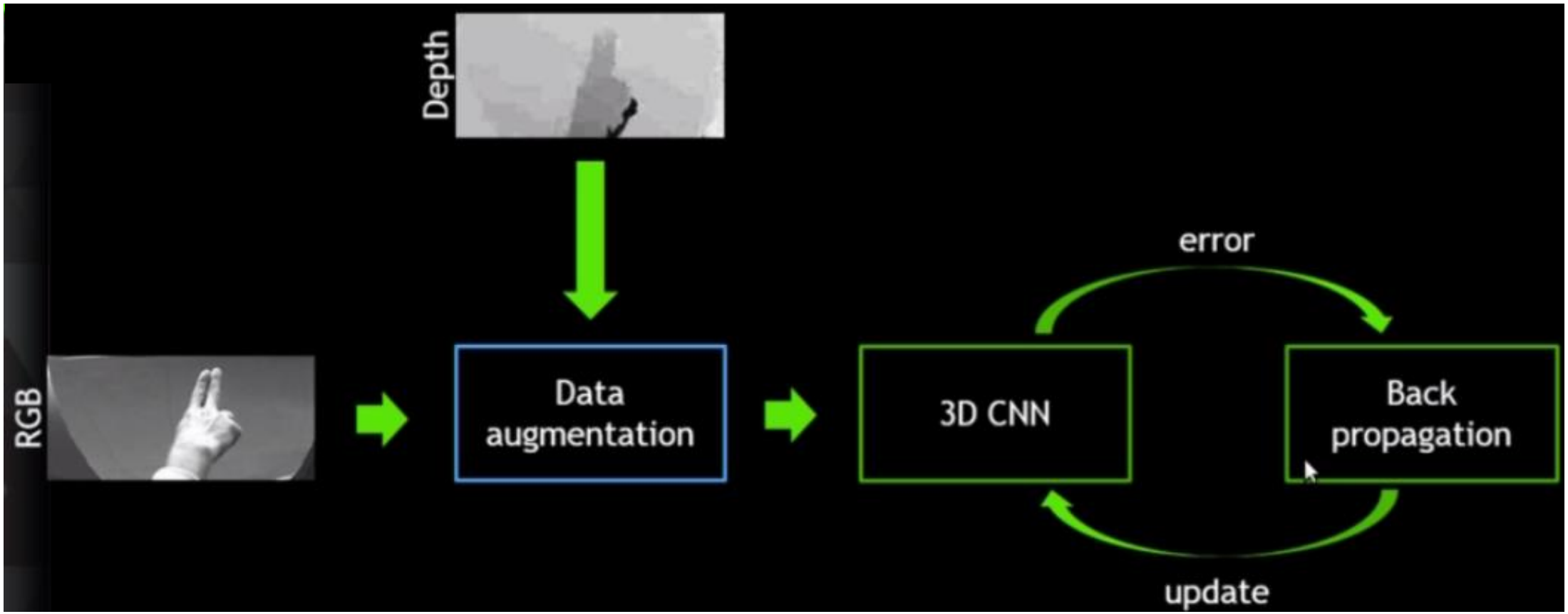
セグメントジェスチャー認識



一度目の結果

	HON4D[1]	HOG[2]	3D-CNN
Testing set	58.7%	64.5%	48.3%
Training set			99.9%

データ・オーグメンテーション

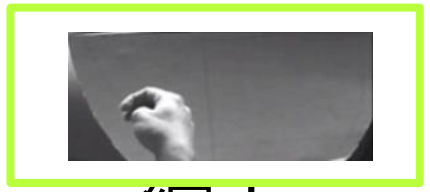


データ・オーグメンテーション

Spatial geometric Transformation

元データ

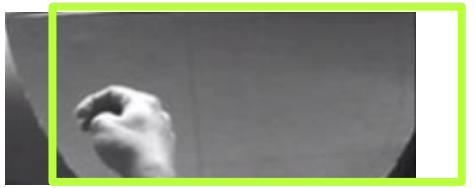
拡大



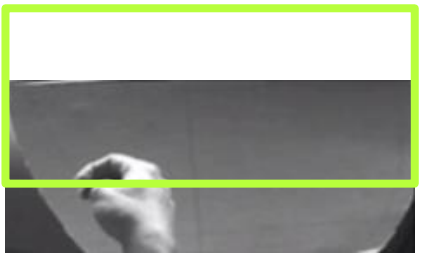
縮小

左回転

右回転



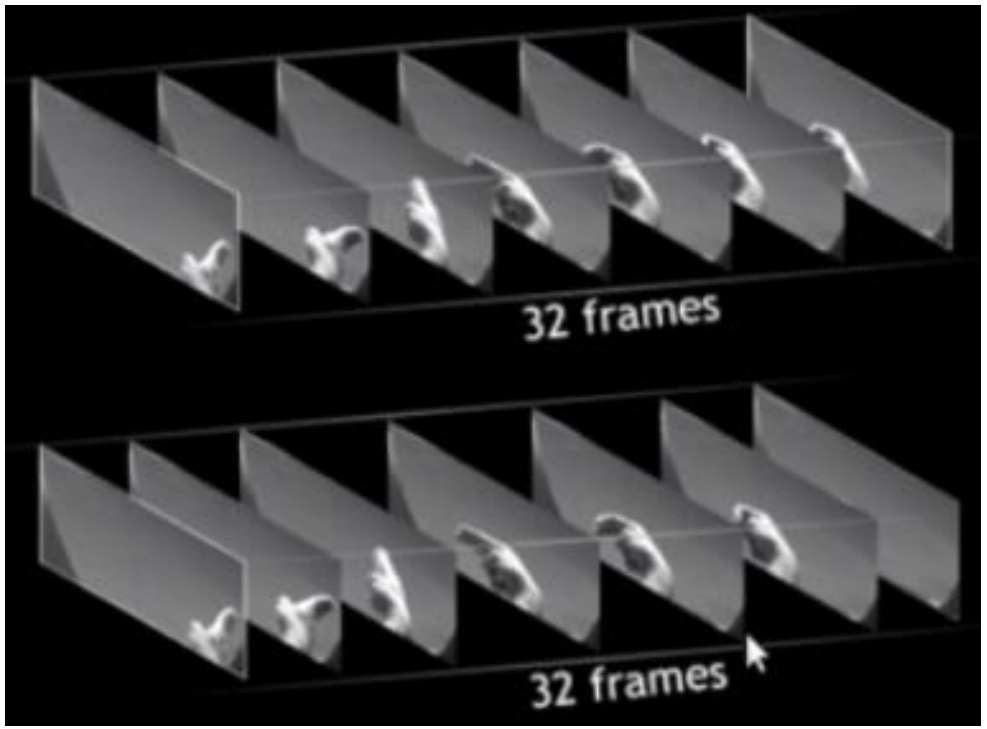
左右移動



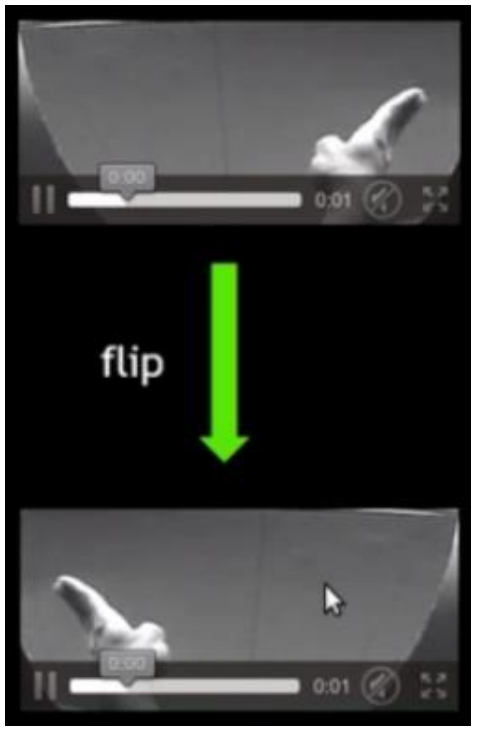
上下移動

データ・オーグメンテーション

Temporal augmentation/Generating new training data

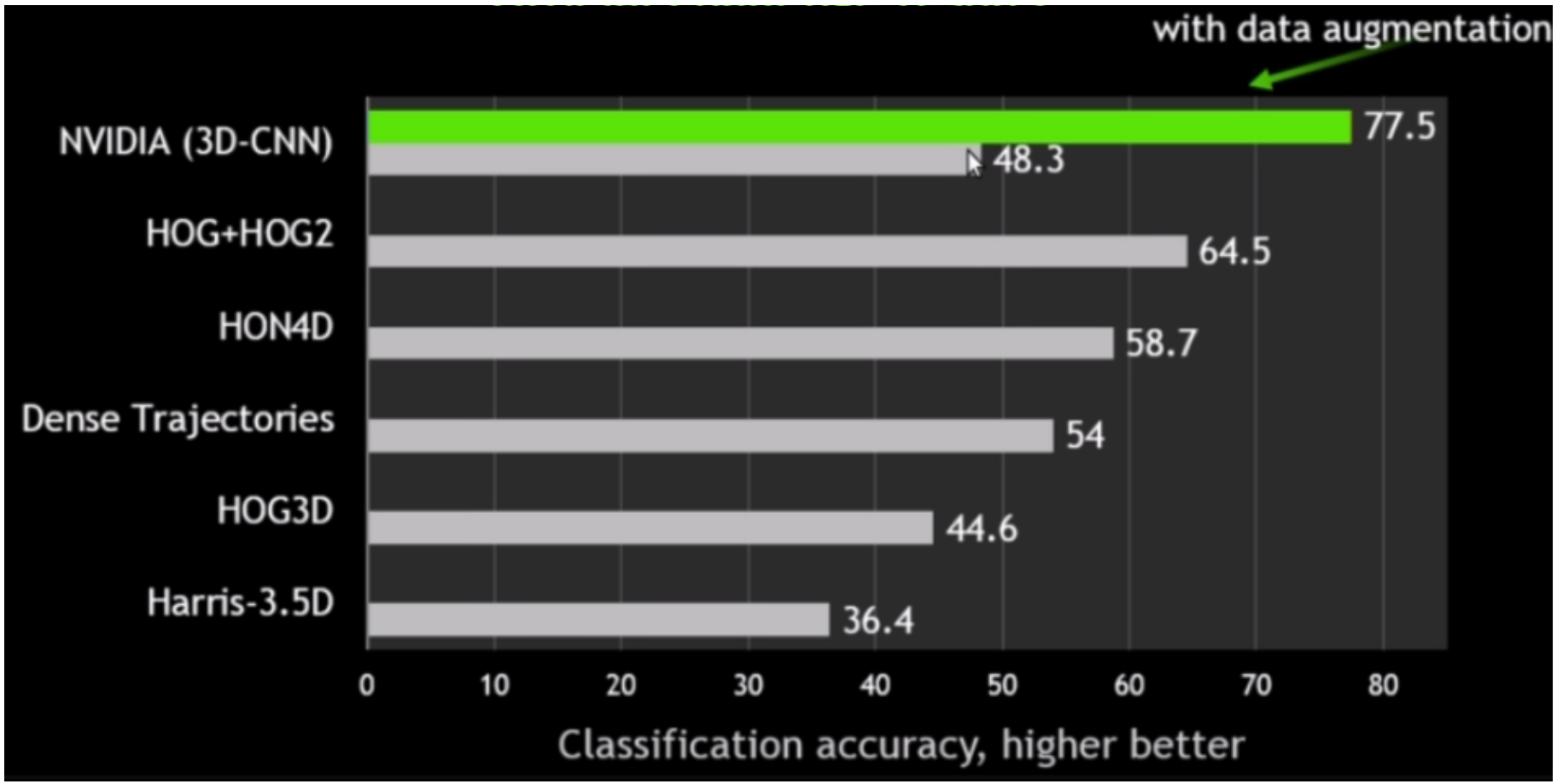


時間方向にフレームをずらす

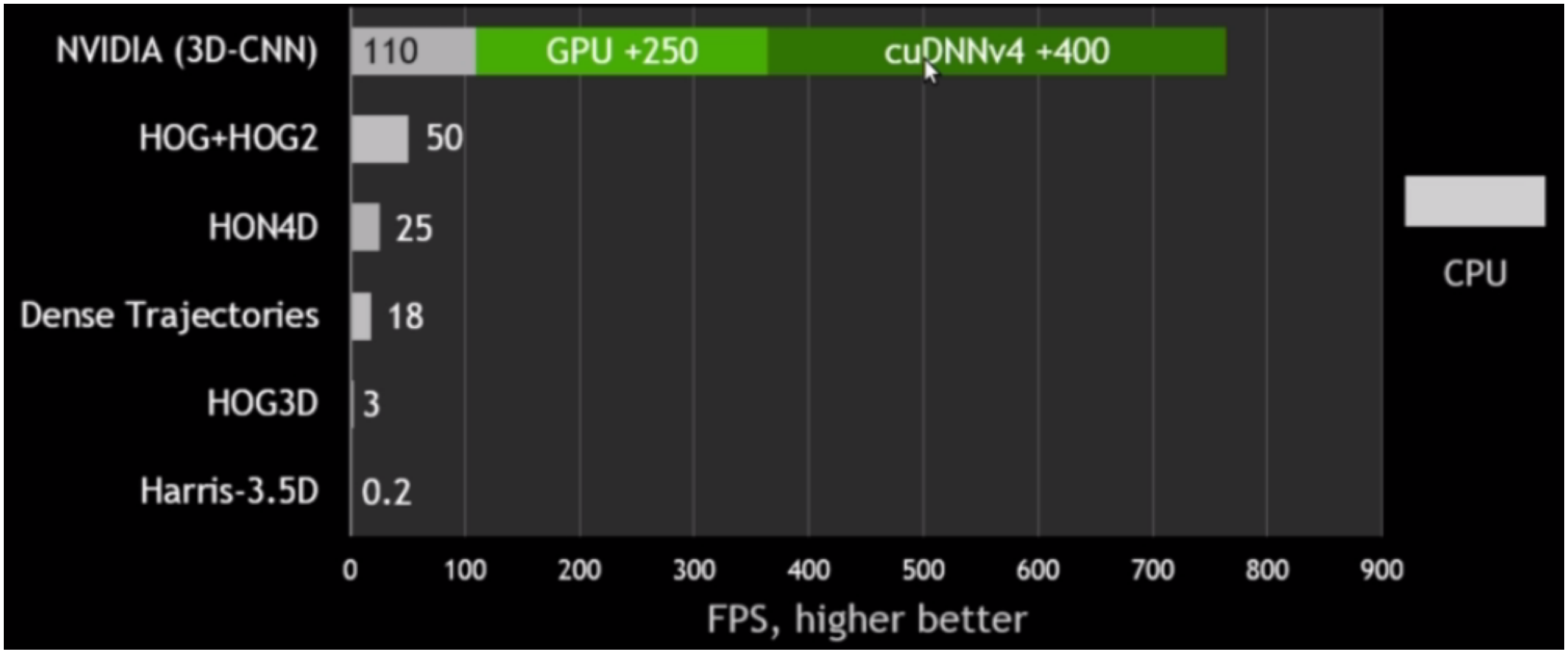


フリップ

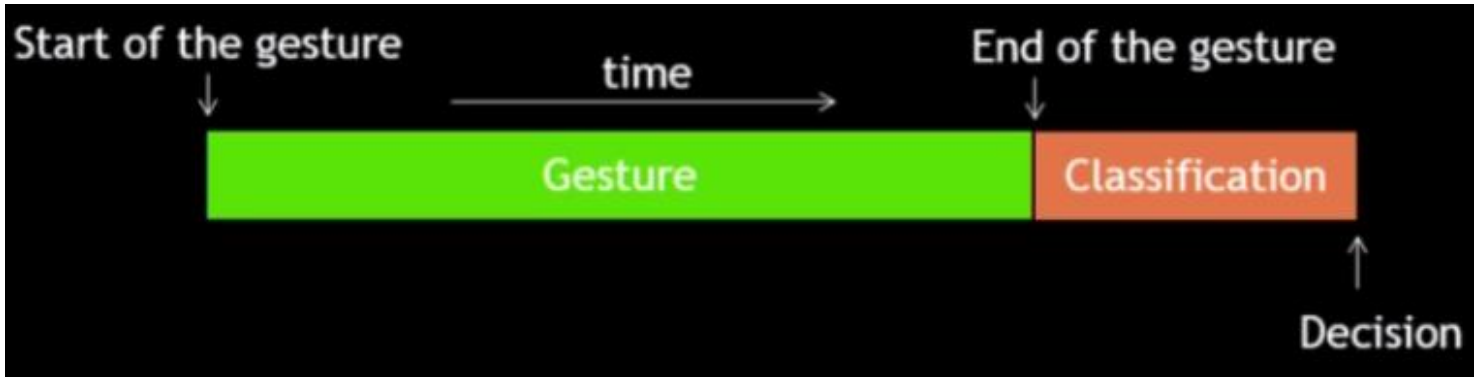
公式の結果

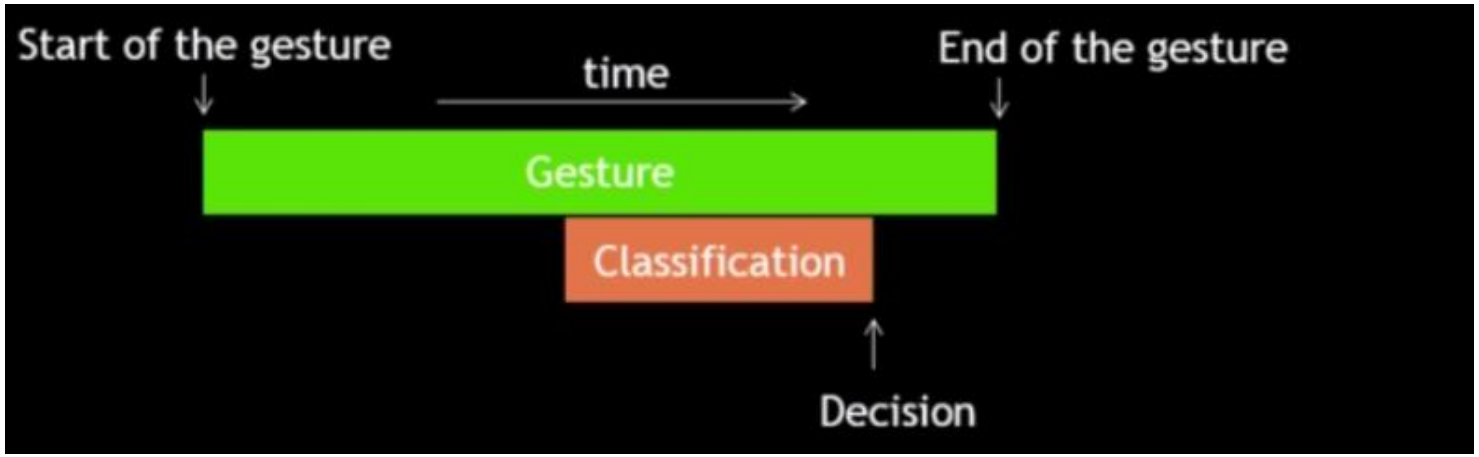

with data augmentation
NVIDIA (3D-CNN)
77.5
48.3
HOG+HOG2
64.5
HON4D
58.7
Dense Trajectories
54
HOG3D
44.6
Harris-3.5D
36.4
0
10
20
30
40
50
60
70
80
Classification accuracy, higher better

認識速度


NVIDIA (3D-CNN)
110
GPU +250
cuDNNv4 +400
HOG+HOG2
50
HON4D
25
Dense Trajectories
18
HOG3D
3
Harris-3.5D
0.2
CPU
0
100
200
300
400
500
600
700
800
900
FPS, higher better

