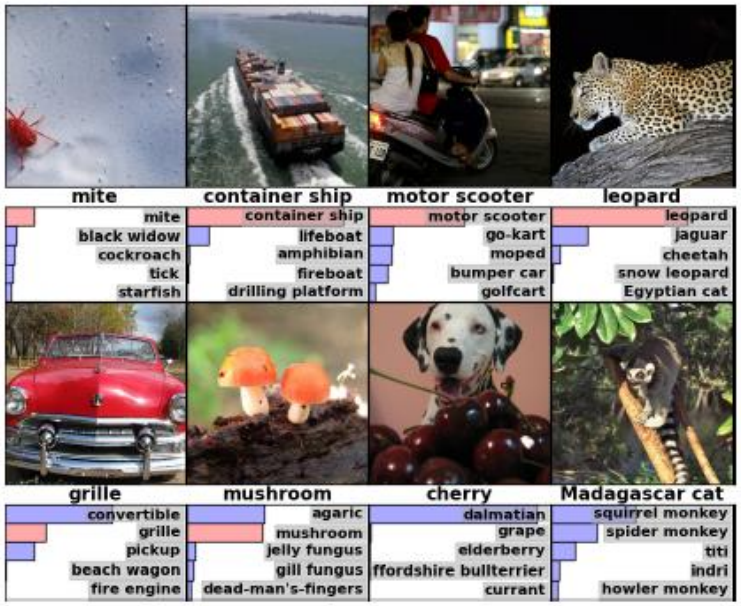
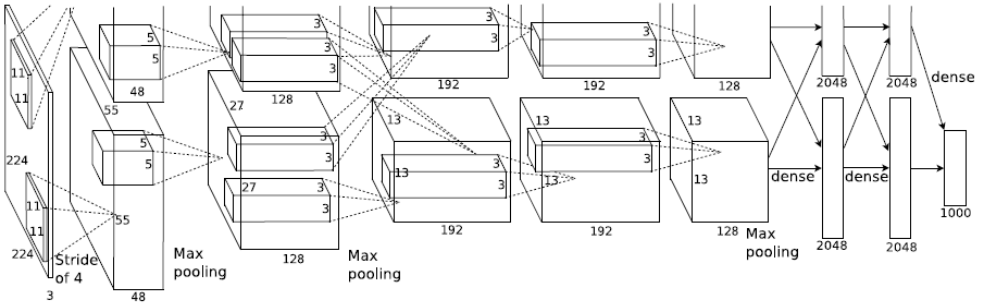
ニューラルネットワークを用いた動画像内の物体認識

