解　説

深層学習による画像認識

Image Recognition by Deep Learning

藤吉弘亘*　山下隆義*　*中部大学
Hironobu Fujiyoshi* *and* Takayoshi Yamashita*　*Chubu University

1. はじめに

1990 年代後半に入ると汎用コンピュータの進化に伴い，大量のデータを高速に処理できるようになったことから，画像から画像局所特徴量と呼ばれる特徴量ベクトルを抽出し，機械学習手法を用いて画像認識を実現する手法が主流となった．機械学習は，クラスラベルを付与した大量の学習サンプルを必要とするが，ルールベースの手法のように研究者がいくつかのルールを設計する必要がないため，汎用性の高い画像認識を実現できる．2000 年代になると，画像局所特徴量として Scale-Invariant Feature Transform (SIFT) や Histogram of Oriented Gradients (HOG) のように研究者の知見に基づいて設計した特徴量が盛んに研究されていた．このように設計された特徴量は handcrafted feature と呼ばれる．そして，2010 年代では学習により特徴抽出過程を自動獲得する深層学習 (Deep learning) が脚光を浴びている．handcrafted feature は，研究者の知見に基づいて設計したアルゴリズムにより特徴量を抽出・表現していたため最適であるとは限らない．深層学習は，認識に有効な特徴量の抽出処理を自動化することができる新しいアプローチである．深層学習による画像認識は，一般物体認識のコンテストで圧倒的な成績を収めており，様々な分野での利用が進んでいる．本解説では，画像認識分野で深層学習がどのように適用され，どのように解かれているのか，また深層学習の最新動向について解説する．

2. 画像認識における問題設定

一般物体認識は制約のない実世界シーンの画像に対して，計算機がその中に含まれる物体を一般的な名称で認識することである [1]．従来の機械学習 (ここでは深層学習が注目される前の手法と定義する) では，図 1 に示すように入力画像から直接一般物体認識を解くことは難しいため，この問題を画像照合・画像分類・物体検出・シーン理解・特定物体認識の各タスクに細分化して解いてきた．以下では各タスクの定義と各タスクに対するアプローチを紹介する．

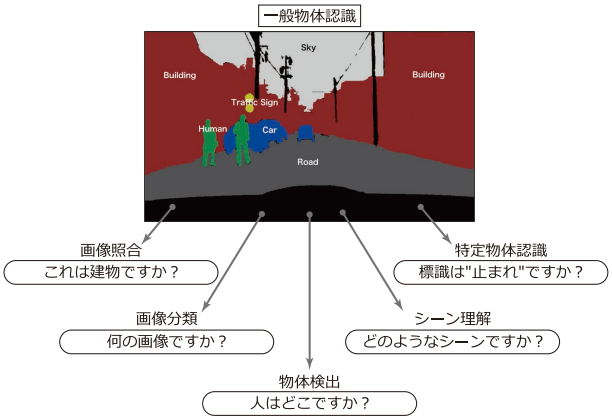


図 1　一般物体認識の細分化．

2.1 画像照合

画像照合は，画像中の物体に対して，登録パターンと同一であるか否かを照合する問題である．照合は登録パターンの特徴ベクトルと入力パターンの特徴ベクトルの距離計算を行い，距離値が小さければ同一，そうでなければ否と判定する．指紋照合，顔照合，人物照合では，本人と他人を判定するタスクとなる．深層学習では，同一人物のペア画像間の距離値が小さく，他人の画像との距離値が大きくなるようなロス関数 (Triplet loss function) を設計し，ネットワークを構築することで人物照合問題を解いている [2]．

2.2 物体検出

物体検出は，あるカテゴリの物体が，画像中のどこにあるかを求める問題である．実用化された顔検出や歩行者検出はこのタスクに含まれる．顔検出は Haar-like 特徴量 [3] と AdaBoost，歩行者検出は HOG 特徴量 [4] と Support Vector Machine (SVM) の組み合わせが広く利用されている．従来の機械学習では，あるカテゴリに対応した 2 クラス識別器を学習し，画像内をラスタスキャンすることで物体検出を実現している．深層学習による物体検出では，複数のカテゴリを対象とするマルチクラスの物体検出を一つのネットワークで実現することができる．