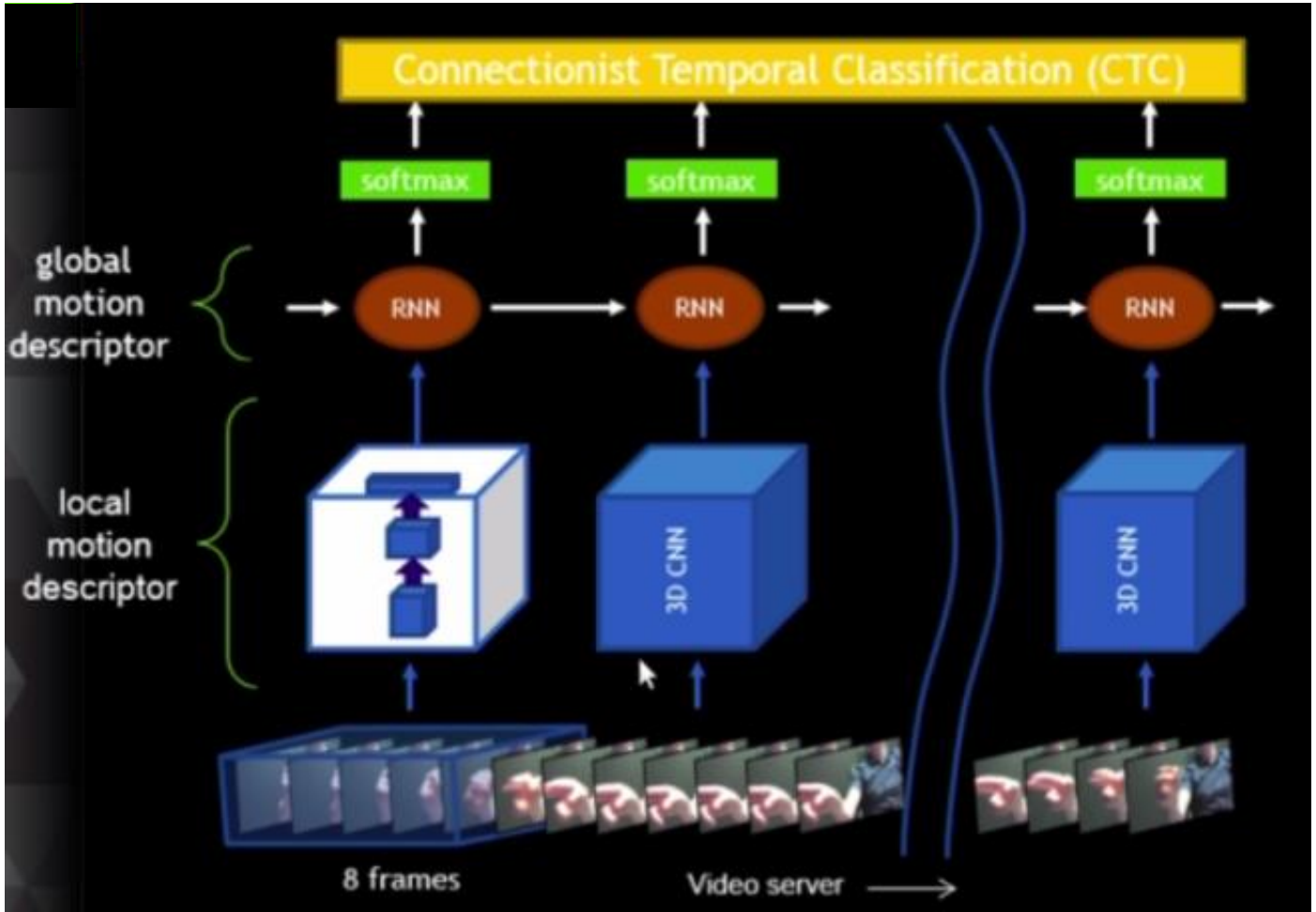
認識の遅延


Start of the gesture
time
End of the gesture
Gesture
Classification
Decision


Start of the gesture
time
End of the gesture
Gesture
Classification
Decision

オンライン・ジェスチャー認識

R3DCNN



INTELLIGENT VIDEO ANALYSIS SYSTEM BASED ON GPU AND DISTRIBUTED ARCHITECTURE

Shiliang Pu Executive Vice President, Hikvision Research Institute

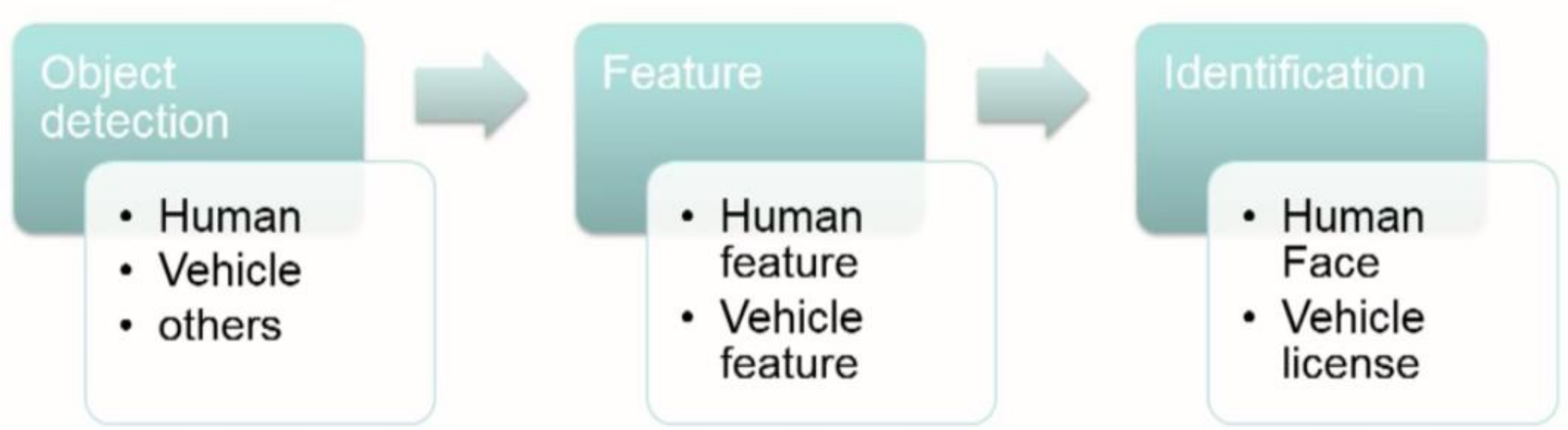
監視カメラが抱える問題

高解像度化　VS　ストレージ

複雑さ　VS　精度

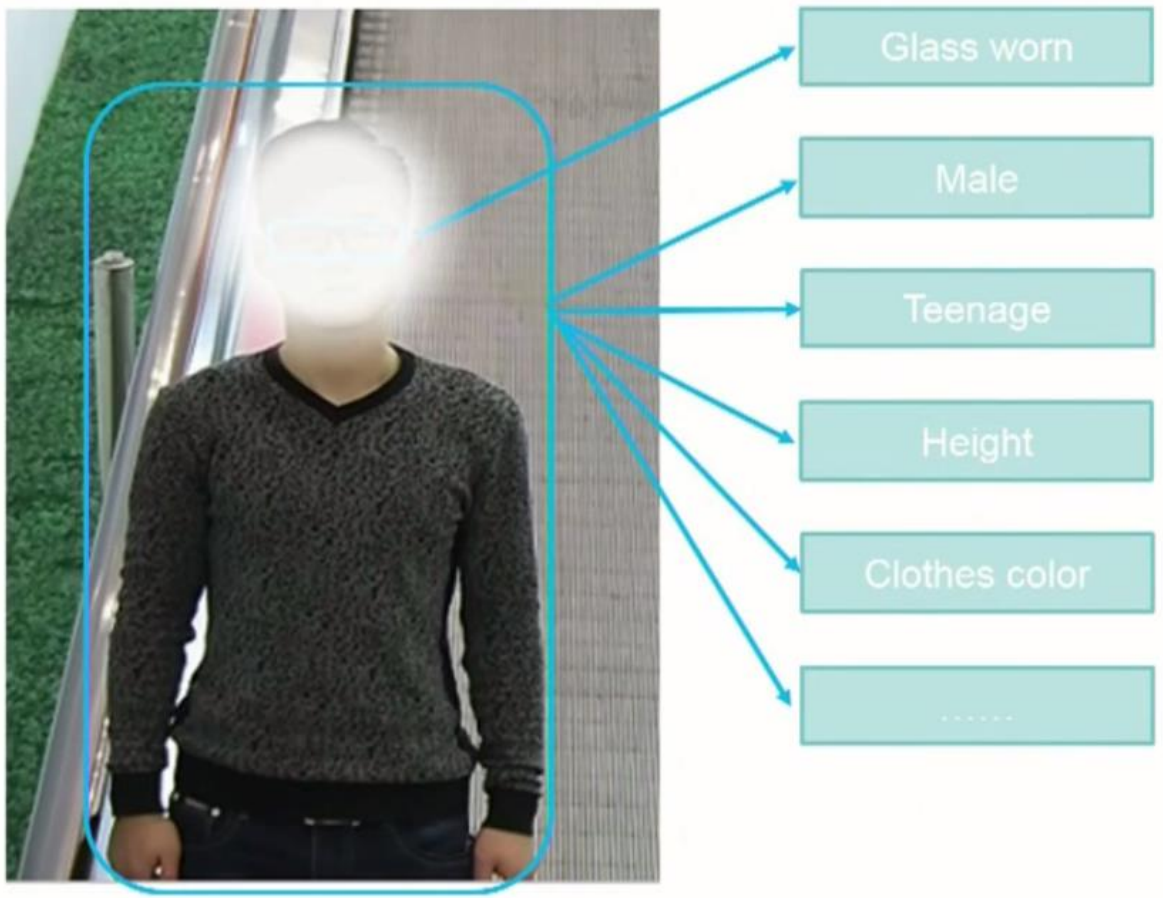
大量のデータ　VS　効率

監視カメラ分析システム



認識例

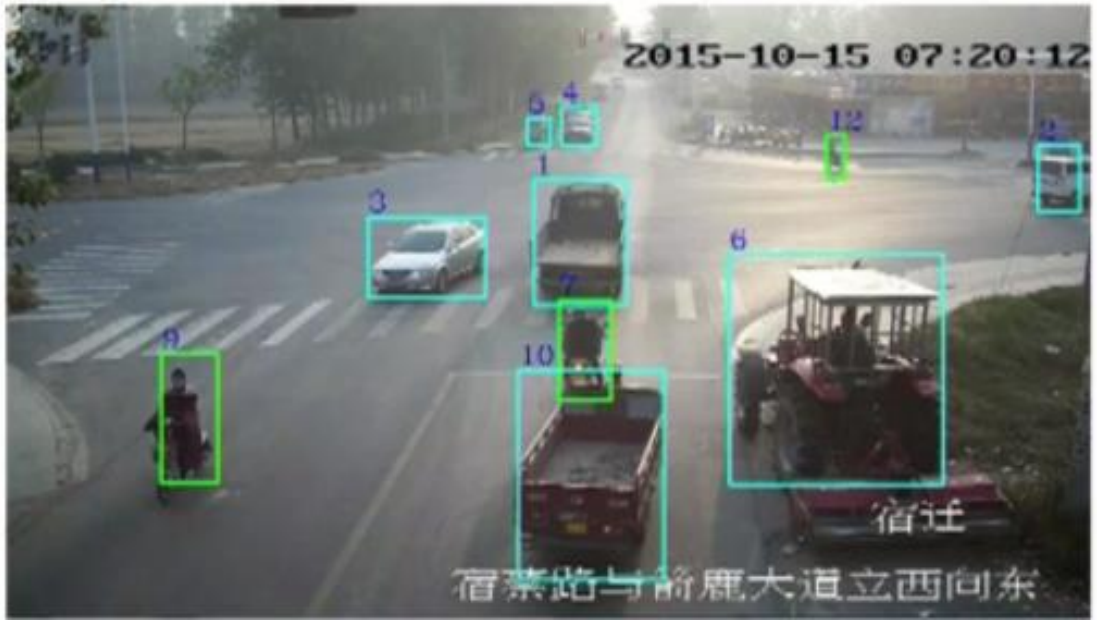

Sun blade closed
Phone calling
Safe Belt
Car
White
Ford Fiesta
皖A？？66R


Glass worn
Male
Teenage
Height
Clothes color
……

複雑なシーンコンテンツは従来のアルゴリズムでは難しい

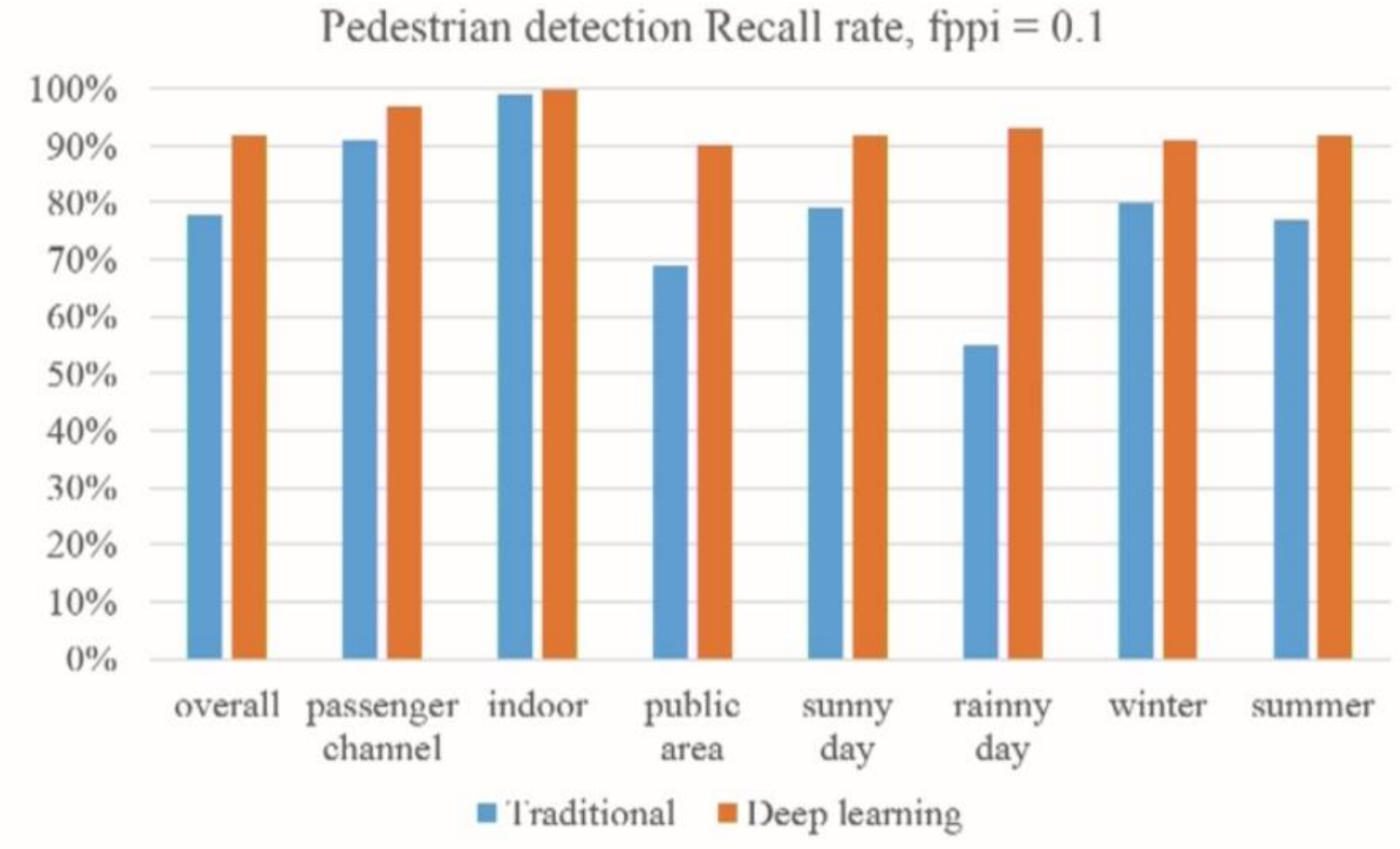
Deep Learningによる飛躍的な認識率改善

従来のアルゴリズム



Deep Learning

Deep Learningによる認識率向上



認識が難しい対象物

Safe belt not fastened

Phone calling

Clothes type

Riding

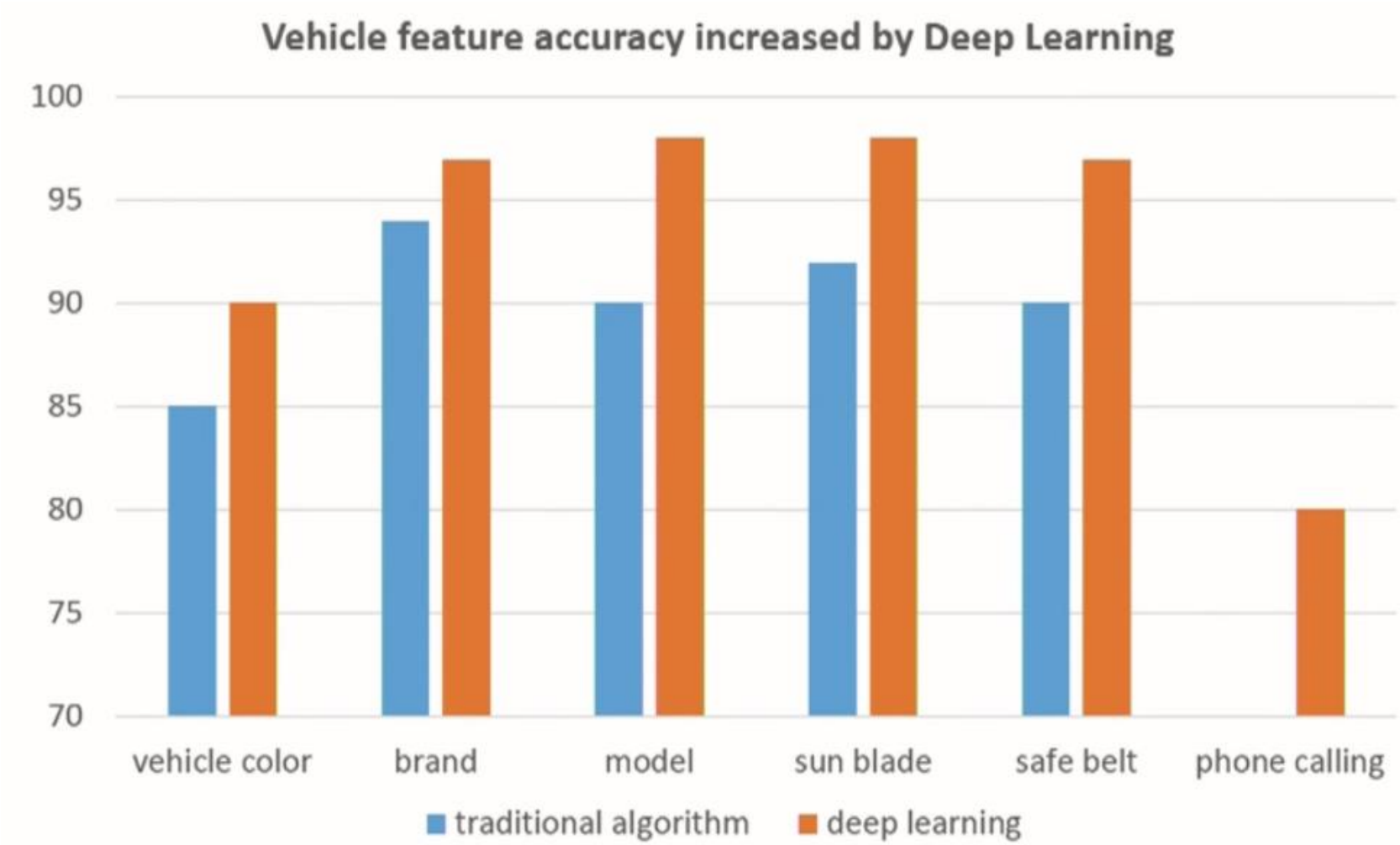
Hanging bag

Mask

backpack

Hat

自動車の特徴における認識率向上



顔認識の例

対象車両の特定

VQA: VISUAL QUESTION ANSWERING

Aishwarya Agrawal Ph.D. Student, Virginia Tech

VQA

Visual Answering Questions

静止画について自然言語の自由回答質問を与え、自然言語の回答を生成する

