

令和 3 年 6 月 18 日現在

機関番号：12608

研究種目：基盤研究(C)（一般）

研究期間：2018～2020

課題番号：18K11348

研究課題名（和文）マルチタスク学習を行う深層学習器のアーキテクチャの開発

研究課題名（英文）Development of deep neural network architecture for multitask learning

研究代表者

川上 玲（Kawakami, Rei）

東京工業大学・情報理工学院・特任准教授

研究者番号：90591305

交付決定額（研究期間全体）：（直接経費） 3,400,000円

研究成果の概要（和文）：申請者は、本事業において、マルチタスク学習（MTL, Multitask Learning）を行う深層学習器、特に、物体の検出と意味領域分割、及び、物体の追跡と検出を行うMTLの開発を行ってきた。タスクの組み合わせやデータセットの選定を行い、MTLを実現する交差接続の提案と改善を行い、MTLによる汎化性能の向上について確認した。交差接続の畳み込みをリカレントニューラルネットワークで代替し、時系列データを処理するネットワークの設計に取り組み、従来手法からの性能改善を達成した。

研究成果の学術的意義や社会的意義

マルチタスク学習を行う深層学習器は今後、人に変わって認識を行うAIにおいて広く普及することになるだろう。本事業はそのマルチタスクの基本的な組み合わせにおいて、設計を吟味し、全般的な性能改善と汎化性能の向上に関する先駆的な結果を得た。一般的にマルチタスク学習では組み合わせても結果が向上するとは限らず、世界的にも様々な試みが提案されている中で、先駆的な結果を得たことは学術的にも社会的にも有意義である。

研究成果の概要（英文）：In this project, the applicant has developed a deep learner that performs multitask learning (MTL, Multitask Learning), in particular, MTL for the combination of object detection and semantic segmentation, and object tracking and detection. We have selected task combinations and data sets, proposed and improved cross-connections to realize MTL, and confirmed the improvement of generalization performance by MTL. The convolution of the cross-connections was replaced by a convolutional recurrent neural network, and a network for processing time-series data was designed, achieving performance improvement over conventional methods.

研究分野：コンピュータビジョン

キーワード：深層学習 マルチタスク学習 ニューラルネットワーク 物体検出 物体追跡 領域分割

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

## 1. 研究開始当初の背景

マルチタスク学習 (MTL, Multi-task Learning) とは, シングルタスク学習 (STL, Single Task Learning) により獲得する特徴量を複数のタスクで共有することで, 各タスクにおける性能を向上させる学習を差す. 人間は, たとえば「テニスをする」というタスクの学習に, 歩く, 走る, 跳ぶ, ボールを投げる, 物を掴む, ラケットを振る, ボールの軌跡を予測する, などのタスクにおける経験を活かしており, これらは同一のタスクではないが, 互いに関連がある. 多くのタスクを学習することで複雑なタスクが実行可能となり, また学習の過程で様々なタスクに適用可能な一般化された表現 (Representation) を獲得するはずである.

MTL の利点は実際に古くから研究され, この語を冠した論文は数多く提案されてきた. MTL は獲得する表現の汎化性能を向上させるので, タスクが似ているほど STL よりも優れた性能を示す. (MTL は, あるタスクの学習器を別のタスクに適用させる転移学習や, 同じタスクであるがデータのドメインが異なるドメイン適応の文脈で言及されることも多い. しかし, 本稿では複数のタスクが実行できる学習器に焦点を絞る.) MTL の応用はゲノミクス, 自然言語処理, コンピュータビジョン (CV) と多岐に渡るが, CV における応用に限れば, 多クラス認識, 領域分割と物体検出, 物体追跡と検出, 法線推定とエッジ分類, 属性推定とクラス認識, 人物の姿勢推定と行動認識, 顔検出と姿勢・年齢・ランドマーク推定, 画像分類と審美的品質推定, などがある. ほとんどが潜在的な情報の階層構造を利用する研究である.