原稿受付　20xx 年 x 月 xx 日
キーワード：深層学習，画像分類，物体検出，シーン理解
*〒 487-8501 愛知県春日井市松本町 1200
*1200 Matsumoto-cho, Kasugai, Aichi 487-8501, JAPAN

2.3 画像分類

画像分類は，画像中の物体があらかじめ定義したカテゴリの中で，どのカテゴリに属するかを求める問題である．従来の機械学習では，画像局所特徴量をベクトル量子化し，画像全体の特徴をヒストグラムで表現する Bag-of-Features (BoF) [5] というアプローチが用いられてきた．その後，画像分類における特徴表現には，特徴ベクトルがもつ情報をより豊かに表現するフィッシャーベクトル [6] や，特徴量を表現するためのメモリ使用量の削減を目的とした Vector of Locally Aggregated Descriptors (VLAD) [7] が提案された．一方，深層学習は画像分類タスクと相性が良く，2015 年には 1,000 クラスの画像分類タスクにおいて，人間の認識性能を超える精度を実現したことで話題となった。

2.4 シーン理解 (セマンティックセグメンテーション)

シーン理解は画像内のシーン構造を理解する問題である．中でも画像中の画素単位で物体カテゴリを求めるセマンティクセグメンテーションは，従来の機械学習では解くことが難しいとされきた．そのため，コンピュータビジョンの最終問題の一つとして捉えられきたが，深層学習の適用により解決できる問題であることが示され始めている．

2.5 特定物体認識

特定物体認識は，固有名詞を持つ物体に属性を付与することで，特定の物体カテゴリを求める問題であり，一般物体認識問題のサブタスクと定義されている．特定物体認識は，画像から SIFT [8] 等で特徴点を検出し，登録パターンの特徴点との距離計算と投票処理により実現されている．ここでは機械学習が直接利用されていないことから，深層学習の適用がされていなかったが，2016 年に提案された Learned Invariant Feature Transform (LIFT) [9] は，SIFT の各処理過程を深層学習で置き換えるように学習し，性能向上を実現している．

以降では，画像分類，物体検出，シーン理解 (セマンティックセグメンテーション) のタスクに焦点を当て，深層学習の適用とその動向について述べる．

3. 画 像 分 類

大規模画像認識のコンペティション ILSVRC (ImageNet Large Scale Visual Recognition Challenge) における 1,000 クラス画像分類問題 (120 万枚の画像から 1,000 クラスのカテゴリ識別を行う classification タスク) では，2012 年以降に深層学習によるアプローチが上位となった．このような大規模画像認識で圧倒的な成績を収めた深層学習手法が Convolutional Neural Network (CNN) [10] である．ILSVRC2012 で優勝したトロント大学が構築した CNN である AlexNet [11] は，図 2 に示すように畳み込みは 5 層，全結合は 3 層で，最終層は 1,000 カテゴリに対応したユニット 1,000 個の構成となっている．図 3 は 1,000 クラスの物体画像を大量に学習した際の AlexNet の 1 層目の畳み込みフィルタを可視化した画像である．方向成分を持つエッジやテクスチャ，カラー情報を抽出する様々なフィルタを学習により自動的に獲得していることがわかる．

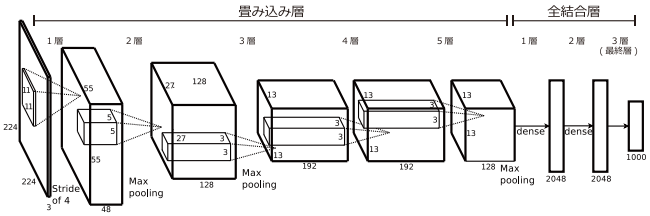


図 2 AlexNet の構造.

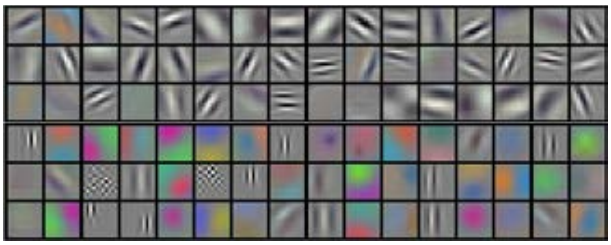


図 3 AlexNet における重みフィルタの可視化例 (文献 [11] より引用).

