

## U-Net モデルを用いた加速度センサデータの超解像 Super-Resolution for acceleration sensor data using U-Net model

武仲 紘輝<sup>†</sup> 長谷川 達人<sup>†</sup>  
Koki Takenaka Tatsuhito Hasegawa

### 1. はじめに

スマートフォンやウェアラブルデバイス等の加速度センサデータからユーザの行動を推定する行動認識に関して、機械学習を用いた研究は数多く存在し、深層学習を用いた研究も増加している。しかし、多くの研究が同一機器から取得できる加速度センサデータを利用して行動分類を行っており、サンプリング周波数等の異なる加速度センサデータが存在する状況を考慮した研究は少ない。また、深層学習のモデルが表現を獲得するには大量の学習データを必要とする。大量の学習データの収集には様々な機器での計測の実施が効率的である。一方で、このようにして収集された加速度センサデータはサンプリング周波数等の条件が混在するという問題を包含している。サンプリング周波数が混在した加速度センサデータを用いて機械学習モデルの訓練を行った場合、推定精度が低下する恐れがある。したがって、条件が異なる場合においても頑健に動作する行動認識手法が望まれる。条件の異なるセンサデータを利用し、高い推定精度を得られる手法というのは明らかになっていない。

機械学習を用いた行動認識分野において、サンプリング周波数等の条件の異なる加速度センサデータを利用した状況下における、推定精度の向上手法を開発することを本研究の目的とする。サンプリング周波数の異なる加速度センサデータに対して超解像技術を適用することでサンプリング周波数を統一する手法を新たに提案し、行動認識精度の向上を目指す。行動認識のベンチマークデータセット HASC[1]を用いて、様々なアップサンプリング手法と比較評価実験を行い、超解像の精度の比較を行う。

### 2. 関連研究

音声分野の超解像に関する研究として Kuleshov らは U-Net 構造[2]を基にしたモデル[3]を提案している。彼らは音声データに対して画像の超解像手法を用いた。ただし、本研究で用いる加速度センサデータのように負の値が存在する場合は対象としていない。

Stisen ら[4]は、様々なスマートフォンで測定した加速度センサデータを利用した行動認識に取り組んでいる。この研究では、各デバイスのセンサ特性の差異や、CPU 使用率における影響を示している。その影響を軽減するためにセンサ特性が似ている機種ごとにクラスタリングし、クラスターごとに学習を行う手法を提案している。本研究は計測時の情報のみを取り扱うため、その点でこの研究と異なる。

室内位置推定研究と建物構造推定研究を主な目的とした室内歩行センシングコーパス HASC がある。このコーパスは行動認識に用いることのできる加速度センサデータを含んでいる。様々なスマートフォンやセンサデバイス等の機

<sup>†</sup> 福井大学大学院工学研究科

Graduate School for Engineering, University of Fukui

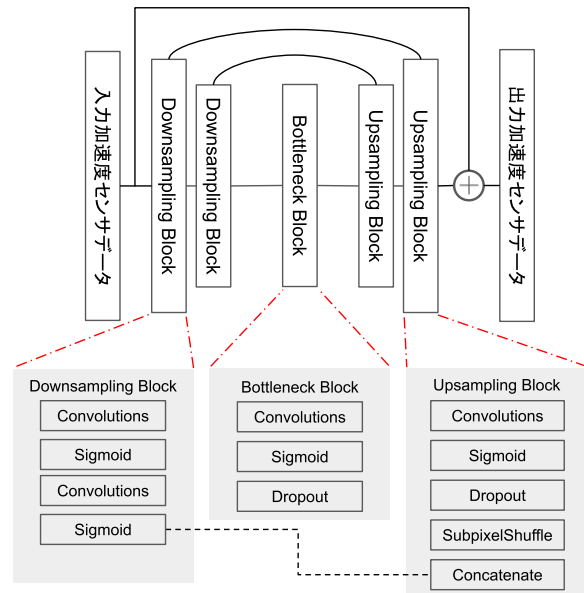


図1 提案モデルの構成。

器が混在した環境下でデータが計測されており、サンプリング周波数等といった計測条件に関しても様々である。本研究では基本行動認識のベンチマークとして利用する。

画像分野において深層学習を用いた超解像手法は活発に研究されており SRCNN[5] や SRGAN[6]、またそれぞれの手法を改良したモデル等多くのモデル[7,8,9]が提案されている。一方、画像の超解像手法をセンサデータに適応させた研究は、筆者が調べた限り見つからない。

### 3. 提案手法

Kuleshov らのモデル[3]を参考に、加速度センサデータを超解像するモデルを提案する。モデルは U-Net 構造を基にしており、Downsampling ブロックと Bottleneck ブロック、Upsampling ブロックの3つのブロックからなる。そして、Downsampling ブロックと Upsampling ブロックの間にスキップコネクションがある。モデルの特徴として、Upsampling ブロックに Subpixel Shuffling を用いている点があげられる。このレイヤは画像分野のアップサンプリングを行うことができる Subpixel [8]を音声データや加速度センサデータなどの1次元データに適応させたレイヤである。しかし、このモデルは音声データでの利用を想定しており、出力できる値の範囲が限られている。そこで我々はこのモデルを参考にし、加速度センサデータに対応できるようなモデルを提案する。

実験に使用する機械学習モデル(図1)は参考にしたモデルと同様に Downsampling ブロック、Bottleneck ブロック、そして Upsampling ブロックの3つのブロックから構成された U-Net 構造をしている。各ブロックは図1のように上から

