Vol.2021-NL-248 No.20
Vol.2021-CVIM-226 No.20
2021/5/21

個人行動とグループ構成の同時学習によるグループ行動認識

中谷 千洋[1,a)]　千藤 滉平[1]　浮田 宗伯[1,b)]

概要：本研究では，シーン全体で起きているイベントを理解するためのグループ行動認識の精度向上のために「個人行動とグループ構成の同時学習によるグループ行動認識」を提案する．同時学習により「個人行動認識」と「グループ構成」という類似度の高い二つのタスクの情報を共有することで，一方の推定ミスをもう一方の推定結果によって修正することが可能である．同時学習をしない場合に比べ，同時学習をした場合のグループ行動認識の精度が向上したことから，グループ行動認識における同時学習は有効であると示された．また，提案手法は最新手法と組み合わせることができる汎用的な手法であり，提案手法と最新手法を組み合わせることで最新類似法と比べて最高性能を達成した．

1. はじめに

人の行動認識は，コンピュータビジョンにおいて重要な話題の一つである．個人行動認識の研究は盛んに行われており成熟しつつあるのに対して，グループ行動認識の研究は発展途上である．本研究で扱うチームスポーツ映像におけるグループ行動認識は，様々なスポーツでの戦術解析やシーン検索などに応用可能である．

グループ行動認識においては，確率変数間の依存関係をグラフ表現したグラフィカルモデル [1], [2], [3], [4] がよく使われていた．グラフィカルモデルの多くは，まず個々の選手の行動を認識し，その個々の選手の行動や位置情報を基にグループ行動を認識する．このような段階的な手法（図 1 左）では，個人行動認識の誤りが続くグループ構成やグループ行動認識に悪影響を与えることがある．このような認識ミスを減らすため，提案手法 (図 1 右) では，個人行動とグループ構成を同時学習することで相互の認識精度を高め，最終的にグループ行動認識の精度向上を目指す．

本研究の貢献は，以下の 3 点である．

- 「個人行動」と「グループ構成」の同時学習を提案し，グループ行動認識の精度を向上させることを実験で確認した．
- 提案手法の入力の 1 つである個人行動認識結果は，任意の個人行動認識手法の結果から変換可能な形式である．その有用性を様々な個人行動認識手法を使うことで検証した．

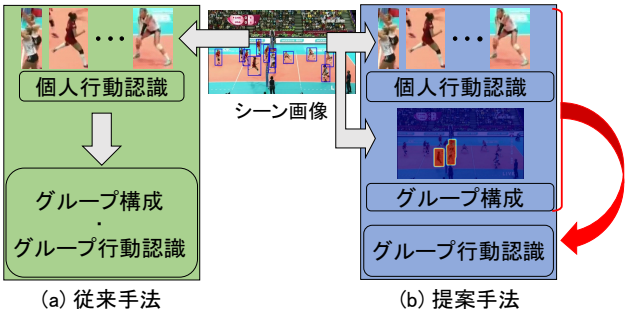

図 1　グループ行動認識手法の比較．(a) 従来の段階的なグループ行動認識手法．(b) 提案する同時学習によるグループ行動認識手法．

- 提案手法を既存グループ行動認識手法とアンサンブルすることで，最新類似法と比べ最高性能を達成した．

2. 関連研究

2.1 個人行動認識

個人行動認識は各人の行動クラスを分類する手法である．Two-Stream Convnet (TSC) [5], [6], [7], [8] は，RGB 画像とフロー画像を用いて各人の個人行動を認識する．TSN [9] は，動画フレームからスパースに数フレームごとの塊を抜き出し，それぞれの塊を使って個人行動認識をした後，その結果を混合する．これにより，TSN は計算コストを減らしつつ効果的に動画全体を学習している．本研究では，パラメータ数が比較的少なく，精度が高い TSC [6] と TSN [9] を個人行動認識のネットワークとして使う．

2.2 グループ構成

本研究におけるグループ構成とは，各シーンの理解において重要な役割を果たしている選手の集合である．従来法

[1] 豊田工業大学
Toyota technological Institute, Nagoya, Aichi, 468-8511, Japan
[a)] sd18064@toyota-ti.ac.jp
[b)] ukita@toyota-ti.ac.jp