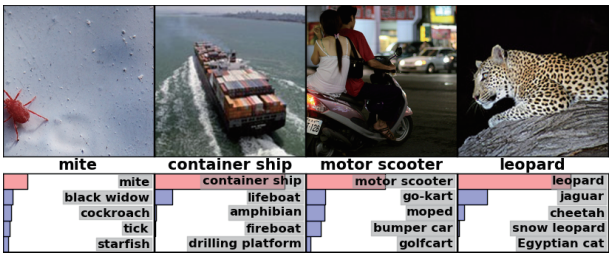


図 4 AlexNet による画像分類結果 (文献 [11] より引用).

2014 年に提案された VGG [12] は 19 層，GoogLeNet [13] は 22 層と畳み込み層を多段にして深くすることで分類性能の向上が図られた．層がさらに深くなると，途中で勾配が 0 となり逆伝播ができなくなる勾配消失問題が発生する．2015 年に発表された ResNet [14] はバイパスを通して逆伝播させるルートを作り，152 層の超深層な構造を実現した．これにより，ResNet のエラー率は 3.56%まで改善した．同様のタスクを人間が行った場合，そのエラー率は 5.1%であり，深層学習のアプローチは人間の能力と同等以上の認識性能を獲得したことで，大きな話題となった．図 4 に AlexNet による画像分類の結果を示す．

4. 物 体 検 出

従来の物体検出は，識別器をラスタスキャンするアプローチであったが，深層学習では物体候補領域を別手法で抽出し，CNN で物体の多クラス分類を行う Region Proposal によるアプローチが多い．これにより，CNN が得意とする分類問

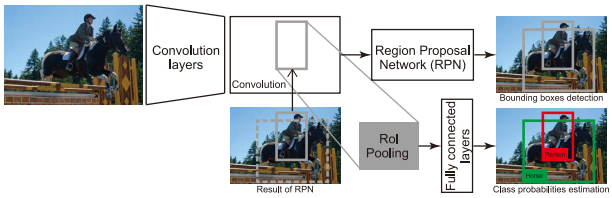


図 5 Faster R-CNN の構造.

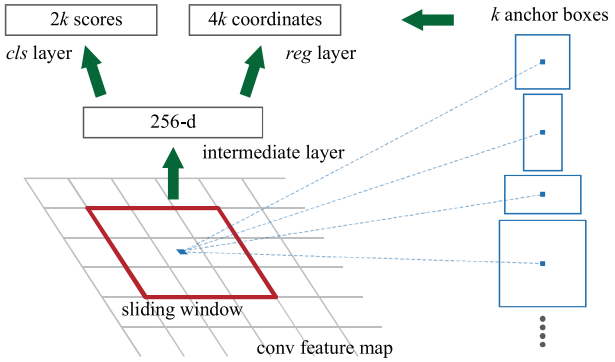


図 6 アンカーによる走査 (文献 [17] より引用).

題として扱うことができる．Regions with Convolutional Neural Network (R-CNN) [15] は，Selective search [16] で検出した物体候補領域を AlexNet (または VGGNet) に入力して物体検出する手法である．Selective search は，様々なしきい値で色やテクスチャが類似する領域を繰り返しグルーピングしていくことで物体候補を大まかにセグメンテーションし，物体候補領域を検出する．CNN を用いた物体検出を実現し，かつマルチクラスの検出ができるようになった点は画期的であるが，Selective search は物体候補領域を求める際に繰り返し領域の統合を行うため，非常に計算コストが高い．

Faster R-CNN [17] は，図 5 のように Region Proposal Network (RPN) を導入し，物体候補領域の検出とその領域の物体クラスの認識を同時に行う．はじめに，入力画像全体に対して畳み込み処理を行い，特徴マップを得る．RPN では，得られた特徴マップに対して検出ウィンドウをラスタスキャンして物体を検出する．ラスタスキャンする際に，RPN ではアンカーという検出方法を導入している．アンカーは，図 6 のように注目領域を中心に k 個の決められた形状の検出ウィンドウを当てはめてラスタスキャンする方法である．アンカーにより指定した領域を RPN に入力し，物体らしさのスコアと入力画像上の検出座標を出力する．また，アンカーで指定した領域は別の全結合ネットワークにも入力され，RPN で物体と判定された際に物体認識を行う．これらの Region Proposal 手法により，アスペクト比の異なる多クラスの物体検出が可能となった．

