

Quad-directional LSTM を用いた音楽音響信号修復とその評価

谷口亮輔¹, 干場功太郎¹, 中臺一博^{1,2}

¹ 東京工業大学工学院システム制御系

² (株) ホンダ・リサーチ・インスティテュート・ジャパン

1 はじめに

本稿では、深層学習の位置手法である LSTM (Long Short-Term Memory) を用いた音楽音響信号修復について報告する。一般的に、深層学習を用いて性能の高いモデルを学習するためには、大量のデータが必要である。実際に音楽音響信号修復に LSTM を適用した場合、学習データ量の不足によって、情報が比較的スパースである高域の学習が十分でなくなり、修復性能が劣化する。この問題に対し、これまで我々は入力信号に対して高域を強調する周波数フィルタを用いて、その解決を試みた [1]。また、その拡張として、順方向の時系列のみではなく、逆方向の時系列情報を考慮した BLSTM (Bi-directional LSTM) を用いた拡張を行った [2]。さらに、より詳細な修復のために時間方向のみではなく、周波数方向の系列情報を考慮することが可能な QLSTM (Quad-directional LSTM) への拡張を行った [3]。本稿では、これらの手法を実際の欠損に対して用いる方法を提案し、その修復性能を評価する。

2 深層学習を用いた音楽音響信号修復

2.1 フィルタ内包型 LSTM

LSTM は、系列データである音楽音響信号を扱うのに適した深層学習手法である。特徴としては、系列方向に対して再帰構造を持つことが挙げられる。また、内部に情報を保持するメモリセルと、各種入出力にゲートを持つ。これらによって、学習の際に、誤差の消失や異常な増加に対する影響を受けにくく、長期の依存関係を扱うことができる。しかし、通常の構成のまま用いると、音楽音響信号に対しては、低音域に対して情報が比較的スパースな高音域の学習が十分に行われない。そのため、周波数フィルタを内包し、事前に高音域を強調することで学習を効率的に行うことができるフィルタ内包型 LSTM [1] を今回用いた。

2.2 BLSTM

単一の LSTM を使用する場合、単一方向 (順方向) のみの系列しか考慮できない。そこで、逆順の系列を入力として処理を行う LSTM を合わせて用いることにより、順逆両方向 (Bi-directional) での系列を考慮できるようにした BLSTM による音楽信号修復を提案した [2]。順方向と逆方向それぞれのフィルタ内包型 LSTM を独立に学習して用い、出力の平均を取ることで BLSTM を実現した。

2.3 QLSTM

音楽音響信号は、時間方向の系列データであるとともに、周波数方向にも関係性を持つ二次元の系列データである。そのため、周波数方向を考慮することができる BLSTM をあわせて用いることでより修復性能の向上が期待できる。この二次元四方向での構成を行ったものが QLSTM である。今回の評価では、時間方向、周波数方向それぞれを独立に学習し、出力を平均を取ることで統合を行う構成 [3] である。

3 実際の欠損修復を想定した評価実験

楽曲データ (サンプリングレート 16 kHz) に対し、フレーム長 512 サンプル、シフト長 128 サンプル、ハミング窓を窓関数として用いた STFT を行い、振幅スペクトルを得る。各楽曲での最大振幅値を用いて正規化を行う。学習では、時間方向 LSTM は、連続する 3 フレームを入力し、次のフレームの振幅スペクトルを予測する学習をする。周波数方向 LSTM では周波数ビン f ごとに、前後 3 フレーム ($x_{t,f} \sim x_{t+2,f}$, $x_{t+10,f} \sim x_{t+12,f}$) の計 6 フレーム部分を入力とし、入力に挟まれる 6 フレーム ($x_{t+3,f} \sim x_{t+9,f}$) 部分を予測するように学習を周波数方向に行った。すべてのモデルに対して学習では 6 曲の Jazz 楽曲を用いた。欠損は、複数フレームにわたってすべての周波数ビンを 0 にすることで作成した。欠損のフレーム数 N は 2, 4, 6 の三種類を用いた。欠損開始フレームを t_m とすると、 $t_m \sim t_m + N - 1$ が欠損フレームとなるので、この部分を提案手法によって復元する。 t_m を 1~1000 まで変化させて、その都度修復した結果を並べたものを修復結果として、欠損のない元データとの比較を行った。なお、周波数方向 BLSTM については構成上 6 フレーム分を一度に出力するため、6 フレームの欠損を与えた場合の出力結果を、2, 4 フレームの欠損を与えた場合にも用いる。