MTL のより近年の動向では, たとえば CV における既存の MTL では, 全体で共有する部分とタスク固有の部分 (たとえば CNN における隠れ層でどこを共有すべきか) は, 実験による試行錯誤で決定しており, 設計時の統一的な指針が明らかでない. Misra らは, 二つのタスクを行う CNN を用意し, 各々の中間層の出力の線形和をとり, これをタスクで交差させて学習する手法を提案した [1]. しかし, 同じ層構造の二つの CNN を前提としており, また交差させるユニットは中間層に対する線形な重みを学習するに留まる. CV における MTL は Misra らの研究を除き, 異なるタスクでも同一の画像で学習することが前提で, 異なる形式の学習データを同時に学習することはできない. また, MTL のための統一的なフレームワークは依然として明らかでない.

[1] I. Misra et al. “Cross-stitch networks for multi-task learning.” CVPR, 2016.

## 2. 研究の目的

本事業では, 二つの MTL の開発を通して, MTL を行う深層学習器のアーキテクチャを模索する. 二つのマルチタスクとは, (1) 物体の検出と意味領域分割, 及び, (2) 物体の追跡と検出, である.

第一段階として, (1) で着目する「物体検出と意味領域分割」における MTL の性能を, STL と比較して向上させることに着手する. Caltech Pedestrian (人検出) と Cityscapes (意味領域分割) などの既存のベンチマークの組み合わせで数パーセント以上の性能向上が達成できれば有効性が確認できる.

第二段階として, (2) で着目する「物体の追跡と検出」では, (1) のマルチタスク学習の構造を時系列処理へ展開し, その構造が, STL や STL の組合せと比較して数パーセント以上の性能向上を達成することを検証する.

### 3. 研究の方法

物体検出と意味領域分割では、図1のようなネットワーク構造を適用し、二つのデータセットを学習させる。既存のMTLでは下層をすべて共有し上層をタスクごとに分離させるか、図2のように中間層を線形和で共有する[1]しかない。一方、図1では、畳込層を経由して情報が共有され、別のタスクに情報を伝達するか否かを交差接続 (Cross-Unit) の畳込層が判断する。交差して挿入された畳込層により、入力画像解像度やチャンネル数の不整合、学習時のランダム性の問題が回避できる。