2016 年には新たなマルチクラスの物体検出アプローチとして，Single Shot の手法が注目されている．これは，画像

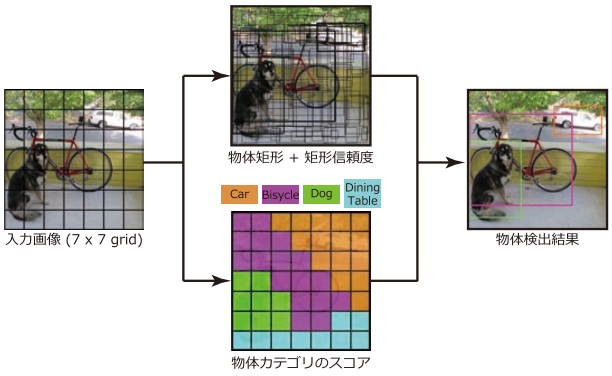


図 7 YOLO の構造.

中をラスタスキャンせず，画像全体を CNN に与えるだけで，複数の物体を検出する手法である．その代表的な手法である YOLO (You Only Look Once) [18] は，図 7 に示すように 7×7 のグリッドで分割した局所領域毎に物体矩形と物体カテゴリを出力する手法である．まず，入力画像を畳み込みおよびプーリング処理を通して特徴マップを生成する．得られた特徴マップ ($7 \times 7 \times 1024$) の各チャンネルの位置 (i, j) が入力画像のグリッド (i, j) に対応した領域特徴となる構造であり，この特徴マップを全結合層に入力する．全結合層を通して得られた出力値は各グリッド位置における物体カテゴリのスコア (20 カテゴリ) と 2 つの物体矩形の位置，大きさ，信頼度となる．そのため，出力層のユニットはカテゴリ数 (20 カテゴリ) に 2 つの物体矩形の位置，大きさ，信頼度 ($(x, y, w, h,$ 信頼度$)$ $\times 2$) を加算し，グリッド数 (7×7) を乗算した数となる．

YOLO は，大まかに定義したグリッドに沿って物体検出をしているため，物体のスケールの変化に弱い．Single Shot Multi-Box Detector (SSD) [19] は，図 8 のように各畳み込み層から物体矩形と物体カテゴリのスコアを出力させることで，スケール変化に頑健な物体検出を実現している．SSD では，入力層に近い層でスケールの小さな物体の矩形を検出し，出力層に近い層ではスケールの大きい物体の矩形を検出している．入力層に近い特徴マップほど，プーリングによる特徴マップの縮小の影響を受けにくい．SSD は，特徴マップの位置毎に物体矩形と物体カテゴリを出力していく．よって，別のネットワークを介して物体カテゴリを推定する必要がないため，高速に物体検出をすることができる．図 9 に SSD による物体検出の結果を示す．

5. セマンティックセグメンテーション

コンピュータビジョン分野において，セマンティックセグメンテーションは難易度の高いタスクであり，長年研究されてきた．そして，高精度なセマンティックセグメンテーションを実現するには時間がかかると考えられていた．しかしながら，他のタスクと同様に，深層学習よる手法が提

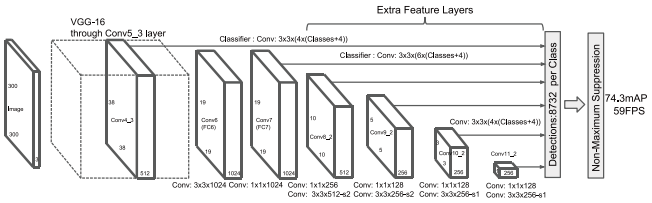


図 8 SSD の構造 (文献 [19] より引用).