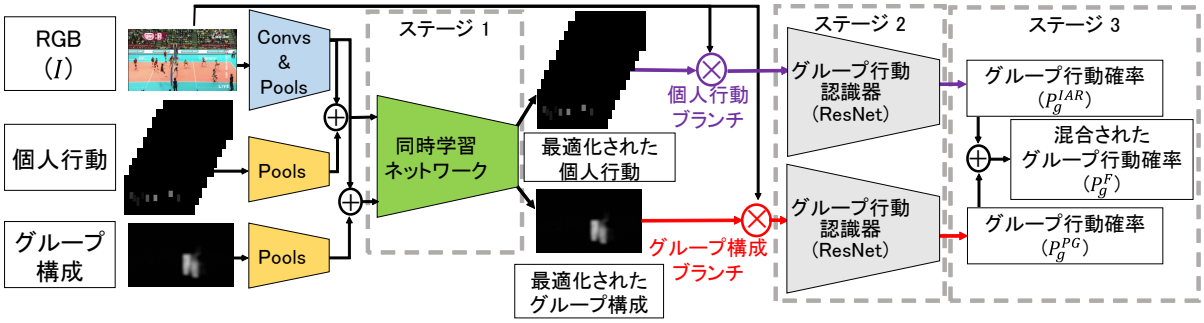

図 2 提案するグループ行動認識のための同時学習ネットワーク．3 種類の入力 (RGB 画像，個人行動とグループ構成のヒートマップ画像) が三つのステージから構成されるネットワークに与えられ，グループ行動認識をする．

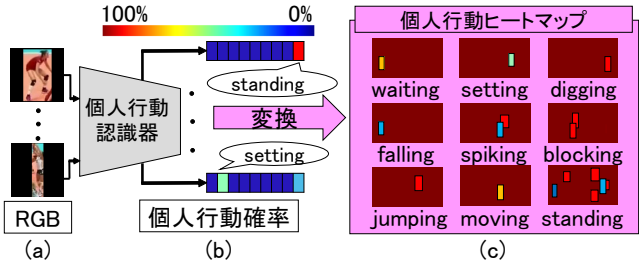

図 3 個人行動認識のヒートマップによる表現．これらの個人行動ヒートマップは図 2 に示されるネットワークに与えられる．この図では，ヒートマップ中の確率値を寒色から暖色に変化する色で可視化しているが，実際に代入される値は 0 から 1 の間の実数値である．

では，人物トラッキングを用いて，人の位置や移動方向の類似度からグループ構成をしていた [10]．単純な位置情報のみに基づくこのような手法は，人の動きをいくつかのパターンに分類できる場合は効果的である．しかし，チームスポーツなどにおける複雑にグループ構成が変化するシーンへの適用は難しい．よって本研究では，単純な位置情報に加えて人の画像特徴量や背景情報を用いてグループ構成をする手法 [11], [12] を使う．

2.3 グループ行動認識

グループ行動認識とは，シーン中の複数人によって構成される集団行動を認識することである．従来のグループ行動認識としては，主に MRF [1] や，AND/OR グラフ [2], [3]，階層的モデル [4], [13] のようなグラフィカルモデルを用いた手法が多い．グラフィカルモデルは難しい状況を表現できるが，認識対象人数が動的に変化する環境に適用するためには難しさがある．スポーツ中継映像では，複数プレイヤーが遮蔽によって一時検出できなかったり，カメラパン，チルト，ズームによってカメラ視野が大きく変動するため，認識対象人数の動的な変化は不可避である．

2.4 同時学習

現在では様々なタスクにおいて同時学習がされている．SPFTN [14] は物体検出とセグメンテーションを同時学習している．Bagautdinov *et al.* [15] はバレーボールシーンにおいて複数人の検出，個人行動認識およびグループ行動認識を同時学習している．提案手法でも，個人行動とグループ構成を同時学習することで相互の認識精度を高め，グループ行動認識の精度向上を目指す．

