エッジAIの実現問題

～弊社の軽量化技術：CompactAI～

1

会社紹介

「"雑草魂"で、岩手から世界に貢献」

2

会社紹介

本社　：　　　岩手県盛岡市（マリオス10F）
社員数　：　　エンジニア13名・総務5名

ベトナム　：　　ホーチミン市
社員数　：　　エンジニア15名・総務1名

3

会社紹介

課題を解くのが好き　　　　　　　　独習

どんな社員？

今まだやったことが
ないことをチャレンジ　　　　最先端の技術に興味

4

３つの柱

画像処理

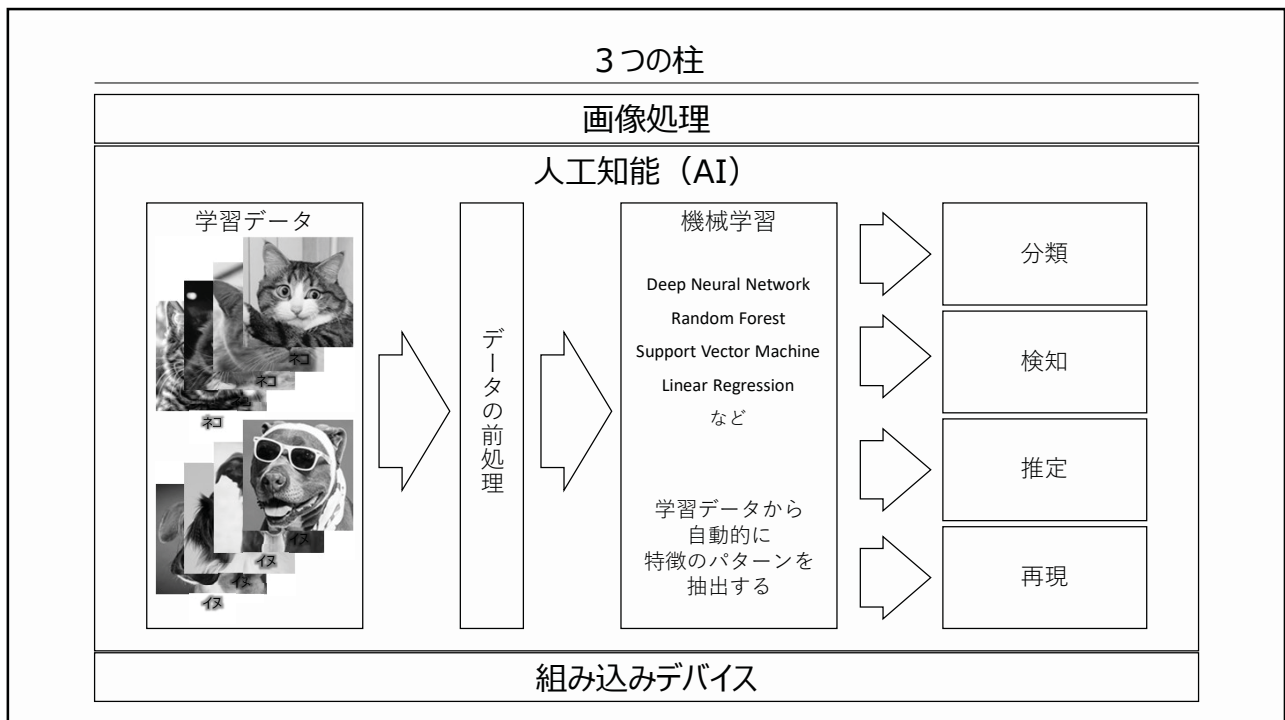人工知能（AI）

組み込みデバイス

5

３つの柱

画像処理
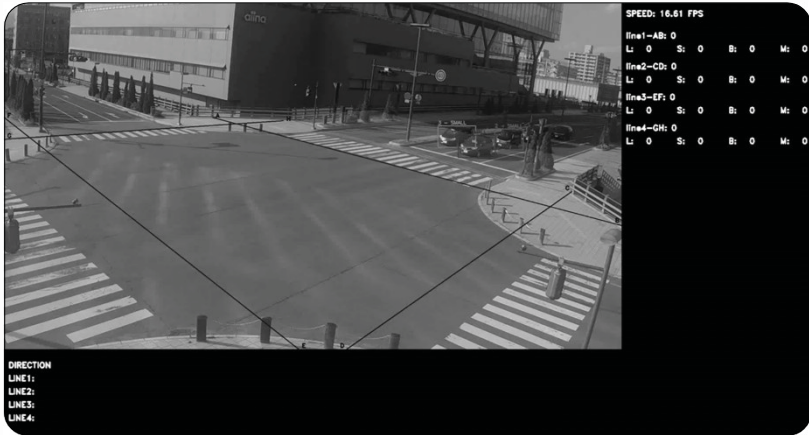画像鮮明化ソリューション：LuxEye

人工知能（AI）

組み込みデバイス

6

３つの柱

画像処理
画像解析、3D化、コンピュータービジョン

d=2

人工知能（AI）

組み込みデバイス

7

３つの柱

画像処理

人工知能（AI）

学習データ

データの前処理

機械学習

Deep Neural Network
Random Forest
Support Vector Machine
Linear Regression
など

学習データから自動的に特徴のパターンを抽出する

分類

検知

推定

再現

ネコ
イヌ

組み込みデバイス

8

３つの柱

画像処理

人工知能（AI）

組み込みデバイス

9

３つの柱

画像処理

人工知能（AI）

組み込みデバイス

10

３つの柱

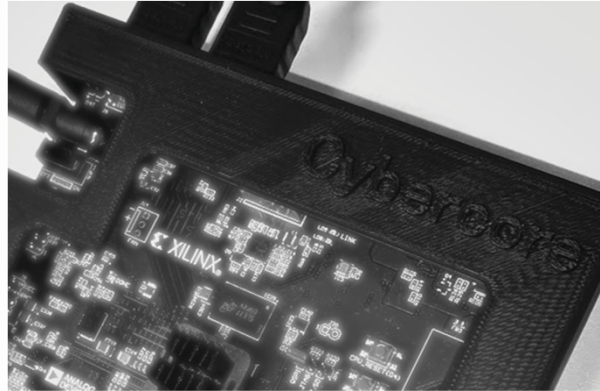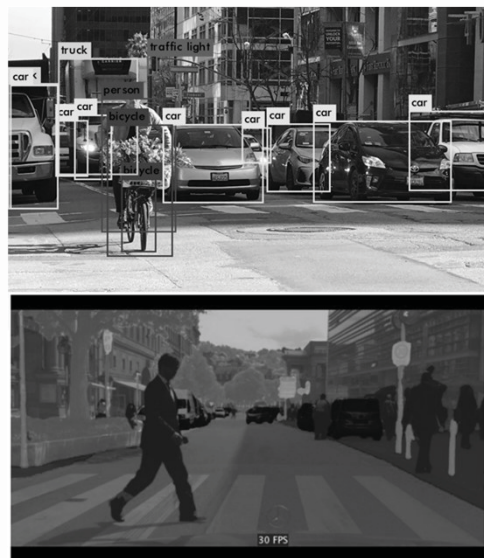| 画像処理 |
| --- |
| 人工知能（AI） |

組み込みデバイス
LuxEye（FPGA版）

11

AIの現状

- 分類
  - Noisy Student
  - ResNeXt-101

- 検知
  - Cascade Mask R-CNN
  - EfficientDet-D7 + AA

- セグメンテーション
  - HRNetV2 + OCR
  - DeepLabV3Plus + SDC

12

AIの現状

・推定（シーケンス）
　・GPT-2（テキスト生成）
　・WMT2014（翻訳）

・画像生成：
　・NCSN
　・COCO-GAN

In a shocking finding, scientist discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English.

The scientist named the population, after their distinctive horn, Ovid's Unicorn. These four-horned, silver-white unicorns were previously unknown to science.
Now, after almost two centuries, the mystery of what sparked this odd phenomenon is finally solved.

Dr. Jorge Pérez, an evolutionary biologist from the University of La Paz, and several companions, were exploring the Andes Mountains when they found a small valley, with no other animals or humans. Pérez noticed that the valley had what appeared to be a natural fountain, surrounded by two peaks of rock and silver snow.