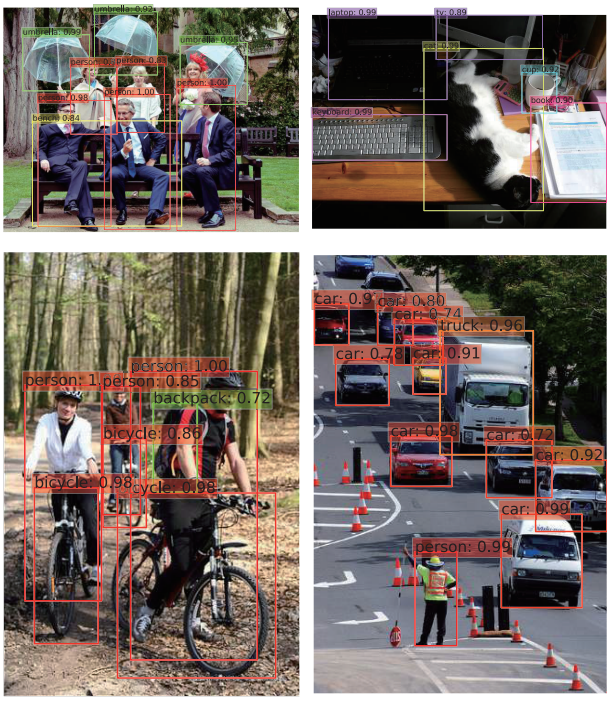


図 9 SSD による物体検出結果 (文献 [19] より引用).

案され，従来手法を上回る性能を達成している．CNN が注目された 2012 年に，3 層構造の CNN により得られた特徴マップとスーパーピクセル手法を組み合わせた手法が提案された [20]．この手法では，複数のネットワークと別の手法との統合が必要であり，複雑な処理を必要とする．

Fully Convolutional Network (FCN) [21] は CNN のみを用いて end-to-end で学習およびラベリングが可能な手法である．FCN の構造を図 10 に示す．FCN は，全結合層を有しないネットワーク構造となっている．入力画像に対して，畳み込み層およびプーリング層を繰り返し行うことで，生成される特徴マップのサイズは小さくなる．元の画像と同じサイズにするために，特徴マップを最終層で 32 倍に拡大処理し，畳み込み処理を行う．これをデコンボリューションと呼ぶ．最終層は，ラベンリングしたい各クラスの確率マップを出力する．確率マップは各画素におけるクラスの存在確率となるように学習している．このように特徴マップの拡大を行うと，粗いセグメンテーション結果となる．そこで，中間層の特徴マップを統合して用いることで高精度化を図っている．一般的に CNN の中間層の特徴マップは入

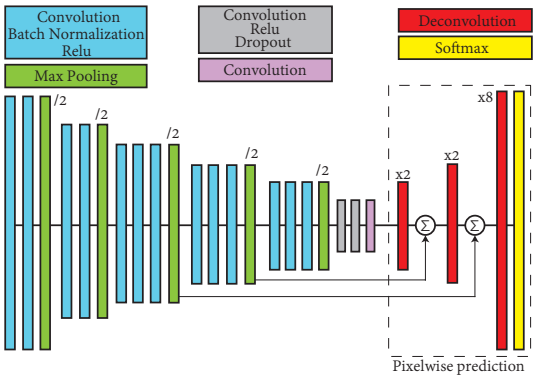
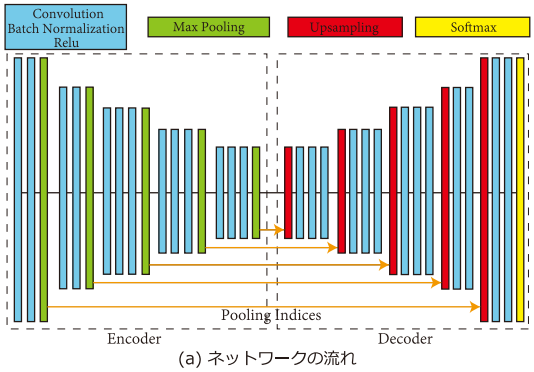


図 10 Fully Convolutional Network (FCN) の構造.

