Vol.2019-CVIM-217 No.10
2019/5/30

領域に応じた超解像倍率選択によるスケール不変な物体検出

秋田 和俊[1,a)]　吉田 智樹[1,b)]　Muhammad Haris[1,c)]　浮田宗泊[1,d)]

概要：本研究では，カメラ画像による物体検出において，画像中に小さく写っている遠方の物体を含めた様々なスケールの物体を同時に検出する手法を提案する．このようなスケールに依存しない物体検出は，自動運転において快適かつ安全な制御を可能にすることができるなど，重要かつ多方面に渡る応用例が考えられる一方で，遠方物体の検出は解像度の低さから非常に難しく，更なる発展が求められている．本研究では，このようなスケールに依存しない物体検出を実現するため，超解像技術を用いる手法を提案する．超解像技術とは，画像の空間解像度を高める技術であり，この超解像により遠方物体を拡大することにより検出が可能になることが期待される．しかし，超解像をそのままカメラ画像全体に適用すると，超解像の不完全さによって別の物体と似た特徴が現れることがあり，誤検出が発生するという問題がある．本研究では，このような誤検出を抑制するため，画像全体のシーン構造などの情報から画像の各領域に適応する最適な超解像倍率を推定する手法を提案する．この手法により，各領域を推定された倍率で超解像を行い，物体検出を行う．これにより誤検出を抑制することができ，この手法を利用しない場合と比較してmAP を約 1.2 %向上（23.95 %から 25.14 %）させることに成功した．

キーワード：物体検出，スケール不変，超解像

1. はじめに

近年，畳み込みニューラルネットワーク（CNN）の発展に伴い，物体検出やセグメンテーションなどの画像認識技術の性能が急速に向上しており，自動運転などへの応用が期待されている．しかし，遠方物体，すなわち低解像な物体の検出は依然として難しく，例えば自動運転車への応用では，多くの場合，遠方物体の検出はミリ波レーダや LiDAR によって行われている．これらのレーダは，100 m以上遠方の物体の検出にも成功しており [1], [2]，自動運転車用の遠距物体検出センサとして注目されている．しかし，電磁波の反射を利用するという特性上，電磁波が反射しない物体は検出ができないことに加え，形状しか認識できず物体の種類の特定が難しい，高価であるなどの欠点がある．カメラ画像はこれらのレーダと違い物体のテクスチャなどの認識が可能であるため，物体の種類の特定に長け，道路標識や白線の認識なども可能である．したがって，カメラ画像でも遠方物体の検出が可能にすることで，ミリ波レーダなどとの組み合わせによる高度な自動車制御の実現や，その他様々な応用が期待できる．

低解像な物体の検出を可能にする手法として，画像の空間解像度を高める超解像技術を用いる手法が提案されている [3]．超解像により画像を拡大した後に物体検出を行うことによって，低解像な物体を検出可能にするものである．この手法により，低解像化した画像においても，高い検出精度を維持することに成功している．この超解像技術を利用し，画像を複数の超解像倍率で拡大した後に検出を行うことにより，近傍物体と遠方物体を同時に検出可能であるスケール不変な物体検出が実現できると考えられる．しかし，超解像による拡大を適用すると，図 1 のように人として誤検出をしてしまうような領域が発生する場合がある．このような誤検出は，超解像が不完全であることに起因するボケや歪みなどが原因であると考えられる．人がいるはずのない領域に対して過剰に高い倍率で拡大を行うことで，人として認識されてしまうような特徴が現れてしまうのである．このため，画像全体に対して様々な超解像倍率による拡大を適用して検出を行うと，図 2 のように誤検出が多く発生してしまうという問題点がある．

このような誤検出を抑制する方法として，各領域に適用するべき超解像倍率を適切に選択するという手法が考えられる．遠方の人物がいるはずがない箇所には超解像を適用しない，もしくは低い超解像倍率でのみ拡大し，一方で遠

[1] 豊田工業大学
Toyota Technological Institute
[a)] sd19401@toyota-ti.ac.jp
[b)] sd19455@toyota-ti.ac.jp
[c)] mharis@toyota-ti.ac.jp
[d)] ukita@toyota-ti.ac.jp

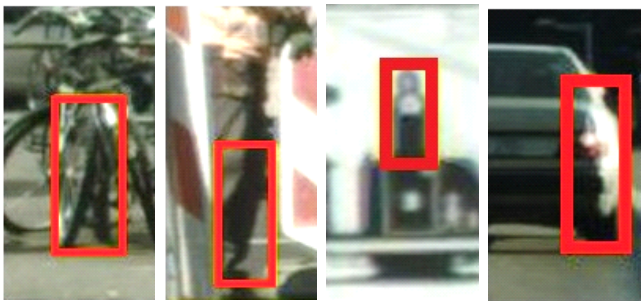

図 1　超解像が原因となり発生する誤検出の例．左から自転車の一部，人の足，広告の一部，車の一部を人として誤検出している．このような誤検出は，超解像が不完全であるために超解像画像に発生するボケや歪みなどが原因であると考えられ，人がいるはずのない領域に対して過剰に高い超解像倍率により拡大を行った場合に多く発生する．