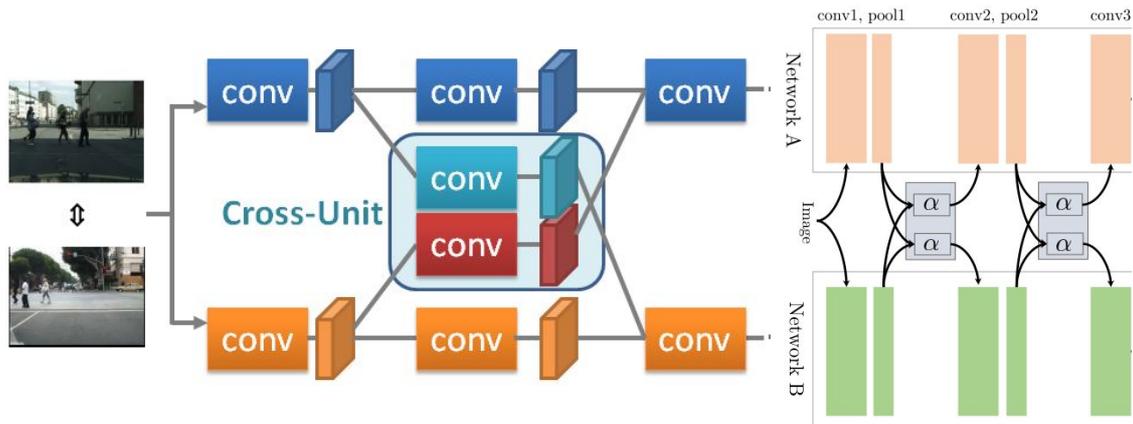


図1 MTLを行うネットワークの構造

図2 [1]の構造

「追跡と検出」は、図3のようなRNNによる構造に基づく。時系列で毎フレーム検索ウィンドウの中でターゲットの位置を追跡しつつ、その物体が何であるかを画像分類する。追跡を行う特徴量と、分類を行う特徴量は、狭い意味でマルチタスク学習になっている(ただし、位置情報の誤差は逆伝播されない。これは、位置情報が追跡に十分なほどアノテーションされていないからである)。追跡では時系列処理を扱うため、図1の交差接続による情報伝達を Conv LSTM (Convolutional LSTM) で代替している。Conv LSTM は時系列の画像から局所的な特徴(物体の変形や加速)を学習し、追跡と分類に利用する。図3の Conv LSTM が図2の畳込層の時系列への拡張であると捉えれば、図1と図3は二つのタスクの情報伝達を間の Conv 層が担っており、似た構造をしている。

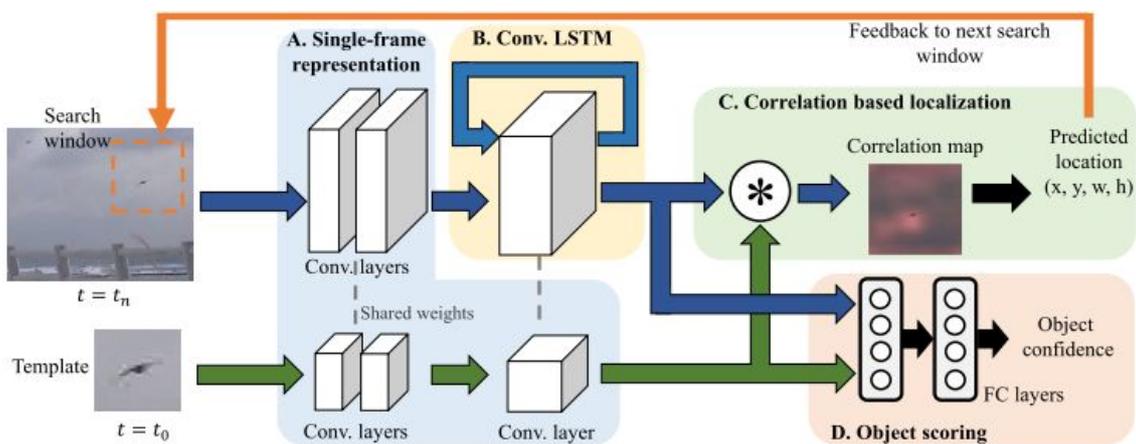


図3 物体の追跡と検出を同時に行うアーキテクチャ

### 4. 研究成果

物体検出と意味領域分割の実験では、意味領域分割において性能向上が見られなかったものの、物体検出では Caltech Pedestrian での (Average) Miss Rate が約 2 % 向上した。詳細な結果を Table 1 に示す。また、これを未学習の KITTI というベンチマークに適用したところ、提案手法が全ての既存手法を上回る結果を得た。Caltech Pedestrian はバウンディングボックスごとの、人か人でないかの 2 クラス分類であるのに対し、Cityscapes は 30 クラスほどの画素ごとの分類(意味領域分割)であるため、CityScapes のタスクの方がより複雑な情報を持ち、これから Caltech Pedestrian のタスクの学習器が恩恵を受けたと考えられる。また、未学習のデータに適用した際に最も性能が良かったことから、比較手法の中ではマルチタスク学習の恩恵を提案手法が最も受けることができたと考えられる。

Table 2 に示すように、鳥検出と領域分割の実験では、マルチタスク学習の効果により、領域分割の性能が向上した。このデータは領域分割のアノテーションが十分でないため、マルチタスク学習の恩恵を容易に受けることができたと考えられる。一方、鳥検出器を未学習の苦前のデータセットに適用したところ、人検出の場合と同様、提案手法が他手法を大きく上回る結果を得た。これらの結果は、申請者の主著で国際会議に採択された。

