Attention Branch Network を導入した運転手の 2 次元姿勢推定に関する研究

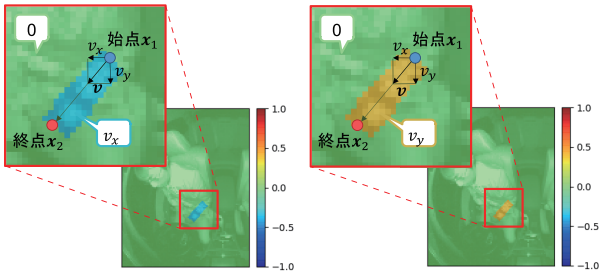
TP17010 奥野薫子　　指導教授：山下隆義

1. はじめに

自動運転はレベル 0 から 5 までの 6 段階あり，特に加速，操舵，制動の動作をシステムが行うレベル 3 および 4 の実用化が期待されている．自動運転レベル 3 および 4 では，緊急時以外の運転はシステムが行い，緊急時は運転手が運転する．このシステムの実現には，運転手の状態を常にモニタリングし，把握する必要がある．運転手の状態を把握するためには，運転手の姿勢推定が重要な要素技術となる．Cao らが提案した OpenPose[1] は，人体の関節間の関連を表す Parts Affinity Fields と呼ばれる補助タスクを導入することで，高精度な姿勢推定を実現した．しかし，OpenPose は深いネットワーク構造のため，計算コストが膨大という問題点がある．本研究では，特定の領域に注目した Attention 情報を導入することにより，深いネットワーク構造が不要な姿勢推定手法を提案する．

2. 関節間の関係を利用した姿勢推定

関節間の関係性を利用した手法として OpenPose がある．この手法では，深いネットワーク構造と，Parts Affinity Fields と呼ばれる 2 点の関節間の領域情報と方向情報を含んだ表現を補助タスクとして学習することで高精度な姿勢推定を可能にしている．図 1 のように始点 $\boldsymbol{x}_1$ から終点 $\boldsymbol{x}_2$ までの単位ベクトル $\boldsymbol{v} = (\boldsymbol{x}_2 - \boldsymbol{x}_1)/\|\boldsymbol{x}_2 - \boldsymbol{x}_1\|$ としたとき，Parts Affinity Fields のベクトル表現は図 1 のようになる．このように関節の位置情報と関節間の関連を表すベクトル情報を学習することで，関節の関連性を捉えることができる．また，OpenPose は複数のネットワークを直列に繋いで深いネットワーク構造にすることで，注視領域がボトムアップ的に広がり，他関節への誤認識を改善する．



(a) x 方向　　(b) y 方向

図 1: Parts Affinity Fields

3. 注視領域と Attention 機構

注視領域とは，物体認識や顔属性認識などにおいて，認識のためにネットワークが着目している画像領域のことである．注視領域は特徴マップの応答値をもとに抽出することができる．ここでは図 2 のように，注視領域を重み付きマップ $\boldsymbol{W}$(Attention map) として特徴マップ $\boldsymbol{F}$ に掛け合わせて重み付き特徴マップ $\boldsymbol{F}'$ を算出する機構を Attention 機構と言う．Attention map を作成する段階で学習に必要と思われる特徴を含むように学習する手法が必要となる．しかし，Attention 機構には特徴マップ $\boldsymbol{F}$ に対して重み付

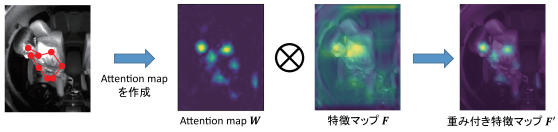


図 2: Attention 機構

けする際に，重みが 0 の部分の特徴が消失する問題がある．そのため，Attention Branch Network(ABN)[2] では，重み付けする際にバイパスを取り入れることで，重みが 0 の部分の特徴を残したまま重みが 0 以外の部分の特徴を強調する．

$$\boldsymbol{F}' = (1 + \boldsymbol{W}) \cdot \boldsymbol{F} \tag{1}$$

4. 提案手法

本研究では，Attention 機構を導入することで，トップダウン的に着目する領域を限定させるネットワーク構造を実現する．これにより，深いネットワーク構造の利点を代用でき，計算コストを削減できる．また，Attention map の作成において，より学習に有効な Attention map の獲得のため，PAFs に代わる補助タスク (Parts map) を学習する．