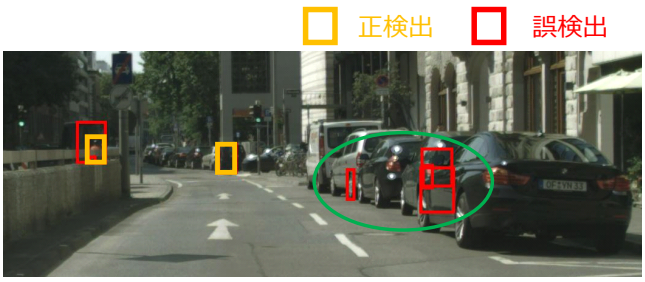

図 2　画像全体をそのまま拡大した場合の検出結果．近傍の車が写っている領域に被るように人がいるという検出結果を出力している．しかし，緑丸で示す領域の例であれば，近傍の車が写っている領域では遠方物体は遮蔽されて見えないはずであることが人間の目で見れば判断できる．このように，シーン構造を認識することで誤検出を抑制できる可能性がある．

方人物が写っている可能性がある場所には高い倍率での拡大を行い検出をすることで，誤検出を抑制するのである．各領域に適応するべき超解像倍率を推定する手段として，画像のシーン構造を利用することが挙げられる．図 2 の例では，緑丸で囲まれた領域は近傍の車が写っているため，遠方人物は遮蔽されて見えるはずがないと認識でき，拡大を行わない，もしくは低い超解像倍率でのみ検出を行えば良いということが判断可能である．このように，注目領域に物体が写っているかどうかや，注目領域周辺の見え方はどのようになっているのかなどのシーン構造によって，適用するべき超解像倍率を推定できると考えられる．事実，物体検出においてこのようなシーン構造を利用することの有効性もいくつかの研究で示されている [4], [5].

本研究では，超解像の利用がスケール不変な物体検出に有効であることを示すと同時に，シーン構造から各領域に適用するべき超解像倍率を推定可能であるという知見に基づいて，倍率選択ネットワークと呼ぶ CNN を構築し，その推定結果に応じた倍率で対応領域を拡大し検出を行うことで，超解像が要因となる誤検出を抑制する手法を提案する．

2. 関連研究

本節では，提案手法で用いる超解像技術と物体検出技術についての動向を紹介する．

2.1 （単一画像）超解像

超解像とは，低解像画像から高解像画像を復元する技術を指し，超解像により復元された画像を超解像画像と呼ぶ．古いデバイスによって撮影された解像度の低い過去の画像や映像を，現在のデバイスに合わせて高解像化する，監視カメラ映像を高解像化してより精度の高い人物同定を可能にするなど，様々な応用が可能な技術である．超解像の手法は様々あるが，本研究では単一画像による超解像を単に超解像と呼ぶ．近年，Dong ら [6] により，超解像に対して CNN の手法が有効であることが示され，以後の研究によって，CNN を用いることでより高倍率な超解像画像が低い復元誤差で生成可能になった [7], [8], [9], [10]．しかし，超解像技術の進展が進むに従って，復元誤差が低い超解像画像が必ずしも視覚的な美しさの向上や物体検出などのタスクにおいて有効ではないことが指摘され [3], [11], [12]，復元誤差に代わる新たな損失関数が提案された．Haris ら [3] は，物体検出に超解像を利用する場合，物体検出ネットワークの学習における損失を超解像ネットワークにも逆伝播させる End-to-end 学習を行うことによって，復元誤差では劣るものの物体検出に有効な超解像画像を生成する超解像ネットワークを学習することに成功した（TDSR）．本研究では，この TDSR を利用することで，遠方の低解像な物体を検出する．

2.2 スケール不変物体検出

物体検出とは，画像中から物体の位置と種類を特定する技術である．Girshick ら [14] により，物体検出に CNN を用いることで性能向上が可能であることが示されると，数多くの研究によって高速かつ高精度な CNN による物体検出手法が提案された [15], [16], [17]．しかし，これらの手法では 1 つのスケールの特徴量マップしか参照せず，幅広いサイズの物体を同時に検出するというスケール不変な物体検出は難しかった．そこで，スケール不変な物体検出を行う手法として，様々なスケールの特徴量マップを CNN により抽出し，各特徴量マップで独立に検出を行う Signle Shot Multibox Detector（SSD）と呼ばれる手法が提案された [13]．これにより，様々なサイズの物体を同時に検出することが可能になった．一方で遠方の小さな物体は解像度の低さや畳み込み回数の少なさから，物体の検出に必要な特徴が十分に抽出できず，検出は依然として難しい問題であった．この問題を解決する手法として，大きく分けて 2 つのアプローチが考えられる．1 つは畳み込み回数の少

