DEIM2022 J31-3

モバイル端末上でのプライバシに配慮した画像認識モデルを構築する手法の提案と実装

近藤　華† 山口　実靖†† 神山　剛††† 小口　正人†

† お茶の水女子大学　〒 112–8610 東京都文京区大塚 2-1-1
†† 工学院大学　〒 163–8677 東京都新宿区西新宿 1-24-2
††† 長崎大学　〒 852–8521 長崎県長崎市文教町 1-14
E-mail: †hana-k@ogl.is.ocha.ac.jp , oguchi@ogl.is.ocha.ac.jp, ††sane@cc.kogakuin.ac.jp, †††kami@nagasaki-u.ac.jp

あらまし　近年，深層学習技術を用いた画像認識技術は様々な問題で利用されており，Android などのスマートフォンを始めとするモバイル端末においても顔認識，ジェスチャー認識などの用途で活用されている．それと同時に，モバイル端末内のプロセッサの高性能化が進んでいる．また，モバイル端末の高性能化が進むことで，プライバシの保護のためモバイル端末に保存した画像データを外部に送信せずに，それらの画像データから得られた情報を用いた画像認識を行う仕組みが求められるようになると考えられる．そこで本研究では，Android 端末の高性能プロセッサを利用し，端末上のみで精度の高いパーソナライズ化された画像認識モデルを構築する手法を提案する．
キーワード　深層学習，機械学習，Android, ファインチューニング

Proposal and Implementation of a Method for Building Privacy-Protected Image Recognition Models on Mobile Devices

Hana KONDO†, Saneyasu YAMAGUCHI††, Takeshi KAMIYAMA†††, and Masato OGUCHI†

† Ochanomizu University, 2–1–1 Otsuka, Bunkyou-ku, Tokyo 112–8610 Japan
†† Kogakuin University, 1–24–2 Nishi-shinjuku, Shinjuku-ku, Tokyo 163–8677 Japan
††† Nagasaki University, 1–14 Bunkyo-cho, Nagasaki-shi, Nagasaki 852–8521 Japan
E-mail: †hana-k@ogl.is.ocha.ac.jp , oguchi@ogl.is.ocha.ac.jp, ††sane@cc.kogakuin.ac.jp, †††kami@nagasaki-u.ac.jp

1 はじめに

近年，深層学習技術や画像認識技術が Android などのスマートフォンを始めとするモバイル端末で一般的に利用されるようになった．顔認識，物体認識，ジェスチャー認識などが例として挙げられる．

また，モバイ端末の高性能化も進んでいる．ハイエンド端末では，CPU のみならず GPU や NPU など行列計算を始めとした高負荷な処理を行うことが可能な高性能プロセッサが搭載されるようになった．加えて，モバイル端末に搭載される保存領域も増加傾向にある．これによりユーザはモバイル端末内に以前より多くのデータを保存することが可能になった．

上記のような理由から，モバイル端末内で端末内のデータを使用し端末内専用の画像認識システムが作られるようになった．現状この仕組みを実現する手法として，転移学習技術を用いる手法 [1] や端末のデータを外部のサーバに送信し、演算させた結果を受け取る手法 [2] [3] が存在する．しかし，これらの手法にはそれぞれ精度があまり高くない，データの内容を外部のサーバに送信するという欠点が存在する．データの内容を外部のサーバに送信したくない場合の例としては，プライバシの観点から画像認識モデル作成時に学習時に端末内にしか保存されていない個人情報が含まれた画像を入力データとして使用したい場合や，そのような個人情報が含まれた画像を入力データとして学習させたモデルを外部のサーバに送信したくない場合が考えられる [4].