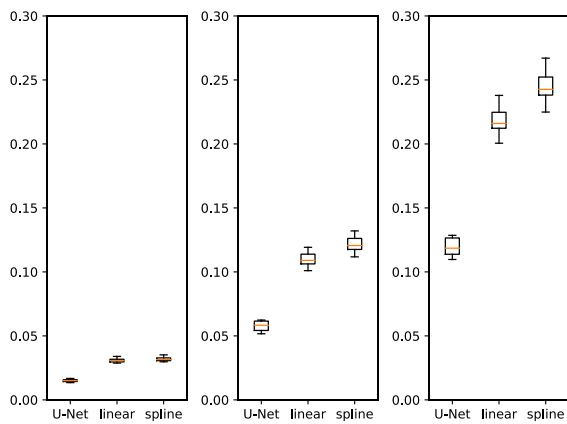


図 2 提案手法，線形補間，spline 補間のアップサンプリング精度 (MSE) の比較結果 (左から，2，4，8 倍のアップサンプリング)。

順にレイヤが接続されており，Downsampling ブロックと Upsampling ブロック間にスキップコネクションがある。このスキップコネクションは図 1 中の点線で示されているように Downsampling ブロックの出力と Upsampling ブロックの Concatenate レイヤに結合している。また，各ブロックの活性化関数を ReLU 関数から Sigmoid 関数に変更した。この変更により加速度センサデータに存在する負の値に対応することが可能となった。

## 4. 比較実験

### 4.1 実験設定

復元精度を比較するために加速度センサデータの測定条件に過度な差が存在しないように，使用する HASC データセットの加速度センサデータの内，Apple 製品の 100Hz のデータを使用する。また，使用する加速度センサデータに関して全ての被験者の内 7 割 (354 人) を訓練用，残り 3 割 (153 人) を検証用としてランダムに設定し，これを 10 セット用意する。

加速度センサデータは 100Hz から 50Hz，25Hz，12.5Hz の 3 種類にダウンサンプリングして実験を行う。それぞれ 2，4，8 倍のアップサンプリングを行うことで 100Hz にし，復元精度の比較を行う。比較を行う手法は線形補間，spline 補間，そして提案モデルによる手法とし，復元精度の指標は平均二乗誤差 (Mean Squared Error; MSE) とした。

モデルを訓練する際，損失関数は MSE を使用し，最適化メソッドは Adam を使用した。加速度センサデータは 1024 のサイズに分割し訓練を行い，エポック数を 40，バッチサイズを 32 と設定した。

### 4.2 結果

実験結果を図 2 に示す。この図は各手法の比較結果を表しており，各グラフは左から 2，4，8 倍の倍率でアップサンプリングを行った時の，10 セットの試行結果の MSE を箱ひげ図で示している。加速センサによる行動認識を対象データとしている特性上，100Hz と 50Hz には大きな差がなかったため，2 倍に超解像した場合，どの手法も MSE が小さい。また，どの超解像倍率においても提案手法が最も

誤差の小さい結果となった。線形補間と spline 補間では，提案手法と比較した場合の差ほど大きな差は見られない結果となった。さらに，倍率を大きくするほど誤差が顕著になるが，提案手法は線形補間と比べ，MSE をおよそ半分に低減することに成功している。

興味深い結果として，12.5Hz を提案手法で 100Hz に復元した際の MSE と，25Hz を従来手法で 100Hz に復元した際の MSE がほぼ同程度 (0.12) になっていることがある。これは，提案手法によって 12.5Hz のデータを 25Hz のデータと同様の情報量に復元できている可能性を示唆する。

## 5. まとめ

本実験では，超解像手法を用いた加速度センサデータのサンプリング周波数の統一のため，U-Net 構造を基にした機械学習モデルの提案した。そして，加速度センサデータに対する各種補間方法における補間精度の比較を行った。結果として提案手法が MSE を最も小さくでき，線形補間と比べると MSE がおよそ半分に低減できた。

今後の課題として，画像分野における超解像手法をセンサデータに適用すること，補間を行った加速度センサデータに対して行動認識の精度の調査を行うことがあげられる。

### 謝辞

本研究の一部は立石科学技術振興財団の研究助成 (A) によるものである。ここに謝意を表す。

### 参考文献

- [1] N. Kawaguchi, N. Ogawa, Y. Iwasaki, K. Kaji, T. Terada, K. Murao, S. Inoue, Y. Kawahara, Y. Sumi, N. Nishio, "HASC Challenge: Gathering Large Scale Human Activity Corpus for the Real-World Activity Understandings", Proceedings of the 2nd Augmented Human International Conference (AH2011), 27(2011)
- [2] O. Ronneberger, P. Fischer, T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, LNCS, Vol.9351: 234-241, 2015
- [3] V. Kuleshov, S. Z. Enam, S. Ermon, "Audio Super-Resolution using Neural Nets", Workshop track - ICLR 2017
- [4] A. Stisen, H. Blunck, S. Bhattacharya, T. S. Prentow, M. B. Kjergaard, A. Dey, T. Sonne, M. M. Jensen, "Smart devices are different: Assessing and mitigating mobile sensing heterogeneities for activity recognition". In Sensys, 2015
- [5] C. Dong, C. C. Loy, K. He, X. Tang, "Learning a Deep Convolutional Network for Image Super-Resolution", in Proceedings of European Conference on Computer Vision (ECCV), 2014
- [6] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, W. Shi, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network", 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 105-114, doi: 10.1109/CVPR.2017.19.
- [7] C. Dong, C. C. Loy, X. Tang, "Accelerating the Super-Resolution Convolutional Neural Network", Proc. of European Conference on Computer Vision (ECCV), 2016
- [8] W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network", Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016
- [9] Y. Chen, Y. Tai, X. Liu, C. Shen, J. Yang, "FSRNet: End-to-End Learning Face Super-Resolution with Facial Priors", 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 2018, pp. 2492-2501, doi: 10.1109/CVPR.2018.00264