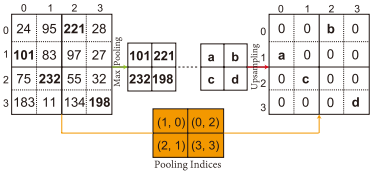
力層に近いほど詳細な情報を捉えている．プーリング処理によりこれらの情報が統合され，細かな情報が欠落している．物体認識においては，これらの詳細な情報は不要であるが，セメンティクセグメンテーションのタスクでは重要な情報である．そこで，FCN は，ネットワークの途中の特徴マップを最終層で統合する処理を行う．FCN はこの統合に用いる特徴マップのサイズにより FCN-32s，FCN-16s，FCN-8s とある．FCN-8s では，3 回目にプーリングした特徴マップと，4 回目にプーリングした特徴マップを最終層の入力に加える．このとき，全ての特徴マップのサイズを 3 回目にプーリングした特徴マップに合わせるために，4 回目にプーリングした特徴マップを 2 倍拡大し，最終層手前の特徴マップを 4 倍拡大する．これらの特徴マップをチャネル方向に連結してコンボリューション処理を行い，元画像と同じサイズのセグメンテーション結果を出力する．

FCN は中間層の特徴マップを記憶しておく必要があり，メモリ使用量が大きい．SegNet [22] [23] は，中間層の特徴マップを記憶する必要がないエンコーダ・デコーダ構成をしている．図 11(a) のように SegNet のエンコーダ側では畳み込み処理とプーリング処理を繰り返し行う．一方，デコーダ側では，エンコーダ側で生成された特徴マップをデコンボリューション処理で拡大し，元の画像サイズのセグメンテーション結果を出力する．これらの処理において，図 11(b) のようにエンコーダ側のプーリングは選択された位置を記憶しておき，デコーダ側で特徴マップ拡大する際に対応する位置にのみ値を挿入する．これにより，中間層の特徴マップを利用せずに，詳細な情報を復元することができる．

PSPNet [24] は，エンコーダ側で得られた特徴マップを拡大する際に，複数のスケールで拡大する Pyramid Pooling Module によりスケールの異なる情報を捉えることができる．Pyramid Pooling Module は，図 12 のようにエンコーダ側で元画像に対して縦および横のサイズがそれぞれ 1/8 に縮小された特徴マップを 1×1，2×2，3×3，6×6 でプー



(a) ネットワークの流れ



(b) エンコーダ・デコーダ部におけるプーリング処理

図 11 SegNet の構造.

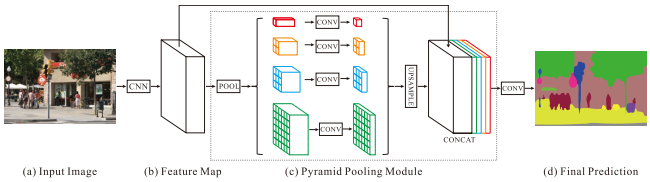


図 12 PSPNet の構造 (文献 [24] より引用).

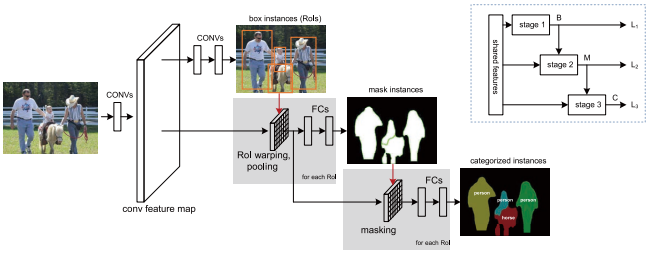


図 13 インスタンスセグメンテーションの構造 (文献 [27] より引用).

図 14 SegNet によるセマンティックセグメンテーション例 (文献 [23] より引用).