What is the mustache made of?

AI System

bananas

cloudcv.org/vqa/?useVoice=1&listenAnswer=1

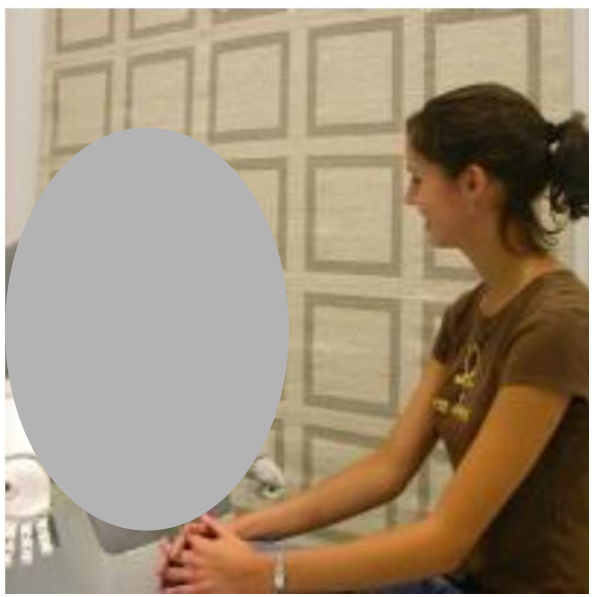
使用用途

視覚障害者の補助

通りを渡っても安全ですか？

監視カメラ

赤いシャツを着た男性が乗り去った車の種類は？



ロボットとの会話

ノートPCは2階の寝室にある？

VQA データセット

Tsung-Yi Lin *et al.* "Microsoft COCO: Common Objects in COntext." ECCV 2014.
http://mscoco.org/

MSCOCOの画像データ

amazonmechanical turk
Artificial Artificial Intelligence

What color are her eyes?
What is the mustache made of?

How many slices of pizza are there?
Is this a vegetarian pizza?

Is this person expecting company?
What is just under the tree?

Does it appear to be rainy?
Does this person have 20/20 vision?

自由回答形式の質問
複数選択肢がある質問

VQA データセット

25万点以上のイメージデータ（MSCOCO＋5万のイラストデータ)

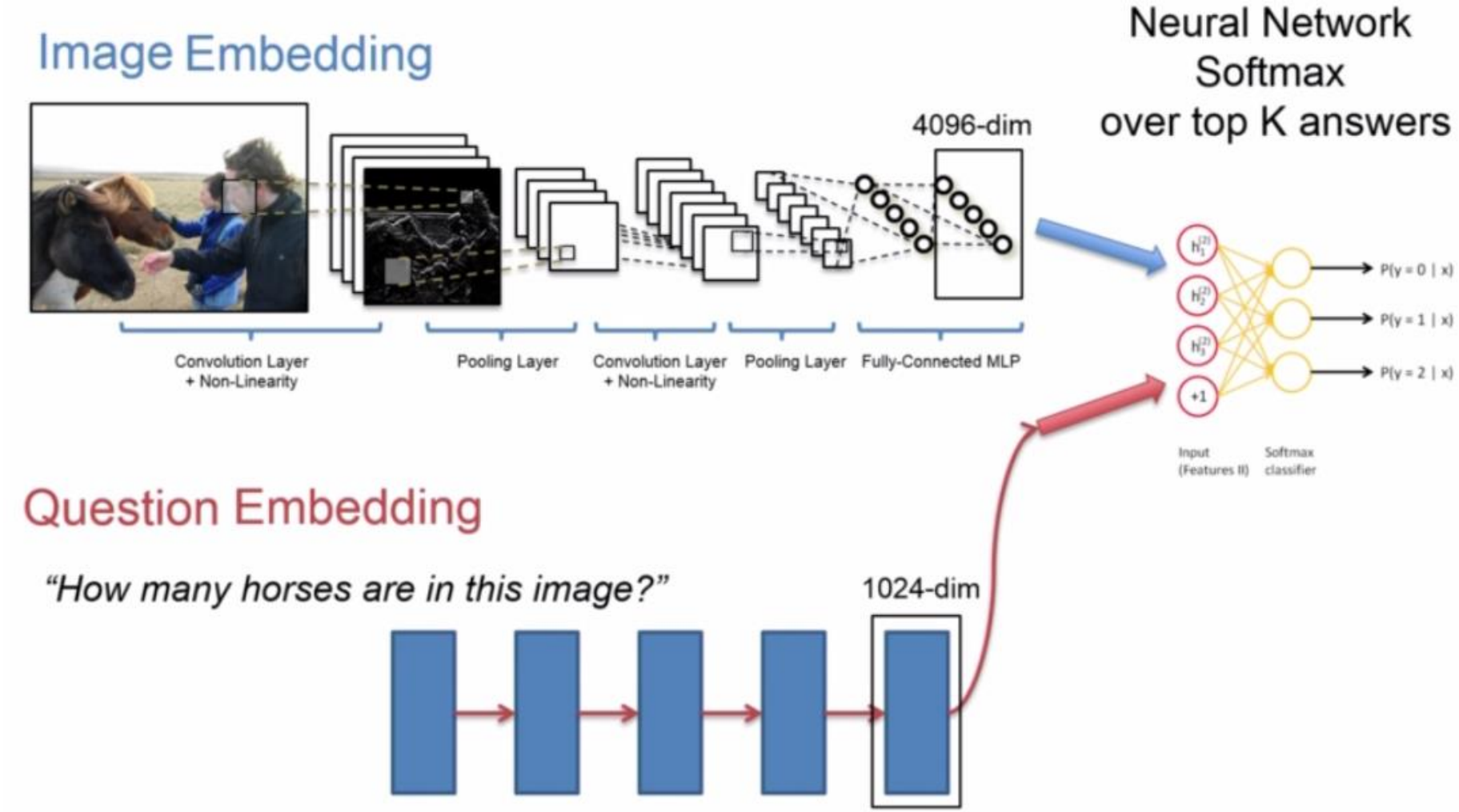
75万の質問（3質問/イメージ)

1000万の回答

データセットはこちら

http://www.visualqa.org/

２チャンネル　VQAモデル



精度の指標

$$\mathrm{Acc}(ans) = \min\left\{\frac{\#\text{humans that said } ans}{3}, 1\right\}$$

1940. COCO_train2014_000000012015

Open-Ended/Multiple-Choice/Ground-Truth

Q: WHAT OBJECT IS THIS

Ground Truth Answers:

(1) television	(6) television
(2) tv	(7) television
(3) tv	(8) tv
(4) tv	(9) tv
(5) television	(10) television

Q: How old is this TV?