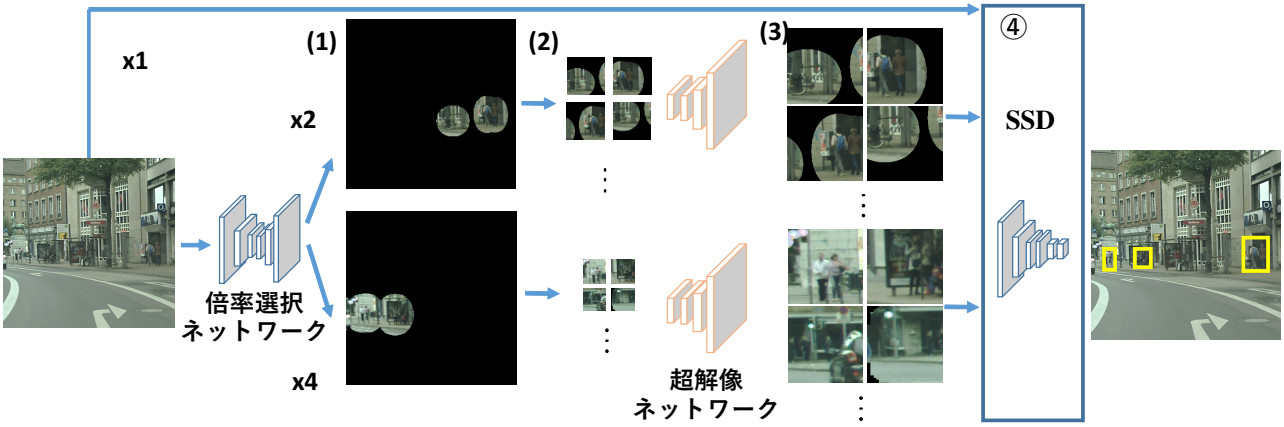

図 3　提案するネットワークの概略図．（1）まず，入力画像から倍率選択ネットワークにより 2 倍もしくは 4 倍に拡大して検出する領域をそれぞれ推定する．（2）次に，超解像により拡大した後の画像サイズが 300 × 300（SSD[13] の入力サイズ）になるようなサイズで元画像をパッチ分割する．このとき，拡大するべき確率が低い領域からはパッチを生成しないようにすることで，拡大の必要がない領域は超解像を適用した検出を行わないようにする．（3）こうして得られたパッチを超解像ネットワークにより拡大し，（4）拡大されたパッチ及び元画像に対しそれぞれ SSD により物体検出を行い，その結果を NMS により統合することで，最終的な出力を得る．

なさを改善するアプローチである．Liu ら [18] は，畳み込みによりよる特徴量抽出がよく行われた小さいスケールの特徴量マップを再び拡大することで，スケールが大きく，かつ検出をするに十分な特徴量を持つ特徴量マップを生成することで，小さな物体の検出を可能にした．また，Lin ら [19], [20] は，畳み込みの浅い層での特徴量マップと拡大によって得られた特徴量マップ対応するサイズ間で結合させることで，さらなる精度向上に成功した．現在，スケール不変な物体検出ではこのようなアプローチが広く用いられており [21], [22]，最新の研究においてもこのアプローチに基づいた手法が提案されている [23]．もう 1 つは解像度の低さを改善するアプローチである．Haris ら [3] は，超解像を利用することにより，低解像画像においても高い検出精度を維持することに成功している．Zhao ら [24] は，超解像画像の復元誤差ではなく，物体検出ネットワーク中で畳み込みによって抽出された超解像画像の特徴量マップが高解像画像と一致するように学習を行う手法を提案している．

現在，超解像を用いてスケール不変な物体検出をする研究は前例がない．そこで本研究では，超解像を利用したアプローチを取ることにより，スケール不変な物体検出に対する超解像の有効性を示すことを目的する．

3. 提案手法

本研究では，画像に様々な超解像倍率での拡大を適用することでスケール不変な物体検出を行い，その上で超解像により発生する誤検出を抑制する手法を提案する．提案するネットワークの概略図を図 3 に示す．提案手法では，初

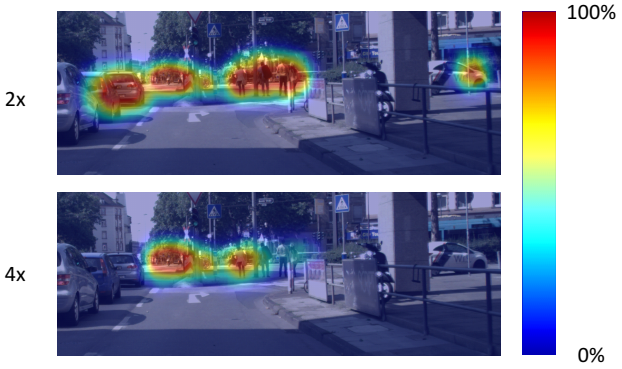

図 4　倍率選択ネットワークにより推定されたヒートマップの例．対応する倍率で拡大するべき領域である確率を推定している．遠方と思われる部分にのみヒートマップが発火しており，近傍に対しての過剰に高い超解像倍率による拡大を回避することができる．

めに各領域に適応するべき拡大倍率を推定するネットワーク（倍率選択ネットワーク）により，RGB 画像から各領域に適応するべき超解像倍率を確率的に推定する（図 4）．次に，超解像による拡大の結果，画像サイズが 300 × 300（SSD[13] の入力サイズ）になるようにパッチ分割を行い，超解像ネットワークにより拡大して SSD へ入力する．このとき，拡大するべき確率が低い領域からはパッチを生成しないようにすることで，拡大の必要がない領域は超解像を適用した検出を行わないようにしている．