総合人間学部　認知情報学系　神谷研究室
中村　優太

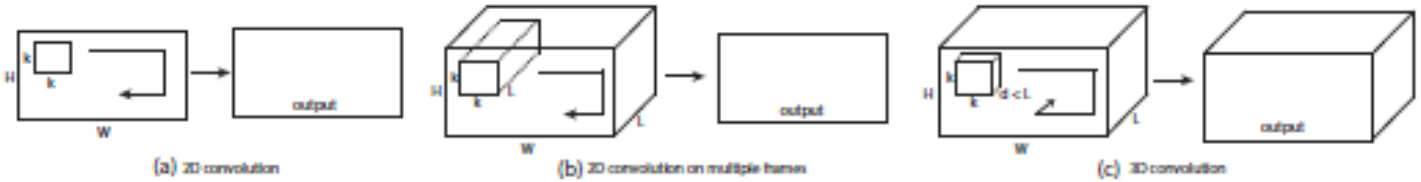
ニューラルネットワークとは

- 機械学習の手法の一つで，特に画像認識の分野で飛躍的な成果を生み出してきた手法





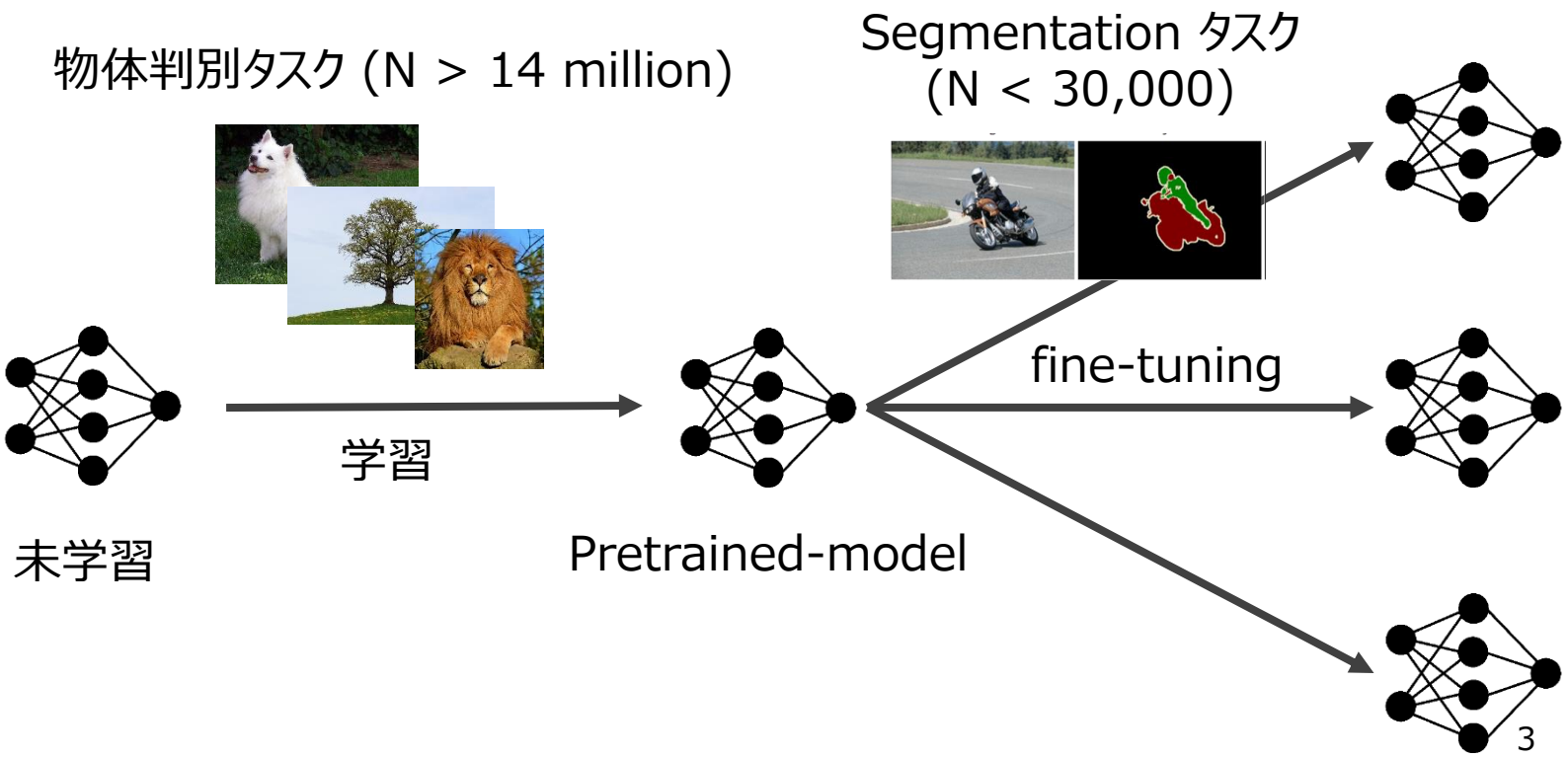
(Krizhevsky Hinton, 2012)



(Tran, Bourdev, Fergus, Torresani, Paluri, 2015)

ニューラルネットワークのfine-tuning

- ニューラルネットワークを訓練するには、大量のデータが必要
- 学習済みのモデルを元に、ターゲットとするタスクを学習する fine-tuningが広く用いられている.



動画中の物体判別タスクの学習

- 動画中の物体判別タスクにおけるfine-tuningを複数の条件で行い、fine-tuningの特性を検証
- 学習するデータの特性とタスクの特性がfine-tuningに与える効果を検証

CNNアーキテクチャ	学習済みタスク
2次元畳み込み (静止画用)	静止画中の物体判別タスク
3次元畳み込み (動画用)	静止画中の物体判別タスク
3次元畳み込み (動画用)	動画中の動詞判別タスク

