

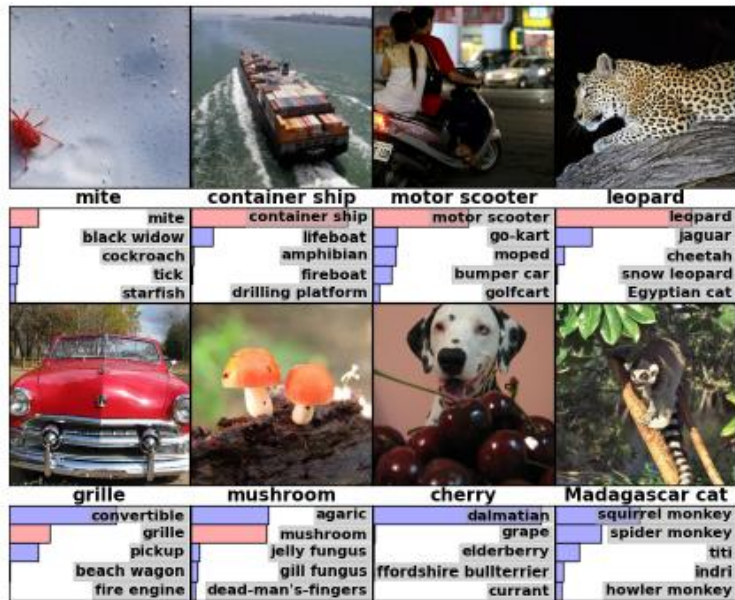
ニューラルネットワークを用いた 動画像内の物体認識

総合人間学部 認知情報学系 神谷研究室

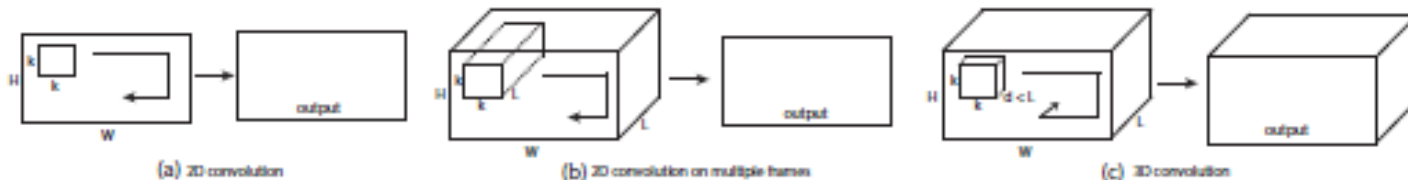