4.1 Attention 機構を導入した姿勢推定

提案手法は 4 つの役割から構成されており，その中で関節の位置と補助タスクを学習する 2 つのブランチを持つ．提案手法のネットワークを図 3 に表す．図 3 より赤色の領域で強調した部分 (Feature Extractor) は，入力画像の特徴抽出するネットワークである．ここから，特徴マップ $\boldsymbol{F}$ を生成する．次に，緑色の領域で強調した部分では，Branch1 および 2 から Attention map$\boldsymbol{W}$ を作成する．この Attention map は，トップダウンな特徴として理想的な重みを作成するために，図 4 のように関節間の領域に着目した Parts map と関節の位置を推定する Heat map を連結したマップを畳み込む．次に，青色で強調した部分は，ABN に導入されているバイパスを取り入れた Attention 機構である．最後に，黄色で強調した部分は Attention map により重み付けされた特徴マップ $\boldsymbol{F}'$ を入力として関節位置と補助タスクを推定する．このように特徴マップに重み付けすることにより，OpenPose のような深いネットワークを浅くする．また図 4 より，Parts map は始点から終点までに特定のパーツがあるかないかを表現するバイナリマップである．これは，PAFs のようなベクトル表現を簡略化したものと見なすことができる．

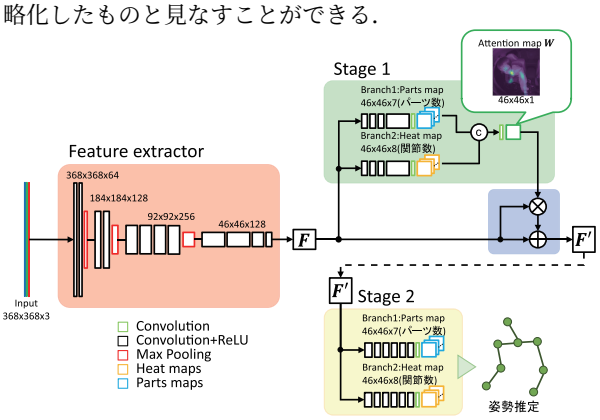


図 3: 提案手法のネットワーク

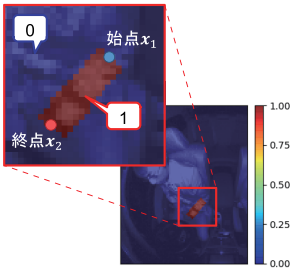


図 4: Parts map

4.2 学習誤差

本研究では，姿勢推定で用いる Parts map および Heat map を式 (2) のような二乗平均誤差 $E_{Branch_{t,x}}$ にて算出する．このとき，$\boldsymbol{L}_{x,n}$ は x 番目のブランチの教師マップ，$\boldsymbol{O}_{t,x,n}$ は t ステージにおける x 番目のブランチの出力マップ，N はマップの枚数を表す．

$$E_{Branch_{t,x}} = \sum_{n=1}^{N} \|\boldsymbol{L}_{x,n} - \boldsymbol{O}_{t,x,n}\|_2^2 \tag{2}$$

最終的に用いる学習誤差 E を式 (3) に示す．

$$E = \sum_{t=1}^{T} \sum_{x=1}^{X} E_{Branch_{t,x}} \tag{3}$$

5. 評価実験

提案手法の有効性を評価実験により確認する．

5.1 実験概要

本実験では OpenPose (stage6) と提案手法 (stage2)，3 つの実験用ネットワーク (stage2) について比較する．比較手法のネットワーク構造を表 1 と図 5 に示す．表 1 と図 5 より，① は同時学習する補助タスク，② は次ステージの入力マップの処理方法を表す．次ステージの入力マップの処理方法は，Attention 機構と複数のマップを連結する処理 (concat) で比較する．なお，事前学習モデルは用いない．最適化手法は MomentumSGD，学習率は 0.001，バッチサイズは 10，学習回数は 30 万 iteration とする．

表 1: 比較手法の補助タスクと次ステージの入力マップの処理

手法	①	②
OpenPose	PAFs	concat
位置情報	なし	concat
位置情報	なし	Attention
位置+パーツ領域方向情報	PAFs	Attention

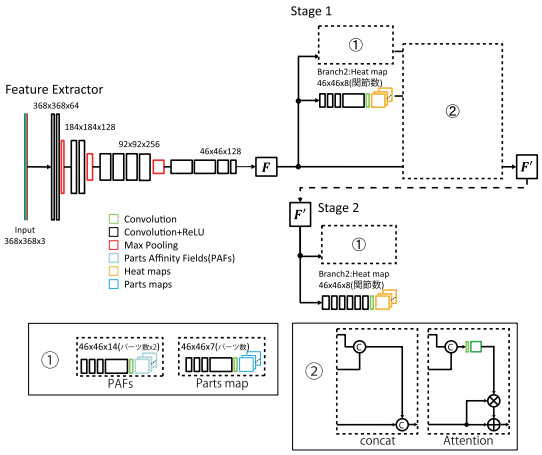


図 5: 比較手法のネットワーク構造

5.2 実験結果

本実験での姿勢推定は，教師信号と出力値との誤差 (ユークリッド距離) を算出し，誤差がしきい値以下であれば検出成功とする．図 6 に姿勢推定の評価結果を示す．ここで，縦軸が検出率，横軸がしきい値である．