Pérez and the others then ventured further into the valley. "By the time we reached the top of one peak, the water looked blue, with some crystals on top," said Pérez.
Pérez and his friends were astonished to see the unicorn herd. These creatures could be seen from the air without having to move too much to see them – they were so close they could touch their horns.

While examining these bizarre creatures the scientists discovered that the creatures also spoke some fairly regular English. Pérez stated, "We can see, for example, that they have a common 'language,' something like a dialect or dialectic."

Dr. Pérez believes that the unicorns may have originated in Argentina, where the animals were believed to be descendants of a lost race of people who lived there before the arrival of humans in those parts of South America.

While their origins are still unclear, some believe that perhaps the creatures were created when a human and a unicorn met each other in a time before human civilization. According to Pérez, "In South America, such incidents seem to be quite common."

However, Pérez also pointed out that it is likely that the only way of knowing for sure if unicorns are indeed the descendants of a lost alien race is through DNA. "But they seem to be able to communicate in English quite well, which I believe is a sign of evolution, or at least a change in social organization," said the scientist.

13

AIの現状

・推定（シーケンス）
　・GPT-2（テキスト生成）
　・WMT2014

・画像生成：
　・NCSN
　・COCO-GAN

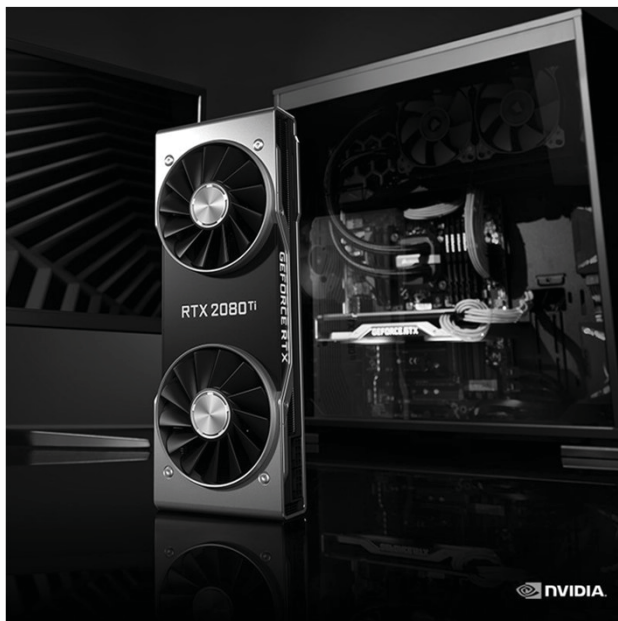https://www.thispersondoesnotexist.com/

14

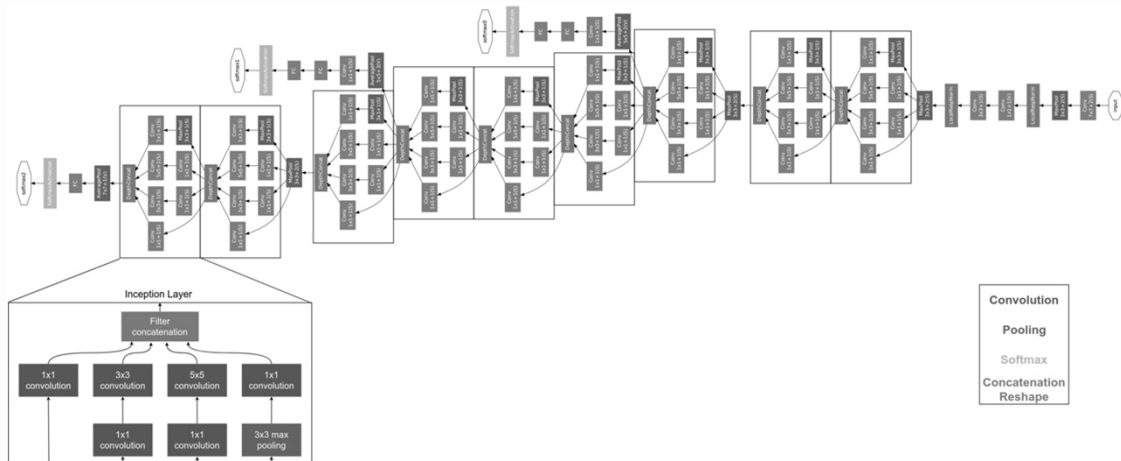## AIの現状
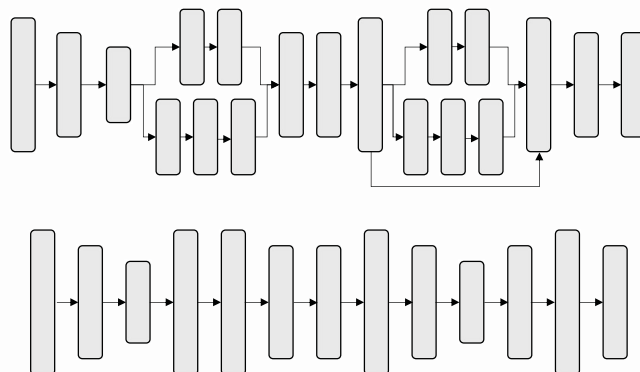
学習ハードウェア

