7S-01

# 深層学習を用いた低コストハンドジェスチャ家電操作システムの開発

伊藤　優太†　　渡辺　悠真‡　　ラシキア　ジョージ¶

中京大学

## 1. はじめに

非接触型インタフェースは、近年の新型コロナウイルス蔓延の影響もあり注目を集めている。中でもハンドジェスチャは手話に代表されるように複雑な意思疎通が可能であり、スマートウォッチ操作や医療現場でのモニター操作など多くの場面で活用されている．しかしこれらは、電気信号を認識する装置やkinectやleap motion等のセンサなど専用の機器が必要となっている．ハンドジェスチャ認識（HGR）には古典的な画像処理を利用したHaar特徴やHOG特徴等の手法[1] [2] [3]が存在する。これら提案された HGR 手法は手動で作成した特徴量と分類に基づいているため，複雑環境下では満足なパフォーマンスは得られない。他にも画像ソースのばらつき，照明の不均一性などにより，パフォーマンスは安定しない。HGRシステムで最も重要なことは，優れたロバストな特徴表現の選択であるため，手動による特徴設計は難しい。さらに，認識できるジェスチャ数が少なく，新しいジェスチャの追加も手間がかかり，できないこともある。

近年、深層学習(DL)に基づく HGR が数多く提案されてきた。多くは手の姿勢認識を目的とした静的ジェスチャ[4] [5]と手の動きを対象とした動的ジェスチャ[6] [7]がある。静的ジェスチャはジェスチャ数が限られており不便であり、動的ジェスチャは RNN や３DCNN といったモデルを使用しており処理が重いためコストの高い GPU 等の機器が必要である問題がある。今回安価なハードウェアのみで実用的なシステム開発を行うことが目的である為、静的ジェスチャの追跡によるシンプルな手の動きに対応する手法のハイブリッドHGRに注目する。

小林氏が提案した MobilenetV3Small-YOLOv3（MNSY）を用いたハイブリッド HGR 家電操作システム[8]は YOLOv3 の backbone を MobilenetV3Small に置き換え、Head 部分に Depthwise Separable Convolution を適応したモデルでパラメータ削減による軽量化と高速化を可能にした。しかし、左右動作を行う際に手の向きが変わりやすいため、左右動作の認識率の問題があった。

そこで本システムのモデル出力後のデータ処理により認識率の改善を提案する。具体的には移動距離に応じてモデルのしきい値の調整を行うことで追跡時の手の向きの変化による認識率低下の問題を改善した。また、手の側面の画像を新たに撮影し、モデルの再学習を行った。新たに提案されたモデルについて Fine Tuning を行い、速度と精度の面で比較を行った。

「Development of low-cost hand gesture deep learning system for home appliance control」

†「Yuta Ito・Chukyo University」

‡「Yuma Watanabe・Chukyo University」

¶「George Lashikia・Chukyo University」

## 2. 提案手法

**〈2･1〉 提案モデル**　本システムの目的は実用的な軽量モデルの開発である。ハイブリッド HGR を行うために物体検出ニューラルネットワークを利用する。手の側面を含むデータセットを作成し、MNSY を含め近年発表された最先端軽量モデル（MobilenetV3Small+SSD, YOLOv5, YOLOv7）を Fine-tuning し、パフォーマンス比較を行った。また、システムの面では、始めの認識位置からの移動距離を元にしきい値に対してコサイン アニーリングを適応することで、精度の安定性を図る。

**〈2･2〉 学習**　モデルを学習するために Creative Senz3d Dataset[9] [10]をもとに手の画像を集めた。Creative Senz3d Dataset に含まれていない，人差し指、中指、薬指を立てた３の姿勢の手を自作で集めた。また暗い環境でのデータや手の側面の画像も少ないためこちらも自作で集めた。クラスは 12 種類にし，画像の枚数は合計約 36,000 枚を用意した。学習データに 8 割，検証データに 1 割，テストデータに 1 割を使用した。画像のアノテーションは，自作のアノテーションツールを作成して行なった。ニューラルネットワークの作成にはTensorFlow と Pytorch を使用した。

**〈2･3〉 モデルの評価**　実験を PC 上で行った。ハードウェアは Intel corei7 10700K，OS は Windows11 を使用した。今回は認識速度，認識精度，についてモデルの比較を行った。YOLOv7 はモデルの標準バージョンは精度が高いが，パラメータ数が膨大であるため高速に実行することのできる YOLOv7-tiny を比較対象とした。同じ理由で YOLOv5 は軽量モデルである YOLOv5s を、YOLOv3 は軽量モデルである YOLOv3-tiny を選択し、MobileNetV3 + SSD の backbone には MobileNetV3 Small を選択した。

精度評価に mean Average Precision (mAP)の 0.5-0.95 を使用した。これは COCO Object Detection Challenge の評価指標であり、IoU の値を 50～95%まで 5%刻みで変化させ認識率を計算し，その結果を平均したものが最終的な mAP となる。実行速度を評価する指標として一般的に使われる frames per second（fps） を使用した。得られた実験結果を Table１に示す。結果、MNSY は 84.75 と高い精度で 18.69fps と精度と速度のバランスが良いため，システムに MNSY を利用した。また、低スペックデバ