・Attention 機構の有用性

図 6 より，しきい値が 4pixel のとき，連結処理を用いた位置情報は 69%，Attention 機構を用いた位置情報は 74%と 5 ポイント精度が向上した．

・導入した補助タスク (Parts map) の有用性

図 6 より，しきい値が 4pixel のとき，Attention 機構を用いた位置+パーツ領域方向情報は 70%，提案手法 (Attention 機構を用いた位置+パーツ領域情報) は 74%と 4 ポイント精度が向上した．これより，Attention 機構とパーツ領域情報ともに姿勢推定の精度向上に有効であることが分かる．

・OpenPose と提案手法の姿勢推定精度の比較

図 6 より，OpenPose は 73%，提案手法 (Attention 機構を用いた位置+パーツ領域情報) は 74%と 1 ポイント精度が向上している．また，図 7 に Heat map の可視化と姿勢推定例を示す．姿勢推定例は，赤色が教師信号，緑色が出力値，部位の位置を示す点の半径は 1pixel となっている．図 7 の Heat map の可視化より，(a) は注視領域のが広く推定しているが，左手の Heat map が右手に大きく誤認識している．(b) は (a) より注視領域が狭く，左手の Heat map は左手を認識している．これは，姿勢推定例の結果からも同じことが言える．

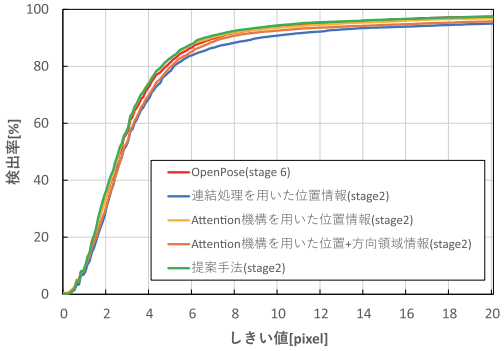


図 6: 姿勢推定の評価

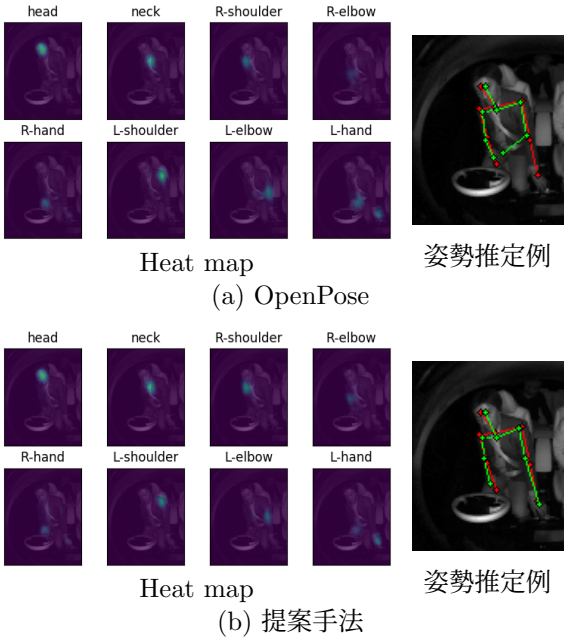


図 7: Heat map の可視化と姿勢推定例

・OpenPose と提案手法の計算コストの比較

図 8 に，GPU(GeForce GTX 1080 Ti) を使用した場合の画像 1 枚当たりの処理時間 [msec] を示す．メモリ使用量は OpenPose と提案手法それぞれ 178MB と 64MB であったことから約 $\frac{1}{3}$ に削減できた．また，図 8 より OpenPose と提案手法はそれぞれ 4.7msec，2.7msec であり，約 $\frac{1}{2}$ に削減し，リアルタイム処理が可能となった．

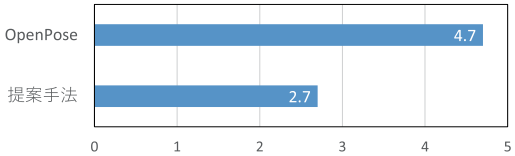


図 8: 画像 1 枚当たりの処理時間 [msec]

6. おわりに

本研究では，Attention 機構の導入により推定精度の向上と Heat map の注視領域の縮小化に成功した．今後は精度向上のため Heat map の改善について検討する．

参考文献

[1] Z.Cao, *et al.*, “ Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields ”, CVPR, 2017.

[2] 福井 宏 等, “ Global Average Pooling の特性を用いた Attention Branch Network ”, MIRU, 2018.

研究業績

[1] K. Okuno, *et al.*, “ Body Posture and Face Orientation Estimation by Convolutional Network with Heterogeneous Learning ”, IWAIT, 2018.

(他 4 件)