**Table 1:** Results of detection (MR) and segmentation (IoU) of pedestrians on the Caltech, Cityscapes, and KITTI.

	Caltech MR	City IoU	KITTI MR
Single-task (RPN)	21.47	-	55.29
Single-task (VGG16-PSPNet)	-	<b>76.68</b>	-
Single-task cross-connected (Det)	22.69	-	60.25
Single-task cross-connected (Seg)	-	76.24	-
Share 1	22.15	75.23	50.85
Share 2	22.40	75.31	N/A
Share 3	23.33	75.47	N/A
Share 4	23.21	75.39	N/A
Cross-stitch	22.37	75.39	52.06
Cross-connected (proposed)	<b>19.38</b>	75.33	<b>48.61</b>

**Table 2:** Results of detection (MR) and segmentation (IoU) for birds in the Kinki and Tomamae datasets.

	Kinki		Tomamae
	MR	IoU	MR
Single-task (RPN)	18.59	-	41.35
Single-task (VGG16-PSPNet)	-	33.89	-
Single-task cross-connected (Det)	21.44	-	34.13
Single-task cross-connected (Seg)	-	34.65	-
Share 1	17.74	34.07	44.06
Share 2	18.57	34.39	42.02
Share 3	<b>16.95</b>	34.78	39.56
Share 4	24.45	31.74	41.34
Cross-stitch	18.54	34.01	40.84
Cross-connected (proposed)	18.88	<b>35.45</b>	<b>30.36</b>

本事業の第二段階として、着目した物体検出と追跡について、申請者が平成 28~29 年度で作成した鳥の動画や、既存のドローンの飛行動画に適用し、他手法と性能を比較した。その結果、分類を学習した特徴抽出器が追跡にも効果的に働き、既存の一般的な追跡器よりも追跡精度が高く、かつ、単純な検出器、追跡器と分類器の組み合わせ、空間情報を棄却する LSTM と追跡器の組み合わせ、などよりも、提案手法が高い分類(検出)精度を示した。結果の詳細を図 4 に示す。図 4 の RCN と太字で書かれた濃青と緑が提案手法である。左側の検出の評価では左下にカーブがあるほどよく、右側の追跡の評価では右上にカーブがあるほどよい。提案手法が比較された既存手法を上回っている。提案手法による性能改善は中程度のサイズの物体で顕著であった。これは、データにおける、学習と追跡のしやすさの兼ね合いによると考えられる。これらの結果は arxiv 版への投稿が完了している。

これらの研究を進める中で、検出やマルチタスク学習、時系列処理の派生研究が、国際会議や海外雑誌に採択され、豊富に知見を重ねることができた。3 年間の事業を通し、海外雑誌への採択が 3 件、国際会議への採択が 3 件、arxiv への投稿が 1 件、また、国内会議での発表が 9 件ある。