また，深層学習技術の発達により，汎用的な深層学習モデルをベースとし更に限定したタスクや類似したタスクに対応したモデルを構築する手法として，転移学習のみならずファインチューニングが考案された．ファインチューニングは転移学習よりも計算コストが高いもののより良い精度のモデルを構築す

ることが可能である [5] [6].
そこで，本稿ではモバイル (Android) 端末内のみでファインチューニングを行いモデルの精度を高める，プライバシに配慮した画像認識モデルを構築する手法の提案を行う．本稿では特に，入力画像のデータ量が少ない場合でもファインチューニングを行えるかどうかに着目する．これは，限定モバイル端末上で扱えるデータ量は増加傾向にあるとはいえ，端末サイズが制限されない PC と比べると扱えるデータ量は遥かに小さいためである．

2 関連技術

2.1 Android

Android は Google 社が開発したモバイル端末向けのオペレーティングシステムであり，現在世界でトップシェアを誇っている [7]．Android はオープンソースオペレーティングシステムであり，オペレーティングのコードのみならずアプリケーションソフトウェアのソースコードも公開されている [8]．ソースコードのライセンスは，そのほとんどが Apache License 2.0 に準拠しており [9] ソースコードの改変が自由に行える．本研究では，シェアの高い事による実用性の高さやソースコードが公開されている事による実装のしやすさから Android を実行環境とし選択した．

2.2 NNAPI(Android Neural Networks API)

NNAPI とは Android デバイス上で演算負荷の高い機械学習処理を実行するために設計された Android C API である [10]．この API を使用することで，図 1 [11] に示すように演算処理を実行する際に Android 端末上の利用可能なより良い高性能プロセッサを利用することができる．

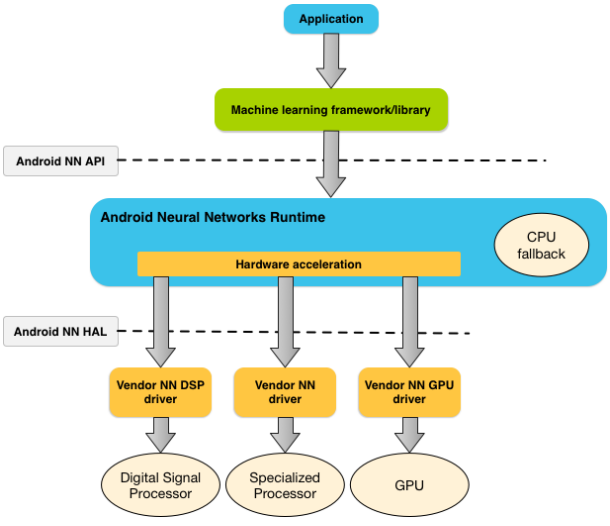

図 1　Android Neural Networks API のシステム アーキテクチャ [11]

2.3 Tensorflow Lite

TensorFlow Lite [12] は TensorFlow のモバイル端末向けの Google の機械学習向けソフトウェアライブラリである．TensorFlow で作成されたモデルは，TFLite Converter [13] を使用することで TensorFlow Lite 形式に変換することができる．その後，変換したモデルをモバイル端末に組み込むことで端末上で推論処理を行うことができる．また．Tensorflow Lite では NNAPI デリゲートの設定を行うことで，NNAPI を使用することができる [14].

2.4 深層学習 (ディープラーニング)

深層学習 (図 5) とは，隠れ層（中間層）の数を増やし階層を深くしたニューラルネットワークを学習することで学習データに含まれる特徴を段階的に深く学習することができる手法である．近年は，深層学習技術により高い精度の画像認識や音声認識が可能になり様々な分野での実用化が進んでいる．

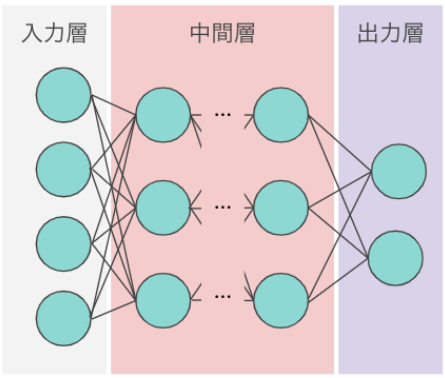

図 2　ディープラーニングのモデル

2.4.1 CNN(畳み込みニューラルネットワーク)

CNN [15] とは深層学習で使用されるニューラルネットワークの 1 つであり，画像認識や音声認識に使用されている．高い精度を誇り，画像認識で使用される深層学習モデルのうち殆どが CNN をベースとしたものになっている．畳み込み層とプーリング層を用い，データのプリミティブな情報を抽出するという特徴がある．