15

AIの現状

推定ハードウェア

16

CompactAI

「CompactAI」：エッジデバイス用のニューラルネットワーク

17

CompactAI

学習済のTeacherネットワーク：
分岐、合流があり、
ハードウェアで実行しにくい
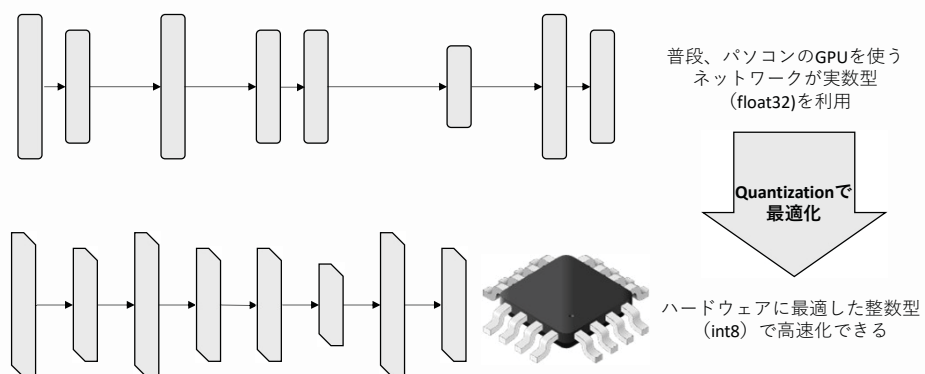
Distillationで
知識移植