Ground Truth Answers:

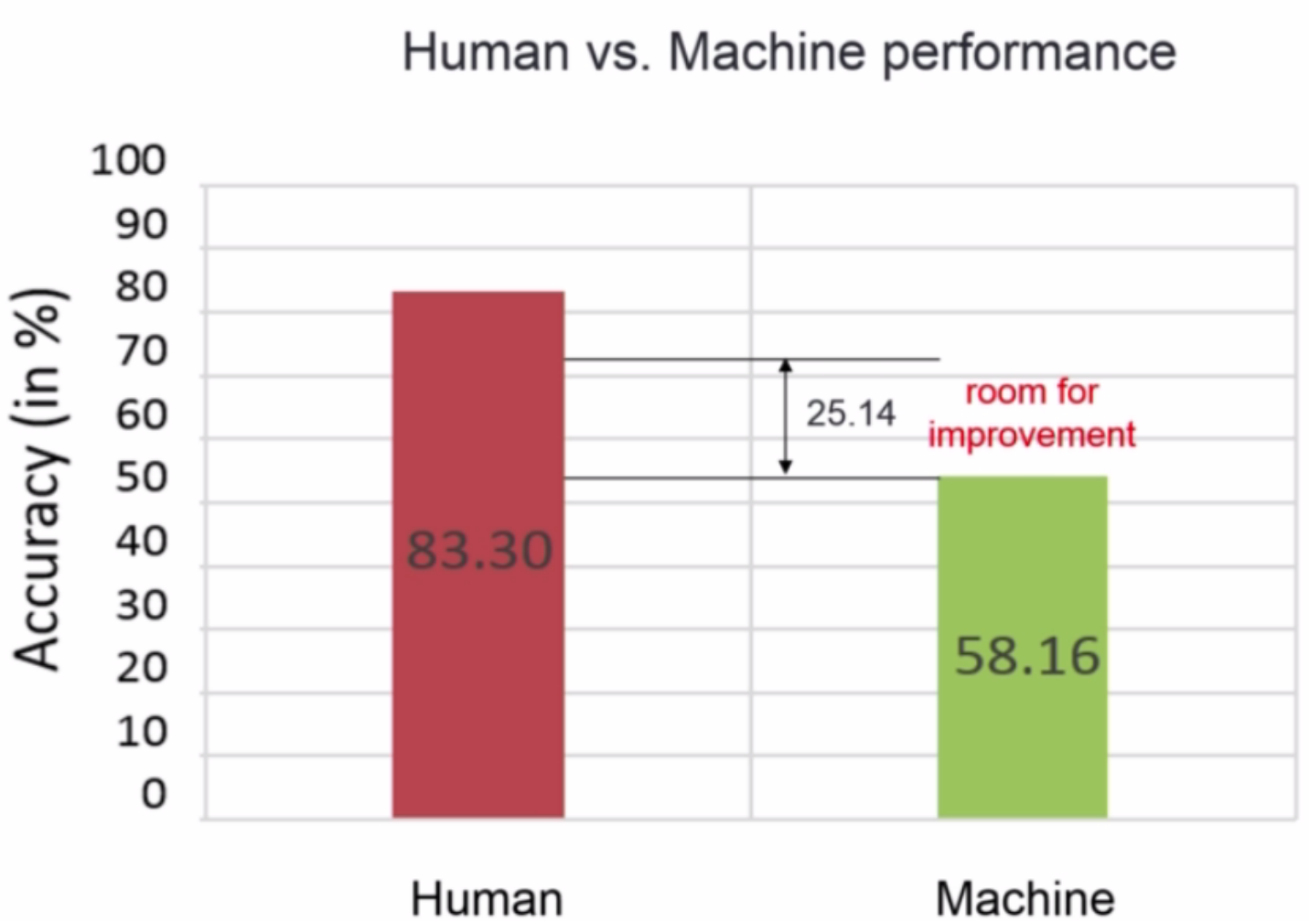
(1) 20 years	(6) old
(2) 35	(7) 80 s
(3) old	(8) 30 years
(4) more than thirty years old	(9) 15 years
(5) old	(10) very old

Q: Is this TV upside-down?

Ground Truth Answers:

(1) yes	(6) yes
(2) yes	(7) yes
(3) yes	(8) yes
(4) yes	(9) yes
(5) yes	(10) yes

自由回答形式の質問問題の精度



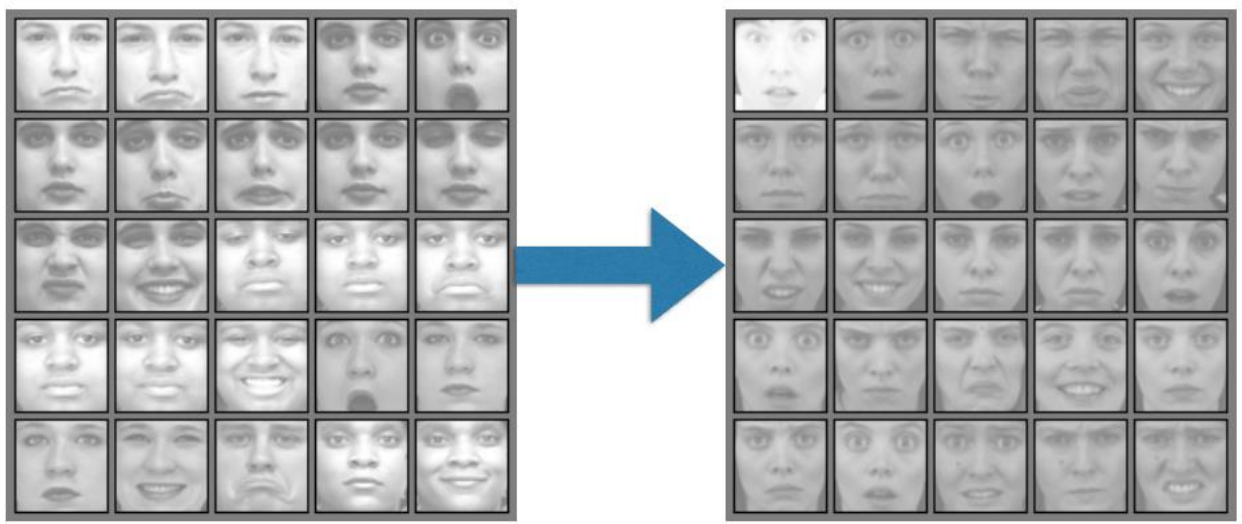
GENERATIVE ADVERSARIAL NETWORKS

Ian Goodfellow Senior Research Scientist, OpenAI

Generative Adversarial Networks

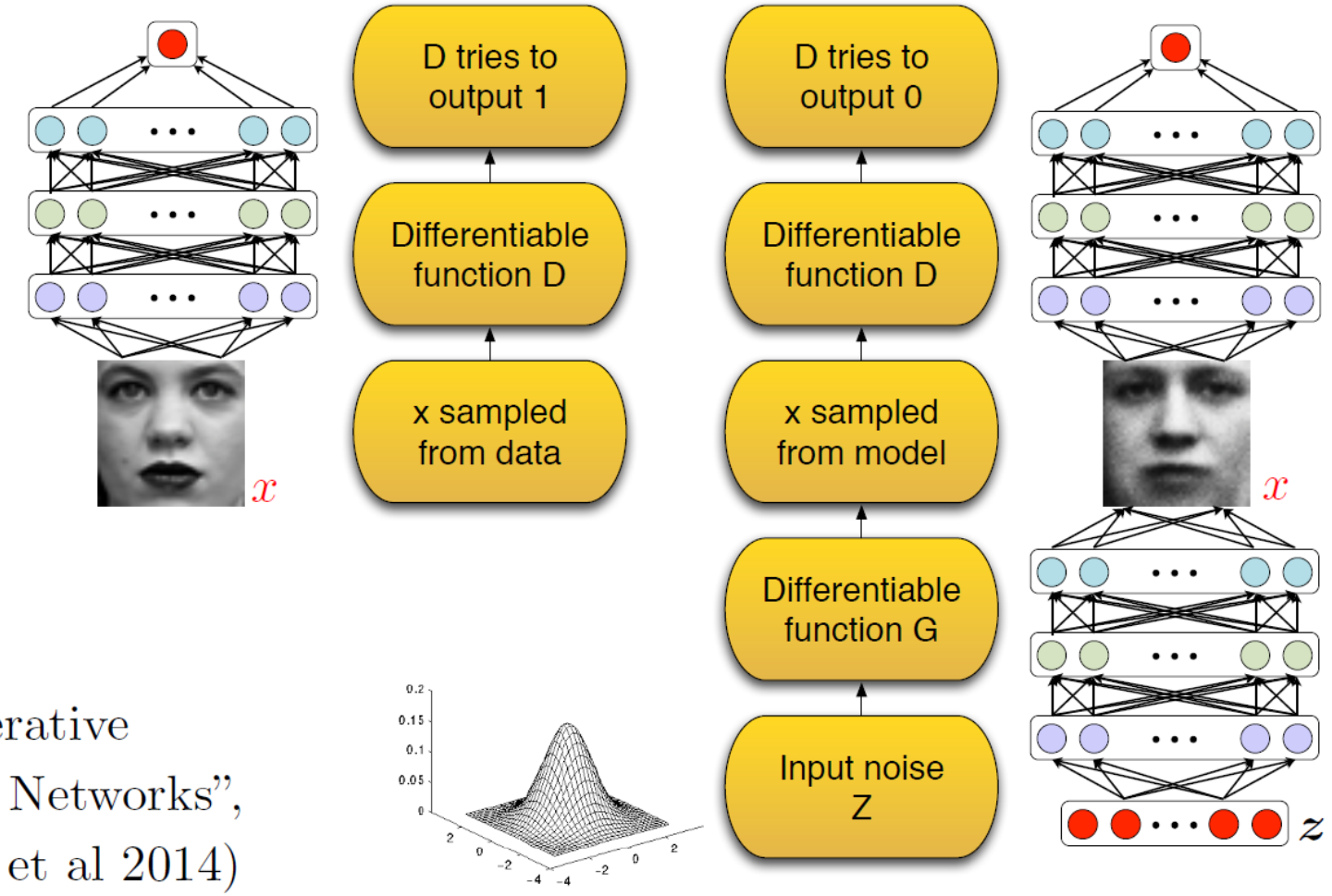
Generative Modeling

- Have training examples: $\boldsymbol{x} \sim p_{\text{train}}(\boldsymbol{x})$
- Want a model that can draw samples: $\boldsymbol{x} \sim p_{\text{model}}(\boldsymbol{x})$
- Want $p_{\text{model}}(\boldsymbol{x}) = p_{\text{data}}(\boldsymbol{x})$



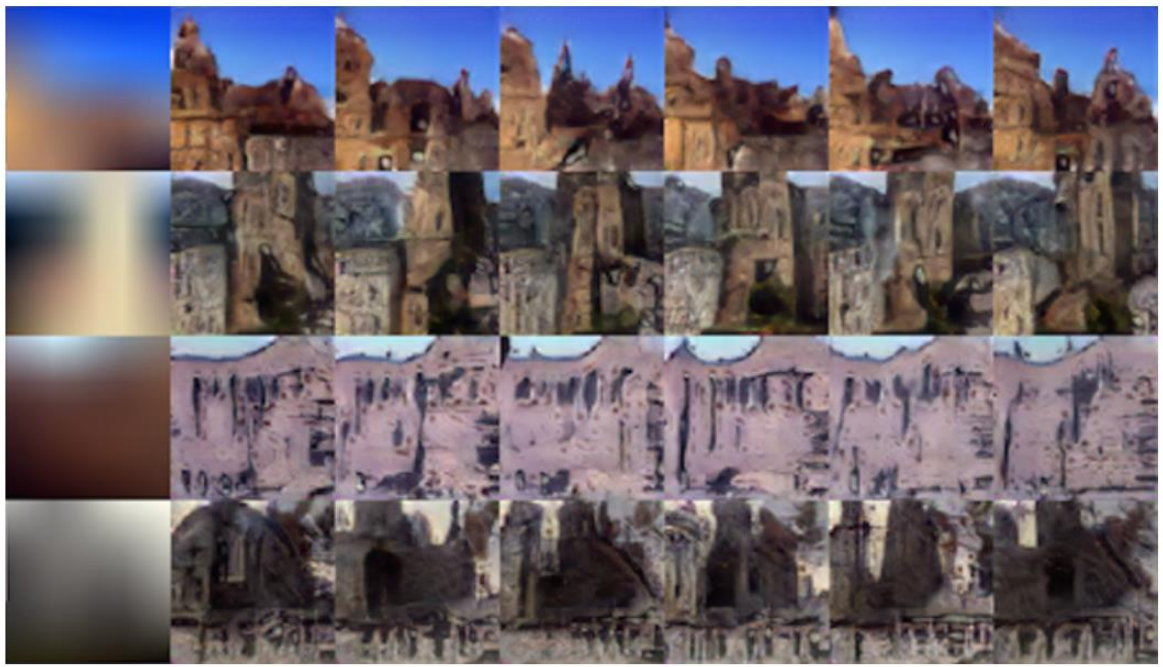
(Images from Toronto Face Database)