リングする．そして，それぞれの特徴マップに対して畳み込み処理を行う．その後，特徴マップを同一サイズに拡大して連結後，畳み込み処理を行い，各クラスの確率マップを出力する．PSPNet は，2016 年に開催された ILSVRC の Scene Parsing 部門で優勝した手法である．また，車載カメラで撮影された Cityscapes Dataset [25] でも高い精度を達成している．

CNN をベースとした手法は，高い精度が得られる一方，ある物体の一部に異なるクラスが混入するような誤ラベリングが起こることがある．特に，このような現象は物体のサイズが大きい場合に起こりやすい．CRFasRNN [26] は，CNN で得られた各クラスの確率マップを入力として，条件付きランダム確率場（Conditional Random Field：CRF）により全体の整合性がとれるように確率マップのチューニングを行う．この手法は，CRF をリカレントニューラルネットワークのように扱い，更新した確率マップを繰り返し入力して徐々に更新するチューニング方法となっている．この手法は， 後処理として行うのではなく，CNN の学習に組み込み，end-to-end で学習することができる．また，確率マップを出力する CNN の構造に依存しないため，様々な手法に導入することが可能である．

セマンティックセグメンテーションは画像の各画素に対してラベリングすることが可能であるが，重なっている物体を区別することはできない．歩行者や自動車などを個別にラベリングするインスタンスセグメンテーションとして，物体検出とセグメンテーションを統合した手法がある [27]．この手法では，図 13 のように Faster R-CNN のように物体候補領域を検出し，その候補領域ごとに物体のマスク画像を生成する．そして，マスク領域内の物体クラスをラベルとしてセグメンテーションを行う．これら一連の処理を end-to-end で行うことで，検出とセグメンテーションを同時に行うことを可能としている．図 14 に SegNet によるセマンティックセグメンテーションの結果を示す．

6. ま と め

本稿では，画像認識のタスクにおいてどのように深層学習が適用されているかを解説した．タスクによってにネットワークモデルや適用方法は異なるが，深層学習で解ける問題は，大量のデータと正確な教師ラベルから写像関数を求めることである．今後は，少量のデータからの学習や，少量の教師ラベル付きデータと大量のラベルなしデータによる半教付き学習の実現が深層学習における課題である．さらには，深層学習がロボットのより良い動作と動作を生成するために必要となる認識過程を同時に獲得するために，

強化学習を含めた end-to-end 学習の実現を大いに期待したい．

参 考 文 献

[1] 柳井 啓司, “一般物体認識の現状と今後”, 情報処理学会論文誌コンピュータビジョンとイメージメディア, vol. 48, no. SIG16, pp. 1–24, 2007.

[2] De Cheng, Yihong Gong, Sanping Zhou, Jinjun Wang and Nanning Zheng, “Person re-identification by multi-channel parts-based cnn with improved triplet loss function”, IEEE Conference on Computer Vision and Pattern Recognition, pp. 1335–1344, 2016.

[3] Paul Viola and Michael Jones, “Rapid object detection using a boosted cascade of simple features”, IEEE Computer Society Computer Vision and Pattern Recognition, 2001.

[4] Navneet Dalal, and Bill Triggs, “Histograms of Oriented Gradients for Human Detection”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 886–893, 2005.

[5] Gabriella Csurka, Christopher R. Dance, Lixin Fan, Jutta Willamowski, and Cédric Bray, “Visual Categorization with Bags of Keypoints”, ECCV Workshop on Statistical Learning in Computer Vision, 2004.

[6] Florent Perronnin, and Christopher Dance, “Fisher Kernels on Visual Vocabularies for Image Categorization”. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2007.

[7] Hervé Jégou, Florent Perronnin, Matthijs Douze, Jorge Sánchez, Patrick Pérez, and Cordelia Schmid, “Aggregating Local Image Descriptors into Compact Codes”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 9, pp. 1704–1716, 2012.

[8] David G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”, International Journal of Computer Vision, vol. 60, pp. 91–110, 2004.

[9] Kwang Moo Yi, Eduard Trulls, Vincent Lepetit, and Pascal Fua, “LIFT: Learned Invariant Feature Transform”, European Conference on Computer Vision, 2016.

[10] Yann LeCun, Léon Bottou, Yoshua Bengio and Patrick Haffner, “Gradient-based learning applied to document recognition”, Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[11] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, “Imagenet classification with deep convolutional neural networks”, Advances in neural information processing systems, 2012.

[12] Karen Simonyan, and Andrew Zisserman, “Very deep convolutional networks for large-scale image recognition”, arXiv preprint arXiv:1409.1556, 2014.