3. 提案手法

我々が提案した三つのステージから構成される同時学習ネットワークを，図 2 に示す．ステージ 1 では 3.1 節，3.2 節で事前に得た個人行動とグループ構成が，同時学習により最適化される．ステージ 2 では，個人行動ブランチとグループ構成ブランチでグループ行動が独立に認識される．ステージ 3 では，二つのブランチで認識されたグループ行動認識結果を混合することで，最終的なグループ行動認識結果が得られる．

3.1 個人行動認識

各人の個人行動クラスとバウンディングボックスがデータセットにより与えられているとき，個人行動認識のためのネットワーク (TSC [6]，TSN [9]) は，図 3 (a) のように画像から人領域を切り取ることでネットワークを学習する．このネットワークは，各人の個人行動クラスの確率ベクトルを図 3 (b) のように予測する．提案手法では，得られたすべての人の個人行動クラスの確率ベクトルを，ヒートマップに変換する．このヒートマップは図 3 (c) のように，個人行動クラスの数だけ生成される．例えば，図 3 (c) に示される “spiking” のヒートマップでは二人の人が検出され，それぞれのバウンディングボックスに対応する領域が確率ベクトルの値に置き換えられている．ヒートマップの各画素には確率値が入るため，値は 0 から 1 の間の実数値となる．このようにヒートマップを使うことで，認識対象人数が動的に変化する環境に適用するには難しさがあるというグラフィカルモデルの問題点を回避している．

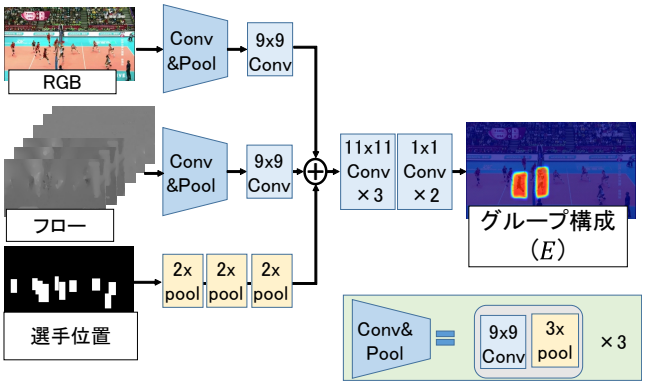

図 4 グループ構成推定ネットワークの構造．このネットワークは 3 種類の入力 (RGB 画像，フロー画像，選手位置画像) からグループ構成を推定する．

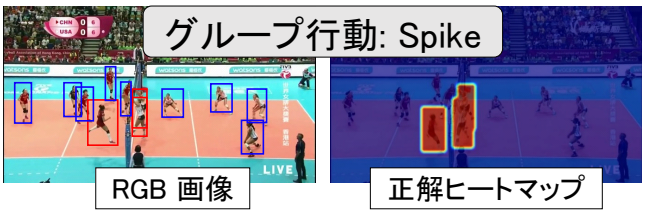

図 5 (左) 同じグループ行動に含まれる人 (この例であれば，スパイクをしている 1 人とブロックをしている 2 人) が，赤いバウンディングボックスで囲まれている．(右) バウンディングボックスから生成された正解ヒートマップ．バウンディングボックスの境界面にはブラーをかけている．

3.2 グループ構成

提案するグループ構成推定ネットワークを，図 4 に示す．このネットワークは，認識対象フレーム t の RBG 画像，$t-1$ から $t+1$ フレームまでの連続 3 フレームの推定フロー画像，フレーム t の選手位置画像を入力とする．選手位置画像は，選手がいる領域といない領域をバイナリで表現しており，各人のバウンディングボックス内部にあるピクセル値は 1 に，そうでないピクセル値は 0 にセットされる．なお，この選手位置画像の生成に使用するバウンディングボックスは SSD [16] により検出する．本ネットワークが推定するグループ構成は，グループ行動に関与する人々が写っている画素が発火するようなヒートマップとして出力される．このヒートマップが推定されるようにネットワークを学習するため，グループ構成の正解ヒートマップは図 5(右) が示すようにアノテーションされる．ヒートマップの画素値は，0 から 1 の実数値を持つ．グループ構成推定ネットワークは，以下の Binary Cross Entropy loss function により学習される．