Generative Adversarial Network

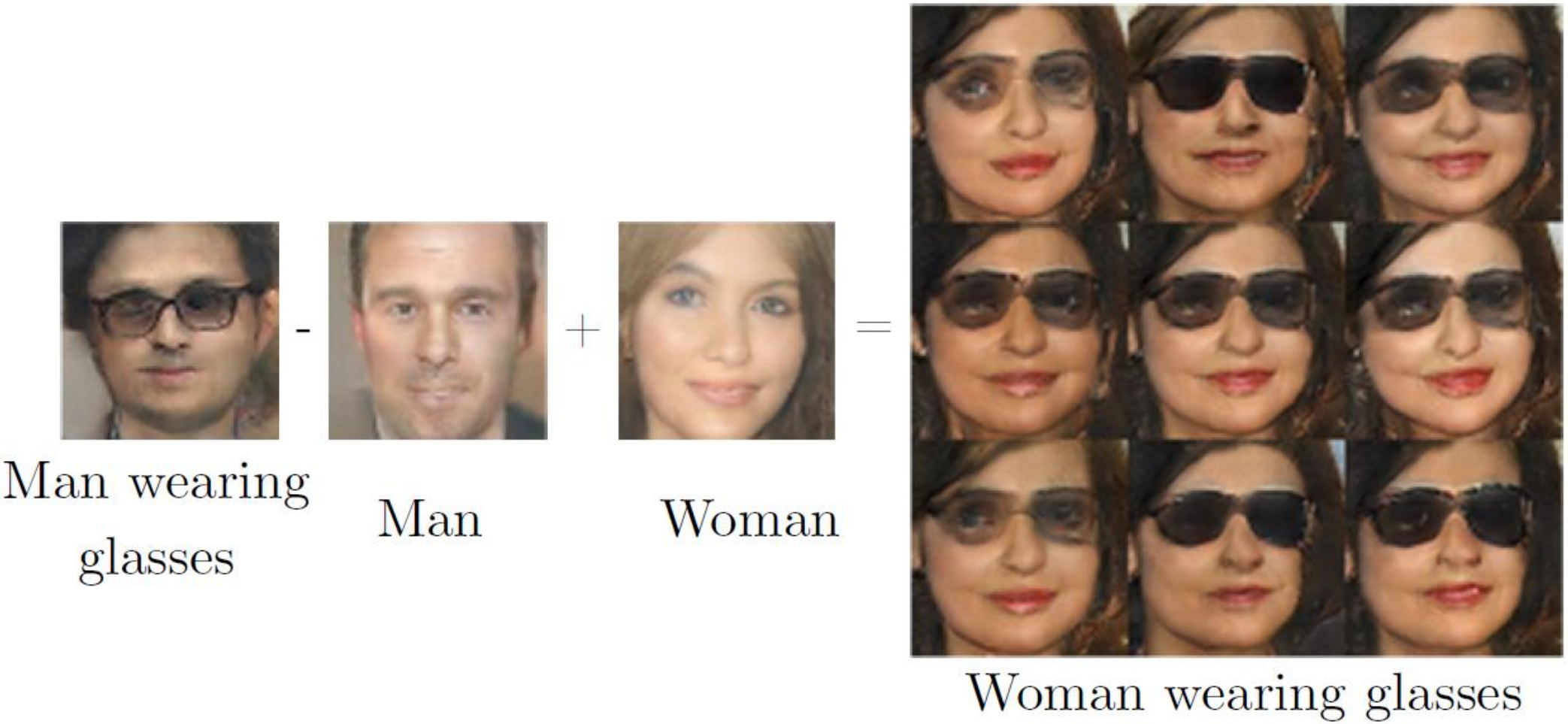


("Generative Adversarial Networks", Goodfellow et al 2014)

LAPGAN/DCGAN



DCGANのベクトル演算性

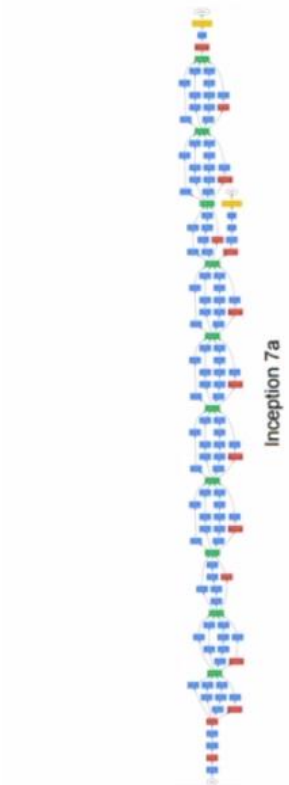


MXNET: FLEXIBLE DEEP LEARNING FRAMEWORK FROM DISTRIBUTED GPU CLUSTERS TO EMBEDDED SYSTEMS

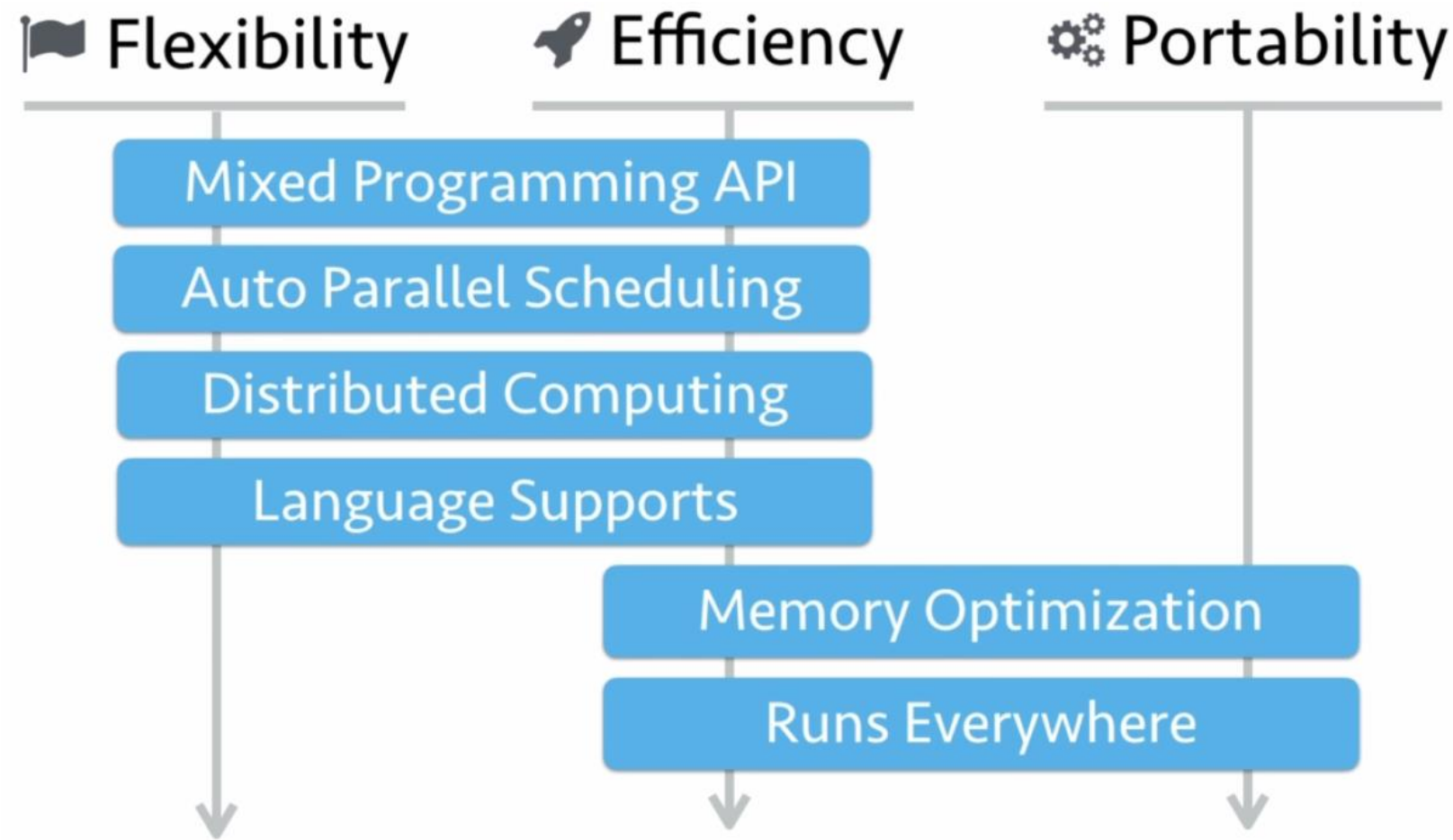
Mu Li Ph.D. Student, Carnegie Mellon University
Tianqi Chen Ph.D. Student, University of Washington

MXNet：分散GPUクラスターから組込みシステムまで

Flexibility　Efficiency　Portability

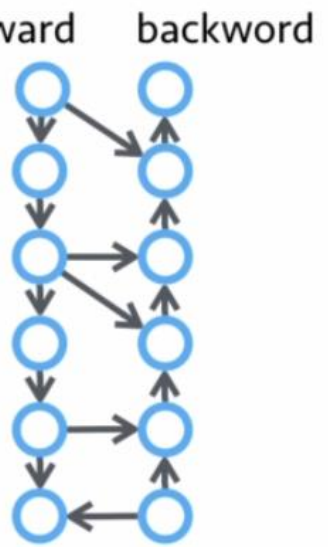


MXNet：分散GPUクラスターから組込みシステムまで



ミックス プログラミング API

Computational Graph
of the Deep Architecture



Needs heavy optimization,
fits **declarative** programs

Updates and Interactions
with the graph

- Parameter update
- Beam search
- Feature extraction ...

$$w = w - \eta \, \partial f(w)$$

Needs mutation and more
language native features, good for
imperative programs

MXNet：両方の実装が可能

Imperative
NDArray API

```
>>> import mxnet as mx
>>> a = mx.nd.zeros((100, 50))
>>> a.shape
(100L, 50L)
>>> b = mx.nd.ones((100, 50))
>>> c = a + b
>>> b += c
```

Declarative
Symbolic Executor

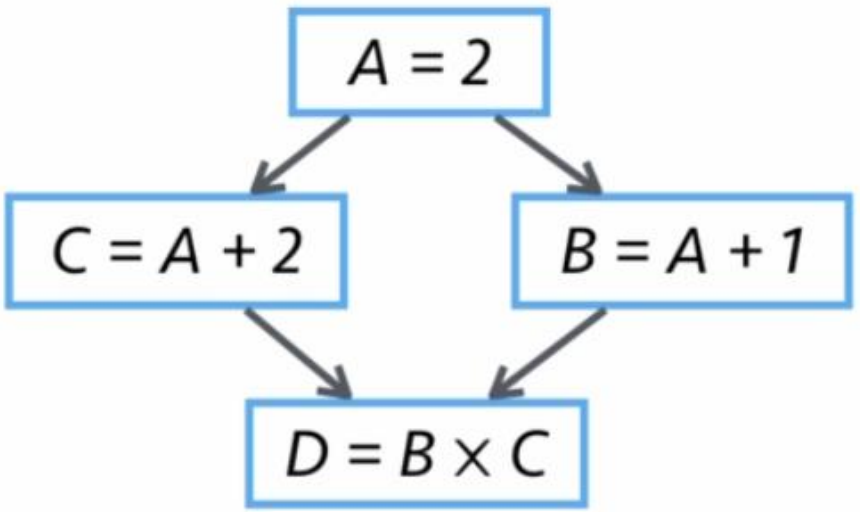
```
>>> import mxnet as mx
>>> net = mx.symbol.Variable('data')
>>> net = mx.symbol.FullyConnected(data=net, num_hidden=128)
>>> net = mx.symbol.SoftmaxOutput(data=net)
>>> type(net)
<class ‘mxnet.symbol.Symbol’>
>>> texec = net.simple_bind(data=data_shape)
```

自動パラレルスケジューリング

Write **serial** programs

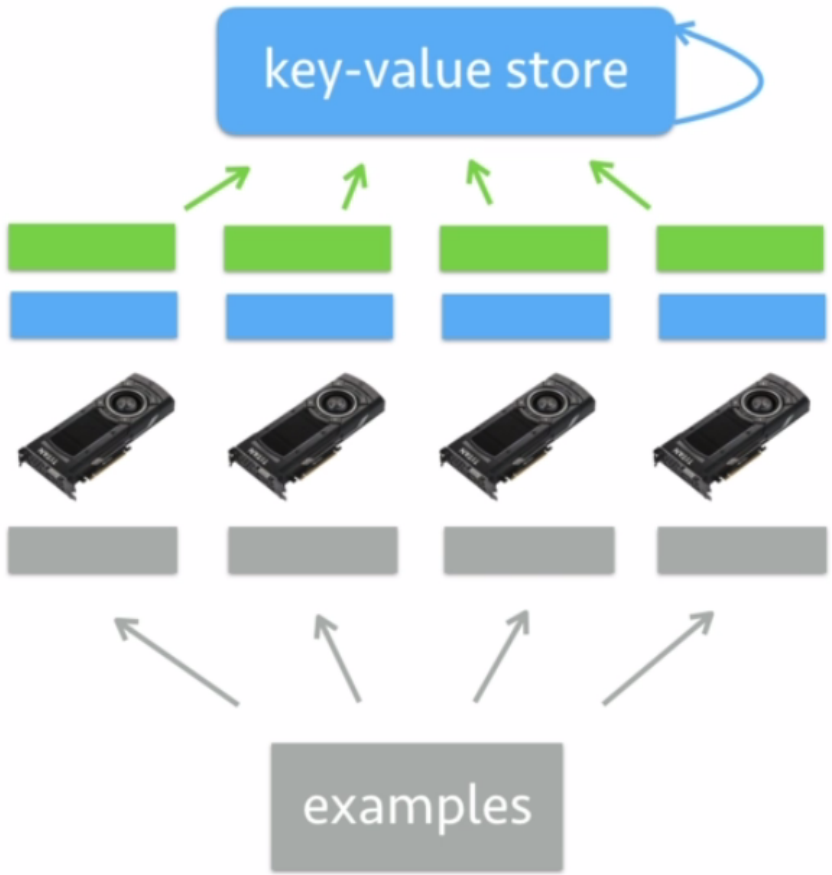
```
>>> import mxnet as mx
>>> A = mx.nd.ones((2,2)) *2
>>> C = A + 2
>>> B = A + 1
>>> D = B * C
```

Run in **parallel**



分散コンピューティング

データ並列



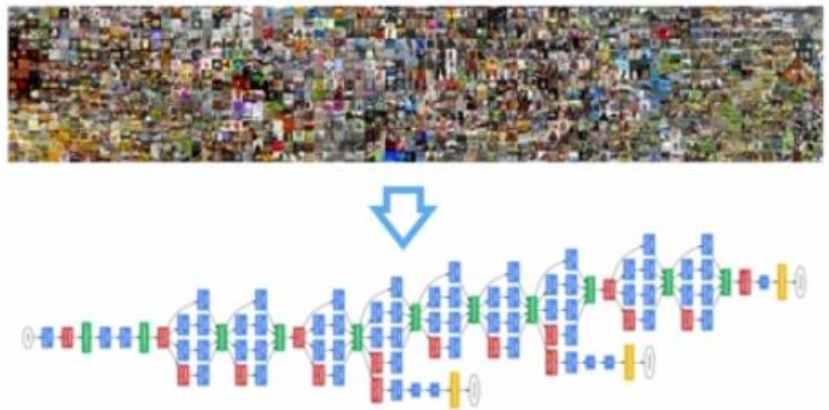
1. Read a data partition
2. Pull the parameters
3. Compute the gradient
4. Push the gradient
5. Update the weight