イスでの高速な実行を実現するため、tensorflow-liteを利用しモデルの最適化を行ったものをシステムに導入した。

| Model | FPS | mAP(0.5-0.95) |
|---|---|---|
| YOLOv7-tiny | 35.21 | 73.75 |
| YOLOv5s | 32.67 | 67.38 |
| MNSY | 18.69 | 84.75 |
| YOLOv3-tiny | 7.30 | 81.26 |
| MobileNetV3+SSD | 30.95 | 75.04 |

**表１：Table 1. Performance Comparison Among Different Models**

**〈2･4〉 提案システム**　本システムに用いる主な部品は，Raspberry Pi4，Pi カメラ、irMagician である。全体構成を図１に示す。

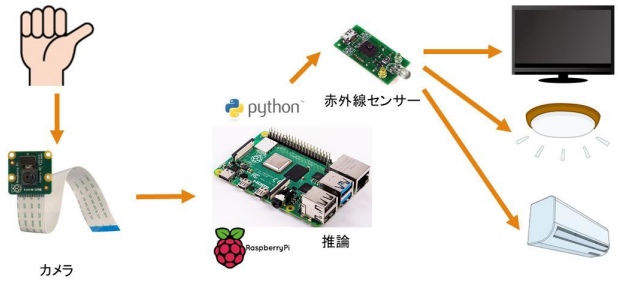


図１：全体構成

本システムはすべての処理を Raspberry Pi 上で完結するため，ネットワーク環境は必要ない。本システムを動作させると，カメラが起動し，撮影された映像のリアルタイム物体認識が行われる。姿勢を認識すれば一定時間同行を追跡し、対応する赤外線信号を出力する。赤外線との対応付けはシステム内の GUI により設定できる。ジェスチャは複数フレームを用いて決定している。誤認識抑制とよりスムーズなジェスチャ認識の為しきい値の調整を行った。最初のしきい値を高い値にして誤認識を抑制し、動的ジェスチャ中はしきい値に対してコサインアニーリングを適応することで移動距離に応じて徐々にしきい値を下げてジェスチャ中の認識はずれ問題の解消を図る。

## 3　今後の展開

今後の展開は MNSY の更なる高速化のためモデルの改良を行う。また最も高速であった YOLOv7-tiny に対してもモデルの改良を行い、各モデルの性能比較を行う。また、システムの評価実験を行い有用性についても調査を行う。

## 4　まとめ

近年，非接触型インタフェース中でも HGR は注目を集めている。そこで，本研究で家電操作システムの開発を行った。提案手法として DL に注目し，安価なデバイスで実行できる様々な最先端モデルを作成し，本システムに最適なモデルを採用した。誰でも気軽に利用できるようにエンドツーエンドリアルタイムシステムを開発した。今後はモデルの更なる高速化のために改良をし、各モデルの性能比較を行う。また完成したシステムに対し評価実験を行い、有用性を調査する。

参考文献

[1] 牛丸太希，佐藤一誠，中川裕志，“3 次元 Haar 特徴量を用いたハンドジェスチャ認識”，研究報告数理モデル化と問題解決 (MPS), 2014.
[2] 山下大輔，間博人，山本泰士，本田雄亮，三木光範，“モバイル端末のアプリケーション利用時における内蔵照度センサを用いたハンドジェスチャ認識手法の提案”，情報処理学会論文誌 vol.59, no.2, pp. 715-722, 2018
[3] H. Lahiani and M. Nejib, “Hand Gesture Recognition Method Based on HOG-LBP Features for Mobile Devices”, Procedia Computer Science, vol. 126, pp. 254–263, 2018.
[4] S. Ameen and S. Vadera, “A Convolutional Neural Network to Classify American Sign Language Fingerspelling from Depth and Colour images”, Wiley Expert Systems, 2016.
[5] V. Adithya and R. Rajesh, “A Deep Convolutional Neural Network Approach for Static Hand Gesture Recognition”, Procedia Computer Science, Elsevier, vol. 171, pp. 2353-2361, 2020.
[6] P. Molchanov, X. Yang, S. Gupta, K. Kim, S. Tyree and J. Kautz, “Online Detection and Classification of Dynamic Hand Gestures with Recurrent 3d Convolutional Neural Network”, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.
[7] O. Köpüklü, N. Köse and G. Rigoll, “Motion Fused Frames: Data Level Fusion Strategy for Hand Gesture Recognition”, IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018.
[8] 小林 優太，ラシキア 城治，“低コストのディープニューラルネットワークベース家電操作用リアルタイム手振り認識システム”，情報処理学会論文.C，vol.141, no.7, pp.822-831，2021.
[9] A. Memo, L. Minto and P. Zanuttigh, “Exploiting Silhouette Descriptors and Synthetic Data for Hand Gesture Recognition”, STAG: Smart Tools & Apps for Graphics, 2015.
[10] A. Memo and P. Zanuttigh, “Head-mounted Gesture Controlled Interface for Human-computer Interaction”, Multimedia Tools and Applications, 2017.