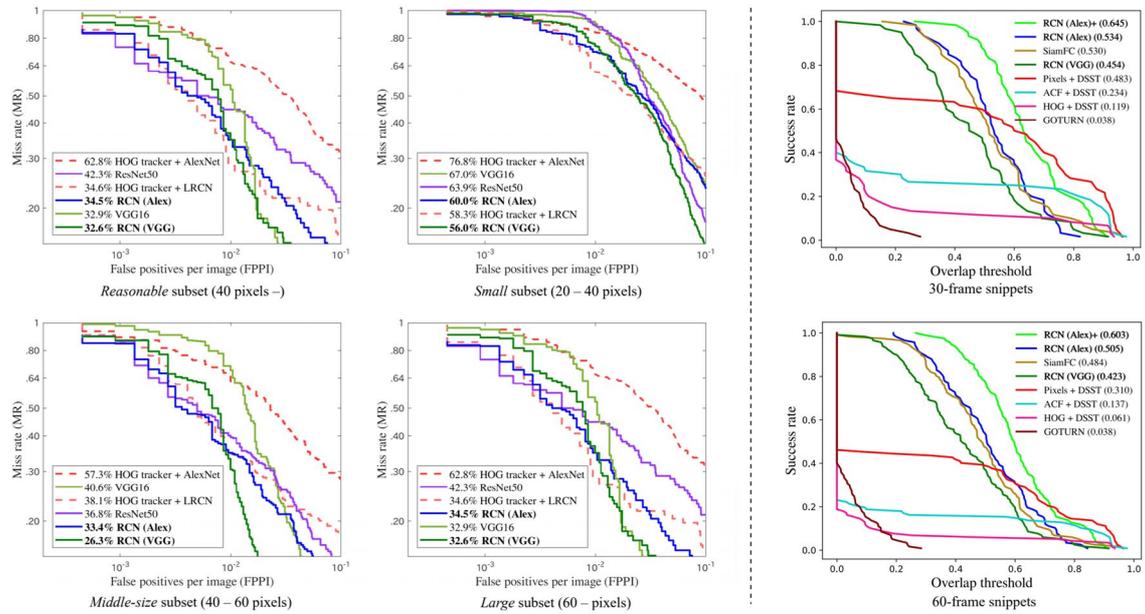


図4 左側に鳥検出の結果を示す。False positive per image に対する Miss rage をプロットしており，左下にカーブがあるほどよい（誤検出が少なく，見落としも少ない）。RCN と太字で書かれた濃青と緑が提案手法である。HOG tracker は既存の追跡器，AlexNet や VGG16 は，既存の検出器，LRCN は全結合層の（畳込層のない）LSTM である。右側に追跡の良さを示す。真値との重なり具合に対する追跡の成功割合をプロットしており，右上にあるほどよい（真値と推定値の重なりの高くても追跡に成功する）。

## 5. 主な発表論文等

〔雑誌論文〕 計8件（うち査読付論文 3件/うち国際共著 4件/うちオープンアクセス 5件）

1. 著者名 Shao Wen, Kawakami Rei, Yoshihashi Ryota, You Shaodi, Kawase Hidemichi, Naemura Takeshi	4. 巻 41
2. 論文標題 Cattle detection and counting in UAV images based on convolutional neural networks	5. 発行年 2019年
3. 雑誌名 International Journal of Remote Sensing	6. 最初と最後の頁 31～52
掲載論文のDOI（デジタルオブジェクト識別子） 10.1080/01431161.2019.1624858	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する
1. 著者名 Wang Yinan, Yoshihashi Ryota, Kawakami Rei, You Shaodi, Harano Tohru, Ito Masahiko, Komagome Katsura, Iida Makoto, Naemura Takeshi	4. 巻 11
2. 論文標題 Unsupervised anomaly detection with compact deep features for wind turbine blade images taken by a drone	5. 発行年 2019年
3. 雑誌名 IP SJ Transactions on Computer Vision and Applications	6. 最初と最後の頁 1～7
掲載論文のDOI（デジタルオブジェクト識別子） 10.1186/s41074-019-0056-0	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する
1. 著者名 川上玲	4. 巻 61
2. 論文標題 OUR Shurijo みんなの首里城デジタル復元プロジェクト	5. 発行年 2020年
3. 雑誌名 情報処理	6. 最初と最後の頁 138～140
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 Yoshihashi Ryota, Trinh Tu Tuan, Kawakami Rei, You Shaodi, Iida Makoto, Naemura Takeshi	4. 巻 10
2. 論文標題 Pedestrian detection with motion features via two-stream ConvNets	5. 発行年 2018年
3. 雑誌名 IP SJ Transactions on Computer Vision and Applications	6. 最初と最後の頁 1--13
掲載論文のDOI（デジタルオブジェクト識別子） 10.1186/s41074-018-0048-5	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

1. 著者名 Seiichiro Fukuda, Ryota Yoshihashi, Rei Kawakami, Shaodi You, Makoto Iida, Takeshi Naemura	4. 巻 5
2. 論文標題 Cross-connected Networks for Multi-task Learning of Detection and Segmentation	5. 発行年 2018年
3. 雑誌名 arXiv:1805.05569v1	6. 最初と最後の頁 1--16
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