分散コンピューティング：実装

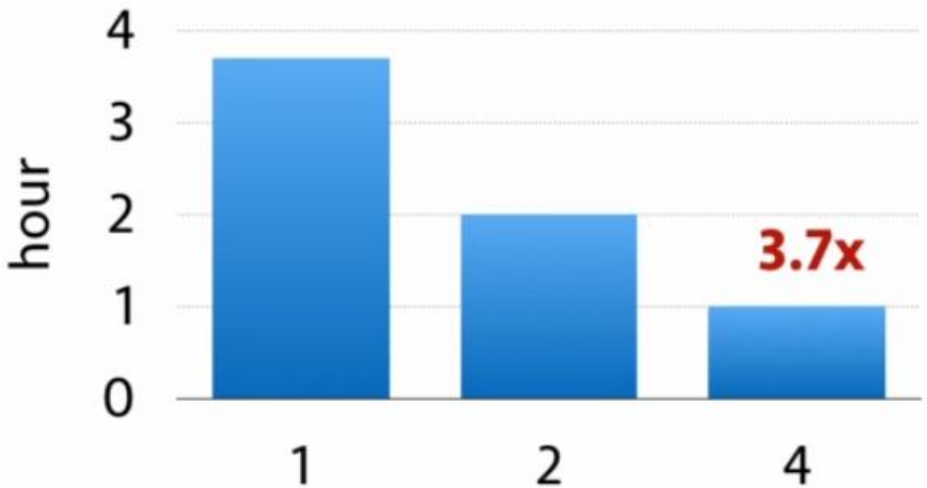
```
% create executor for each GPU
execs = [symbol.bind(mx.gpu(i)) for i in range(ngpu)]
% w -= learning_rate * grad
kvstore.set_updater(…)
% iterating on data
for dbatch in train_iter:
    % iterating on GPUs
    for i in range(ngpu):
        % read a data partition
        copy_data_slice(dbatch, execs[i])
        % pull the parameters
        for key in update_keys:
            kvstore.pull(key, execs[i].weight_array[key])
        % compute the gradient
        execs[i].forward(is_train=True)
        execs[i].backward()
        % push the gradient
        for key in update_keys:
            kvstore.push(key, execs[i].grad_array[key])
```

分散コンピューティング：性能結果

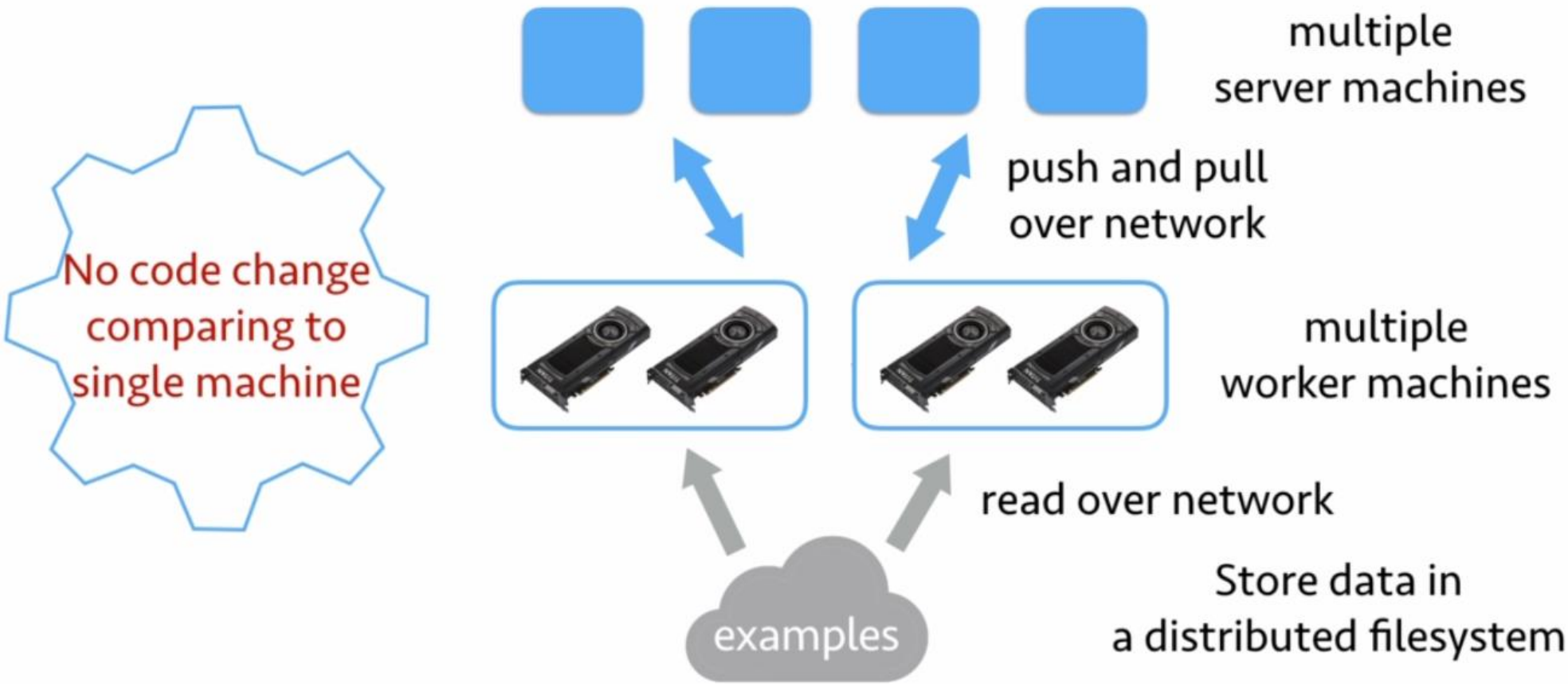
- IMAGENET with 1.2m images and 1,000 classes
- 4 x Nvidia GTX 980
- Google Inception Network



Time for one epoch:

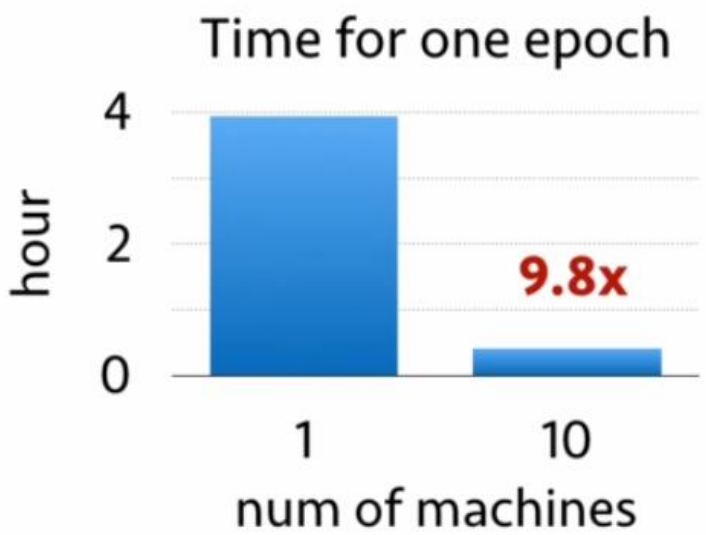


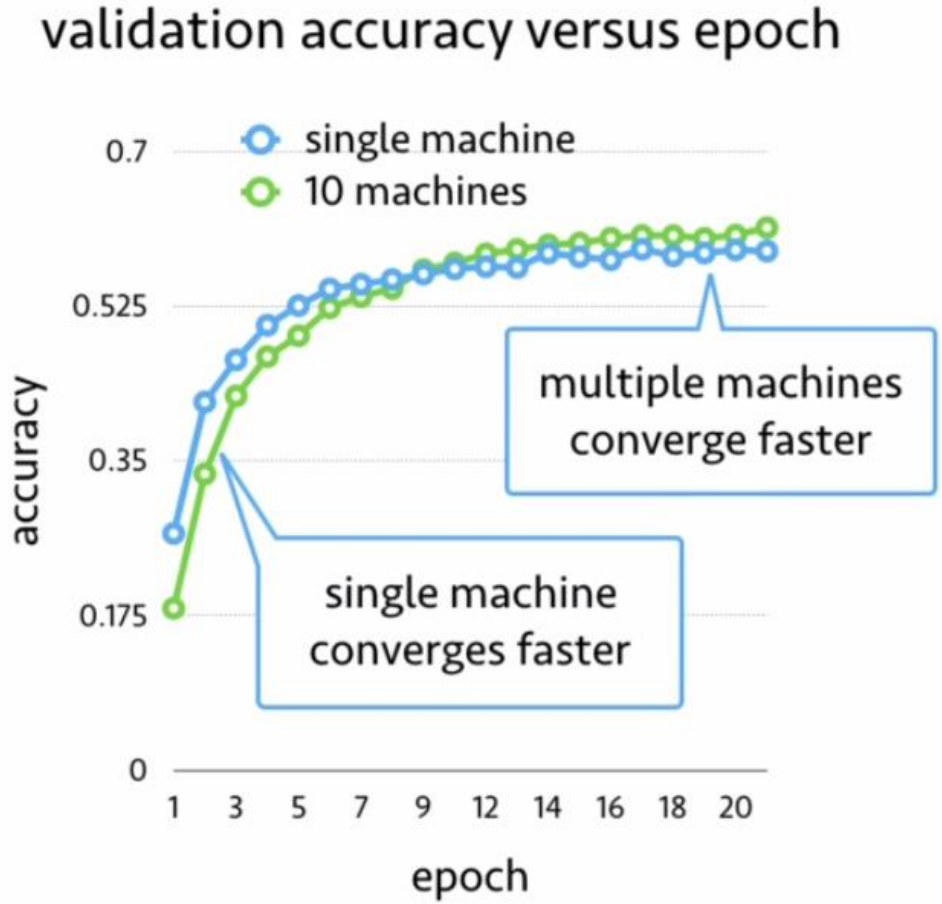
マルチノード 分散コンピューティング



マルチノード分散コンピューティング：性能結果

- ImageNet with 1.2m images and 1,000 classes
- AWS EC2 GPU instance, 4 GPUs per machine
- Google Inception Network





多言語サポート

single implementation of backend system and common operators

performance guarantee regardless which frontend language is used

MinPy：MXNet　Numpy　パッケージ

is the de facto scientific computing package in Python

Great flexibility (500+ operators) but CPU-only

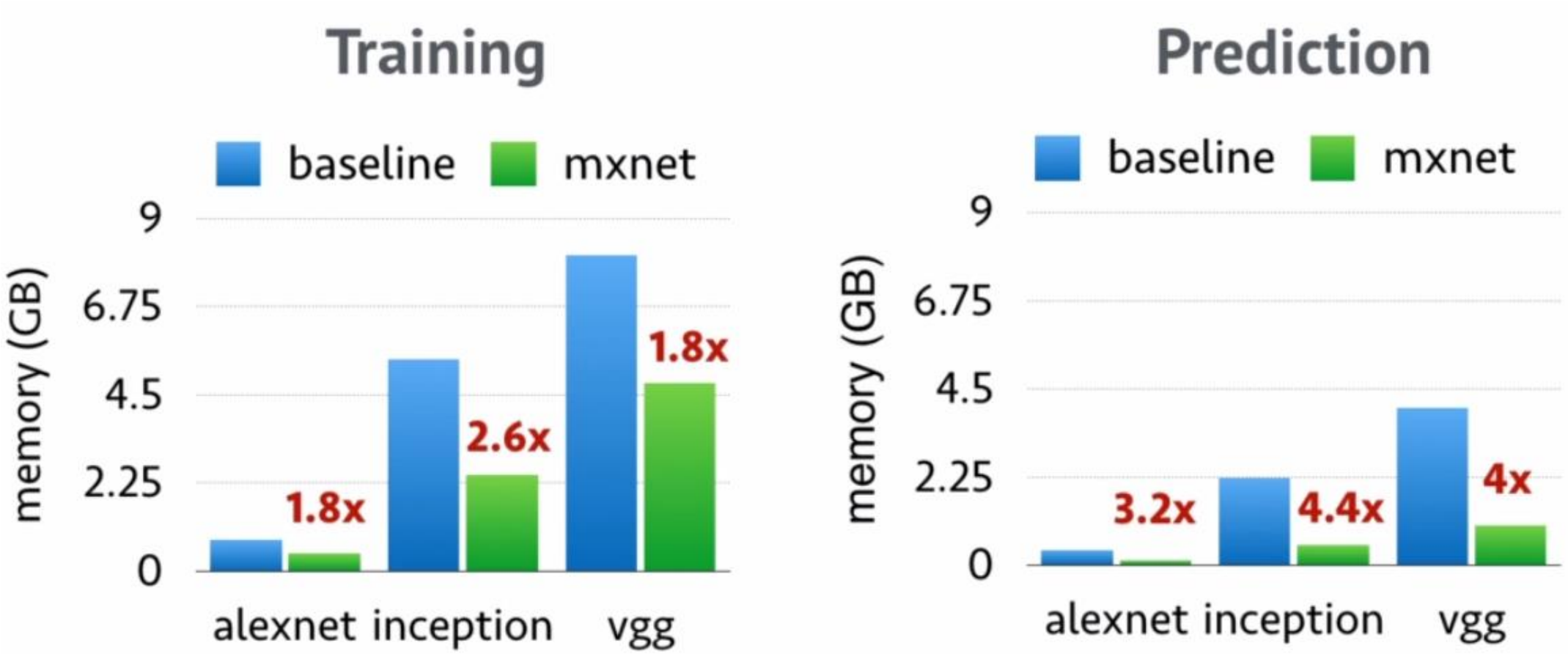
- Native Numpy Integration

```
>>> import numpy as np   ⇨   >>> import minpy as np
```

- Transparent CPU and GPU co-execution

```
>>> x = np.zeros((10, 20)) # call GPU function
>>> y = np.sort(x)         # call CPU function; copy GPU->CPU
>>> z = np.log(y)          # call GPU function; copy CPU->GPU
```

メモリ最適化

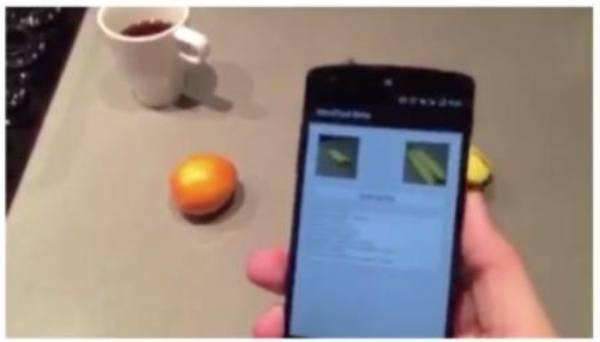

Training
baseline
mxnet
memory (GB)
9
6.75
4.5
2.25
0
1.8x
2.6x
1.8x
alexnet
inception
vgg
Prediction
baseline
mxnet
memory (GB)
9
6.75
4.5
2.25
0
3.2x
4.4x
4x
alexnet
inception
vgg

豊富な動作環境

Beyond

Amalgamation

- Fit the core library with all dependencies into a single C++ source file
- Easy to compile on ...