中村 優太

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



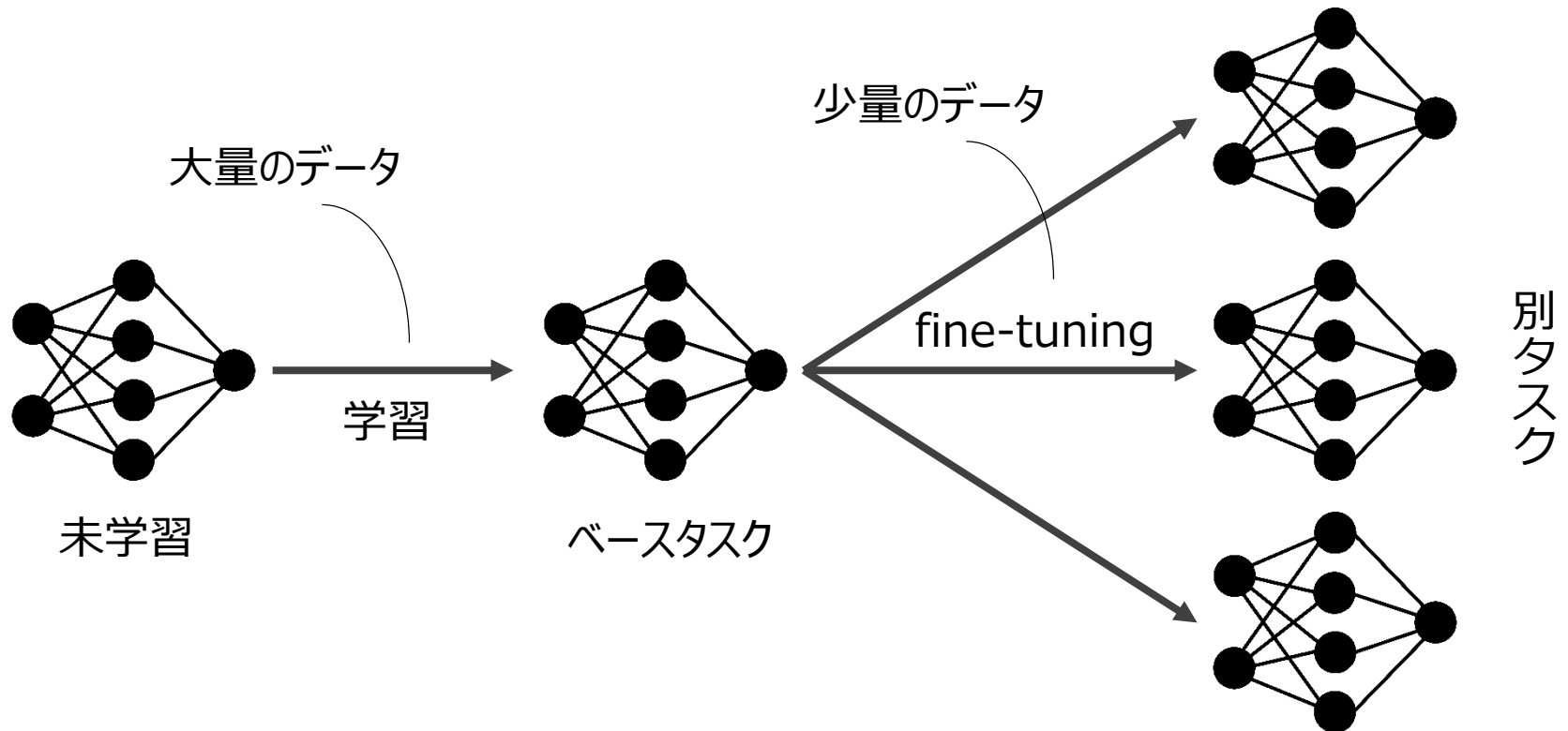
(Krizhevsky Hinton, 2012)