$$\sum_i \left(-G_i log(s(E_i)) - (1-G_i) log(1-s(E_i))\right), \quad (1)$$

E_i と G_i は，それぞれ推定ヒートマップと正解ヒートマップの i 番目のピクセル値を示している．$s(\cdot)$ はシグモイド関数である．

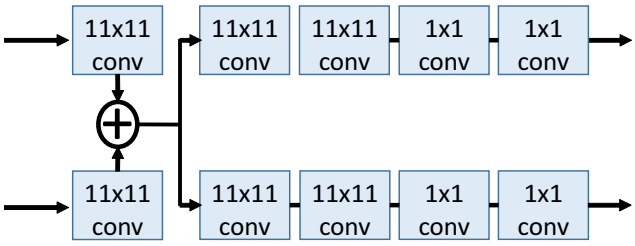

図 6 同時学習ネットワークの構造．このネットワークは個人行動とグループ構成を同時に最適化する．

推論時は，推定ヒートマップの画素値は事前に定義した閾値 0.5 により二値化される．二値化されたヒートマップは以降の処理に渡される．

3.3 グループ行動認識

我々が提案した三つのステージから構成されるネットワークを図 2 に示す．ステージ 1 では，3.1 節で得られた個人行動と 3.2 節で得られたグループ構成が図 6 に示す同時学習ネットワークより最適化される．同時学習ネットワークでは 1 層目の出力の個人行動から得られた特徴量マップとグループ構成から得られた特徴量マップを結合する (図 6 の $\oplus$)．その後，結合された特徴量マップはそれぞれ 2 層目へと与えられる．この特徴量の結合は early fusion とみなすことができるため，認識ミスが互いに修正されることが期待できる．

ステージ 2 では，個人行動ブランチとグループ構成ブランチにおいてグループ行動が独立に認識される．我々はこれら二つのブランチにおける認識ネットワークとして ResNet [17] を使う．この二つのネットワークへの入力には以下の式 (2) から得られる画像 O を使う．

$$O_i = \begin{cases} I_i \times H_i, & H_i > 0.2 \\ I_i \times 0.2, & \text{otherwise} \end{cases} \quad (2)$$

I_i と H_i は，それぞれ RGB 画像 I と同時学習により二つのブランチから得られたヒートマップの i 番目のピクセル値を示している．なお，この処理は図 2 では $\otimes$ と表現される，この処理をすることで，認識ネットワークの入力 O は元の RGB 画像 I の持つグローバルコンテキストを保ちつつ，重要な部分が強調された画像となる．

ステージ 3 では，最終的なグループ行動認識結果がアンサンブルにより求められる．具体的には，アンサンブルは個人行動ブランチとグループ構成ブランチのグループ行動結果 (図 2 において P_g^{IAR} と P_g^{PG} で表現されるグループ行動の確率値) を 7:3 の割合で重み付けすることで実現する．なお，この重み付けの割合は実験から決定した．

我々のステージ 1，2，3 から構成されるネットワークは，以下の二つの損失関数を使って学習される．最初に，式 (1) に示される BCE loss がステージ 1 の同時学習ネットワーク

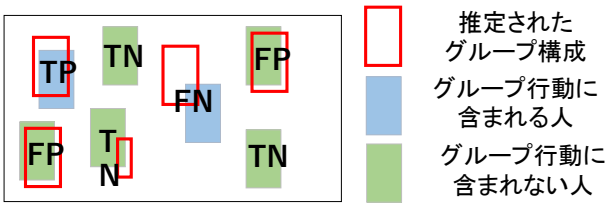

図 7 推定されたグループ構成の評価方法．推定されたグループ構成とグループ行動に含まれる人のバウンディングボックスの IoU が 50%以上であれば TP となり，50%未満であれば FN となる．また，推定されたグループ構成とグループ行動に含まれない人のバウンディングボックスの IoU が 50%以上であれば FP となり，50%未満であれば TN となる．