3.1 倍率選択ネットワーク

倍率選択ネットワークでは，RGB 画像を入力し，図 4 のように各領域が対応する倍率で拡大するべき確率を表

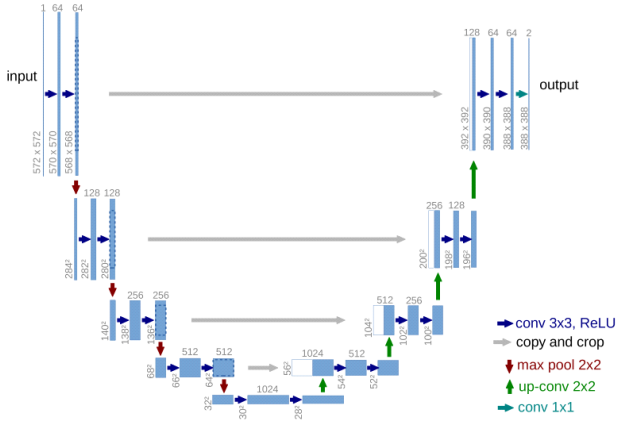

図 5 倍率選択ネットワークの構造として用いる UNet[25] の模式図．一般に，プーリングによる画像サイズの圧縮を行う CNN では，局所的な特徴が失われてしまう．そこで，プーリングによる画像サイズの圧縮が行われていない初期の畳み込み層の出力を，対応するサイズの逆畳み込み層の入力へと結合することによって，プーリングによる局所的特徴の喪失を防ぐ．

すヒートマップを出力する．倍率選択ネットワークの構造には，局所的特徴と大域的特徴の両方を同時に認識可能な構造を持つ UNet[25] を利用する．UNet の構造を図 5 に示す．局所的特徴とは，注目領域に写っている物のテクスチャや遠方に写っている小さな物体の見え方など，画像中の細かな情報を捉える特徴である．局所的特徴により，注目領域に写っている物の種類や小さく写っている物体のようなものの見え方を認識可能である．大域的特徴とは，注目領域の周辺の物体や大きく写っている物体の見え方や，車道や建物の見え方などの，広範囲な画像情報を捉える特徴である．大域的特徴により，注目領域の遠近感や他の物体との位置関係などが認識可能である．このような大域的特徴と局所的特徴の両方を考慮することで，例えば道路が先細りになっていく見え方（大域的特徴）がある部分に小さく物体のようなものが写っている（局所的特徴）領域は，遠方の人物が写っている可能性があるという判断や，注目領域に大きく物体が写っており（大域的特徴），物体の種類が車や建物の壁面である（局所的特徴）場合には，遠方の人物が写っているはずがないという判断が可能になる．

倍率選択ネットワークの学習は，自作のデータセットによって行う．学習データ作成手順を以下に示す．またその模式図を図 6 に示す．

(1) 人以外を含むすべての物体のバウンディングボックスの高さから，その物体が何倍に拡大して検出を行うべきかという振り分けを行う．
(2) 対応する倍率に振り分けられたボックス内部をすべて最大画素値で塗りつぶしたヒートマップを作成する．
(3) ガウシアンフィルタを適用することにより，物体とその周辺にヒートマップが発火するようにする．このとき，物体の中心では画素値が最大となるように調整を行う．

このように学習データを作成することで，実際の画像について遠方もしくは近傍の物体が写っている領域の見え方に応じて画像の各領域に適切な倍率が与えられるため，この学習データにより学習されたネットワークは各領域に適用するべき超解像倍率が推定可能となる．

3.2 超解像による拡大と物体検出

倍率選択ネットワークにより，各領域が対応する倍率で拡大するべき確率が推定される．この確率があるしきい値を上回る領域のみを拡大して検出することで，超解像によって発生する誤検出を抑制するように処理を行う．初めに，元画像を超解像による拡大後に画像サイズが 300 × 300 になるような大きさのパッチにより分割する（例えば 4 倍による拡大を適用する場合には 75 × 75 の画像サイズでパッチ分割を行う）．このとき，パッチ内の 80 %以上の面積が，対応する倍率で拡大するべき確率が高い（一定のしきい値以上）領域で占められるように，かつパッチ同士がオーバーラップをするように設定する．このようにして得られたパッチ毎に超解像による拡大を行い，その後 SSD による物体検出を行うことで，検出結果を得る．このようなパッチ分割を行うのは，SSD の入力サイズが 300 × 300 に固定されているためである．通常の SSD では，入力画像を 300 × 300 に縮小して画像サイズを合わせている．しかし，今回は超解像によって拡大した画像で検出を行う必要性があるので，縮小により画像サイズを合わせると超解像による拡大を無意味にしてしまうため，このような処理を行う．

3.3 検出結果の統合

前述のパッチ分割や様々な超解像倍率の選択により，画像内の同領域で複数の異なる検出結果が得られてしまう．そこで，R-CNN[14] に用いられている Non-Maximum Suppression（NMS）を利用して，推定の確信度が高いバウンディングボックスを残し，重なりの大きいバウンディングボックスを削除することによって，検出結果の統合を行う．NMS の手順は以下のとおりである．

(1) 得られたバウンディングボックスの中から，推定の確信度が最大のものを選択する．
(2) 選択されたバウンディングボックスと，その他すべてのバウンディングボックスとの IoU を計算する．
(3) IoU が一定以上のバウンディングボックスを削除する．
(4) 選択されたことのない残りのバウンディングボックスで 1〜3 の処理を繰り返す．