シンプルなStudentネットワーク：
中間層が減り、分岐が無くす

18

CompactAI

学習により、一部の中間層の
影響が「ゼロ」に近いため、
削減できる

Pruningで
軽量化

シャローネットワーク完成

19

CompactAI

普段、パソコンのGPUを使う
ネットワークが実数型
（float32)を利用

Quantizationで
最適化

ハードウェアに最適した整数型
（int8）で高速化できる

20
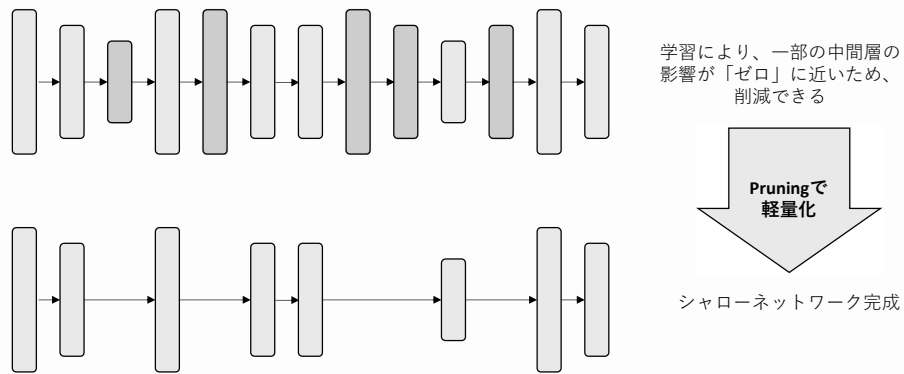
10

CompactAI

事例：顔認証

21

CompactAI
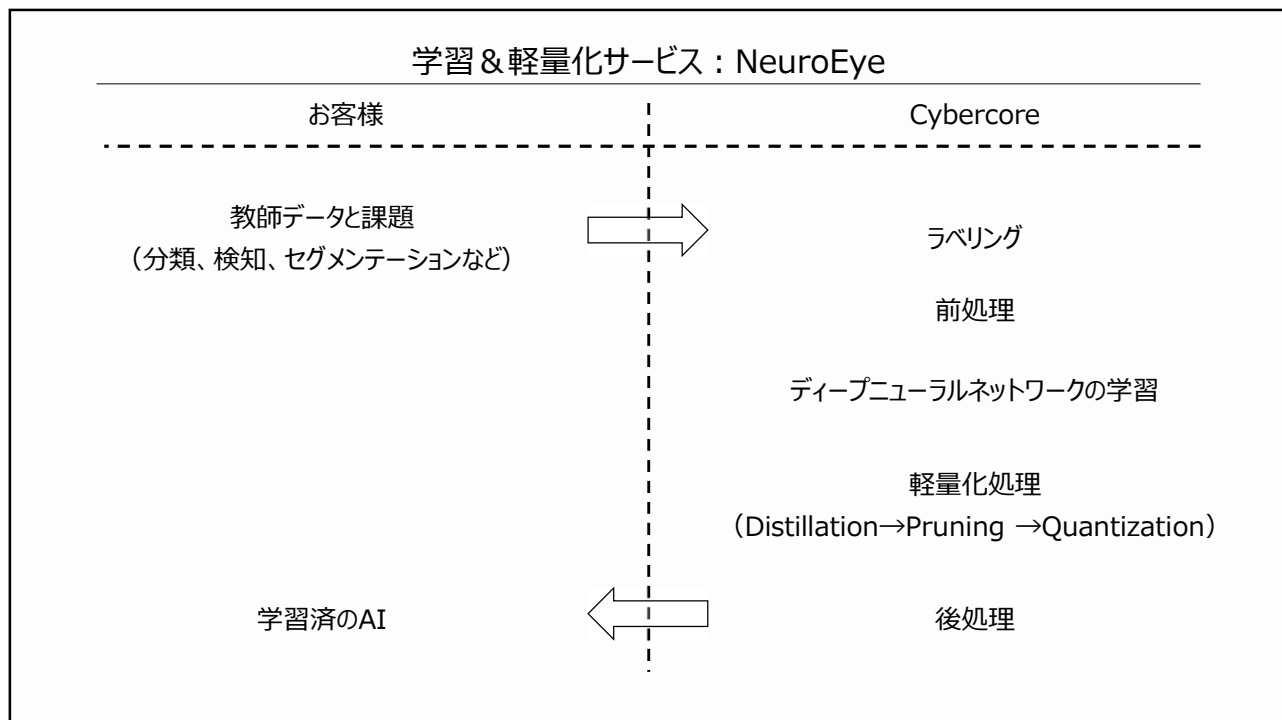
メリット
・使用メモリ、計算量が低い
・中間層が少なく、処理が速い

デメリット
・再学習にて、軽量化処理を繰り返す
・課題により、パイプラインの調整が必要

22

学習＆軽量化サービス：NeuroEye

お客様	Cybercore
教師データと課題 （分類、検知、セグメンテーションなど）	ラベリング
	前処理
	ディープニューラルネットワークの学習
	軽量化処理 （Distillation→Pruning →Quantization）
学習済のAI	後処理

23

CompactAI

・現在：「物体検知」のパイプライン
・将来：各種（セグメンテーション、分類など）のパイプラインも構築の予定

・NeuroEyeのアピール：
+AI専門家の目で教師データを整理
+サブスクリプションモデルで、簡単に追加学習が可能
+オンラインの場合は、自動的に更新が可能

24

25

まとめ

・現代のAIで様々なことができる

・ただし、計算量が多く、高性能のHWが必要

・軽量化技術で、エッジデバイスでも可能になった

・課題により、具体的に軽量化をカスタマイズしなければならない

・要素技術、およびオープンソースを活用したビジネスモデルの可能性

26

ご清聴ありがとうございました

http://www.cybercore.co.jp/

27