2.4.2 ファインチューニング

ファインチューニングとは，汎用的な学習済みモデルをベースモデルとして用い，そのベースモデルに特定の画像認識タスクを行うモデルに対応させるために再学習させる手法のことである．ファインチューニングの手順は，まずベースとする学習済みモデルの入力層と中間層のみを取り出し，新しく出力層を追加する (図 3)．その後，モデルの学習を行うというものである．この学習を行うことを再学習という．再学習の手順は，まず出力層 (図 4 の①) を学習させる．その後，出力層と一部の中間層の重みを再学習させる (図 4 の②) というものである．先に出力層を学習させることで，モデルの過学習を防ぐことができる．ファインチューニングの転移学習との相違点は，転移学習では追加した出力層のみ再学習を行うが，ファインチューニングでは Batch Normalization 層や畳み込み層の再学習を行うという点である．
画像認識タスクにおける転移学習とファインチューニングの精度の比較を行なっている先行研究では，ファインチューニングがより精度が高いことが示されている [5].

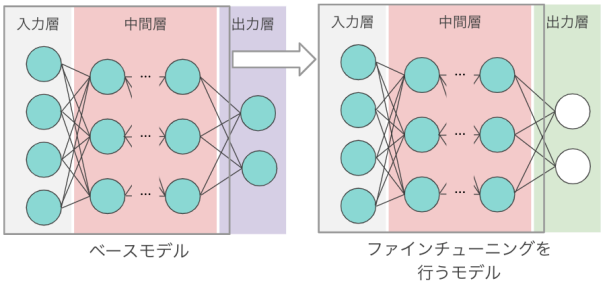

図 3　ベースモデルの入力層と中間層の重みを使用

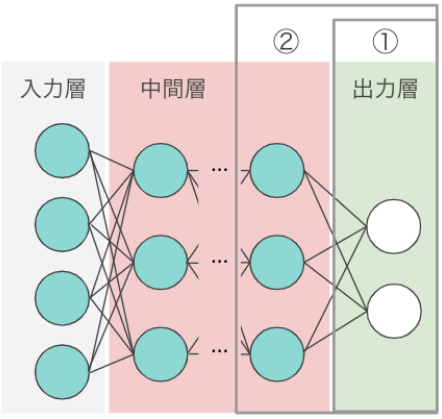

図 4　出力層と中間層の一部を再学習

2.5　MobileNet

モバイルや組み込みアプリケーション向けの画像認識用小型 CNN である [16]．畳み込み層において深さ方向とチャネル方向で分離して演算を行うことで，パラメータ数や計算量を削減している．また，MobileNet の構造は NNAPI で対応可能な層のみで構成されている．このような特性から本研究では MobileNet の構造を使用することにした．

2.6　ImageNet

ImageNet は画像認識用の大規模データセットである．本研究では．ベースモデルとしては ImageNet [17] で学習済みの MobileNet を使用することにした．これは先行研究にもあるように MobileNet は ImageNet で学習することで汎用的な画像認識モデルとして良い精度を出すことが確認されている [16] からである．

3　実装方法の検討

3.1　モデルの構造

ファインチューニングのベースモデルとしては MobileNet [16] を ImageNet [17] で学習したものを使用する．また，追加する出力層は，Global Average Pooling 層，Dropout 層，全結合層で構成する．

3.2　対象とする画像認識タスク

本稿では，犬と猫の判別を行う画像分類タスクを行う．また再学習に使用するデータとしては大きさ 160 × 160 の画像を各種類 100 枚ずつ使用する．100 枚と少ないのは再学習に使用する全てのデータはモバイル端末内のデータであり，モバイル端末内で負荷なく保存可能なデータ量でなければならないためである．