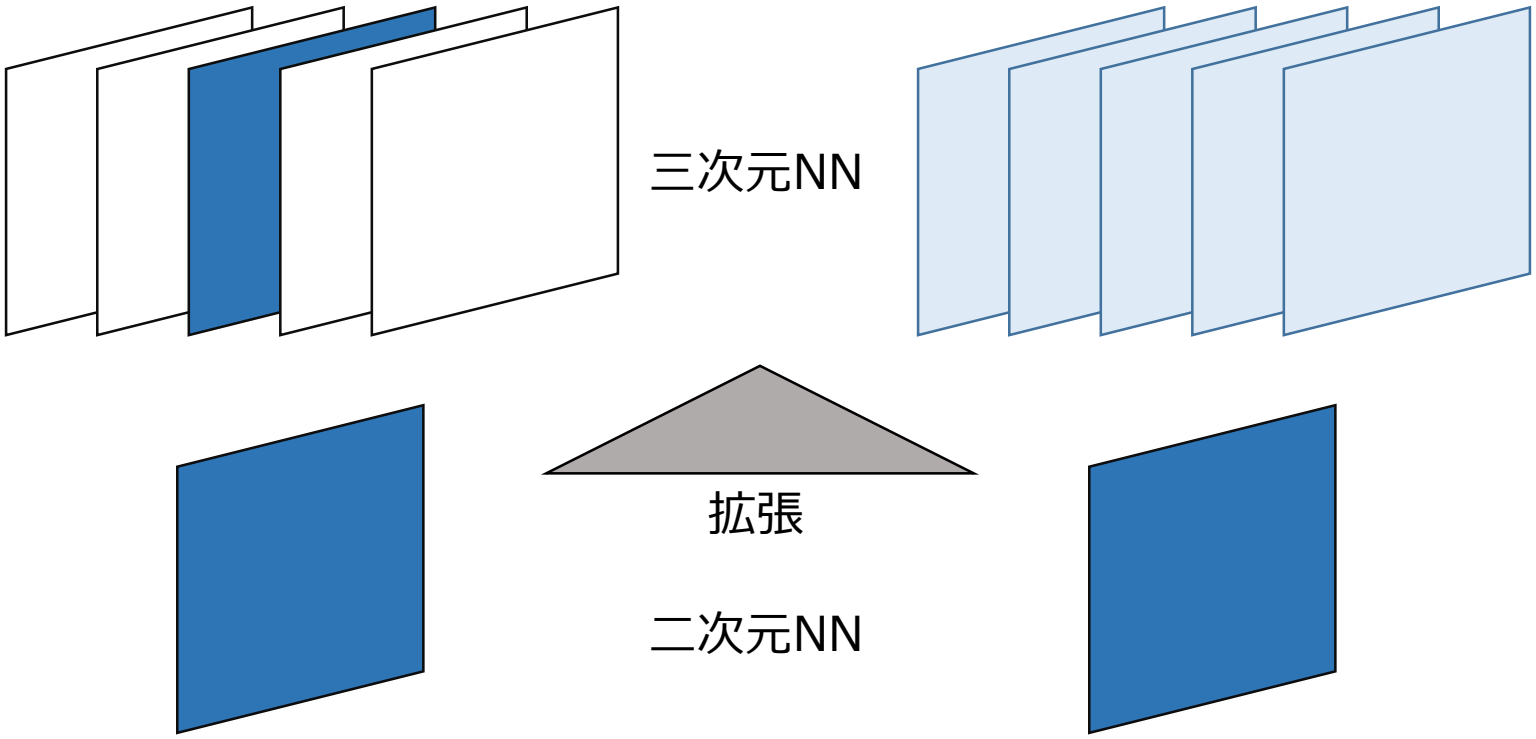
検証に用いたニューラルネットワーク

- 2次元CNNは ResNet50 (He, Zhang, Ren, Sun, 2015)のアーキテクチャ
- 3次元CNNは ResNet50 を3次元に拡張したアーキテクチャ
- 静止画中の物体判別タスクとして、ImageNet (Jia Deng, et al., 2009) の1000クラス物体判別タスク
- 動画中の動詞判別タスクとして Kinetics (Kay, Carreira, Simonyan, 2017) の動詞判別タスク

CNNアーキテクチャ	学習済みタスク
2次元畳み込み (静止画用)	静止画中の物体判別タスク
3次元畳み込み (動画用) 中心化拡張	静止画中の物体判別タスク
3次元畳み込み (動画用) 平均化拡張	静止画中の物体判別タスク
3次元畳み込み (動画用)	動画中の動詞判別タスク

ニューラルネットワークの拡張

- 静止画像判別のためのニューラルネットワークを、動画用に拡張する技術 I3D (Kay, Carreira, Simonyan, 2017) を用いて3次元に拡張した。



検証の手続き

前述のネットワークを用いて、動画中の物体判別タスクを行い、マルチラベル判別タスクの成績を比較した

- データ: Moments In Timeデータセット (Monfort, et al., 2018) を元に, 1動画に含まれる物体を複数ラベル付けしたデータを自作
- ネットワーク: 二次元画像判別NN,中心化拡張NN, 平均化拡張NN, 三次元動詞判別NNを使用
- 学習: 全てのNNにおいて、最適化手法としてSGD with Momentum, 学習率0.01を用いて学習
- 評価: 各カテゴリ毎に,予測結果からAUCを算出