この NMS により，同領域に現れる複数の検出結果のうち，最も確信度の高い検出結果だけを残すことが可能になる．

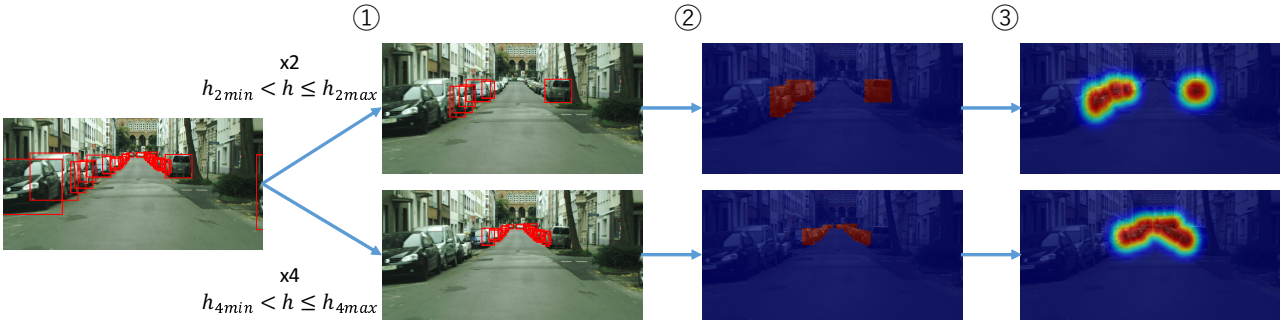

図 6 倍率選択ネットワークの正解データ作成法．(1) 全物体のバウンディングボックスの高さにより拡大すべき倍率に振り分け，(2) 振り分けたバウンディングボックス内部を最大画素値ですべて塗り潰す．その後，(3) ガウシアンフィルタをかけることにより，物体とその周辺にヒートマップが発火するように設定する．このとき，バウンディングボックス中心の確率は 100 %となるようにガウシアンフィルタを適用した結果を調整する．

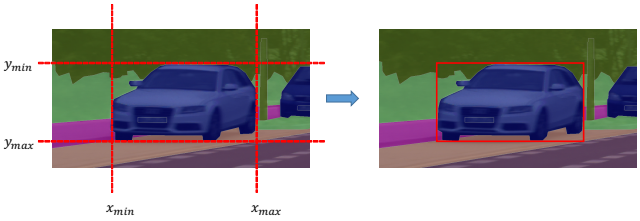

図 7 CityScapes Dataset からのバウンディングボックス算出方法．インスタンスセグメンテーションのデータから，対象物体が存在する x,y 座標の範囲を計算し，バウンディングボックスへと変換する．

3.4 実験条件

最適倍率選択ネットワークの構造には，Ronneberger により提案された UNet[25] の構造をそのまま利用し，最終層の出力チャネルのみ 2 に変更した．これは，今回の倍率選択ネットワークでは，2 倍にして検出するべき領域と 4 倍にして検出するべき領域の 2 種類を推定するため，出力層が 2 つ必要であったためである．また，超解像ネットワークと物体検出ネットワークには，TDSR[3] に倣い，それぞれ DBPN[10] と SSD[13] を利用した．なお，今回は検出対象を人のみに限定して実験を行う．

倍率選択ネットワークの学習は，損失関数として Binary Cross Entropy（BCE），最適化手法には Stochastic Gradient Descent（SGD）を用い，学習率は初期値を 0.1 とし，50epoch の間バリデーションロスが下がらなかった場合に学習率を 1/2 にするよう設定し，学習率が低下してもバリデーションロスが下がらなくなるまで学習を続けた．また，重み減衰は 5e-4，バッチサイズは 4 に設定した．データ拡張として，ランダムに左右反転を加えている．なお，使用メモリの都合上，学習及び検証時には入力画像と出力画像が元画像サイズから 1/2 に縮小した．また，倍率選択ネットワークの学習データ作成では，2 倍に拡大するべき物体のバウンディングボックスの高さは 60 ピクセルから 160 ピクセル，4 倍に拡大するべき物体のバウンディングボックスの高さは 80 ピクセル以下と設定した．利用するガウシアンフィルタは，分散を 30 に設定した．検証時には，拡大するべき確率を表したヒートマップの値が 25 %以上の領域は対応倍率で拡大するべき領域であると設定した．出力結果統合のための NMS では，IoU が 45 %以上のボックスを削除している．

学習及び検証用のデータセットには CityScapes Dataset[26] を利用した．CityScapes Dataset は，遠方の小さく写っている物体にも正確にアノテーションされているため，遠方物体の検出に適している．なお，CityScapes Dataset はインスタンスセグメンテーション用のデータセットであるが，本研究においてはインスタンスセグメンテーションのアノテーションデータから，対象物体の存在する x,y 座標の最小値及び最大値を用いることでバウンディングボックスを算出して利用している．バウンディングボックスの算出方法を図 7 に示す．なお，バウンディングボックスの算出を行うのは，CityScapes Dataset で定義されている human と vehicle のグループに属するクラスの物体のみを対象とし，construction や nature などのグループに属するものは背景として扱う．