4　評 価 実 験

4.1　実　験　1

本実験では，少数の再学習用データのみを使用するという条件下にて，ファインチューニングが転移学習と比べ精度向上に貢献可能かの確認を行った．

4.1.1　実 験 概 要

この予備実験はモバイル端末上ではなく PC 上で行った．PC 上で実験を行った理由は，PC 用のプログラムは現在 Tensorflow 等のライブラリが存在し，比較的容易に実装が可能なためである．使用したデータセットは”Dogs vs. Cats” dataset [18] であり，このデータセットのうち Dog(犬) の画像 100 枚，Cat(猫) の画像 100 枚を再学習用データとして使用し，再学習用データとは別の Dog の画像 500 枚，Cat の画像 500 枚を精度測定に使用した．バッチ数は 4 に設定した．また，学習回数は出力層の学習で 10 エポック，出力層と再学習する層の学習で 10 エポックとした．

転移学習のみを行う手法 (方法 1) とファインチューニングを行う手法 (方法 2) を比較する実験を行った．それぞれ図 5 に矢印と点線で示している層から出力層までの再学習を行い，再学習後のモデルにおける精度を測定した．図 5 における A の層，B の層とは MobileNet(図 6) の A の層と B の層に対応している．

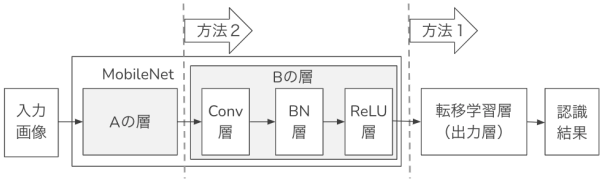

図 5　検証実験時のモデルの概要

4.1.2　実 験 結 果

実験結果を表 1 に示す．

表 1　再学習する層を変化させた際の精度の比較

	方法 1	方法 2
精度 [%]	91.00	95.99

方法 1 と方法 2 の際の精度と損失の変化はそれぞれ図 7 と図 8 となった．

方法 1，方法 2 ともに学習は殆ど収束しており，収束するまで学習した際も方法 1 よりも方法 2 の方が精度が高くなることが確認できる．

この結果より，再学習用のデータが各種類 100 枚と少ない枚数でも，方法 2 の手法である畳み込み層も再学習するファインチューニングを行う手法により精度の良いモデルを構築可能であることが示された．

	Type / Stride	Filter Shape	Input Size
A	Conv / s2	3 × 3 × 3 × 32	224 × 224 × 3
	Conv dw / s1	3 × 3 × 32 dw	112 × 112 × 32
	Conv / s1	1 × 1 × 32 × 64	112 × 112 × 32
	Conv dw / s2	3 × 3 × 64 dw	112 × 112 × 64
	Conv / s1	1 × 1 × 64 × 128	56 × 56 × 64
	Conv dw / s1	3 × 3 × 128 dw	56 × 56 × 128
	Conv / s1	1 × 1 × 128 × 128	56 × 56 × 128
	Conv dw / s2	3 × 3 × 128 dw	56 × 56 × 128
	Conv / s1	1 × 1 × 128 × 256	28 × 28 × 128
	Conv dw / s1	3 × 3 × 256 dw	28 × 28 × 256
	Conv / s1	1 × 1 × 256 × 256	28 × 28 × 256
	Conv dw / s2	3 × 3 × 256 dw	28 × 28 × 256
	Conv / s1	1 × 1 × 256 × 512	14 × 14 × 256
	5× Conv dw / s1	3 × 3 × 512 dw	14 × 14 × 512
	Conv / s1	1 × 1 × 512 × 512	14 × 14 × 512
	Conv dw / s2	3 × 3 × 512 dw	14 × 14 × 512
	Conv / s1	1 × 1 × 512 × 1024	7 × 7 × 512
	Conv dw / s2	3 × 3 × 1024 dw	7 × 7 × 1024
B	Conv / s1	1 × 1 × 1024 × 1024	7 × 7 × 1024
	Avg Pool / s1	Pool 7 × 7	7 × 7 × 1024
	FC / s1	1024 × 1000	1 × 1 × 1024
	Softmax / s1	Classifier	1 × 1 × 1000

図 6　MobileNet の構造 [19]

