

## 胸部X線画像の finetune による判別

近藤正紀<sup>1)</sup>

1) 新潟医療福祉大学 医療情報管理学科

【背景・目的】 機械学習におけるディープラーニングが成功するためには与えるデータの量と質が無視できない。与えるデータ量が十分でないと学習不足となり有用な性能が得られず、また、データに偏りがあると汎化性能に影響が出る。そして、適用分野によっては大量のデータを用意することが困難である。この問題の解決策の一つに転移学習や finetune 提案され、成果も出始めている。

本稿では finetune を用いて事前学習済みモデルが胸部X線画像に対して示す挙動を解析し、読影システム構築の基礎を得ることを目的とする。

【方法】 本稿では NIH Clinical Center の Wang らが 2017 年に公開した、匿名化された 30,805 名の相異なる患者から得た 112,120 枚の胸部X線画像データを使用した。

Finetune はある領域で学習済みのモデルを他の領域のデータセットで再学習する手法であり、比較的少ない学習データでも高い精度が得られることが多く、学習時間が短くなるというメリットもある。各X線画像は14種類の異常陰影のラベル付けがなされ、最大で9個のラベルを付けられた画像も存在するが、本研究では単一ラベルのみを扱うことし、複数ラベルの画像を除外し、枚数が少ないラベルの画像はその他として一括した。また、不鮮明で学習困難と考えられる画像を目視で除外した。一連のデータクレンジング後、15歳以上の成人のみを抽出したところ男性 30,194 枚、女性 22,659 枚が得られたため、finetune の利点を活かすことと現有計算資源との兼ね合いから、本稿では成人女性の画像を対象とした。これを同一患者を含まない訓練用 19,272 枚とテスト用 3,387 に分割した。

以上のデータを、ImageNet で事前学習済みの VGG16 に与えて finetune を行う。VGG16 は Simonyan と Zisserman が 2014 年の ILSVRC において高い成績を示した 6 ブロック 16 層の CNN である。

なお、本研究で使用した計算資源では 19,272 枚の訓練画像を一括して主記憶に展開できないため、同一の患者を含まない 5 個のサブセットに分割して学習を行った。

本研究は倫理審査の対象外のものであり、関連する利益相反はない。

【結果】 図 1 は比較的成績の良かったブロック 4 以降、すなわち 16 層のうち 8 層目以降を finetune して得た学習曲線の一つである。方法で示したように成人女性の訓練用セットを 5 個のサブセットに分割しそれぞれ 50 エポックずつの学習を行っているが、図 1 ではそれらを一続きの学習としてエポック 0~49 をサブセット 1、以降 50 エポ

ックごとにサブセット 2~5 の学習曲線を示す。訓練データのサブセットの切り替え時（エポック 50、100、150、200）に訓練用セットの正答率が鋸歯状に低下していることがわかる。また、テスト用セットの正答率は概ね直前のサブセットの正答率を引き継ぐ形で上昇しているが、第 4 サブセットで大きく下げ、第 5 サブセットでは最大正答率 0.4195 を記録した後、徐々に低下した。

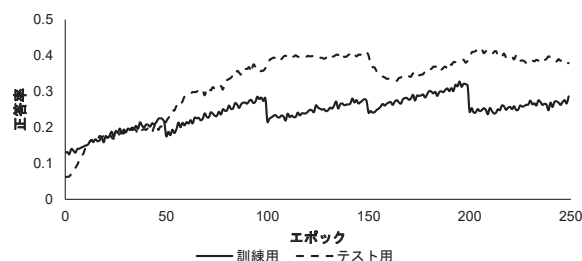


図 1 学習曲線

【考察】 事前学習済みの VGG16 に finetune を行った結果は、テスト用データに対する正答率の最大値は成績が良いものであっても 0.42 と程度と成績が良いとは言えない状態であった。

正答率が低い最大の要因は事前学習に使用した ILSVRC2012 と胸部X線画像が似ていない、すなわち『モデルが見たことが無い』画像ばかりであったという点と考えられる。画像の異なり具合と finetune の成否については未だ不明点が多いが、ILSVRC2012 にはヒトの画像が含まれていないため、『入力層に近い層は一般的な特徴を抽出する』という finetune が成立する条件が満たされていない可能性がある。また、ILSVRC2012 に合わせて 1024x1024 の画像を 224x224 に縮小したことも要因と考えられ、さらに再学習によって入力と出力の関係が大きく乱された結果『安定性と可塑性のジレンマ』が表面化した可能性もある。

正答率向上を図るには、データクレンジングを機械化して再実行した後、画像を縮小しない学習が効果的と考えられるが、事前学習済みの VGG16 を用いるとしても、この方法は finetune の定義から外れる可能性があり、finetune の限界の見極めとともに今後の課題である。また、COVID-19 と他の異常陰影の判別実験が各所で行われていることに倣い、異常陰影ラベルの分類ではなく異常陰影の有無のみを判別する方法も考えられる。

【結論】 事前学習済みのモデルの finetune による胸部単純X線画像の分類を試みたが、ImageNet で事前学習をした VGG16 という限定した範囲では芳しい結果を得られなかった。finetune や転移学習は高精度の画像分類を行う際に有効であろうと期待される手法であるが、適用領域を慎重に選択しないと満足いく結果を得るのは難しいという示唆を得た。