

多目的自律飛行ドローンシステムのためのLBCNNを用いた人物認識方式 Human Recognition for Resuce Drone by Local Binary Convolution Neural Network

宮里 太也¹⁾ 上原 和加貴¹⁾ 長山 格²⁾
Miyazato Takaya Uehara Wakaaki Nagayama Itaru

1 研究概要

自律飛行型ドローンは、有人ヘリコプターや有人飛行機に比べて機動性が高く、二次被害の可能性を防止できる等の利点から、災害救助時の支援活動や情報収集などにおいて幅広い活用が期待されている。しかし、そのためには高精度な人物自動検出・人物自動認識機能を持つことが必要であるが、上空（俯瞰視点）からの人物認識は容易ではなく、現在、研究が進められている。[1]

一方、画像認識・検出において、Convolution Neural Network（以下、CNN）はすでに欠かすことのできない手法である。この手法は、クラス分類の問題において、人間が網羅できない多数の学習パラメータを制御することで1000クラス以上を分類できるという汎用性を持つ。しかし、この多数の学習パラメータの調整の困難さ、分類時の抽出特徴が不明瞭ということが課題としてあげられる。これらの問題を解決するために、Felixらは畳み込み層としてLBP（Local Binray Pattern）を利用したLocal Binary Convolution Neural Network（以下、LBCNN）を提案している。[3]

我々は、LBPの特徴としてエッジ情報の強調と明度不変性を持つことに注目し、俯瞰視点からの人物検出問題に対してLBCNNは通常のCNNよりも有効である可能性があると考えた。

そこで、本研究ではINRIA Parson Datasetを用いてLBCNNとCNNの2つの手法の精度比較を行い人物検出に対する性能を検証した結果を報告する。本研究の最終目標は、俯瞰視点からの新しい人物検出手法を提案し、自律飛行型ドローンの性能向上を図ることにある。

2 関連研究

CNNは学習パラメータの膨大さから、モデルの計算量を削減する研究が行われている。[4]その中で、FelixらはLBPフィルタを畳み込み層として活用するLBCNNを提案した。このLBCNNとCNNの畳み込み層の処理の違いを図2に示す。図2の上部分がLBCNNによる畳み込みの処理で、下部分がCNNの畳み込み処理の例である。 X_i は一つ前の層の特徴マップを示しており、 X_{i+1} は次層の特徴マップを示している。

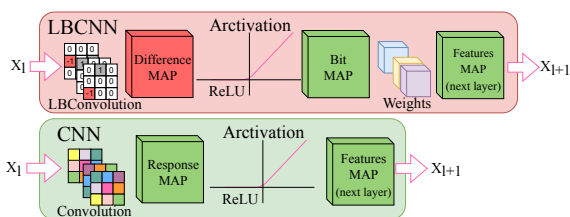


図1 LBCNNとCNNの畳み込み層の違い

- 1) 琉球大学大学院理工学研究科情報工学専攻
- 2) 琉球大学大学院理工学研究科知能情報システムコース

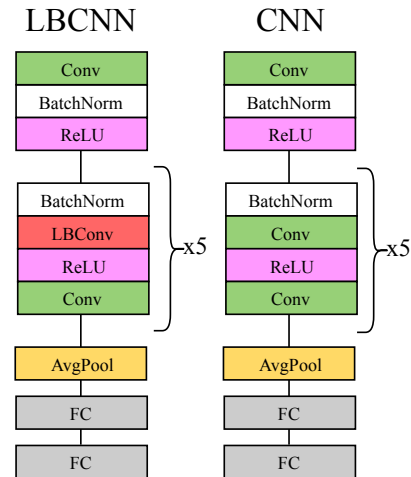


図2 比較するモデルの構造

この図2より、LBCNNは特徴マップに対して前処理としてLBPフィルタを適用したと捉えることができる。LBPはエッジ情報の強調と明度不変性を持つ特徴量である。そこで、本研究ではLBCNNはこの2つの特徴を保持しつつ学習が可能かを調査する。そして、レスキュードローンへ応用するために、人物認識に対してCNNよりも有効性があるか比較検証を行う。

3 比較実験

本研究では、INRIA Person Datasetを用いて2つの問題を検証する。第一の問題は、「LBPフィルタの形状強調により、人の形状を把握しやすい。そのため、人の認識の精度が高くなる。」であり、第二の問題は、「LBPフィルタの明度不変性により、CNNよりも明度に頑健性がある」である。

まず、第一の問題に対しては、CNNとLBCNNの人物認識の汎用性を比較検証することで評価する。評価方法として、4分割交差検定を用いてaccuracyの比較を行った。これにより、認識精度を評価する。

第二の問題に対しては、ホールドアウト法を採用し、テストデータのみ明度を変動させたデータを作成して、それに対するaccuracyの比較を行った。これにより明度に対する頑健性を評価する。

本研究で比較したモデルの構造を図2に示す。図2の左側はLBCNNのモデルであり、右側はCNNのモデルである。両方のモデルは、中央のブロック部を4回同じ処理を繰り返す。実装には、フレームワークとしてpytorchを使用した。

そして、学習の条件は、batch size数は16、入力サイズを224x224サイズ、学習回数は50epoch、誤差関数には交差エントロピー関数を使用した。また、最適化関数は確率的勾配降下法(SGD)を選択し、学習率は0.01と