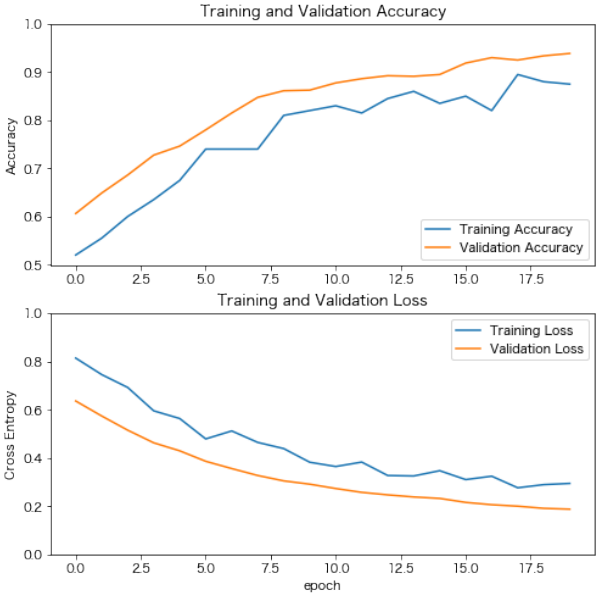

図 7　方法 1 の精度と損失関数の変化

4.2　実　験　2

本実験では，4.1 の方法 2 の手法を採用した際に，再学習を行わない層での演算を Android 端末上で行う際に NNAPI を使うことで実行速度の向上が見られるかどうかの確認を行った．

4.2.1　実 験 概 要

TensorFlow により再学習を行わない層のみのモデルを作成し，そのモデルを TensorFlow Converter を使用して TensorFlow Lite 形式に変換した．その後，そのモデルの実行速度のベンチマークを測定した．

ベンチマークの測定には，ベンチマークアプリケーション [20] を使用した．また，ベンチマークの測定を行った端末は，Pixel4，Pixel5，POCO F2 Pro の 3 種類である．それぞれの端末の性能は表 2 に示す．

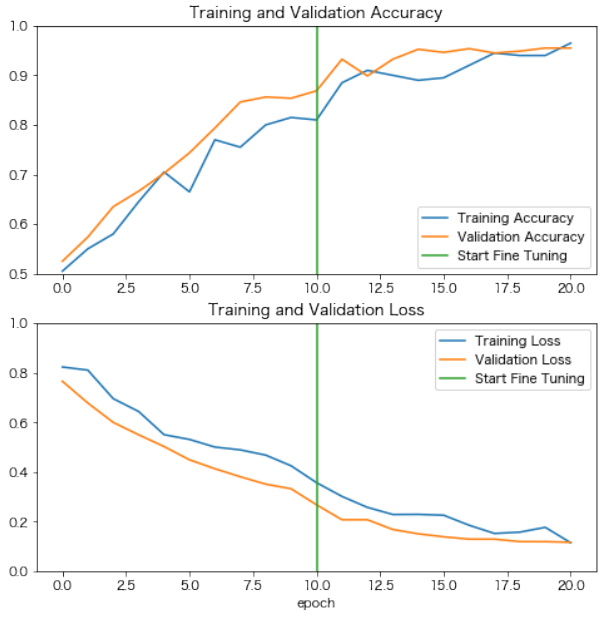

図 8　方法 2 の精度と損失関数の変化

表 2　実験 2 の際に使用した端末の性能

Model number	Pixel4	Pixel5	POCO F2 Pro
OS	Android 12	Android 11	Android 10
SoC	Snapdragon 855	Snapdragon 765G	Snapdragon 865
Memory	6 GB	8 GB	8 GB

4.2.2　実 験 結 果

実験結果を表 3 に示す．

表 3　ベンチマークの測定結果

Model number	Pixel4	Pixel5	POCO F2 Pro
NNAPI 未使用時	16203.4	16697.2	13664.8
NNAPI 使用時	5660.83	13826.7	6593.78