3.5 実験結果

提案手法の検証のため，Liu ら [13] により学習済みの SSD と，Haris ら [10] により学習済みの TDSR での検出結果との比較を行った．なお，SSD では元画像を直接入力して得られた結果を，TDSR では倍率選択を行わず，画像全体をパッチ分割し SSD に入力した結果で比較を行う．SSD，TDSR，提案手法による検出結果の一部を図 9 に示す．また，各手法における検出性能を mAP により評価した結果を表 1 に示す．なお，ここでは正解データとの IoU が 50 %以上の検出結果を正検出として扱っている．ただし，一つの物体に対して複数の検出結果が得られている場合には，最も IoU の高いものだけを正検出として，他の検出結果は誤検出として扱うものとする．図 9 からわかる

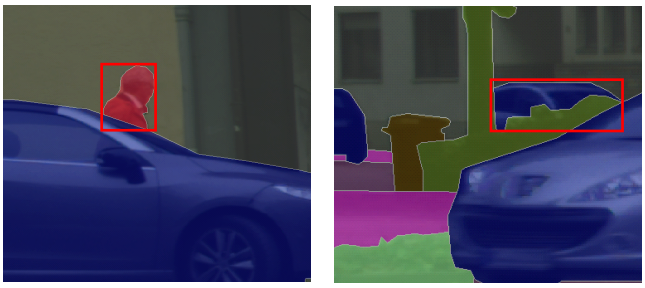

図 8　算出されるバウンディングボックスの高さが遮蔽によって小さくなってしまっている物体の例．本来は，遮蔽されている部分も考慮して全身を囲うようなバウンディングボックスが与えられるべきであるが，セグメンテーション用のデータセットは遮蔽を表現することはできないため，このようなバウンディングボックスが算出されてしまう．

ように，SSD では検出できなかった遠方の物体も，TDSR 及び提案手法では検出できるようになっており，物体検出において超解像の利用が有効であることがわかる．また，TDSR と提案手法を比較すると，誤検出が抑制されている．一方で，TDSR では検出できていたが，誤検出では検出ができなくなってしまったものも発生した．結果として，TDSR と提案手法では，mAP は 1.2 %の向上にとどまった．

検出ができなくなってしまった要因として，倍率選択ネットワークの学習データ作成が適切ではないことが挙げられる．現在，学習データはバウンディングボックスの縦方向のピクセル数（高さ）から，人が設定したしきい値によって拡大すべき倍率を決定している．このしきい値の設定が最適ではなく，適切に超解像倍率が選択できていない可能性がある．また，CityScapes Dataset はインスタンスセグメンテーション用のデータセットであり，一部が遮蔽されている物体は図 8 のようにバウンディングボックスの高さが実際よりも小さくなるものがある．本来は，遮蔽されている部分も考慮して全身を囲うようなバウンディングボックスが与えられるべきであるが．セグメンテーション用のデータセットでは遮蔽を表現することができないため，このようなバウンディングボックスが算出されてしまう．倍率選択ネットワークの学習データ作成においては，バウンディングボックスの高さに応じて物体が存在する領域を何倍で拡大するべきかの振り分けを行っている．そのため，このバウンディングボックスから倍率選択ネットワークの学習データを作成すると，全身を囲うようなバウンディングボックスと比較して高さが小さくなり，本来適用するべき超解像倍率よりも高い倍率で検出するべき領域であるという誤った正解ラベルが与えられてしまい，学習に悪影響を及ぼす可能性が高い．さらに，現在の学習データ作成では，高さにより何倍で拡大するべきかの振り分けをされたバウンディングボックス内を最大画素値で塗りつぶした後にガウシアンフィルタを適用するという作成方法

表 1　従来手法と提案手法での物体検出の mAP．なお，TDSR は倍率選択ネットワークを導入せず，画像全体をそのまま図 3 の（2）へと入力した場合の結果である．SSD と比較して遠方物体検出が可能になり，TDSR と比較して誤検出が抑制されたことにより，提案手法が最も良い mAP を得られた．

	SSD	TDSR	提案手法
mAP	6.9	23.95	25.14

を取っているため，物体が写っている部分とその周辺のみしか拡大して検出するべき領域が存在しないという設定になっている．本来は，物体が認識されていない領域に対しても，シーン構造から遠方であると判断できる場合は拡大して検出をするべきという学習をする必要性がある．

4. まとめと今後の課題

本研究では，画像に対して様々な超解像倍率による拡大を適用するようなスケール不変な物体検出手法を提案し，また超解像を利用した場合に発生する誤検出を抑制するために，画像の各領域に適用するべき超解像倍率を推定するネットワークを導入する手法を提案した．実験から，超解像を利用した物体検出手法は従来手法から大きく性能を向上させることが可能であることが分かった．また，超解像による誤検出を抑制するために，シーン構造を利用し，領域ごとに適応する超解像倍率を適切に選択する倍率選択ネットワークの導入が有効であることが判明した．一方で，超解像を適用する領域が限定されることにより，検出ができなくなるものも発生してしまった．これは，倍率選択ネットワークの学習データの作成方法が適切ではない可能性があることが原因として挙げられる．