データ例

Car, Man

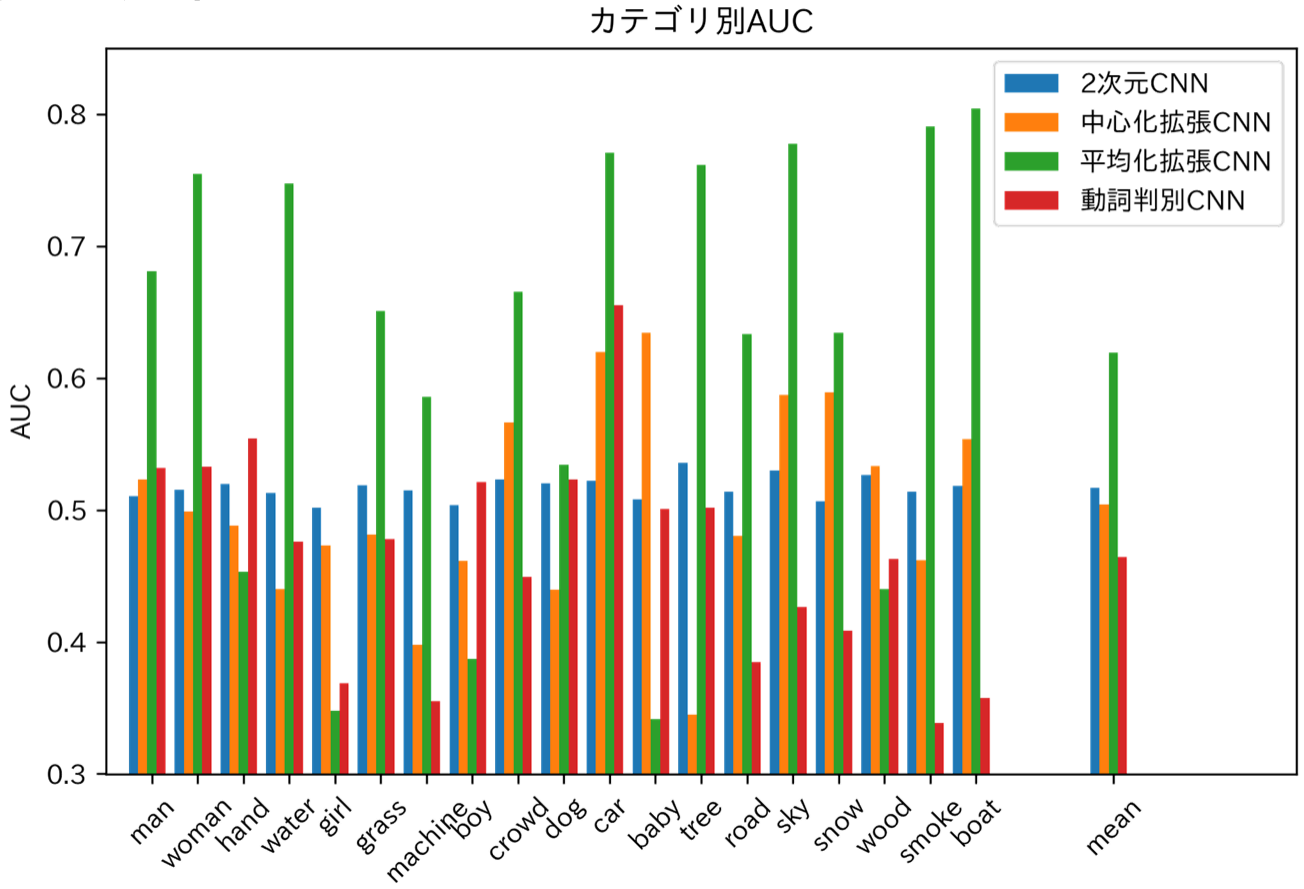
Water, Boat

Man, Baby

Crowd

結果

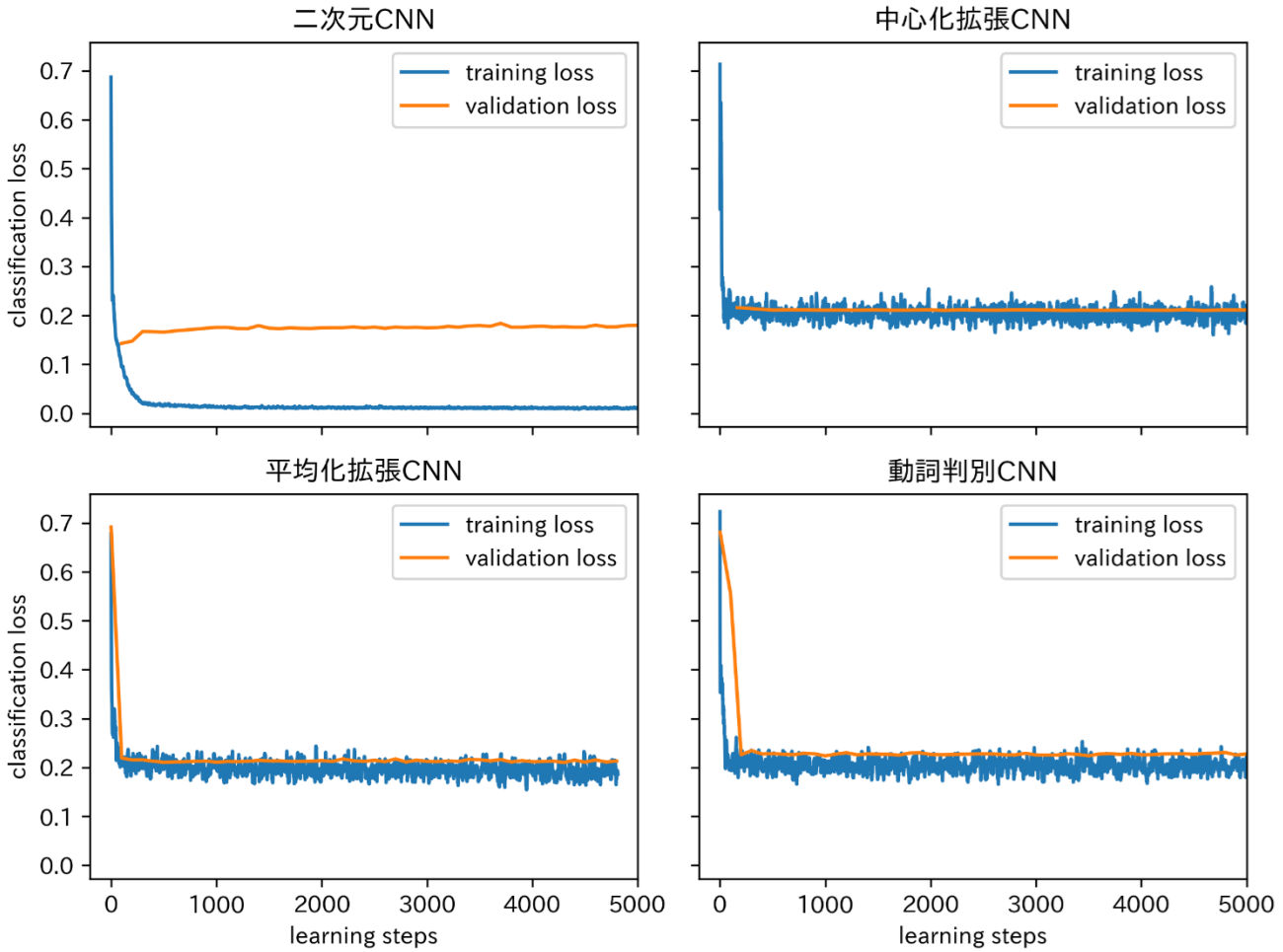
- 各ネットワークからfine-tuningした際の、物体認識タスクの成績は以下のようであった。



平均化拡張による三次元ニューラルネットワークにおいてのみ、成績が高い

二次元画像判別NNの考察

- 二次元の画像判別ネットワークを用いた場合には、過学習の傾向が強くみられた。



三次元画像判別NNの考察

- 三次元のNNの中では、平均化拡張のニューラルネットワークのみが学習に成功した

- その要因の仮説として、今回のような限られたデータを用いたfine-tuningにおいては、ニューラルネットワークの重みの変化量が少なくてもタスクを学習できる性質が必要で、平均化拡張のネットワークがその性質を満たしていたと考えられる

- 学習済みのネットワークの重みを検証する必要がある

まとめ

- 動画中の動詞判別タスクにおけるfine-tuningの手法について比較を行った
- 画像判別用のニューラルネットワークをそのまま動詞判別に利用するとデータ量が少ない場合、過学習に陥りやすいことが明らかになった
- 動詞判別タスクにおいては、動画を扱える三次元ニューラルネットワークを用いることで過学習は避けられるものの、学習が困難になった
- データ量が少ない場合には、平均化拡張により学習済み画像判別ネットワークを三次元に拡張したネットワークによるfine-tuningが有利であることがわかった