の二つの出力に対して使われる．次に，cross-entropy loss が P_g^F で示されるステージ 3 の最終的なグループ行動認識結果に対して計算される．これら二つの損失関数を使うことでステージ 1 において個人行動認識とグループ構成の精度を高めつつ，グループ行動認識の精度を向上させることを可能となる．

4. 実験結果

提案手法の評価にはバレーボールデータセット [18] を使用した．このデータセットは，55 のビデオから抽出された 4830 のシーケンスにより構成される．4930 のシーケンスのうち 3493 シーケンスを学習用に，1337 シーケンスをテスト用に使用した．各シーケンスの中央フレームは 9 種類の個人行動クラス (waiting，setting，digging，falling，spiking，blocking，jumping，moving，and standing) によりアノテーションされている．また，各シーケンスは 8 種類のグループ行動クラス (right set，right spike，right pass，right pass，right winpoint，left set，left spike，left pass，and left winpoint) によりアノテーションされている．各フレームは元の画像サイズより小さい 576 × 324 にリサイズして使用した．

本研究では GPU として「GeForce GTX TITAN X」を使用した．同時学習ネットワークの学習には，三つの GPU を用いて約 8 時間かかる．同時学習ネットワークのテストには，一つの GPU を用いて約 1 時間かかる．

個人行動認識とグループ行動認識の評価には，Multi-class Classification Accuracy (MCA) と Mean Per Class Accuracy (MPCA) [19] を使った．グループ構成の評価には，IoU，Precision，Recall，and F-measure を使った．我々の実験では，IoU の閾値は 50%とした．図 7 に示すように IoU の閾値に基づき TP，FP，TN，FN を求め，TP，FP，TN，FN から Precision，Recall，F-measure を計算する．

4.1 個人行動認識

最初に，表 1 に個人行動認識のクラスごとの結果を示す．同時学習の結果 (表 1 の 3 行目) は提案手法により最適化された TSN の個人行動認識結果である．同時学習によって，アンダーハンドパスの動作である “digging” とスパイクのフェイント動作を含んでいる “jumping” の精度が向上している．“digging” はグループ行動クラスの “Pass” に，“jumping” はグループ行動クラスの “Set” によく見られる個人行動であるから，グループ行動認識のために個人行動が最適化されることで “digging”，“jumping” の精度が向上したと考えられる．

続いて，表 2 に個人行動認識を MCA，MPCA で評価した結果を示す．上から 6 行は，3 種類の異なる入力 (RGB 画像，フロー画像，RGB 画像+フロー画像) を使った場合の TSN，TSC の個人行動認識結果を示す．一番下の行は，提案手法である同時学習を使った場合の最も良い個人行動認識結果を示す．同時学習を使った場合の最も良い個人行動認識結果は，提案手法を構成する個人行動認識手法として入力に RGB 画像とフロー画像を用いる TSN を使うことで得られた．

MCA は，同時学習なしで入力に RGB 画像を用いる TSN (表 2 の 1 行目) による個人行動認識結果が最も良かった．一方，MPCA は提案した同時学習での個人行動認識結果 (表 2 の 7 行目) が最も良かった．今回使用したデータセットでは，それぞれの個人行動クラスの人数にばらつきがある．例えば，“standing” ラベルの数は全体の約 69%を占める．よって，個人行動を全て “standing” と認識しても MCA は約 69%の精度が出てしまうため，このようなクラス不均衡な状況では MPCA が MCA より信頼性が高い指標であるとわかる．

4.2 グループ構成

表 3 に同時学習をする場合としない場合のグループ構成推定の結果を示す．3 種類すべての評価指標において，同時学習をする場合の結果は同時学習をしない場合の結果 [11] を上回っている．

4.3 グループ行動認識

まず，提案した同時学習によるグループ行動認識結果を表 4 に示す．同時学習をしないグループ行動認識では事前に得た個人行動認識とグループ構成の結果をステージ 1 を通さずに，直接ステージ 2 に与える．3 種類すべての手法において，同時学習をする場合の結果が同時学習をしない場合の結果を上回った．

次に，図 8 に，グループ行動認識結果の混同行列を示す．この混同行列は列がグループ行動認識の正解ラベル，行が推定ラベルを示している．同時学習をしなかった場合 (図 8 左) と同時学習をした場合 (図 8 右) を比較すると，“Pass” の認識精度が向上していることがわかる．“Pass” の認識精度が向上した理由としては，同時学習によってグループ行動 “Pass” に現れる個人行動 “digging” の精度が

表 1 個人行動認識のクラスごとの結果 (%) の比較.