BlindTool by Joseph Paul Cohen, demo on Nexus 4

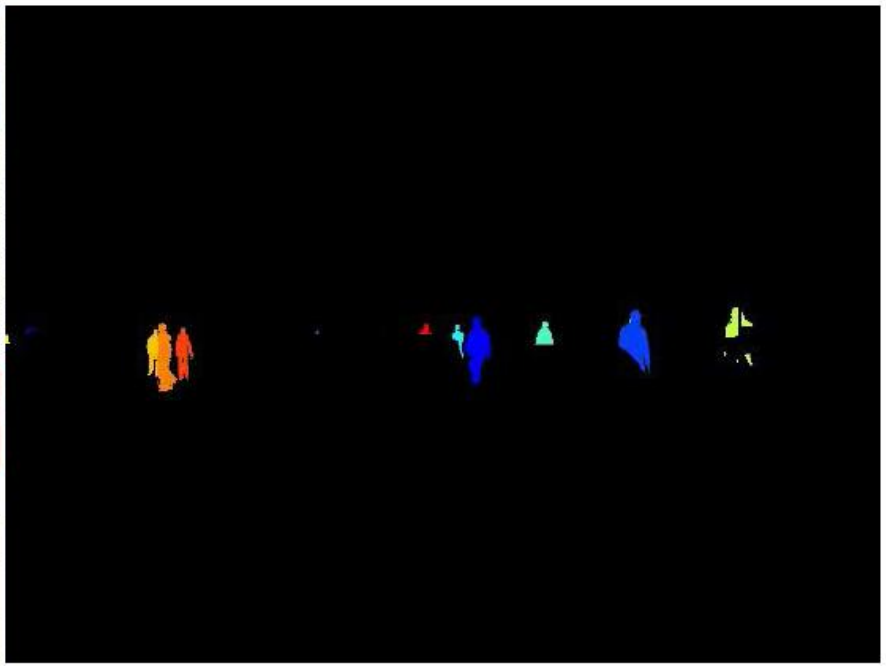
Runs in browser with Javascript

TRAINING MY CAR TO SEE: USING VIRTUAL WORLDS

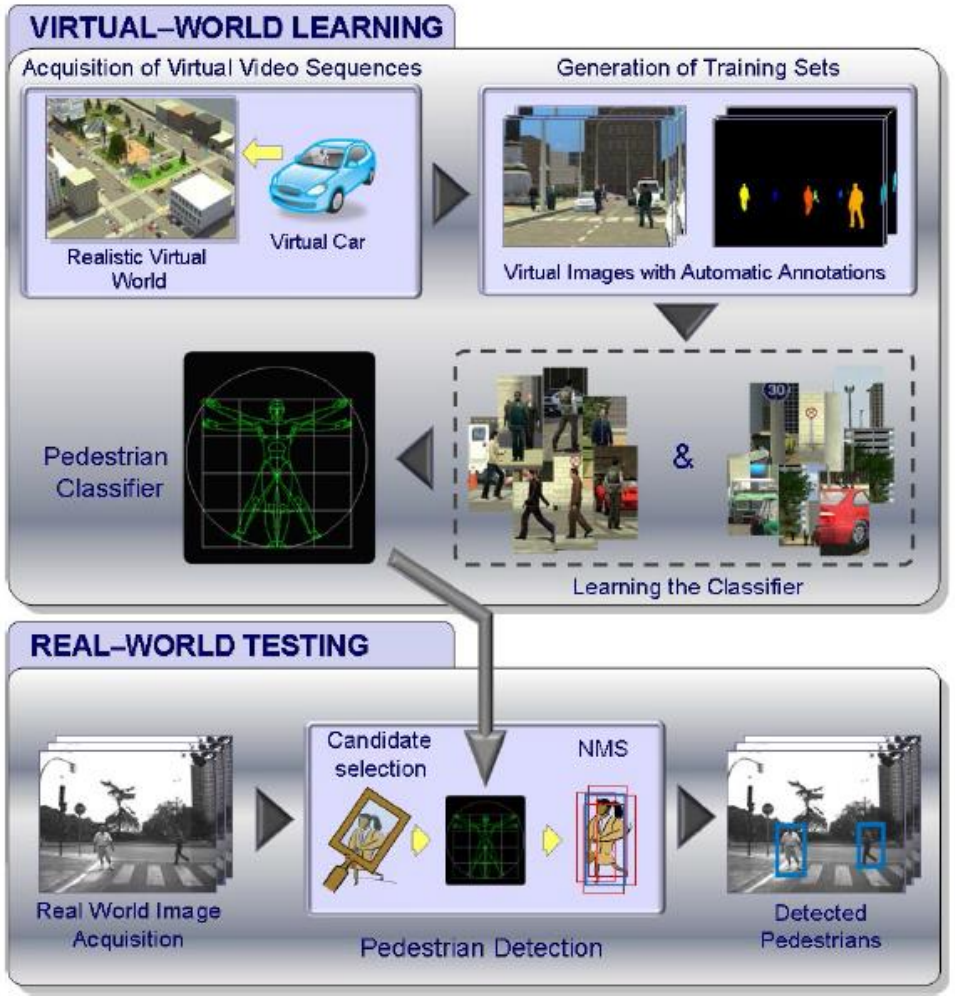
Antonio M. López Principal Investigator & Associate Professor, Computer Vision Center & Universitat Autònoma de Barcelona

車の認識

仮想世界が使用できる？

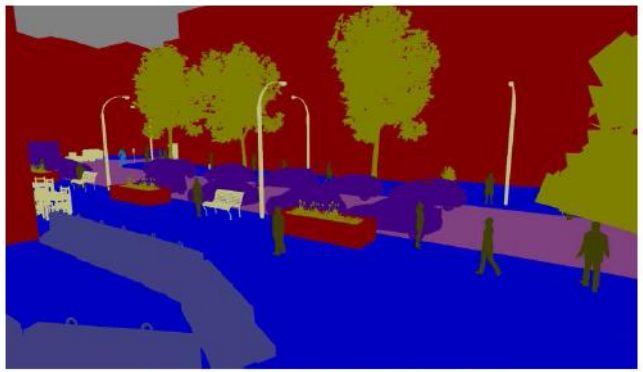


自動注釈付けのための仮想世界



自動注釈付けのための仮想世界

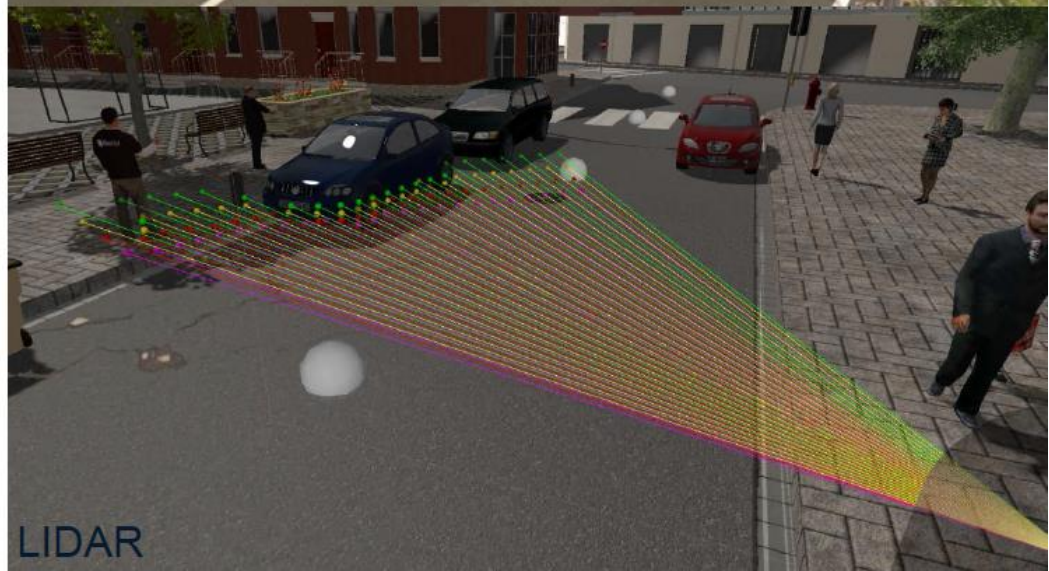
RGB



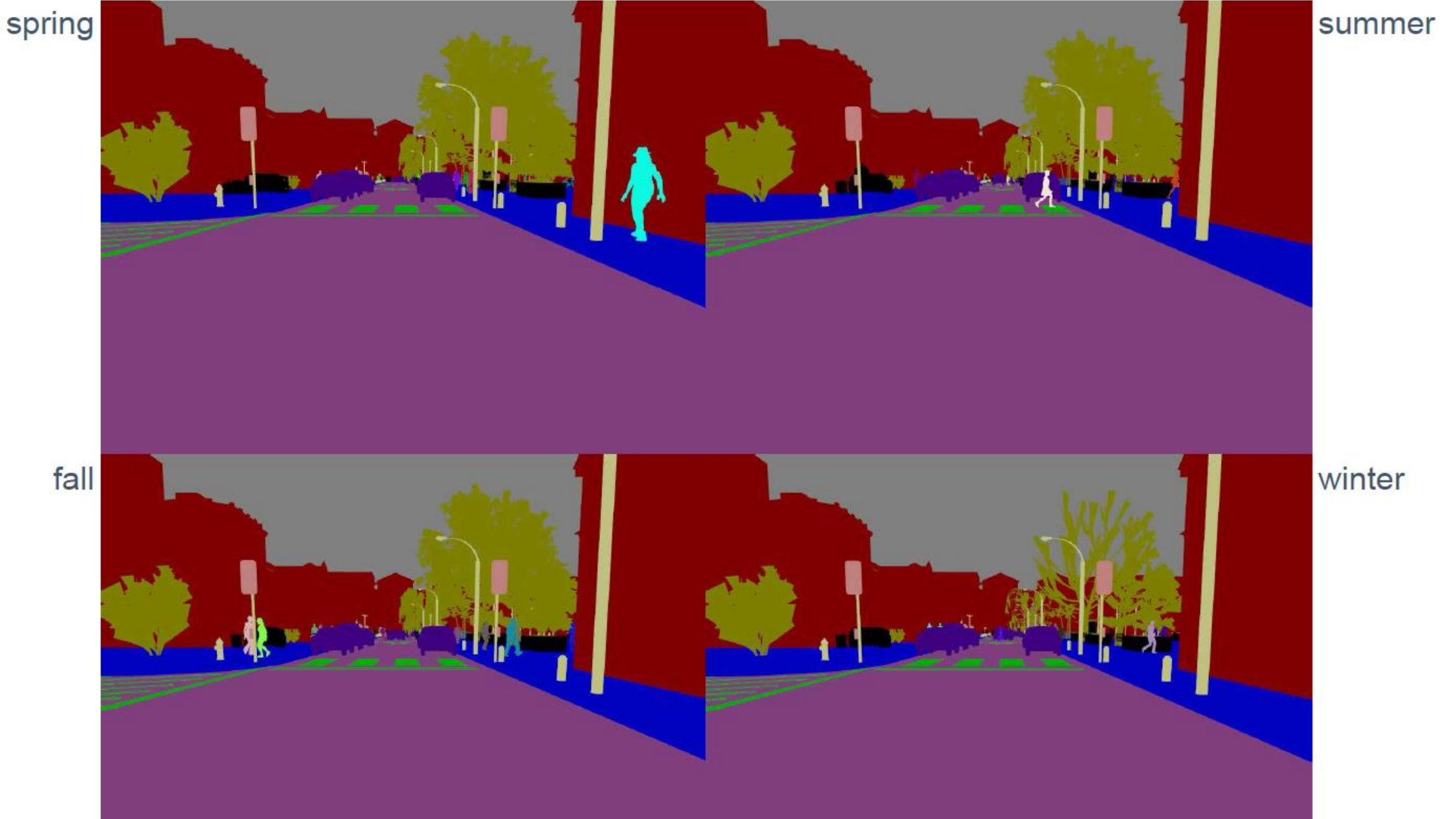
Semantic Segmentation

Depth Map


CLEARANCE 13'-9"
93
90
Quincy
Worcester
CLEARANCE 13'-9"
93
90
Quincy
Worcester


LIDAR


spring
summer
fall
winter

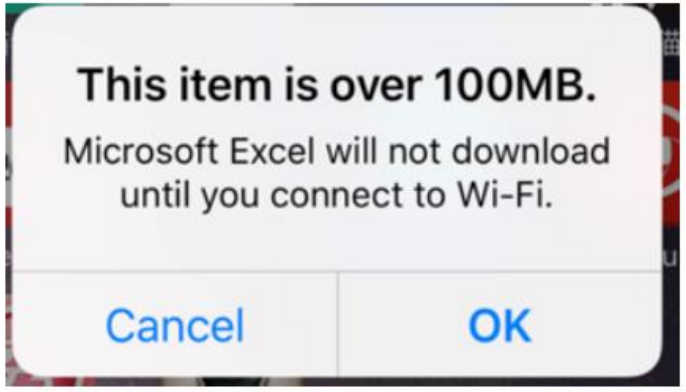

spring
summer
fall
winter

DEEP COMPRESSION AND EIE: DEEP NEURAL NETWORK MODEL COMPRESSION AND EFFICIENT INFERENCE ENGINE

Song Han PhD student, Stanford University

課題

App developers suffers from the model size



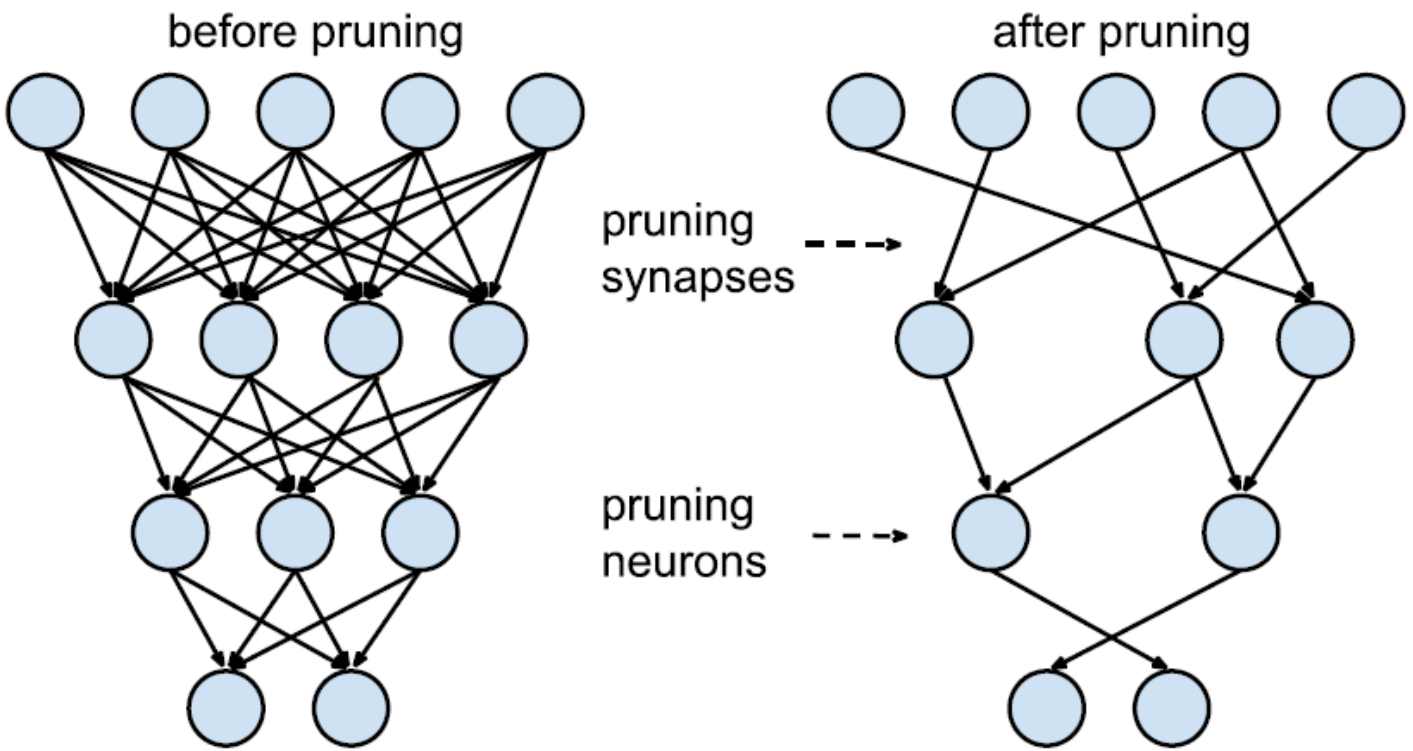
"At Baidu, our #1 motivation for compressing networks is to **bring down the size of the binary file**. As a mobile-first company, we frequently update various apps via different app stores. We've **very sensitive to the size of our binary files**, and a feature that increases the binary size by 100MB will receive much more scrutiny than one that increases it by 10MB." —Andrew Ng