3.5 節で述べた倍率選択ネットワークの学習データ作成方法の問題点を解決する手法として，3D シミュレータを用いたデータセット作成 [27] や，イメージインペインティング [28] を利用するものが挙げられる．

3D シミュレータを用いたデータセット作成では，シミュレータ上に人やその他物体を配置した後に，特定の視点からの画像を生成することによってデータセットを作成する手法である．シミュレータ上では物体の位置は既知であるため，遮蔽を考慮したバウンディングボックスを与えることが可能かつ，遠方物体まで完璧な精度で正解データを得ることができる．これにより，遮蔽によりバウンディングボックスが小さくなってしまう問題を解消しつつ，遠方まで正確なアノテーションが与えられたデータセットを作成することができる．

イメージインペインティングは，画像の情報が欠落している箇所を，周辺の情報から補間・復元する技術である．これを用いることで，文字が重ねて表示してある画像などから，文字を自然に削除するなどの応用が可能である．今回，倍率選択ネットワークの学習データでは，物体が写っている部分とその周辺のみしか拡大して検出するべき領域

が存在しないという設定になってしまっていることが問題であった．そこで，イメージインペインティングを用いて，ヒートマップ作成後に元画像から遠方の物体を削除することにより，物体が写っていない領域であってもヒートマップを発火させるような学習データを作成可能になる．

これらのデータセット作成法を用いることによって，倍率選択ネットワークの問題点を解消し，さらなる物体検出精度向上を目指す予定である．

参考文献

[1] 小河昇平，福永貴徳， 山岸傑，山田雅也，稲葉敬之：自動運転支援向け 76GHz 帯高分解能レーダ (特集 次世代通信への挑戦)，SEI テクニカルレビュー， No. 192, pp. 8–13 (2018).

[2] Hecht, J.: Lidar for self-driving cars, *Optics and Photonics News*, Vol. 29, No. 1, pp. 26–33 (2018).

[3] Haris, M., Shakhnarovich, G. and Ukita, N.: Task-driven super resolution: Object detection in low-resolution images, *arXiv preprint arXiv:1803.11316* (2018).

[4] Pan, J. and Kanade, T.: Coherent object detection with 3D geometric context from a single image, *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2576–2583 (2013).

[5] Yan, J., Zhang, X., Lei, Z., Liao, S. and Li, S. Z.: Robust multi-resolution pedestrian detection in traffic scenes, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3033–3040 (2013).

[6] Dong, C., Loy, C. C., He, K. and Tang, X.: Image super-resolution using deep convolutional networks, *IEEE transactions on pattern analysis and machine intelligence*, Vol. 38, No. 2, pp. 295–307 (2016).

[7] Dong, C., Loy, C. C. and Tang, X.: Accelerating the super-resolution convolutional neural network, *European Conference on Computer Vision (ECCV)*, Springer, pp. 391–407 (2016).

[8] Kim, J., Kwon Lee, J. and Mu Lee, K.: Accurate image super-resolution using very deep convolutional networks, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1646–1654 (2016).

[9] Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D. and Wang, Z.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1874–1883 (2016).

[10] Haris, M., Shakhnarovich, G. and Ukita, N.: Deep back-projection networks for super-resolution, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1664–1673 (2018).

[11] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z. et al.: Photo-realistic single image super-resolution using a generative adversarial network, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4681–4690 (2017).

[12] Sajjadi, M. S., Scholkopf, B. and Hirsch, M.: Enhancenet: Single image super-resolution through automated texture synthesis, *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4491–4500 (2017).

[13] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y. and Berg, A. C.: Ssd: Single shot multibox detector, *European Conference on Computer Vision (ECCV)*, Springer (2016).

[14] Girshick, R., Donahue, J., Darrell, T. and Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580–587 (2014).

[15] Girshick, R.: Fast r-cnn, *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448 (2015).

[16] Ren, S., He, K., Girshick, R. and Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks, *Advances in neural information processing systems*, pp. 91–99 (2015).

[17] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A.: You only look once: Unified, real-time object detection, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788 (2016).

[18] Fu, C.-Y., Liu, W., Ranga, A., Tyagi, A. and Berg, A. C.: DSSD: Deconvolutional single shot detector, *arXiv preprint arXiv:1701.06659* (2017).

[19] Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B. and Belongie, S.: Feature pyramid networks for object detection, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2117–2125 (2017).

[20] Lin, T.-Y., Goyal, P., Girshick, R., He, K. and Dollár, P.: Focal loss for dense object detection, *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988 (2017).

[21] He, K., Gkioxari, G., Dollár, P. and Girshick, R.: Mask r-cnn, *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969 (2017).

[22] Zhou, P., Ni, B., Geng, C., Hu, J. and Xu, Y.: Scale-transferrable object detection, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 528–537 (2018).

[23] Zhao, Q., Sheng, T., Wang, Y., Tang, Z., Chen, Y., Cai, L. and Ling, H.: M2Det: A Single-Shot Object Detector based on Multi-Level Feature Pyramid Network, *arXiv preprint arXiv:1811.04533* (2018).

[24] Zhao, X., Li, W., Zhang, Y. and Feng, Z.: Residual Super-Resolution Single Shot Network for Low-Resolution Object Detection, *IEEE Access*, Vol. 6, pp. 47780–47793 (2018).