手法	wating	setting	digging	falling	spiking	blocking	jumping	moving	standing
TSN	50.1	83.7	42.6	84.5	88.2	87.3	0.06	60.1	94.7
TSC	43.7	84.5	33.6	73.7	89.1	85.8	0.00	48.4	96.0
同時学習	51.7	82.1	47.1	80.0	85.9	86.7	15.7	62.5	63.2

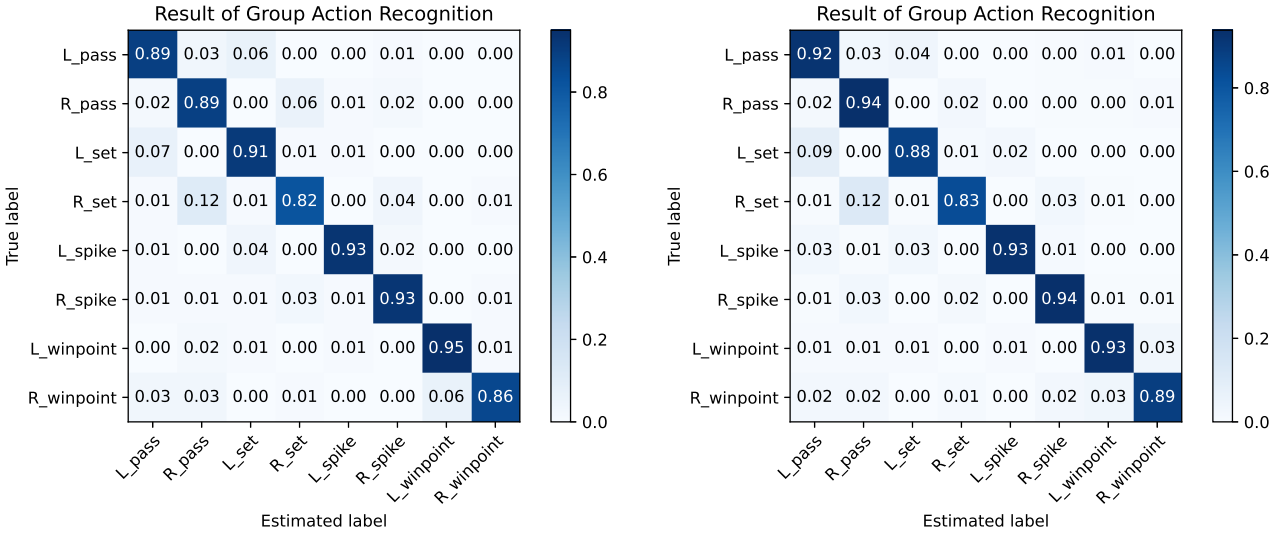

図 8 グループ行動認識結果の混同行列．列がグループ行動認識の正解ラベル，行がグループ行動認識の推定ラベルを示す．(左) 同時学習なしの場合のグループ行動認識結果．(右) 同時学習ありの場合のグループ行動認識結果.

表 2 個人行動認識結果 (%) の比較．提案した同時学習により個人行動認識の精度が向上した.

手法	入力	MCA	MPCA
TSN	RGB 画像	**85.1**	66.3
TSN	フロー画像	82.0	55.6
TSN	RGB 画像+フロー画像	84.9	62.7
TSC	RGB 画像	83.4	57.4
TSC	フロー画像	80.3	44.2
TSC	RGB 画像+フロー画像	82.9	50.4
同時学習 (提案手法)	RGB 画像，フロー画像	84.4	**67.2**

表 3 グループ構成推定結果 (%) の比較．3 種類すべての評価指標において同時学習ありの場合が同時学習なしの場合の結果を上回った.