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

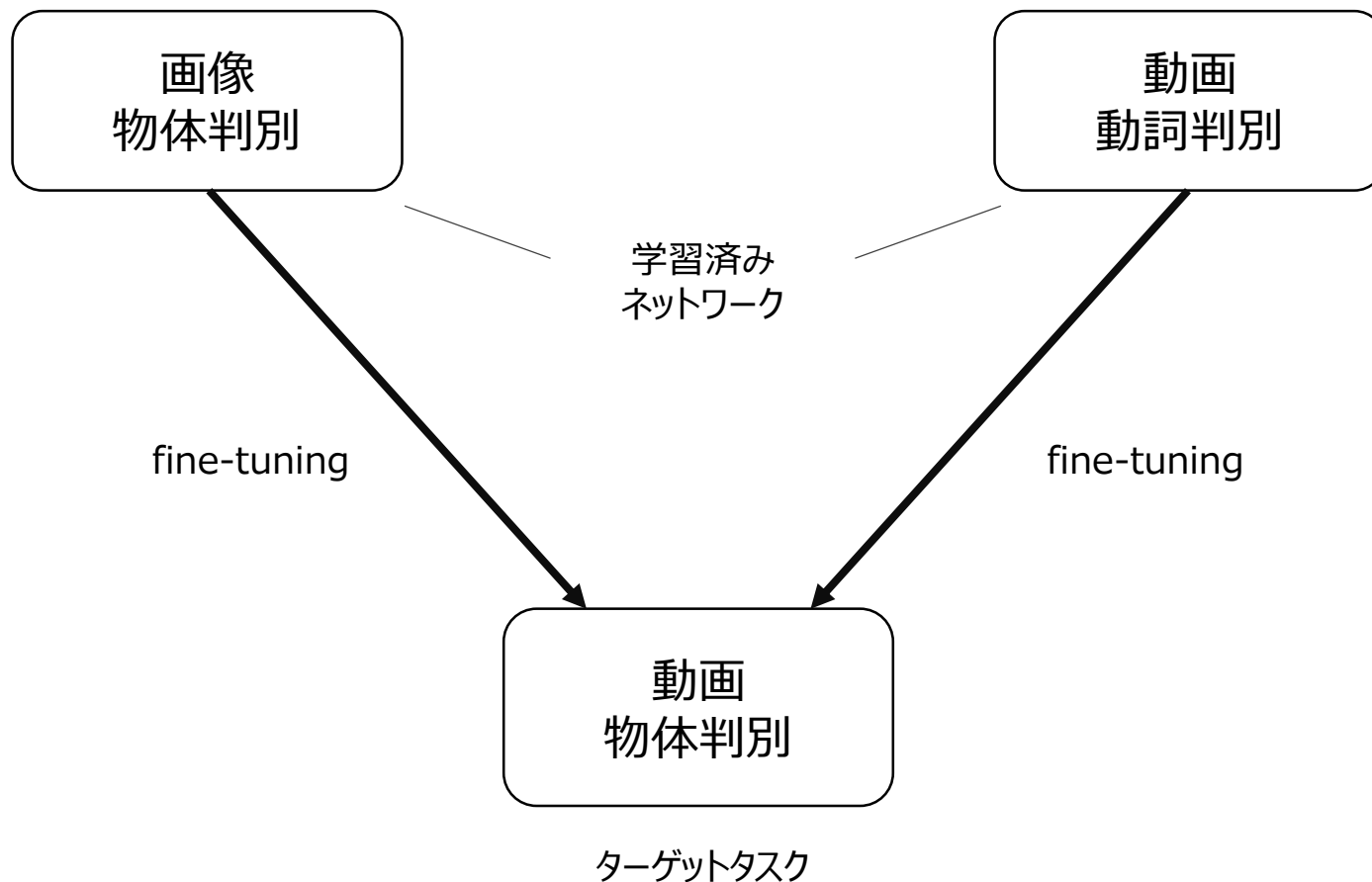
ニューラルネットワークのfine-tuning

- ニューラルネットワークを訓練するには、大量のデータと計算資源が必要
- 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