4 比較結果

Table 1 に各モデルからの出力をとり、SDR (Signal-to-DistortionRatio) を算出した結果を示す。また、このうち、6 フレーム欠損の場合の修復結果を Fig.1 にしめす。なお、SDR は式 1 で表され、 s は正解、 y は出力であり、値が大きいほど歪が少ないことを示す。

$$SDR = 10 \log_{10} \frac{\sum s^2}{\sum (s-y)^2} \quad (1)$$

単方向 (順方向, 逆方向) LSTM, BLSTM, QLSTM については、欠損フレーム数が少ない方が修復性能が高いことがわかる。また、欠損フレーム数にかかわらず、単方向 (順方向, 逆方向) LSTM, BLSTM, QLSTM の順で修復性能が高くなっていることがわかる。これにより、その周辺の情報を考慮すれば考慮するほど性能が向上するといえよう。一方、欠損フレーム数が 6 の場合、周波数方向 BLSTM が最もよい修復性能を示し

Evaluation of musical audio signal restoration using Quad-directional LSTM

Ryosuke Taniguchi¹, Kotaro Hoshiba¹, Kazuhiro Nakadai^{1,2}

¹ Department of Systems and Control Engineering, School of Engineering, Tokyo Institute of Technology

² Honda Research Institute Japan Co., Ltd.