同時学習	Precision [%]	Recall [%]	F-measure [%]
なし [11]	81.3	79.6	80.4
あり	81.3	**80.6**	**80.9**

向上したこと (表 1 の 3 行目) が挙げられる.

表 5 に提案手法と最新手法 (HDTM [18]，CERN [20]，SSU [15]，StagNet [21]，ARG [22]，and CRM [23]) の比較を示す．提案手法は HDTM，CERN，SSU，StagNet を上回る一方で，ARG と CRM より劣っている．この理由として，CRM に関しては特徴量抽出で 3DCNN を使うこと

表 4 提案手法のグループ行動認識結果 (%)．P_g^{PG}，P_g^{IAR}，P_g^F は 図 2 の右端に示されているグループ行動の確率値である.

手法	同時学習	MCA	MPCA
P_g^{PG}	なし	85.4	86.0
P_g^{IAR}	なし	87.7	87.5
P_g^F	なし	89.7	89.9
P_g^{PG}	あり	85.6	86.1
P_g^{IAR}	あり	89.4	89.1
P_g^F	あり	**90.7**	**90.6**

で，ARG に関しては選手の関係性を求めるために GCN (Graph Convolutional Network) を使うことで提案手法より精度が高いと考える.

3.3 節で述べたように，提案した同時学習は様々な個人行動認識手法と結合することができる．今回は提案手法の精度向上のために，ARG による個人行動認識結果を提案手法における個人行動の入力として与えた．そのグループ行動認識結果 (表 5 の 2 行目) は提案手法を上回る一方，ARG と CRM と比べるとまだ劣っている．さらなる提案手法の精度向上のために，より多くのグループ行動認識結果をステージ 3 におけるアンサンブルのために使った.

3-fusion 手法 (表 5 の 3 行目) では，提案手法の個人行動認識ブランチとグループ構成ブランチでの結果に加え，ARG のグループ行動認識結果を混合した．5-fusion 手法

表 5 最新グループ行動認識手法との比較．提案手法は他の個人行動認識手法の結果を使うことや，グループ行動認識手法とのアンサンブルにより精度が向上した．

手法	MCA	MPCA
提案手法	90.7	90.6
ARG の個人行動認識結果を使用	90.8	90.9
3-fusion	93.0	93.3
5-fusion	**93.3**	**93.6**
HDTM [18] (CVPR2016)	81.9	82.9
CERN [20] (CVPR2017)	83.3	83.6
SSU [15] (CVPR2017)	89.9	-
StagNet [21] (ECCV2018)	89.3	-
ARG [22] (CVPR2019)	92.6	-
CRM [23] (CVPR2019)	93.0	-

(表 5 の 4 行目) では，3-fusion 手法と ARG の個人行動認識結果を使用する手法 (表 5 の 2 行目) の個人行動認識ブランチとグループ構成ブランチでのグループ行動認識結果を混合した．5-fusion 手法はすべてのグループ行動認識の最新手法を上回った．

5. まとめ

我々はグループ行動認識の精度向上のための個人行動認識とグループ構成の同時学習を提案した．実験結果から，同時学習により個人行動とグループ構成が最適化されグループ行動認識の精度が向上することを確認した．また，提案手法は既存のグループ行動認識手法とのアンサンブルにより精度が向上した．

今後の課題として，提案手法を構成する個人行動認識，グループ構成，グループ行動認識手法の更なる精度向上が挙げられる．特に，本研究ではグループ構成は教師あり学習で学習したが，グループ構成のためだけに大量のアノテーションデータを準備するのは困難である．よって，教師なし学習によるグループ構成は重要な問題であると言える．