1. 著者名 川上玲	4. 巻 59
2. 論文標題 5分で分かる!? 有名論文ナナム読み Geoffrey E. Hinton, Simon Osindero and Yee-Whye Teh : A Fast Learning Algorithm for Deep Belief Nets	5. 発行年 2018年
3. 雑誌名 情報処理	6. 最初と最後の頁 946-948
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 川上玲	4. 巻 72
2. 論文標題 第98回 私の研究開発ツール マイクロソフトコグニティブサービス	5. 発行年 2018年
3. 雑誌名 映像情報メディア学会誌	6. 最初と最後の頁 423-426
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 谷合竜典, 川上玲	4. 巻 59
2. 論文標題 CVPR 2018 参加報告	5. 発行年 2018年
3. 雑誌名 情報処理	6. 最初と最後の頁 1044-1046
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計8件（うち招待講演 0件 / うち国際学会 2件）

1. 発表者名 Ryota Yoshihashi, Wen Shao, Rei Kawakami, Shaodi You, Makoto Iida, and Takeshi Naemura
2. 発表標題 Classification-Reconstruction Learning for Open-Set Recognition
3. 学会等名 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (国際学会)
4. 発表年 2019年

1. 発表者名 Rei Kawakami, Ryota Yoshihashi, Seiichiro Fukuda, Shaodi You, Makoto Iida, Takeshi Naemura
2. 発表標題 Cross-connected Networks for Multi-task Learning of Detection and Segmentation
3. 学会等名 IEEE International Conference on Image Processing (国際学会)
4. 発表年 2019年

1. 発表者名 田口航平, 邵文, 川上玲, 川瀬英路, 苗村健
2. 発表標題 ドローンによる牧場空撮画像における乳牛の個体識別
3. 学会等名 画像センシングシンポジウムSSII
4. 発表年 2019年

1. 発表者名 児玉悠斗, 川上 玲, 坂内博子, 苗村健
2. 発表標題 てんかん診断に向けた神経細胞画像における アミノ酪酸受容体の異常運動の検出
3. 学会等名 画像の認識・理解シンポジウム (MIRU2019)
4. 発表年 2019年

1. 発表者名 王亦楠, 川上玲, 原野徹, 伊藤昌彦, 駒込桂, 飯田誠, 苗村健
2. 発表標題 風車ブレード画像における損傷検知のための教師あり・教師無し学習の比較
3. 学会等名 風力エネルギー利用シンポジウム
4. 発表年 2019年

1. 発表者名 吉橋 亮太, チン トゥアントゥー, 川上 玲, 尤 少迪, 飯田 誠, 苗村 健
2. 発表標題 動き表現の深層学習に基づく小物体の検出と追跡
3. 学会等名 画像の認識・理解シンポジウム (MIRU2018)
4. 発表年 2018年

1. 発表者名 王亦楠, 吉橋亮太, 川上玲, 尤少迪, 原野徹, 伊藤昌彦, 駒込桂, 飯田誠, 苗村健
2. 発表標題 風車ブレードのドローンによる空撮画像における異常検知手法の検討
3. 学会等名 風力エネルギー利用シンポジウム
4. 発表年 2018年

1. 発表者名 森脇健太, 吉橋 亮太, 川上 玲, ユウ シャオディ, 苗村 健
2. 発表標題 GANを用いた単一画像からのハイダイナミックレンジ画像の生成
3. 学会等名 情報処理学会CVIM研究会
4. 発表年 2018年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

Research  
<http://nae-lab.org/~rei/research/>  
CORSR: Classification-Reconstruction Learning  
<https://nae-lab.org/~rei/research/cosr/>  
Cattle detection and counting in UAV images  
<http://bird.nae-lab.org/cattle/>  
Rei Kawakami  
<https://nae-lab.org/~rei/research/research.html>

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------