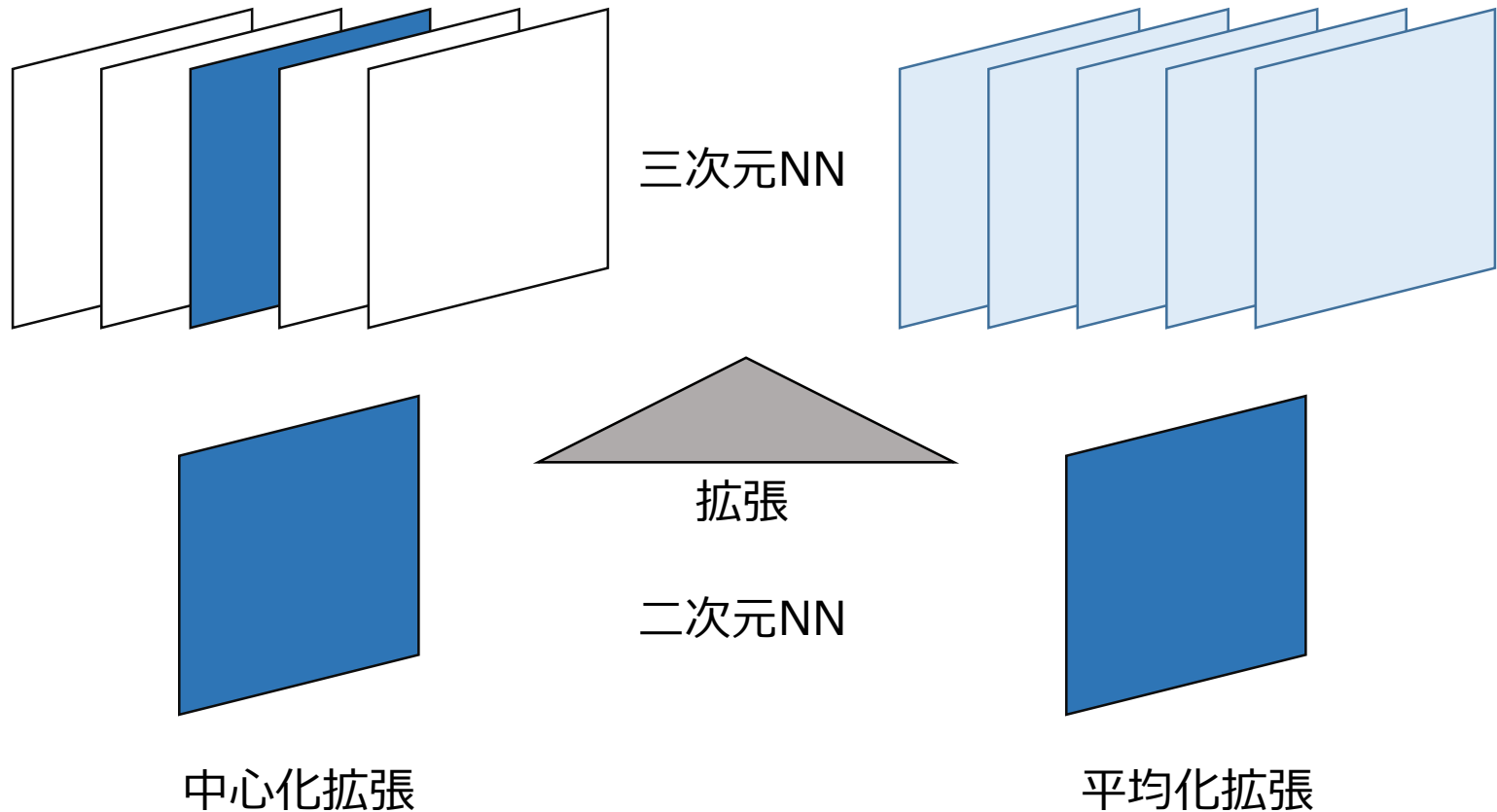


検証に用いたニューラルネットワーク

- 二次元画像判別NN:
ImageNetの1000クラス物体判別タスクを学習した畳み込みニューラルネットワーク (ResNet50)
- 三次元に拡張した二次元画像判別NN:
上述のNNを、動画を直接扱えるように時間方向に拡張した畳み込みニューラルネットワーク
- 三次元動詞判別NN:
Kineticsデータセットを用いて動画中の動詞判別を学習した畳み込みニューラルネットワーク