参考文献

[1] Zhenhua Wang, Qinfeng Shi, Chunhua Shen, and Anton van den Hengel. Bilinear programming for human activity recognition with unknown MRF graphs. In *CVPR*, 2013.
[2] Mohamed R. Amer, Dan Xie, Mingtian Zhao, Sinisa Todorovic, and Song Chun Zhu. Cost-sensitive top-down/bottom-up inference for multiscale activity recognition. In *ECCV*, 2012.
[3] Wongun Choi and Silvio Savarese. A unified framework for multi-target tracking and collective activity recognition. In *ECCV*, 2012.
[4] Tian Lan, Yang Wang, Weilong Yang, Stephen N. Robinovitch, and Greg Mori. Discriminative latent models for recognizing contextual group activities. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(8):1549–1562, 2012.
[5] Karen Simonyan and Andrew Zisserman. Two-stream convolutional networks for action recognition in videos. In *NIPS*, 2014.
[6] Limin Wang, Yuanjun Xiong, Zhe Wang, Yu Qiao, Dahua Lin, Xiaoou Tang, and Luc Val Gool. Temporal segment networks: Towards good practices for deep action recognition. In *ECCV*, 2016.
[7] Christoph Feichtenhofer, Axel Pinz, and Andrew Zisserman. Convolutional two-stream network fusion for video action recognition. In *CVPR*, 2016.
[8] Laura Sevilla-Lara, Yiyi Liao, Fatma Güney, Varun Jampani, Andreas Geiger, and Michael J. Black. On the integration of optical flow and action recognition. In *GCPR*, 2018.
[9] Limin Wang, Yuanjun Xiong, Zhe Wang, Yu Qiao, Dahua Lin, Xiaoou Tang, and Luc Van Gool. Temporal segment networks: Towards good practices for deep action recognition. In *ECCV*, 2016.
[10] Norimichi Ukita, Yusuke Moriguchi, and Norihiro Hagita. People re-identification across non-overlapping cameras using group features. *Comput. Vis. Image Underst.*, 144:228–236, 2016.
[11] Kohei Sendo and Norimichi Ukita. Heatmapping of people involved in group activities. In *MVA*, 2019.
[12] Kohei Sendo and Norimichi Ukita. Heatmapping of group people involved in the group activity. *IEICE Trans. Inf. Syst.*, 103-D(6):1209–1216, 2020.
[13] Mohamed Rabie Amer, Peng Lei, and Sinisa Todorovic. Hirf: Hierarchical random field for collective activity recognition in videos. In *ECCV*, 2014.
[14] Dingwen Zhang, Junwei Han, Le Yang, and Dong Xu. SPFTN: A joint learning framework for localizing and segmenting objects in weakly labeled videos. *IEEE Trans. Pattern Anal. Mach. Intell.*, 42(2):475–489, 2020.
[15] Timur M. Bagautdinov, Alexandre Alahi, François Fleuret, Pascal Fua, and Silvio Savarese. Social scene understanding: End-to-end multi-person action localization and collective activity recognition. In *CVPR*, 2017.
[16] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott E. Reed, Cheng-Yang Fu, and Alexander C. Berg. SSD: single shot multibox detector. In *ECCV*.
[17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016.
[18] Mostafa S. Ibrahim, Srikanth Muralidharan, Zhiwei Deng, Arash Vahdat, and Greg Mori. A hierarchical deep temporal model for group activity recognition. In *CVPR*, 2016.
[19] Rui Yan, Jinhui Tang, Xiangbo Shu, Zechao Li, and Qi Tian. Participation-contributed temporal dynamic model for group activity recognition. In *ACM MM 2018*.
[20] Tianmin Shu, Sinisa Todorovic, and Song-Chun Zhu. CERN: confidence-energy recurrent network for group activity recognition. In *CVPR*, 2017.
[21] Mengshi Qi, Jie Qin, Annan Li, Yunhong Wang, Jiebo Luo, and Luc Van Gool. stagnet: An attentive semantic RNN for group activity recognition. In *ECCV*, 2018.
[22] Jianchao Wu, Limin Wang, Li Wang, Jie Guo, and Gangshan Wu. Learning actor relation graphs for group activity recognition. In *CVPR*.
[23] Sina Mokhtarzadeh Azar, Mina Ghadimi Atigh, Ahmad Nickabadi, and Alexandre Alahi. Convolutional relational machine for group activity recognition. In *CVPR*, 2019.