表1 実験1:4分割交差検定における accuracy

n 交差目	LBCNN	CNN	Alexnet (参考)
$n = 1$	0.998	0.998	0.984
$n = 2$	0.950	0.789	0.981
$n = 3$	0.975	0.994	0.961
$n = 4$	0.948	0.924	0.979
mean	0.968	0.926	0.976

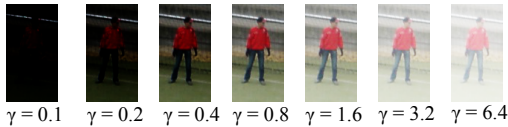


図3 実験2:ガンマ補正で明度を調節した画像

した。

4 実験結果と考察

4.1 実験1:人の認識精度について

はじめに、表1にINRIA Person Datasetにおける4分割交差検定の結果を示す。表1は、左側がLBCNN、真ん中がCNNの各交差時のaccuracyを表している。また、左側に参考データとしてAlexnetによる学習結果も示す。そして、最後の1項目が交差検定の結果(各交差検定の平均値)を示している。最も精度の高い項目は、太字で強調した。

交差結果として、LBCNNの方がCNNよりも、精度が0.042高いことが分かった。しかし、Alexnetが交差検定の結果が0.978と最も高い。

このことから、LBCNNが人認識に対する有効性はあるにはあるが、よくチューニングされたモデルの方がINRIA Person Datasetに対して頑健である。

4.2 実験2:明度不変性について

はじめに、テストデータに前処理としてガンマ補正を施した。パラメータは $\gamma = [0.1, 0.2, 0.4, 0.8, 1.6, 3.2, 6.4]$ としてテストデータの拡張を行った。サンプル画像を図3に示す。

評価としては、ホールドアウト法を採用し、表2に学習終了後のテストデータに対するaccuracyの結果を示す。さらに、図4にepochごとのaccuracy、図5にはepochごとのlossを示す。どちらの図も赤色がLBCNN、緑色がCNN、黄色がAlexnetの結果を表している。図4の曲線は、有効性を差別化するために、accuracyの変化を多項式で近似した曲線である。

表2より、LBCNNが0.962、CNNが0.872であり、LBCNNの方が精度が高いことが分かる。そして、Alexnetは0.710と最も低い精度となった。

図4に示すaccuracyの変化でも、LBCNNが学習の始まりから終わりまでどの手法よりも高い。Alexnetに関しては、25epoch目からaccuracyが減少していることがわかる。これは、図5に示すlossの変化と照らし合わせると、25epoch目以降、トレーニングデータに対して最

表2 実験2:accuracyの結果

モデル	LBCNN	CNN	Alexnet
accuracy	0.967	0.872	0.710

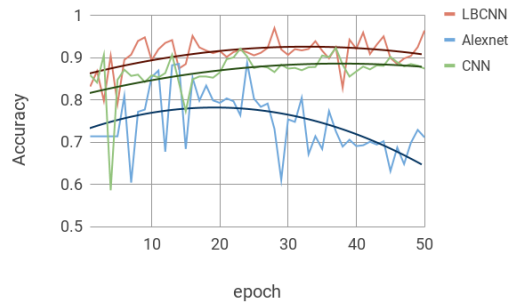


図4 実験2:各エポックの accuracy の変化

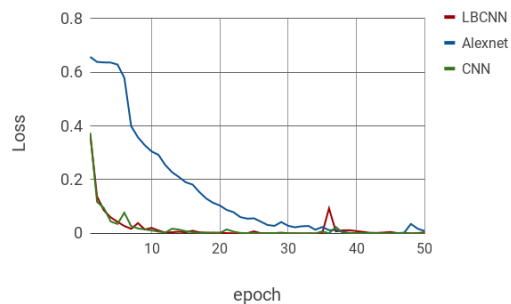


図5 実験2:各モデルの学習曲線

適化されたため、もともと対応していないテストデータに対して過学習が発生したと考えられる。つまり、トレーニングデータとテストデータが依存していると考えられることができる。

一方、LBCNNとCNNの方は近似曲線で見ると、データ同士が依存していない。そのため、明度への不変性を保持していると考えられることができる。

今回の比較実験の結果から、LBCNNは人認識に関しては他の手法と比べて有効性は示せなかったが、明度に対する不変性を持つことがわかった。

5 まとめ

本研究は、レスキュードローンへの応用に向けて、LBCNNとCNNの人物認識精度について比較検証を行った。実験の結果、LBCNNの人物認識に対する有効性を十分に示すことができなかった。しかし、明度に対する不変性を持つことが示唆された。

今後の課題は、このLBCNNを人検出に対しても応用することが挙げられる。

参考文献

- [1] 佐藤 遼次, 越村 俊一, UAV を用いて撮影した光学画像における人検出の精度向上, 地域安全学会論文集 No.27, 2015.11
- [2] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In Advances in Neural Information Processing Systems (NIPS), 2012
- [3] Felix Juefei-Xu, Vishnu Naresh Boddeti, Marios Savvides. Local Binary Convolutional Neural Networks. In the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017
- [4] M. Courbariaux, Y. Bengio, and J.-P. David. BinaryConnect: Training Deep Neural Networks with binary weights during propagations. In Advances in Neural Information Processing Systems (NIPS), 2015