Deep Compression

Deep Compression

- AlexNet: 35x, 240MB => 6.9MB **=> 0.47MB (510x)**
- VGG16: 49x, 552MB => 11.3MB
- With no loss of accuracy on ImageNet12
- Weights fits on-chip SRAM, taking 120x less energy than DRAM

Pruning

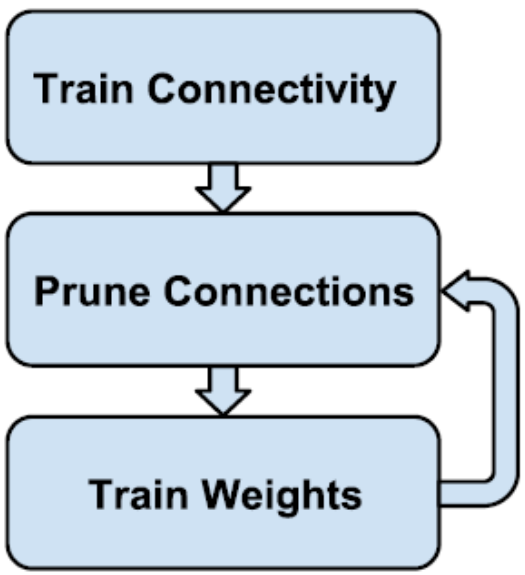

before pruning
after pruning
pruning
synapses
pruning
neurons

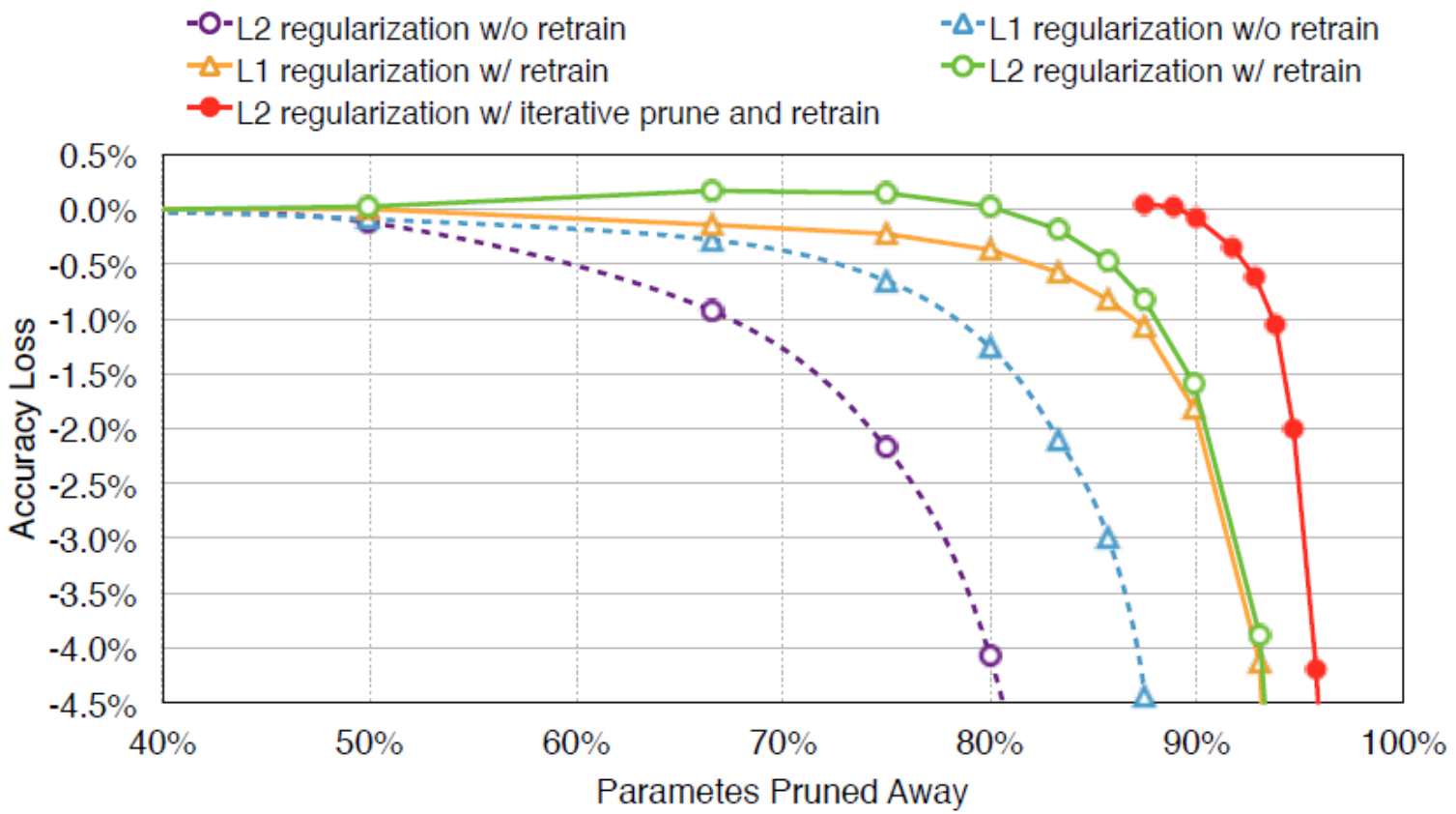
Pruning：背景

Age	Number of Connections	Stage
at birth	50 Trillion	newly formed
1 year old	1000 Trillion	peak
10 year old	500 Trillion	pruned and stabilized

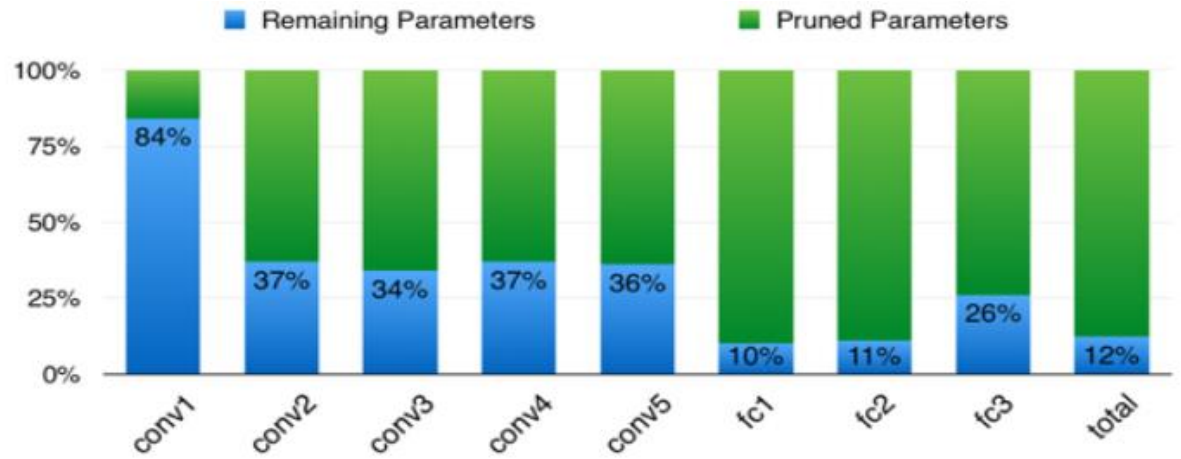
Table 1: The synapses pruning mechanism in human brain development

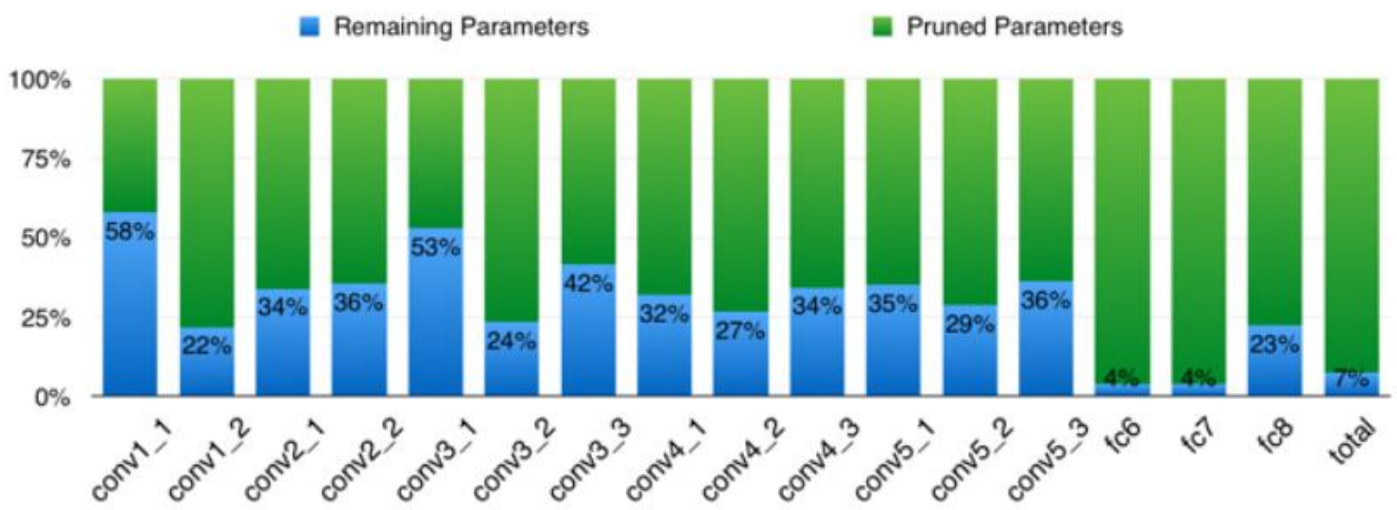
Pruningによる精度変化



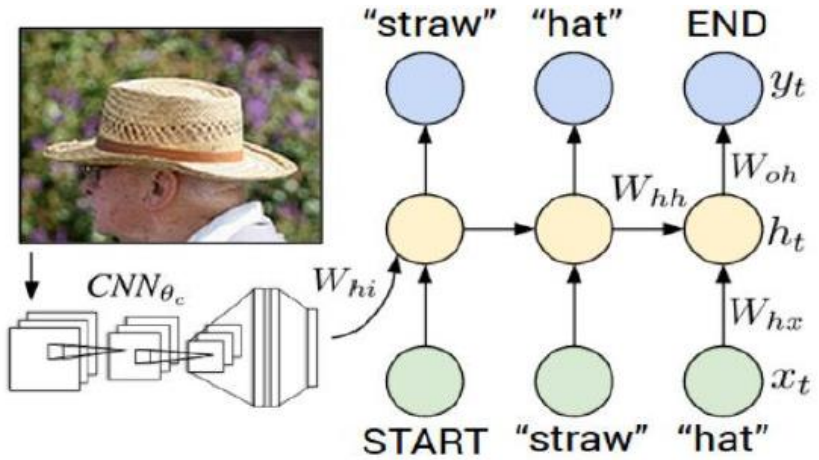


AlexNet & ConvNet



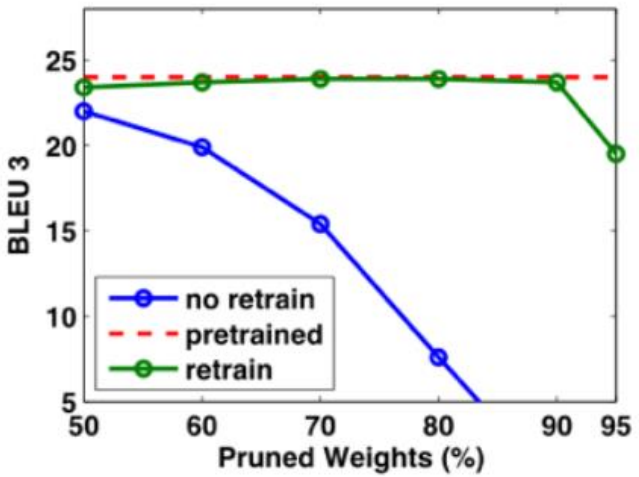


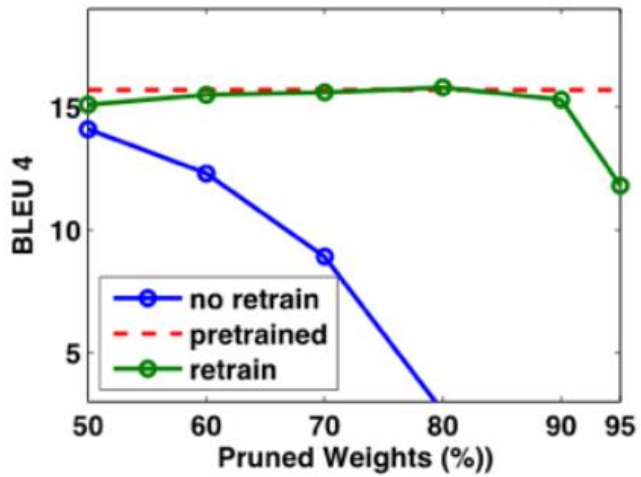
Natural TalkとLSTM



Karpathy, Feifei, et al, "Deep Visual-Semantic Alignments for Generating Image Descriptions"

- Pruning away 90% parameters in NeuralTalk doesn't hurt BLUE score with proper retraining





Natural TalkとLSTM

- **Original**: a basketball player in a white uniform is playing with a ball
- **Pruned 90%**: a basketball player in a white uniform is playing with a basketball

- **Original** : a brown dog is running through a grassy field
- **Pruned 90%**: a brown dog is running through a grassy area

- **Original** : a soccer player in red is running in the field
- **Pruned 95%**: a man in a red shirt and black and white black shirt is running through a field

ディープラーニング相談室

コンサルティング、システムインテグレーションなど各種ご相談に応じます

ディープラーニングのシステム開発にお困りでしたら

DL-HELP@nvidia.com

までお問い合わせください。

内容に応じ、各種パートナー企業様をご紹介します。