[13] Christian Szegedy, Wei Liu, Yangging Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke and Andrew Rabinovich, “Going deeper with convolutions”, IEEE Conference on Computer Vision and Pattern Recognition, 2015.

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun, “Deep residual learning for image recognition”, IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[15] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation”, IEEE conference on computer vision and pattern recognition, pp. 580–587, 2014.

[16] Jasper R. R. Uijlings, Koen E. A. Van De Sande, Theo Gevers and Arnold W. M. Smeulders, “Selective search for object recognition”, International journal of computer vision, vol. 104, no. 2, pp. 154–171, 2013.

[17] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks”, Advances in neural information processing systems, pp. 91–99, 2015.

[18] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi, “You only look once: Unified, real-time object detection”, IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788, 2016.

[19] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg, “SSD: Single shot multibox detector”, European Conference on Computer Vision, pp. 21–37, 2016.

[20] C. Farabet, C. Couprie, L. Najman and Y. LeCun, “Learning hierarchical features for scene labeling”. IEEE transactions on pattern analysis and machine intelligence, vol. 35, no. 8, pp. 1915–1929, 2013.

[21] J. Long, E. Shelhamer and T. Darrell, “Fully convolutional networks for semantic segmentation”. IEEE Conference on Computer Vision and Pattern Recognition, 2015.

[22] V. Badrinarayanan, A. Kendall and R. Cipolla, “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation”, arXiv preprint arXiv:1511.00561, 2015.

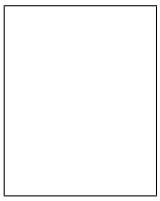
[23] A. Kendall, V. Badrinarayanan and Roberto Cipolla, “Bayesian SegNet: Model Uncertainty in Deep Convolutional Encoder-Decoder Architectures for Scene Understanding”, arXiv preprint arXiv:1511.02680, 2015.

[24] H. Zhao, J. Shi, X. Qi, X. Wang and J. Jia, “Pyramid Scene Parsing Network”, arXiv preprint arXiv:1612.01105,2016.

[25] “The Cityscapes Dataset”, `https://www.cityscapes-dataset.com`

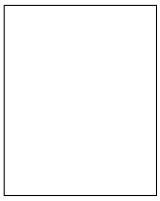
[26] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, P. H. Torr, “Conditional Random Fields as Recurrent Neural Networks”, IEEE International Conference on Computer Vision, pp. 1529–1537, 2015.

[27] J. Dai, K. He, and J. Sun, “Instance-aware semantic segmentation via multi-task network cascades”, IEEE Conference on Computer Vision and Pattern Recognition, pp. 3150–3158, 2016.



藤吉 弘亘（Hironobu Fujiyoshi**）**

1997 年 中部大学大学院博士後期課程修了．1997-2000 年 米カーネギーメロン大学ロボット工学研究所 Postdoctoral Fellow,2000 年 中部大学講師，2004 年より同大学教授,2005-2006 年米カーネギーメロン大学ロボット工学研究所客員研究員．計算機視覚, 動画像処理, パターン 認識・理解の研究に従事．2005 年ロボカップ研究賞, 2009 年 情報処理学会論文誌コンピュータビジョンとイメージメディア優秀論文賞，2009 年 山下記念研究賞,2010・2013・2016 年 画像センシングシンポジウム優秀学術賞，2013 年電子情報通信学会情報・システムソサエティ論文賞，2016 年 画像センシングシンポジウム最優秀学術賞． （日本ロボット学会正会員）



山下 隆義（Takayoshi Yamashita**）**

2002 年 奈良先端科学技術大学院大学博士前期課程修了，2002 年 オムロン株式会社入社, 2011 年 中部大学大学院博士後期課程修了 (社会人ドクター)，2014 年 中部大学講師，人の理解に向けた動画像処理，パターン認識・機械学習の研究に従事．画像センシングシンポジウム高木賞 (2009 年)，電子情報通信学会 情報・システムソサイエティ論文賞 (2013 年)，電子情報通信学会 PRMU 研究会研究奨励賞 (2013 年)，画像センシングシンポジウム最優秀学術論文賞 (2016 年), 同優秀学術論文賞 (2016 年) 受賞．