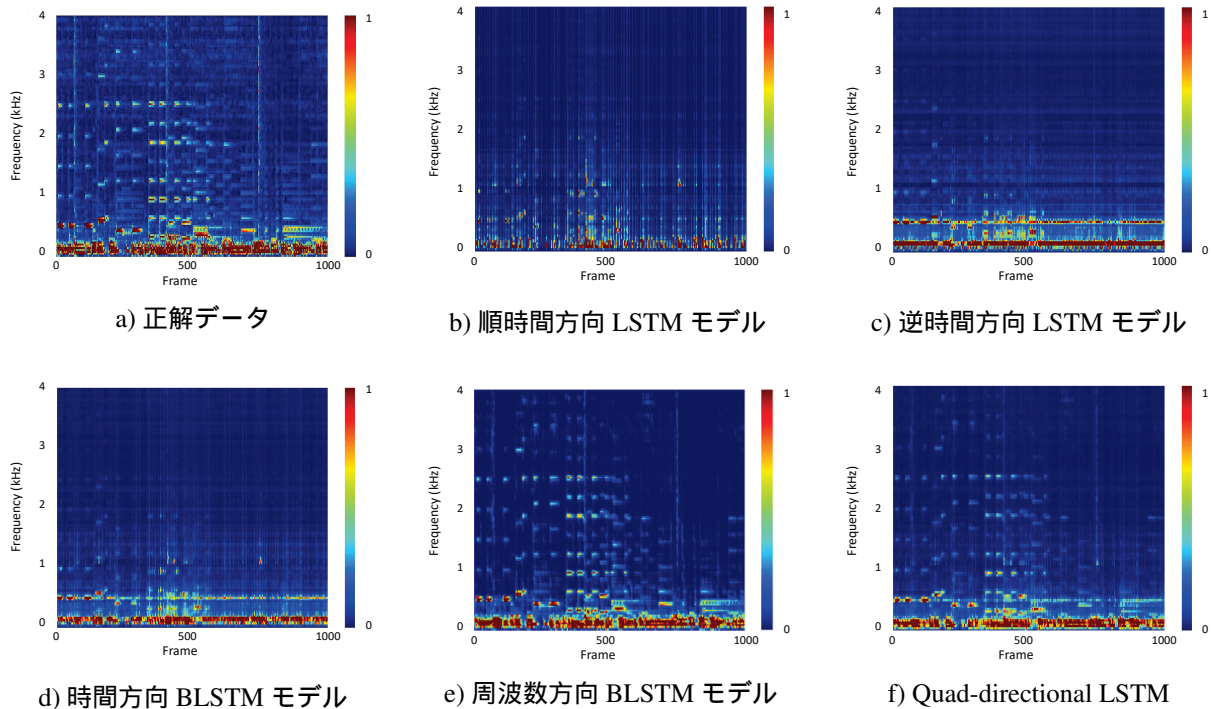


Fig. 1 各モデルによる6フレーム欠損の修復結果と正解データ

Table 1 各モデル出力のSDR

モデル	2 frame	4 frame	6 frame
順時間方向 LSTM	4.24	2.83	1.92
逆時間方向 LSTM	4.33	3.17	2.27
時間 BLSTM	5.35	4.08	3.06
周波数 BLSTM	N/A	N/A	6.51
QDLSTM	6.64	5.99	5.35

ている。さらに、この結果は4フレーム欠損の他の手法の結果よりもよい結果となっている。これは、4フレームつまり40ms離れてしまうと、時間方向の関係の有効性が低下し、周波数方向のそれよりも小さくなってしまふためであると考えられる。また、逆時間方向 LSTM モデルでは500 Hz 付近で全体を通して残るような信号を出力している。これは、減衰していく信号のオンセット部分を学習する際に、信号の強い低域ではうまく学習がされず、オンセットの判断ができなかったためと考えられる。

5 考察

今回用いた方法では、時間方向 BLSTM を構成するために順・逆二方向の LSTM の出力結果を平均をとることで構成している。このような単純な統合でも、単一方向の LSTM よりも性能向上がみられる。これは、平均を取ることで順時間方向 LSTM の偽陰性の誤差が逆時間方向 LSTM の真陽性成分により抑えられる、また、逆に逆時間方向 LSTM の偽陽性信号が順時間方向 LSTM の真陰性の部分により抑えられることも同時に行われるためだと考えられる。これは QDLSTM についても同

様に考えられるが、時間方向の欠損が大きくなりすぎると時間方向 BLSTM より、周波数方向 BLSTM の有効性が大きくなる。このため、これらを統合した QLSTM は、結果として時間方向 BLSTM の性能低下に足を引っ張られた形となってしまう、周波数方向 BLSTM よりも性能が低下してしまった。これを解消するためには、単なる平均ではなく、ニューラルネットワークなどを用いたより高度な統合を行う必要がある。

6 まとめ

本稿では、実際の欠損に対して QLSTM を適用するための方法を提案した。実際の欠損を想定した複数フレームでの修復を行うことで評価を行うことにより、本手法では、一方向 LSTM に対してより歪の少ない修復が可能であることが示された。

謝辞 本研究は、JSPS 科研費 16H02884, 16K00294, 17K00365 および、JST ImPACT タフロボティクスチャレンジの助成を受けた。

参考文献

- [1] 谷口他, “LSTM による音楽音響信号の修復法の提案-周波数フィルタ導入による学習データ量削減の検討”, 第 79 回情報処理学会全国大会, 2017
- [2] 谷口他, “Bi-directional LSTM を用いた音楽音響信号修復法の提案”, 第 35 回 日本ロボット学会学術講演会, 2017
- [3] 谷口他, “Quad-directional LSTM を用いた音楽音響信号修復法の提案”, 第 49 回 人工知能学会 AI チャレンジ研究会, 2017