

## 1. はじめに

骨格データを対象とした動作認識として, Graph Convolutional Networks (GCN) を用いて認識を行う ST-GCN [1] が提案されている. この手法では, 関節座標をノード, 関節の関係性をエッジとしたグラフ構造と見なしている. そのため, 関節座標の特徴だけではなく, 関節間の関係も考慮され複雑な動作を認識できる. 一方で, 人の骨格パターンをあらかじめ定義しているため, 畳み込み処理を適用する範囲が限定される. そのため, 各動作固有の関節の関係性を考慮した特徴を捉えることができない場合がある. そこで本研究では, 動作クラス毎の骨格間の重要度を示す重みの大きさから, 動作認識における最適な接続パターンを獲得する手法を提案する.

## 2. ST-GCN

ST-GCN [1] は, GCN を用いた動作認識手法であり, 骨格データを 2 つのグラフ構造として捉えることで, 従来手法より高い精度を実現している. 2 つのグラフ構造とは, 同一フレーム内の関節を結ぶことで, 関節間の関係を考慮する空間グラフと, フレーム間の同一関節を結ぶことで, 関節の時間的な変動を考慮する時間グラフである. ST-GCN の概略図を図 1 に示す. ここで, グラフの各ノードの特徴  $f_{in}$  における, 空間グラフに対するグラフ畳み込み処理の出力  $f_{out}$  を重み行列  $W$  を用いて式 (1) で定義する.

$$f_{out} = \hat{A} f_{in} W \quad (1)$$

$\hat{A}$  はグラフ構造の接続関係を示す正規化された隣接行列であり, ノード数が  $N$  のとき隣接行列のサイズは  $N \times N$  となる. 隣接行列との乗算により, 接続されたノードとのみ畳み込み処理を行う.

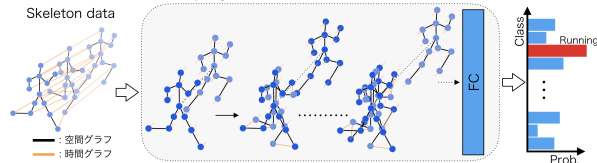


図 1: ST-GCN の概略図

## 3. 提案手法

ST-GCN における空間グラフの畳み込み処理は, 人の骨格における接続パターンでのみ行われる. したがって, 隣接する関節同士でしか畳み込み処理を行わないため, 左右の手の関係や, 手と足との関係などの特徴が考慮されない. また, 動作によって重要となる関節が異なることが考えられる. そのため, 本研究では動作クラス毎の接続パターンを考慮した動作認識手法を提案する.

図 2 に提案手法のネットワークを示す. 提案手法では各動作クラスの特徴を獲得するためにマルチタスクラーニングを導入する. すなわち, それぞれのタスクとして各動作クラスに関する 2 クラス分類を行う. 接続パターンは動作ごとに異なるため, ネットワークの共有部分, 各タスク部分で独立した隣接行列を用いる. また, 動作認識のための重要なエッジを選択するため, 重み行列  $M$  を式 (2) のようにグラフ畳み込み処理に導入する. ノード数  $N$  のとき, 重み行列  $M$  のサイズは  $N \times N$  となる.

$$f_{out} = (\hat{A} \circ M) f_{in} W \quad (2)$$

ここで,  $\circ$  は要素積を示す. 隣接行列  $\hat{A}$  と重み行列  $M$  との要素積をとることで, 各エッジに重みを付与する. 重み行列  $M$  の初期値は  $M_{ij} = 1/N$  とする. また, 隣接行列の初期値を全て 1 とし, 全ノードを接続した状態を初期状態とする. 学習の途中で隣接行列の値を更新することでエッジを削減し, 新たな接続パターンを獲得する. それ以降, 獲得した接続パターンでグラフ畳み込み処理を行う. 各タスクの隣接行列の更新は, 各タスクにある重み行列  $M$  の値をもとに, 重みの大きいエッジ  $K$  本のみに限定する.

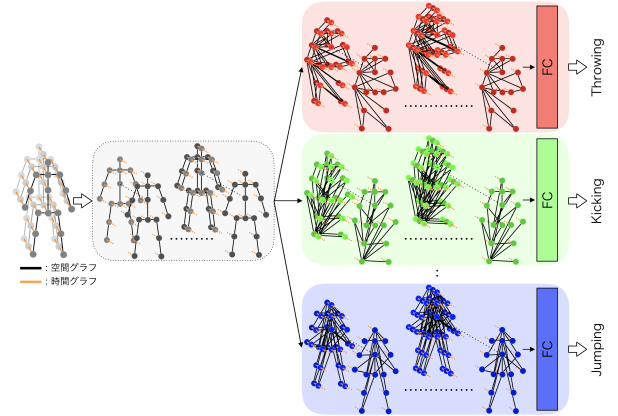


図 2: 提案手法のネットワーク

## 4. 評価実験

提案手法の有効性を評価実験によって検証する.

### 4.1. 実験概要

評価実験には NTU-RGB+D データセット [2] を用いる. 関節数は 25 であり, それぞれ 3 次元座標  $(X, Y, Z)$  を持つ. 60 種類の動作クラスの内, 投げる, 蹴る, ジャンプの 3 つの動作クラスを用いて実験を行う. 各動作のフレーム数は 80 フレームとする. 評価方法は, 学習に 20 人分, 学習に用いなかった 20 人分のデータで評価する. 学習回数は 100epoch とし, 隣接行列の更新は 40, 50, 60, 70epoch 目に行う. このとき, 70epoch 目でエッジの数が 60 本になるようにする. 共有部分では隣接行列を更新しない.

### 4.2. 実験結果

ST-GCN と各動作クラスに関する 2 クラス分類, および提案手法による認識率を表 1 に示す. 表 1 より, 提案手法は全ての動作クラスにおいて認識率が向上した.

表 1: 各動作クラスの認識率 [%]

	投げる	蹴る	ジャンプ
ST-GCN	86.91	90.94	99.63
2 クラス分類	88.36	90.76	99.09
提案手法	<b>96.86</b>	<b>96.98</b>	<b>99.74</b>

獲得した接続パターンを図 3 に示す. 共有部分は, 重みが大きい 60 本のエッジのみを描画している. また, ノードを 5 色に分割し, 分割されたノード方向へのエッジの重みが大きい場合, ノードとエッジの色を同じに描画している. 図 3 より, 共有部分はエッジが体の全体に, 投げる動作では右腕, 蹴る場合は脚にエッジが集中しており, 動作特有の接続パターンを獲得していることが分かる.

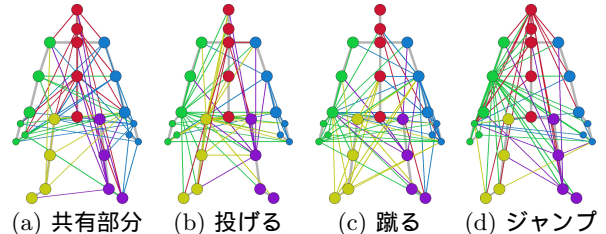


図 3: 獲得した接続パターン

## 5. おわりに

本研究では, エッジの重みを考慮することで, 動作クラス毎の接続パターンを獲得し, 認識率の向上を実現した. 今後の課題として, 他タスクでの実験や, 動作クラス毎の最適なエッジ数を獲得可能な学習方法などが挙げられる.

### 参考文献

- [1] S. Yan, *et al.*, "Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition", AAAI, 2018.
- [2] A. Shahroudy, *et al.*, "NTU RGB+D: A Large Scale Dataset for 3D Human Activity Analysis", CVPR, 2016.