[25] Ronneberger, O., Fischer, P. and Brox, T.: U-net: Convolutional networks for biomedical image segmentation, *International Conference on Medical image computing and computer-assisted intervention*, Springer, pp. 234–241 (2015).

[26] Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S. and Schiele, B.: The cityscapes dataset for semantic urban scene understanding, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3213–3223 (2016).

[27] Richter, S. R., Vineet, V., Roth, S. and Koltun, V.: Playing for data: Ground truth from computer games, *European Conference on Computer Vision (ECCV)*, Springer, pp. 102–118 (2016).

[28] Liu, G., Reda, F. A., Shih, K. J., Wang, T.-C., Tao, A. and Catanzaro, B.: Image inpainting for irregular holes using partial convolutions, *European Conference on Computer Vision (ECCV)*, pp. 85–100 (2018).

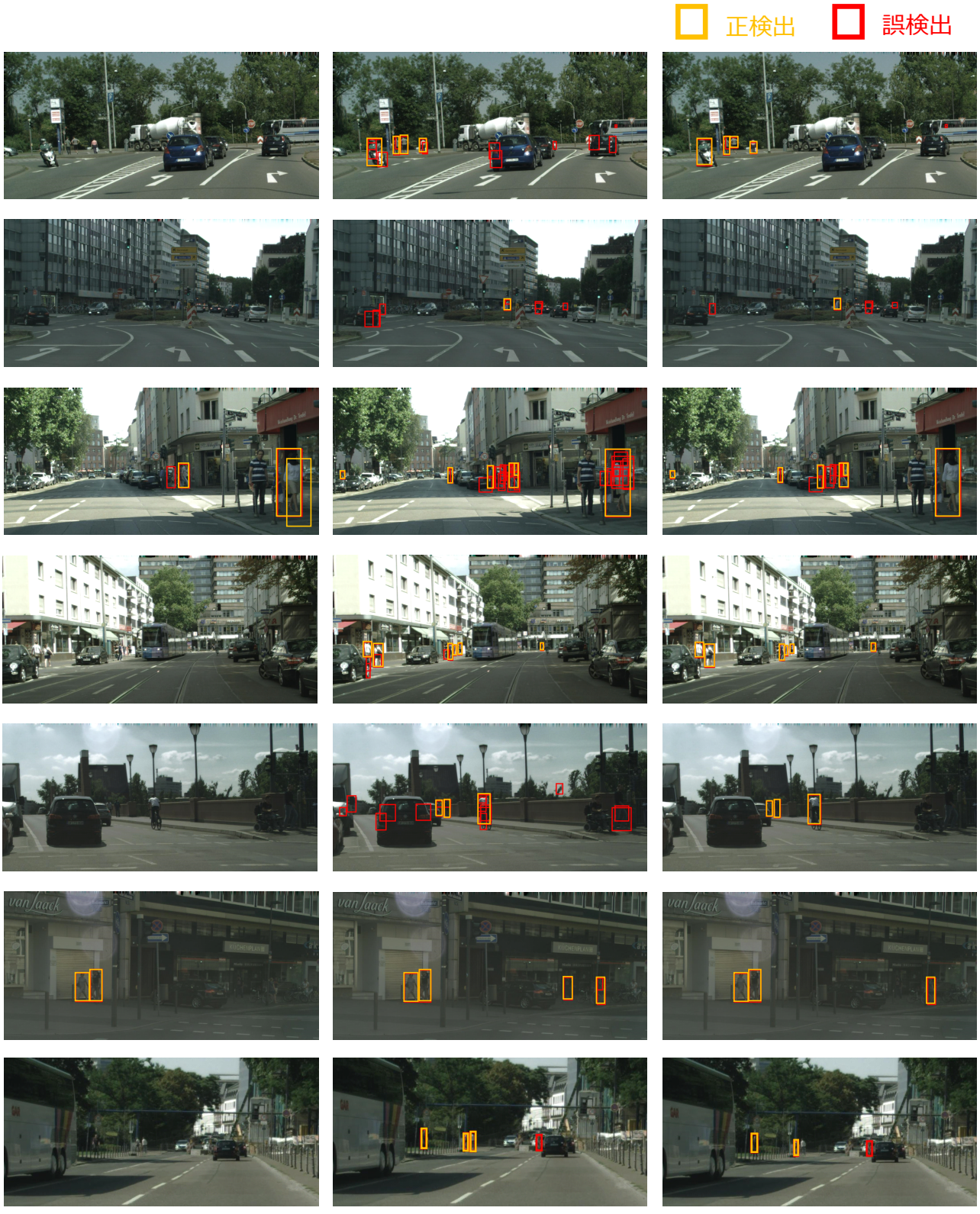

図 9 従来手法と提案手法による検出結果の例．正解データとの IoU が 50 %以上のものを正検出としている．ただし，ひとつの物体に対して複数の検出結果が得られている場合には，最も IoU の高いものだけを正検出とし，残ったものは誤検出として扱う．SSD では検出できていない小さく写っている物体が，TDSR・提案手法では検出できている．また，TDSR で発生している誤検出が提案手法では抑制されている．一方で，TDSR では検出できていたが，提案手法では検出できなくなっているものも見られる．