ニューラルネットワークの拡張

- 画像判別のための学習済みニューラルネットワークを、動画用に三次元に拡張する技術 (I3D) を用いて3次元に拡張した。



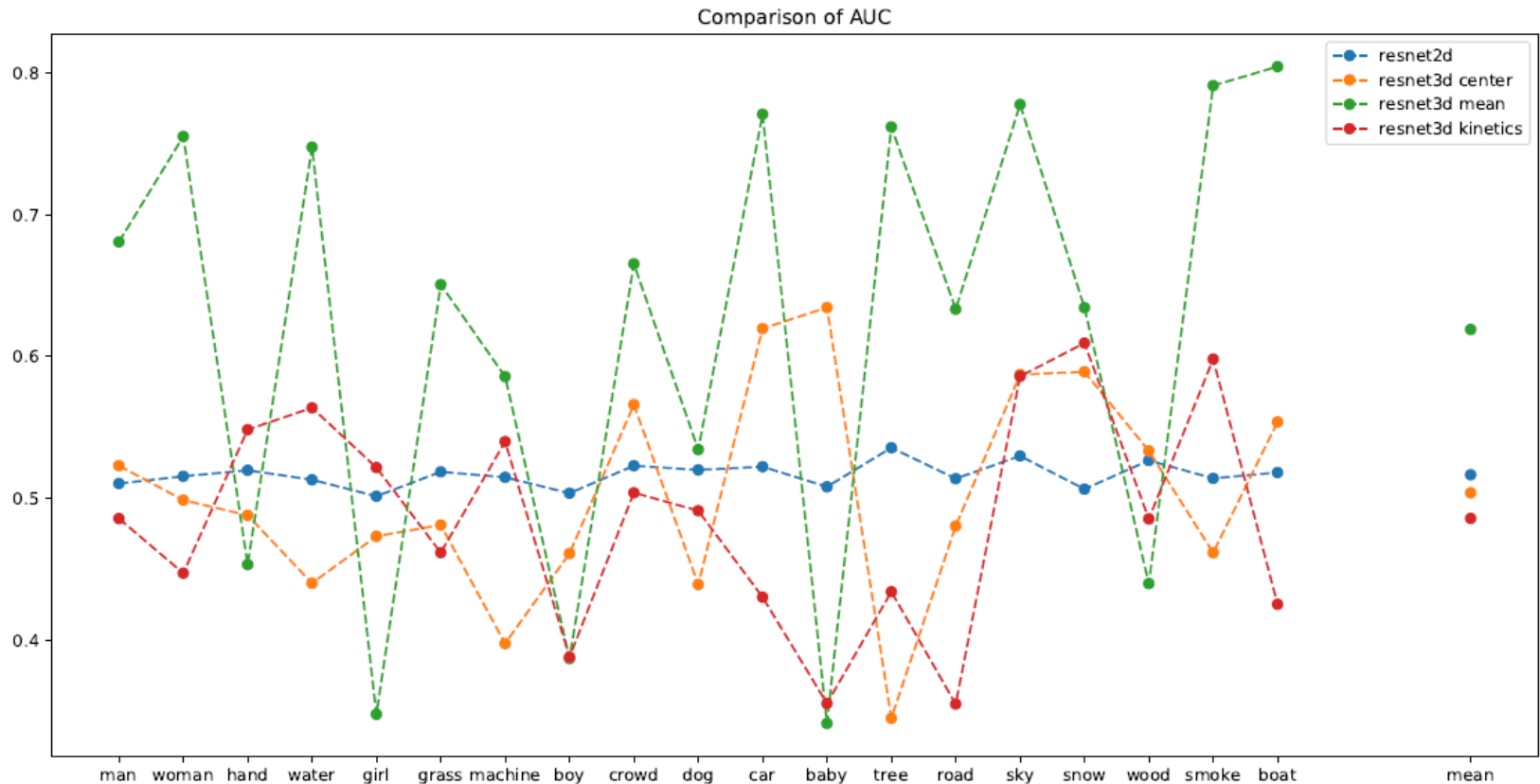
検証の手続き

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、成績を比較した

- データ: Moments In Timeデータセットを元に, 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