いずれの端末でも，NNAPI 使用時の方が未使用時より実行速度が速くなることが確認できた．

この結果に基づき，モバイル端末上で再学習を行わない層の演算を行うモデルの実装を行った．Tensorflow Lite を使用し，NNAPI デリゲートを設定することで NNAPI を使用するようにした．

5　まとめと今後の課題

本稿では，モバイル (Android) 端末向けのプライバシに配慮した精度の良い結果がえられる個人用の画像認識モデルを構築する手法として，モバイル端末上で畳み込み層も再学習するファインチューニングを行う手法を提案した．

また，モバイル端末上で実行可能なモデルの実装において，重みを固定する層では Android Neural Networks API を利用すると学習速度向上することを示した．

今後の課題は，Android 端末上で実行可能なモデルの実装とアプリケーションの作成を行うことである．

モバイル端末上で実行可能なモデルの実装では 4.1 章の方法 2

の手法に従い，再学習を行う部分の実装も行う．また実装後に学習速度の計測を行い，実装方法やモデルの改善を行っていきたいと考えている．

アプリケーションの作成は，上記のように実装したモデルがモバイル端末上で実用可能か調査するために行おうと考えている．

文　　献

[1] Valery, Olivier, Pangfeng Liu, and Jan-Jan Wu. ”Cpu/gpu collaboration techniques for transfer learning on mobile devices.” 2017 IEEE 23rd International Conference on Parallel and Distributed Systems (ICPADS). IEEE, 2017.

[2] Weng, Yu, and Chunlei Xia. ”A new deep learning-based handwritten character recognition system on mobile computing devices.” Mobile Networks and Applications 25.2 (2020): 402-411.

[3] Sachdev, Jayant, Shashank Shekhar, and S. Indu. ”Melanoma screening using deep neural networks.” 2018 3rd International Conference for Convergence in Technology (I2CT). IEEE, 2018.

[4] Chernyshova, Yulia S., Alexander V. Sheshkus, and Vladimir V. Arlazarov. ”Two-step CNN framework for text line recognition in camera-captured images.” IEEE Access 8 (2020): 32587-32600.

[5] Peng, Peng, and Jiugen Wang. ”How to fine-tune deep neural networks in few-shot learning?.” arXiv preprint arXiv:2012.00204 (2020).

[6] Xiang, Qian, et al. ”Fruit image classification based on Mobilenetv2 with transfer learning technique.” Proceedings of the 3rd International Conference on Computer Science and Application Engineering. 2019.

[7] statcounter — https://gs.statcounter.com/os-market-share/mobile/worldwide(最終アクセス 2020/01/05).

[8] Android Open Source Project — https://source.android.com/(最終アクセス 2020/01/05).

[9] Licenses — Android Open Source Project — https://source.android.com/setup/start/licenses(最終アクセス 2020/01/05).

[10] Android Neural Networks API — Android Developers, https://developer.android.com/ndk/guides/neuralnetworks.(最終アクセス 2020/01/05).

[11] [10] の図 1 より転載

[12] TensorFlow Lite — https://www.tensorflow.org/lite/guide(最終アクセス 2020/01/05).

[13] TensorFlow Lite converter — https://www.tensorflow.org/lite/convert(最終アクセス 2020/01/05).

[14] TensorFlow Lite NNAPI delegate — https://www.tensorflow.org/lite/performance/nnapi(最終アクセス 2020/01/05).

[15] Albawi, Saad, Tareq Abed Mohammed, and Saad Al-Zawi. ”Understanding of a convolutional neural network.” 2017 International Conference on Engineering and Technology (ICET). Ieee, 2017.

[16] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. ”Mobilenets: Efficient convolutional neural networks for mobile vision applications.” CoRR, abs/1704.04861, 2017.

[17] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ”Imagenet large scale visual recognition challenge.” International journal of computer vision 115.3 (2015): 211-252.

[18] ”Dogs vs. Cats” dataset, https://www.kaggle.com/c/dogs-vs-cats/data(最終アクセス 2020/01/05)

[19] [16] の Table 1 より転載

[20] パフォーマンス測定 — TensorFlow Lite https://tensorflow.google.cn/lite/performance/measurement(最終アクセス 2020/01/05)