結果

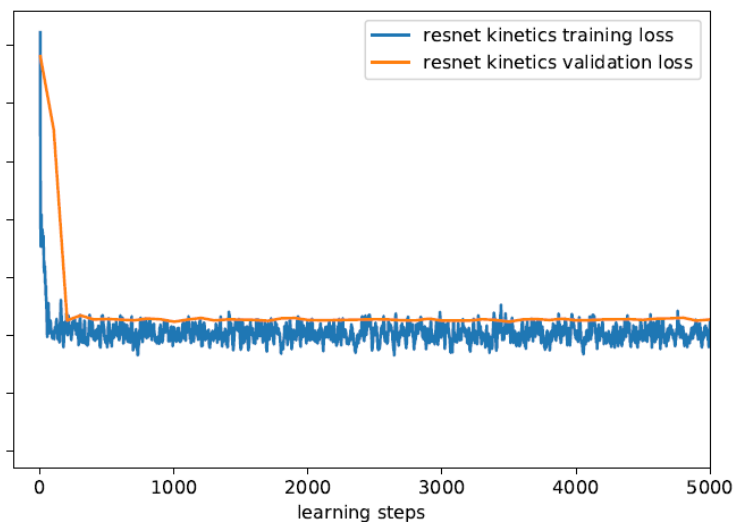
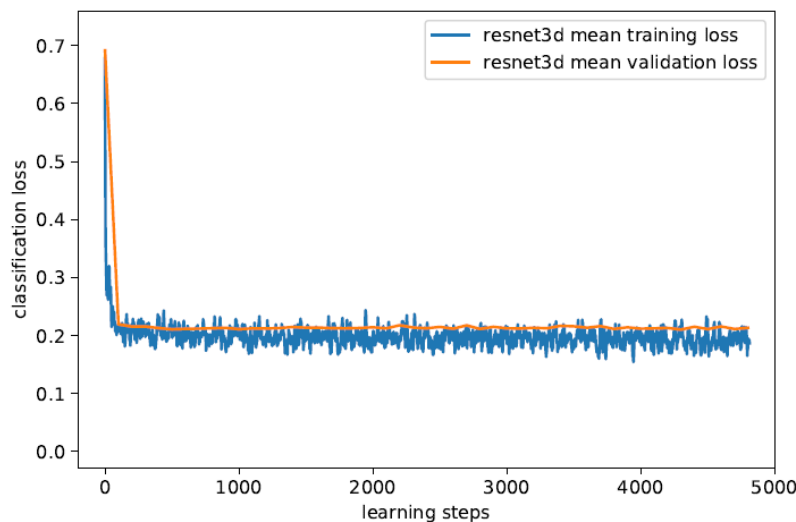
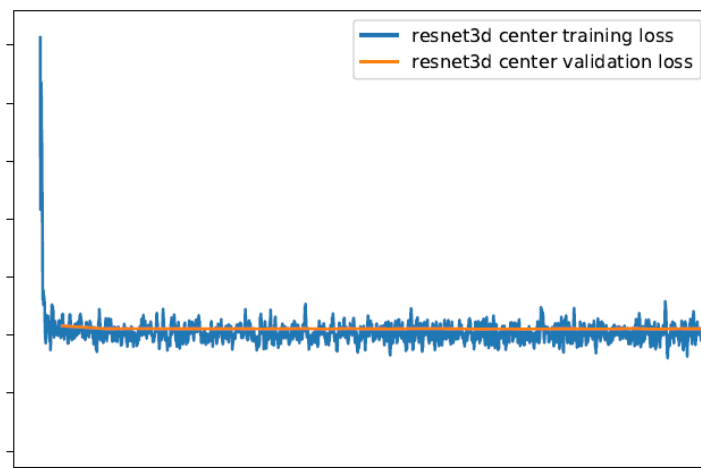
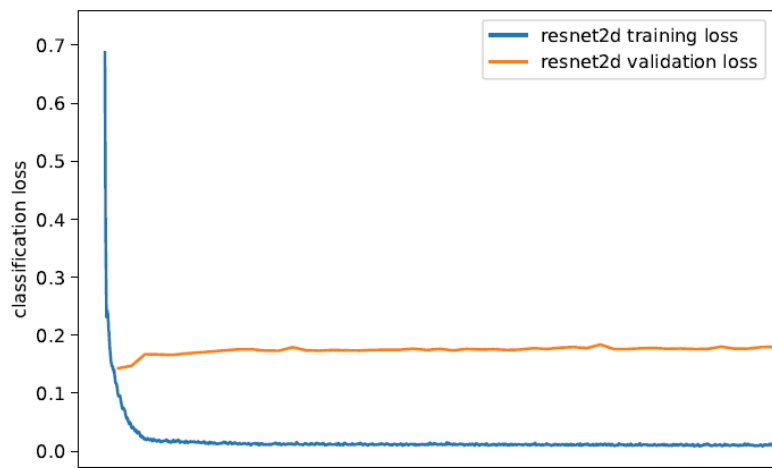
- 各ネットワークからfine-tuningした際の、物体認識タスクの成績は以下のものであった。



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、
fine-tuningが成功した

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